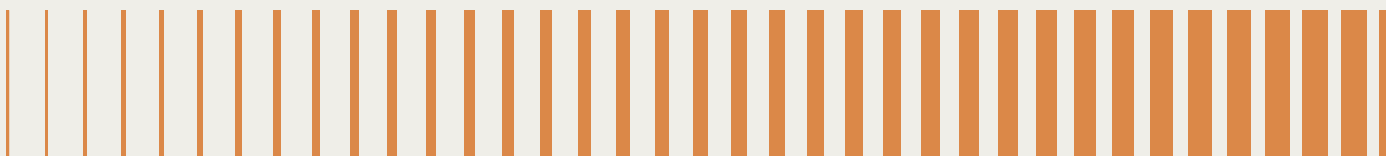


MARCH 2023

# Towards autonomous systems in healthcare



# Executive summary

This output is the second in a series of sector specific deep dives and is part of the National Engineering Policy Centre's (NEPC) project on *Safety and ethics of autonomous systems*. The project set out to explore the potential role of autonomous systems across sectors, and to investigate the opportunities and challenges of developing autonomous systems in healthcare. It also looks at the use of autonomous systems in clinical settings such as hospitals and GP surgeries, as well as the use of these systems in personal devices. We held a series of interviews with experts across the healthcare sector with representation from industry, academia, policymakers, regulators, and the third sector.

If designed and deployed with safety and ethics at the core, implementing autonomous systems across the healthcare sector could improve patient care, shorten hospital stays, lower costs and reduce health inequalities. This report sets out the current technological state of the art for a range of systems, focusing on artificial intelligence (AI) diagnostics, precision medicine, robotics and mobile apps, and explores some of the application-specific challenges and enablers. It includes various case studies to highlight how different applications of automated systems are being used in real healthcare settings and how they have been integrated into clinical workflow.

In healthcare applications, there are currently no systems that are operating autonomously and making informed decisions in complex environments. To enable autonomy in cases where it would significantly improve outcomes, a number of challenges need to be addressed: **safety assurance, regulation, moral responsibility and legal accountability, data governance and interoperability, and skills, culture, and perceptions**. This report considers how these cross-cutting challenges present, and could be addressed, across a range of healthcare applications.

# Introduction

This report is part of the National Engineering Policy Centre *Safety and ethics of autonomous systems* project. As part of a cross-sectoral evidence base on some of the challenges associated with autonomous systems, this report is the **second of a series of sector deep dives** that set out to explore: what is unique about how autonomous systems are developing in each sector; the specific challenges to safe and ethical deployment within sectors; and identification of emerging good practice.<sup>1</sup>

Throughout 2021, the Royal Academy of Engineering held a series of interviews with experts across the healthcare sector with representation from industry, academia, policymakers, regulators, and the third sector. Autonomous systems in healthcare were chosen because of their potential to have a significant impact.

This report explores several questions including:

- **What role might autonomous systems play in healthcare and what are the potential benefits?**
- **How do applications differ?**
- **What are the specific challenges and enabling factors to increasing autonomy for each application?**
- **What are the cross-cutting challenges that need to be addressed across different applications?**

## Increasing automation in healthcare

Automation, autonomy and AI are closely linked concepts that relate to a range of different technologies. In sectors such as transport, the word autonomy has been associated with the embodied technological systems, such as autonomous vehicles which navigate or take action independently. However, these terms also relate to technologies that can underpin either fully autonomous decision-making, or systems that provide advice to human experts. Such systems are generally based on AI and machine learning (ML).

**AI** is a field of science and engineering that uses digital technology to create systems capable of performing tasks commonly thought to require human intelligence.<sup>2</sup> **ML** is a technique within AI that uses algorithms to enable a system to learn from data to provide information or make predictions.<sup>3</sup> **The extent to which these technologies are autonomous depends on the application of these techniques to decision-making.**

AI is the broader set of technologies that to some degree mimic human intelligence or decision-making, with ML referring to systems that learn independently from datasets, and modify their

operation on the basis of such learning. Not all AI and ML systems are autonomous, functioning independently of human operators; however most autonomous systems rely on these techniques to some extent. Some systems are open, continuously learning from new data; others are locked, working on the basis of fixed algorithms derived from programming or learning.

In the healthcare context, technologies may be embedded in physical systems, like surgical robotics; or they may be part of systems that support decision-making, like an AI diagnostic tool that is integrated into medical equipment. In practice, these systems do not tend to operate autonomously. For example, in the case of ML for medical diagnosis and precision medicine, the technologies **typically operate as decision support tools requiring a clinician to confirm the information provided and make the final decision.**<sup>4</sup> **They are not, therefore, operating autonomously.** Furthermore, in these automated systems the model is fixed or 'locked' before the system is deployed, making the system *deterministic* when in use and the system has no self-learning function.

If future technologies allow self-learning to occur during operation, the system can become nondeterministic, where the same input results in multiple different outputs. This means the way the system operates can change during use.

If the system is operating autonomously in a nondeterministic environment, it becomes more difficult to validate and verify the system; this can create questions about how to provide safety assurance for the decisions it makes.

As AI and ML systems are more widely used, evidence of their efficacy and trustworthiness could encourage the development and use of autonomous systems. But the healthcare sector is a safety-critical domain where the role of technology requires deep consideration because of the complex interactions between machines, clinicians, healthcare professionals, and patients. Therefore, increasing the level of autonomy, to give decision-making responsibility to the system, would be a major step change and would require consideration of the complexities around assigning moral responsibility and challenges traditional concepts of ethics. Moral responsibility refers to how far an individual may receive blame or praise for a decision or action made by a system for which they had some control.<sup>5</sup> A multidisciplinary approach is needed to establish appropriate governance over the development and deployment of autonomous systems in the healthcare sector, and this report seeks to identify the challenges that such an approach must address.

## Benefits from increasing automation

Implementing automated systems across the healthcare sector can present opportunities to improve quality of care.<sup>6</sup> Applications such as AI diagnostics and precision medicine offer the potential to predict and detect disease, supporting preventative medicine; and to provide an understanding of the individual that enables improved, more personalised treatment. Effective triaging supported by automated systems can direct individuals to the most appropriate services, making more efficient use of the available care provision. With support for decision-making along the healthcare pathway, bottlenecks can be reduced, shortening hospital stays and lowering care costs. These benefits may be most significant where expertise and resources are already limited, which could help reduce some healthcare inequalities.<sup>7</sup>

The NHS has experienced heightened demand and pressure on its services. In 2022, NHS waiting lists for treatment have reached 6.5 million patients. The staff vacancy rate has increased to 7.9%, up from 5.9% in 2021. Demand is expected to continue growing with people aged 65 and over predicted to represent 24% of the population by 2043.<sup>8,9</sup> Meeting this demand will present challenges. Although there are a number of fundamental factors leading to this pressure on the NHS, if automated systems prove themselves to be trustworthy, obtain regulatory approval, and continue to deliver the benefits outlined above, could increased autonomy be part of the solution to reducing the burden on clinical specialists and the system as a whole?

Increasing the level of autonomy, to give decision-making responsibility to the system, is a step change that requires consideration of the complexities around assigning moral responsibility and challenges traditional concepts of ethics

Applications such as AI diagnostics and precision medicine offer the potential to predict and detect disease, supporting preventative medicine; and to provide an understanding of the individual that enables improved, more personalised treatment

# Realising the systems improvement

To be able to realise the benefits of these technologies, they need to be developed and deployed with a deep understanding of the human function they are augmenting, whether they are doing so in part or in their entirety, and with careful consideration of where there may be unintended consequences. Aside from personal medical devices such as mobile apps, in healthcare settings the system would still be used alongside human clinicians who understand how the system works, and so design for human – machine cooperation is critical. Autonomous systems should only be deployed within a safety envelope which clearly defines the boundaries for safe operation, which would require human oversight at critical points.

Consideration of how these systems can be effectively integrated into the clinical workflow is vital. New training will be required to equip the workforce with the skills to work with these systems but also to challenge, assess, and mitigate the risks that arise from increased autonomy. Alongside training, a significant culture change will be required as roles evolve and change. Both the legal and regulatory framework will need to be in place to allow for the safe and ethical development and deployment of these systems. This will require engagement across society

towards agreement on the responsible and ethical applications of these technologies; as well as the ways to provide assurance that autonomous systems are operating in ways that are trustworthy, safe and ethical.

## **View from a practitioner: Professor Lionel Tarassenko FEng FMedSci**

The increasing use of AI systems in healthcare, for example in speeding up diagnoses from the analysis of medical images or to stratify patients, is a very positive development. However, all the current systems require a human in the loop, from the selection of the data to train the algorithm, to the use of the trained system in the field, when its outputs are reviewed by a human expert. The advances being made through the deployment of ML algorithms in healthcare invariably rely on a human expert being in the decision-making loop. The systems that incorporate these algorithms are trained on carefully curated datasets and tested on independent data. They operate within a regulated framework, the Software and AI as a Medical Device framework, overseen by the UK's Medicines and Healthcare products Regulatory Agency (MHRA).

# Applications

This report looks at a range of applications to explore how the challenges differ depending on the application. This is also impacted by the environment in which the care is being administered, for example, through a robotic device during surgery or use of a mobile app for a remote hospital appointment. There are also different levels of oversight to consider, for example, there will be significant difference between a mobile app that an individual uses at home or a decision support tool that is interpreted by a clinician.

Recognising there are different ways to group these applications, this report has initially focused on the following areas: AI diagnostics, precision medicine, robotics, and mobile apps. For each application the current state of the art, challenges, and enablers are discussed in the table below. The report focuses on these categories as they are currently the areas that offer greatest impact and opportunity because of their current deployment.

The use of automated systems for healthcare administration has intentionally been omitted as this is not so directly linked to patient care. As such, it presents a different set of challenges with respect to patient interaction and potential harm. The use of automated systems for care in domestic settings has also been omitted, but this is a domain which warrants further investigation as technology development and deployment will create unique challenges if they operate unsupervised within people's homes.<sup>10</sup>

This report includes various case studies to highlight how the different applications are being used in real healthcare settings and how they have been integrated into clinical workflow. The examples selected are all automated systems that have received regulatory approval; they are not yet operating autonomously and currently have no self-learning aspects. As argued above, there are benefits to increasing autonomy if automated systems prove themselves trustworthy.

New training will be required to equip the workforce with the skills to work with these systems but also to challenge, assess, and mitigate the risks that arise from increased autonomy

CURRENT STATE OF THE ART	CHALLENGES TO THE DEVELOPMENT AND DEPLOYMENT OF THE SYSTEM	ENABLERS OF AUTOMATION
<p><b>AI diagnostics</b> Diagnosis and screening are the most advanced areas of AI used in healthcare and the use of image-based diagnosis is growing. AI is being used to interpret data in the form of images, and deep learning algorithms are used to model tasks such as medical image analysis through the use of image pattern recognition.<sup>11,12</sup></p> <p>Automated interpretation of images in mammography, retinal imaging, head CT scans, X-Ray imaging and cardiac assessment are the most developed applications.<sup>13</sup></p> <p>ML and deep learning (DL) algorithms are being used to predict the development of conditions and diagnose chronic disease, for example, cancer, cardiovascular disease, and diabetes.<sup>14</sup> These systems are automated rather than autonomous, as the technology functions as a decision support tool in which the information is interpreted by a clinician.</p> <p><b>Case study - CaRi Heart</b> Caristo has developed image analysis technology which uses an algorithm to analyse CT images, in order to detect and quantify inflammation in the heart and its vessels. In doing so it estimates the risk of heart disease. Here, the algorithm serves as a decision support tool, whereby the clinician checks the image that is analysed and then makes treatment decisions.</p>	<p>The black box nature of some AI techniques means that it can be difficult to reach the desired level of transparency and explainability for clinical trust to be established in the technology. Subjective biases may be built into the system. These can be present in the training data from the range of individuals in the dataset, from any conscious or unconscious biases of the data labeller who inputs this data, both from social structures meaning that the selected data can reflect biases in society, and from the way in which the data is collected (ascertainment bias).<sup>15</sup> Individual radiologists may also have biases however the acceptability of human versus machine bias may be different.</p> <p>Consideration must be given to how these systems integrate into wider care pathways, for example in radiology where two trained radiologists review the image could one of these be replaced by a trained ML algorithm? The greatest benefit is only likely to be achieved with wider redesign of the provision of medical imaging services.</p> <p>Public perception of the use of AI in healthcare varies. A 2018 online survey of 2,000 people carried out by the Royal Society for the encouragement of Arts, Manufacturers and Commerce (RSA) and YouGov, found that 74% of people were not familiar with automated decision systems</p>	<p>Access to good quality, comprehensive data; agile regulation; clear ethical principles; and public trust will be key to the development of safe and ethical autonomous systems. As a minimum the dataset needs to be representative of the population it is intended for. Disease prevalence, ethnicity, and socioeconomic factors may vary across the different regions of the UK and beyond. As a result, the dataset may be biased toward the majority population within this setting. It is important that technology developers work in collaboration with clinicians and regulators are aware of the variations in accuracy for different subpopulations.</p> <p>Ideally, the automated diagnostic system should be designed with a certain level of explainability so that the clinician can correctly interpret the information that is being displayed. Without clear traceability of the decision-making process, it can be difficult to assign accountability when patient harm occurs.</p> <p>Better public and professional engagement on the meaning of automated systems is needed. Greater collaboration between the different stakeholders can identify and address risks and build trustworthiness. This can result in greater awareness of the breadth</p>

CURRENT STATE OF THE ART	CHALLENGES TO THE DEVELOPMENT AND DEPLOYMENT OF THE SYSTEM	ENABLERS OF AUTOMATION
<p><b>Precision medicine</b> ML and DL algorithms can also be trained to characterise disease at an individual level and hence deliver personalised treatment. The latter requires the integration large amounts of genetic information, demographic data, blood tests, and vital-sign data from electronic health records.<sup>18,19</sup></p> <p>The definition of precision medicine is broad and this report focuses on accurate disease characterisation which results in personalised and targeted treatment. Using the ML or DL algorithm outputs for decision support, a clinician interprets the information and makes the final treatment decision. Precision medicine can also act much earlier to generate preventative advice based on self-monitoring and/or self-management data.<sup>20</sup></p> <p><b>Case study: AI-enabled precision - e-Stroke</b> Brainomix have developed an algorithm that uses DL to provide interpretation of brain CT scans to help guide treatment. The imaging generates critical information that results in faster treatment decisions in a time-critical situation such as the management of a stroke</p>	<p>being used to aid decisions about healthcare, and 48% opposed the use of these systems.<sup>16</sup> Clinicians may be concerned by being replaced by machines, potentially leading to reduced demand for specialist opinion.<sup>17</sup></p> <p>Quality and representativeness of data can raise questions about the effectiveness of the algorithm; for example, there will be less data from rare cases of a disease, potentially reducing the ability to identify these. Biases have the potential to exacerbate pre-existing health inequalities if these rare cases are more prevalent in under-represented groups than in the majority of the population.</p> <p>It is important for the role of precision medicine to be integrated into the hospital care system. The considerations will be different for screening, prevention and personalised treatment. For example, automated drug delivery could ensure a patient in intensive care receives the necessary levels of pain relief. However, research into safety assurance of such potential devices highlighted the importance of considering human factors to understand the wider implications of replacing typically human tasks (such as selecting the amount of drug to be delivered) with a fully automated machine.<sup>21</sup> When administering drugs, healthcare professionals often maintain an important</p>	<p>of skills required for the development and realisation of this technology.</p> <p>Data that is representative and enriched, to ensure that rare abnormalities can be identified by algorithms, can enable wider use of these systems. Synthetic approaches to data may provide better representation of such rarities; however these datasets will have to be validated against a real dataset.<sup>24</sup></p> <p>Better awareness and education on the interaction between a human clinician and an algorithm-driven machine could build up public trust in automated systems. An improved level of digital literacy for both healthcare workers and patients. would make this easier.</p>



CURRENT STATE OF THE ART	CHALLENGES TO THE DEVELOPMENT AND DEPLOYMENT OF THE SYSTEM	ENABLERS OF AUTOMATION
<p>immediately after it occurs. The consistent interpretation of the brain scans by the algorithm at the same level as a highly trained stroke specialist allows for more patients to receive the optimal treatment, leading to better outcomes. The Brainomix ML algorithm is well integrated into the hospital care system, resulting in improved efficiency.</p>	<p>situational awareness which is vital to the overall care of the individual. Without consideration of the other aspects of the role, there may be a mismatch between the assumed time saved through automation and the reality of continuing to provide good quality care.</p> <p>In this example, communication is happening at different levels: between people, between people and software, between software and AI. It is important the communication is tailored to specific needs. While the RSA survey found that 61% of respondents are concerned that AI does not have the empathy required to make important decisions,<sup>22</sup> when the proposed drug delivery system was tested in practice patients were positive, with them assuming that if the NHS was using it, it must be safe.<sup>23</sup></p>	
<p><b>Robotics</b></p> <p>Robotics in healthcare are often used as a tool to augment human abilities, for instance to assist an interventional surgeon in minimally-invasive procedures.<sup>25</sup> In some areas of surgery robotics enhances the precision and mitigates the effects of human variability.</p> <p>State-of-the-art surgical robots are typically controlled by a surgeon who remotely operates the machine at a console. They are not autonomous; rather both the hardware and software of the</p>	<p>For some treatments, the robotic precision needed to surgically operate on a human body poses a difficult technological challenge and robotic surgery will not be appropriate for all types of surgery.</p> <p>The systems need to integrate effectively into the surgical, rehabilitation, or care home environment, and need to be capable of being used and maintained by different care providers, ideally worldwide.</p>	<p>Clinician confidence in the technology has grown over the last 20 years leading to increased adoption of robotic technology. The first nationwide study to map out the adoption of robotic assisted surgery found that as of 2020, 48 out of 149 acute NHS Trusts (25.9%) were in possession of a surgical robot, and also demonstrated that the prevalence of robotic-assisted procedures continues to rise.<sup>29</sup></p>

CURRENT STATE OF THE ART	CHALLENGES TO THE DEVELOPMENT AND DEPLOYMENT OF THE SYSTEM	ENABLERS OF AUTOMATION
<p>robot perform automated tasks. The output actions for the actual treatment or procedure are carried out by the clinician. The robots do not make any clinical decisions but support the surgeon's decision-making and operating precision. In England, urological procedures make up most (84.2%) of the robotic procedures performed, followed by gynaecological procedures (9.9%), colorectal (4.2%) and general surgery (1.7%).<sup>26</sup></p> <p>Robots are also used as augmenting tools for rehabilitation therapy and as assistive tools to support mobility and decrease dependency.<sup>27</sup></p> <p><b>Case study: surgical robotics - Versius</b></p> <p>Versius is a modular robotic system which is remotely operated by a surgeon from a console. The system enhances treatment for the patient by helping to perform minimal access surgery with enhanced precision and control, which in turn results in improved patient out-comes. Versius integrates well into both the current hospital workflow and operating room.</p>	<p>Public perception and acceptability of robots is low which puts limits on its adoption, because it can be difficult to find patients willing to be operated on with this technology.<sup>28</sup></p>	<p>There is an opportunity for greater engagement between champions of robotics and those who are more hesitant to build a better understanding of the benefits and the safe and ethical use of the technology.</p> <p>Some developers of robotics in the healthcare sector have found the regulatory landscape to be ill-defined and sometimes a barrier to getting a product to market. From a developer's perspective, the safe development of robotics would benefit from more transparency and clearer access to the regulatory pathway information.</p>
<p><b>Mobile apps</b></p> <p>There are 370,000 healthcare apps available, many of which use AI and ML to different degrees to deliver benefits to users such as the management of chronic disease, supporting lifestyle choices, or triaging symptoms</p>	<p>Mobile apps often lack clinical evaluation and poor quality information and gaps in software functionality can pose new risks to patient safety.<sup>31</sup> Platforms with access to apps like Apple and Google do not see it as their</p>	<p>Efforts are underway to build standards, best practice and practical guidance for healthcare apps.<sup>33</sup></p> <p>ORCHA reviews healthcare apps, assessing clinical</p>

CURRENT STATE OF THE ART	CHALLENGES TO THE DEVELOPMENT AND DEPLOYMENT OF THE SYSTEM	ENABLERS OF AUTOMATION
<p>as well as improving access to patient information for clinicians.<sup>30</sup> These apps have no autonomous functionality.</p> <p><b>Case study: ML for natural language processing – Babylon Health</b> The Babylon Health app uses ML for natural language processing to transcribe virtual GP appointments and help write up of doctors' notes. These are then reviewed and edited where necessary by the clinician providing a manual verification step.</p>	<p>responsibility to provide safety assurance are not incentivised to provide such checks.</p> <p>Standard processes of safety assurance break down and a vast majority of healthcare apps fall outside of the remit of effective regulation. There are often no mechanisms in place to monitor use, change or future risks. A lack of regulation of these technologies can result in a lack of awareness of ethical data practices, meaning that apps can impact security and privacy.<sup>32</sup></p> <p>For those apps that are regulated, the current regulatory framework can inhibit rapid innovation, as each software update – for example, to incorporate new features – has to be externally reviewed which is often time-consuming and therefore limits the speed of innovation.</p>	<p>assurance, data privacy, and usability. They have also helped develop a public facing NHS library with healthcare apps approved by NHSX (now the NHS Transformation Directorate). This provides an important level of quality assurance to users.</p>

### Cross-cutting challenges

While the current technologies for each of these applications are not yet operating autonomously in clinical settings, if the technical step change from decision support tools to decision-makers is made, that step presents complications that must be addressed in terms of **safety assurance; regulation; moral responsibility and legal accountability;**

**data and interoperability; and skills, culture, and perceptions.** These challenges are not unique to autonomous systems in healthcare settings, however they may be more acute in the healthcare sector as it is a safety critical domain with complex crossovers between machines, clinicians, healthcare professionals, and patients.

## Safety assurance

Increasing the level of responsibility of technological systems by increasing autonomy and reducing human oversight makes it increasingly challenging to assure safety. This is a critical challenge for healthcare where patient safety is vitally important.

Safety is demonstrated through a safety assurance case written by the developers of the technology and reviewed by the relevant regulator. This provides a consideration of the automated system's function and operational environment to justify confidence or certainty in a system's capabilities, supported by a body of evidence.<sup>34</sup> However, it is not universally agreed that this is enough to demonstrate safety when evidence cannot always be provided for how the system will function in every possible outcome. Real-world testing and runtime verification is important to ensure a system remains within its predicted boundaries. As the risk of potential patient harm increases, a level of transparency is required to demonstrate safe operation and to understand system biases and the potential harm that may be a result of these. Transparency is needed to show how a system works, to explain how the basis of a particular decision has been made and to understand failures. This raises questions about the level of transparency needed and for whom, for example for clinicians to understand the decisions, patients who take on the final risk, maintainers of autonomous systems, or those who investigate failures.

This will become increasingly complicated if we move toward systems that operate autonomously or self-learn, continually improving and adapting their algorithms based on new data. This limits the availability of the evidence that is needed for safety assurance, posing new and unique challenges relating to safety and regulation. This is because the benefits of such systems can only be realised in real-world settings, meaning evidence cannot always be provided in advance for how the system will function in every context as explained above. However, while evidence may always be limited, it is important to embed robust safety assurance practices appropriate to these technologies.

### Case study: ML algorithm – Skin Analytics

Skin Analytics have created a ML algorithm to detect skin cancer. The model is trained using labelled data. Once the system has shown it can reach the desired standard, the data is fixed and locked for a period of time. This process ensures the algorithm remains deterministic until there is a new version release to incorporate the new data. The system is verified and validated after each change. The algorithm is a decision support tool which has a safety net in place with a human clinician double checking the information output.



### How safe is safe enough?

There needs to be dialogue and engagement to consider the question “how safe is safe enough?”, in particular with patient groups and the wider public, as fundamentally patients take on the largest risk. Understanding the level of automation in a system and what its intended function is and where the human and machine interact are key to this process.



## Regulation

The essential role of regulation in this sector is to ensure that medical devices work and reach an acceptable level of safety. Developers can find the current regulatory landscape confusing to navigate and time consuming. It often results in regulation being perceived as a barrier to innovation.<sup>35</sup> There is no single regulatory body with a remit for the whole sector uniting the different regulatory bodies. Developers which are more experienced with medical device regulations tend to find the process more straightforward than new developers or those entering from another sector.

Developer demand for a central guide to help navigate UK regulatory requirements suggests the AI multi-agency advisory service (AI MAAS) being established through the NHSX (now the NHS Transformation Directorate) and AI Lab regulatory programme is welcome. This service was not live at the point of publication, and will require clear signposting, guidance, and advice available during the interim period, to support innovators and enable safe innovation. The AI Lab regulatory programme is also running other collaborative programmes to improve aspects of the regulatory pathway including streamlining the process for technological review, synthetic data, and post-market surveillance.

### Regulating autonomy

Navigating the regulatory landscape is not the only issue; a major challenge is the rapid pace of technology change and its contrast with the long-

term nature of regulations and the much slower pace of regulatory change. External auditing is vital for safety assurance, so cannot be compromised in favour of the speed of innovation. Assurance, and a clear sense of the actions taken in cases of failure, are important for ensuring that these technologies are socially acceptable.

There is work to be done to develop workable regulations that drive good patient outcomes and that are safe and ethical, through a step change from automated to autonomous systems in healthcare. These can be underpinned by good practice standards.

- When issues with the regulations are uncovered, it will be important to amend these rapidly to avoid patient harm.
- There is a need to **maintain international regulatory alignment** where adequate to enable greater market access.
- The requirements for product developers to consider possible bias in the algorithm training process are currently limited. There needs to be **clear guidance on ethical data collection** requirements that are aligned with good information governance practices. If an AI algorithm needs to be retrained on a different dataset than the one on which it was initially trained because of changes in the population or system design, this should require regulators to reassess the system.



**AI multi-agency advisory service (AI MAAS)**

The Health Research Authority (HRA), Care Quality Commission (CQC), Medicines and Healthcare products Regulatory Agency (MHRA) and The National Institute for Health and Care Excellence (NICE) are developing a multi-agency advice service to offer support, information, and advice to technologists. Its aim is to make the pathway clearer, and show which guidance applies when supporting the safe, ethical, and effective development of medical technology, whilst ensuring patient data is protected.

The service is still in its development phase, working on a timeline of approximately two years to reach sustainable implementation of the model, and bodies are carrying out user research to:

- address barriers along the pathway to identify duplications and gaps in the guidance
- address where regulation is ineffective
- establish who the key users of the service are likely to be
- scope the development of a digital solution.

The roles are spread across the different organisations:

- **NICE:** Role of the secretariat, and to provide and produce guidance on cost-effectiveness
- **MHRA:** Remit covers medical devices, work with developers
- **CQC:** Remit covers considerations of safety and care, working with technology adopters
- **HRA:** regulating the use of data collected.

robust safety assurance processes will take some time. These processes are crucial to ensure that a product is safe before it reaches market. Similarly, adaptive algorithms and autonomous systems that self-learn from real-world changes and experiences to improve performance will require specific consideration in future regulations. Real-time adaptation may mean the system performs differently to its premarket assessment.<sup>36</sup> Post-market surveillance will be key to continually assess risk and benefit. The US Food and Drug Administration (FDA) has been progressive in applying this to increasingly autonomous systems.

The **FDA** has approved several AI or ML based medical devices that operate on a locked algorithm. A locked algorithm can be defined as one that provides the same output with each input and does not change with use. In this case, any algorithmic changes that have occurred have required FDA premarket review. A premarket approach is taken if there are any major changes to the algorithm that could significantly affect device performance, or safety and effectiveness.

Medical device regulation has not been designed for adaptive or self-learning algorithms and so requires a different regulatory approach that is total product lifecycle-based - from premarket to post-market development - which allows products to continuously learn while providing effective safeguards and helping to deliver effective patient care. This comes with expectations around transparency as well as the need to collect and monitor real world performance data.<sup>37,38</sup>

The **American Medical Informatics Association (AMIA)** recommends that there should be periodic evaluation of the system, identification of algorithmic shift or drift due to a shift in data, constant review to determine whether bias occurs and continued user education and training. All of which are key to transparency and performance monitoring in a real-world setting.<sup>39</sup>

While the regulatory approval process can be complicated, innovators consider the regulations themselves crucial to the development and release of products.

Given that techniques for reliable verification and validation of autonomous systems are still being developed by standards bodies, establishing

Although self-learning systems are a distant development in healthcare, it is worth exploring the questions and challenges surrounding them to futureproof regulations so they can be updated rather than rewritten. It will be crucial to monitor changes throughout a product's lifespan and this may involve software that is built into the system that makes assessments as it self-learns to ensure that any changes remain safe. There are other impacts to consider such as possible cyberattacks and the resources required for

risk mitigation of this depend on the potential impact on the individual. The regulation of medical devices has not yet fully considered the full impacts of deficient cybersecurity on patients and the delivery of care, and it will be necessary to revisit this and assess whether there is sufficient consideration in the current regulatory framework.<sup>40</sup> Although this is an important consideration, it is not solely applicable to autonomous and self-learning systems, but the adoption of technology more widely.





# Moral responsibility

Automated systems in healthcare maintain a level of human control with a clinician in the loop. Rather than decision-makers, they should be thought of as decision-support tools and assistive tools or as part of the multidisciplinary team which aid the clinician or healthcare professional who is responsible for the final decision. Individuals often operate these systems in complex environments and external factors are also important to consider. For example, the inherent pressures in a clinical environment – such as if a clinician is particularly busy, or the patient’s family are especially demanding – may impact the way a clinician interacts with a machine. While these systems are in place to enhance patient care and increase efficiency, there is a concern of becoming over-reliant on these tools. Like humans, these systems are not infallible, so system errors are inevitable. The combination of these factors can have potentially harmful implications for a patient, and in this case it is important to assign both moral responsibility and legal accountability.

Moral responsibility is distinguishable from legal liability. Moral responsibility refers to how far an individual can be held morally responsible for decisions made by a system for which they had some control, or how far an individual may receive blame or praise for an action.<sup>41</sup> Legal liability refers to how far an individual can be held accountable as defined by law and where legal implications may lie. Autonomous systems in healthcare will challenge traditional practices for both, and legislation in this area has not yet been set out.

Questions that surround moral responsibility require an interdisciplinary approach due to the various social, behavioural, cultural, and organisational issues that can come into play. The complex overlaps between a clinician and machine make assigning moral responsibility for a decision that has been made difficult if patient harm occurs. To help address this, the UKRI-funded project *Assuring Autonomy* aims to develop an interdisciplinary methodology to trace and allocate responsibility of the decisions and outcomes of autonomous systems.<sup>42</sup>

**The level to which the system is explainable, interpretable and transparent will be key to assigning responsibility.**<sup>43</sup> If a clinician is unable to understand and verify a decision made by a machine, it becomes difficult to assign moral accountability in the case of patient harm. Autonomous systems are often making decisions based on statistical probabilities which can have considerable uncertainty, and it is important to have a level of explainability that indicates the assumptions that have been made when the information is presented to the user.

It is also important to make clear the level of control a clinician has over a system decision and where the handover of responsibility is between the human and the machine. There is also the complication that, should a clinician not act on the recommendation of an autonomous system, they may be seen as culpable if this results in a poor outcome for the patient. Concern about litigation

in such a situation may lead clinicians to err towards trusting the system – whereas it may well be better for the clinician to take decisions on their wider knowledge of the patient and the context of their care. **There is a need for a high level of**

**collaboration between all stakeholders – tech developers, system safety engineers, regulators, clinicians, and patients – in order to balance responsibility fairly.**<sup>44</sup>

It is also important to make clear the level of control a clinician has over a system decision and where the handover of responsibility is between the human and the machine



# Data and interoperability

Access to patient data is crucial for the development of autonomous systems in the healthcare sector. This is particularly challenging in this domain because of the sensitivity, possible low quality and inequality of the data, and possible poor access due to information governance and clinical ethics restrictions. Data needs to be findable, accessible, interoperable and reusable (FAIR) in order to support the development of these technologies.<sup>45</sup>

The intersection between interoperability and ethics poses various challenges. The UK data that is needed is held across multiple NHS Trusts and NHS Digital, and beyond. Across these, there is a lot of variation; for example, the prevalence of electronic patient records varies across Trusts. Each NHS organisation has its own Caldicott Guardian responsible for the confidentiality of people's healthcare information. Access to training data can also be difficult due to ethical concerns of patient privacy and data value. Better access to this data would be beneficial to tech developers but the sensitivity of the data requires a cautious approach. For approved researchers from trusted organisations, the Trusted Research Environment service (TREs) for England is an important new service, enabling secure access to datasets and analytics.

**The Grampian Data Safe Haven (DaSH)** facility between NHS Grampian and the University of Aberdeen is an example of secure processing and linking of health data where it is not possible to obtain consent from patients. The facility allows researchers to access but not download data held on different servers, and there are strict controls placed on who can access it, where it is stored, the type of analysis applied and the results that are extracted from the data.<sup>46</sup>

When obtained, the data used to develop a tool can create further challenges. There is a risk that the dataset may only benefit the majority population, so ensuring the validation show no reduction in accuracy for different population groups is vital. This can be achieved by testing on suitable external independent datasets. **Data needs to be representative of the population it is intended for, at a minimum.**

While bias in medical devices has been a challenge for the engineering community, adoption of autonomous systems has the potential to extend the scale of the problem. To counteract this, explainability is needed to understand what key factors drive the decision-making of an algorithm to help identify where and what type of bias may arise. Without this, the technologies may reinforce inequalities and discrimination across the healthcare sector.<sup>47</sup> Developers' unconscious biases can increase unintended bias in the system

It found that 81% of respondents were comfortable providing personal data about themselves to the NHS for the development of healthcare treatments, however there was also uncertainty about data practices and 52% of respondents know little to nothing about how their data is used

and there needs to be greater diversity among the individuals labelling the training data, and consistency of labelling. Where representative data cannot be obtained, the product should not be approved.

There are fundamental issues with the quality of health data as it is primarily collected for the purpose of patient care and not for secondary purposes such as for research or for use as training data. "Cleaning" real world data is important for successful AI model building. When data is anonymised to maintain patient privacy, it inherently lowers the quality of the insights that can be gained from an algorithm. This may or may not matter depending on the questions being asked of the data. Data inequality also poses a challenge as different populations or demographic groups can often be underrepresented in datasets, resulting in a lack of representation of different ethnicities or gender. There are trade-offs between privacy, safety, and how useful the system is. To realise further potential benefits and minimise harm there is a need to access more representative information. The Centre for Data Ethics and Innovation's tracker survey monitors how public attitudes towards the use of data and data-driven technologies change over time. It found that 81% of respondents were comfortable providing personal data about themselves to the NHS for the development of healthcare treatments, however there was also uncertainty about data practices and 52% of respondents know little to nothing about how their data is used.<sup>48</sup> **Therefore,**

**public concerns should be addressed and used to establish trustworthy mechanisms for data sharing.**

It is important to acknowledge the challenge of legacy IT and lack of technical infrastructure across the NHS that needs to be addressed. A large-scale transformation to develop robust data collection systems, encourage data sharing and linkage between trusted providers and provide good and secure access to data, while maintaining the privacy of individuals and communities, is required to be able to realise wider societal benefits.

## **The pitfalls of self-learning in autonomous healthcare systems:**

During the first COVID-19 wave, ML algorithms trained on patient data acquired between February and June 2020 gave accurate predictions of ICU admission for A&E patients testing positive for the SARS-CoV-2 virus. These algorithms perform poorly when applied to the data of today's A&E patients, because the characteristics of A&E patients infected with the SARS-CoV2 virus in 2023 are very different from those of infected patients in the first half of 2020. An autonomous ML algorithm, through self-learning, would have adapted over time to the changes in patient characteristics; however, it would now be generating outputs which were not predictable from its behaviour in 2020 when it was originally trained.

# Skills and culture change

There is a shortage of clinicians across the healthcare sector who are confident working with software-based systems and can effectively work with systems that offer decision-support. Healthcare staff will need these skills to see widespread technology adoption. The Topol Review identified the need to develop a greater awareness of the required capability, provide access to training, and deliver the skills needed for patients and citizens.<sup>49</sup> **It would also be useful to identify champions within hospitals to help provide practical guidance and best practice of using software-based systems.** Dr Tom Lawton, Clinical Head of AI at Bradford Teaching Hospitals is an example of developing leadership in this space and there should be further work to identify leaders through similar roles in hospitals. In doing so he forging a link between innovators and clinicians to harness AI to deliver improved healthcare.<sup>50</sup> As the skills ecosystem develops,

it will allow for easier identification of where skills gaps lie.<sup>51</sup> Clinicians may benefit from the opportunity to work with other academic disciplines, engineers, and developers, to be trained to understand and use the software, AI and robotic technologies. However, the value of this will need to be promoted at all staff levels and given dedicated time due to competing priorities.<sup>52</sup> Although adaptive and self-learning algorithms are a distant development in healthcare, clinicians will also need to be trained to monitor any algorithmic changes. Similarly, clinicians and patients should continue to work with developers to help build an understanding of what makes a useful tool. As the goal of any medical device should be to improve patient care, patient groups, such as those involved in patient advocacy and engagement, also need to be involved in the design process of autonomous systems with their needs reflected throughout and in the final product.

Clinicians and patients should continue to work with developers to help build an understanding of what makes a useful tool

# Perceptions

As well as the culture change needed across the healthcare sector to increase the use of autonomous systems, **the perception of these systems needs to positively build up to increase trust.** The UK's health system has built up a certain level of trust among the public and in society, therefore expectations can be higher in

comparison to other domains. A single case that goes wrong can create backlash to systems that otherwise works successfully. Perceptions may differ depending on the role an individual has in the system. The traditional regulatory safety case could be adapted for different audiences to create an open and accessible dialogue.

PERCEPTIONS	
PATIENTS	<p>Patients will be exposed to the largest risk and the greatest benefit from the deployment of autonomous systems in healthcare. They look to clinicians to deliver treatment and care safely, and so clinicians take on and accept a level of moral responsibility. This becomes more complex when automated systems support the clinician in their decision-making or performs some of the tasks, which may erode trust between clinician and patient. Patients will also require a level of explainability to feel comfortable and safe with the use of automated systems along their care pathway. There is also a need to better frame automated systems in this sector, particularly with regard to apps and personal services as an opportunity for patients to have more agency in their own health and the decision-making.</p> <p>Despite these systems having the potential to increase efficiency, patients may also be concerned that there may be fewer clinicians available with an increased use of automated systems in hospitals.<sup>53</sup> However, such systems may address a severe skills shortage and may help to address waiting times.</p>
GENERAL PUBLIC	<p>The public tend to need a substantial guarantee beyond doubt, for example, clinical trials in large numbers. They will want to know that automated systems are trustworthy, safe, and ethical. There can also be a baseline belief that machine accuracy and reliability is lower than a human's or a low trustworthiness of AI advice.<sup>54</sup> It is therefore important to understand and address misconceptions, manage expectations, ensure the automated systems that are developed are trustworthy and well regulated.</p>



PERCEPTIONS	
<b>CLINICIANS AND HEALTHCARE PROVIDERS</b>	<p>Some clinicians are hesitant to adopt automated systems because of the fear of being replaced.<sup>55</sup> There should be engagement from both users and non-users of the technology to build knowledge and better perception of the benefits. Clinicians also require a level of explainability to understand and have confidence in a system's decision-making and to accept a degree of moral responsibility.</p> <p>Clinicians may also want to a guarantee that the technology is well integrated into the hospital care pathways to ensure that the administration of care is uninterrupted, and that it still allows for an appropriate level of human – patient interaction.<sup>56</sup></p>
<b>HOSPITAL MANAGERS</b>	<p>Hospital managers and administrative staff may want to know that an automated system fits well within the clinical workflow. They will want to ensure that the care system remains uninterrupted with a guarantee that adding the system results in, for instance, shorter hospital stays and lowers costs.</p>
<b>DEVELOPERS</b>	<p>Developers need to be clear that there is a medical need for the products that they are working on, so they are likely to be adopted and produce positive outcomes for the patients. Developers may benefit from a deeper understanding of the expectations and perceptions of all stakeholders in the healthcare system.</p>



## Conclusions

There are complex overlaps between automated systems, clinicians, healthcare professionals, and patients and this report has highlighted some of the resulting challenges, the enablers that can help solve these, as well as the unanswered questions that need to be addressed. It shows that careful introduction and close observation of these systems is needed to judge their effectiveness and to assess whether the evidence of their benefit justifies the use of autonomous systems in healthcare settings.

Such an increase in autonomy would mean that more responsibility is given to a system to make decisions in complex environments. The risks that this poses need to be carefully managed and there must be clear evidence of benefit. To enable greater autonomy in cases where it would significantly improve outcomes, there are particular challenges that would need to be overcome. There must be an acceptable level of transparency to demonstrate safe system operation, as well as

multidisciplinary conversations which include patient engagement groups to decide how safe is safe enough.

Workable regulations need to be developed that drive good patient outcomes and that are safe and ethical. There need to be greater requirements for product developers to consider possible bias in algorithm training and there needs to be clear guidance on ethical data collection that is aligned with good information governance practices. Regulators should reassess the system if an algorithm needs to be retrained on a different dataset than the one on which it was initially trained. It will also be important to maintain international regulatory alignment to enable greater market access.

The complex overlaps between a clinician and machine make assigning moral responsibility for a decision that has been made difficult if patient harm occurs. Therefore, automated systems need

Such an increase in autonomy would mean that more responsibility is given to a system to make decisions in complex environments. The risks that this poses need to be carefully managed and understood to ensure a positive safety benefit

to be explainable, interpretable and transparent as well as a high level of collaboration between all stakeholders to help assign moral responsibility fairly.

Access to data is crucial to the development of autonomous systems in the healthcare sector, however this is challenging because of the sensitivity, possible low quality and inequality of the data. Training data must be representative of population it is intended for at a minimum, and where this is not possible the product should not be approved.

There is a need to improve the skills and digital literacy of healthcare workers and of patients. The sector would benefit from closer collaboration throughout the development and deployment process to ensure that all stakeholders understand the system being developed as well as what would make it useful. It would also be beneficial to identify champions within hospitals to help provide practical guidance and best practice of using software-based systems. It is also important to provide assurance that a system is safe and to prevent erosion of public trust.

Addressing these challenges will enable the use of systems that have the potential to ease the burden on healthcare providers, by providing effective, efficient and targeted care. This is a step towards improved healthcare for all.

It would also be beneficial to identify champions within hospitals to help provide practical guidance and best practice of using software-based systems

# Acknowledgements

The National Engineering Policy Centre would like to thank the following for their contributions:

## Members of the Safety and Ethics of Autonomous Systems Working Group

Professor Nick Jennings CB FREng FIET FBCS (Chair)  
Professor Michael Fisher FBCS FIET  
Professor Muffy Calder DBE FREng FRSE  
Dr Chris Elliot MBE FREng FRAeS  
Professor John McDermid OBE FREng  
Paul Newman FREng  
Andrew Chadwick  
Lambert Dopping-Hepenstal FREng FIET FRAeS  
Professor Marina Jirokta FBCS  
Gordon Meadow FIMarEST  
Dr Robert Merrall FIAgRE  
Dr Sylvain Jamais FIMechE

## Royal Academy of Engineering Staff

Arizona Rodriguez, Policy Advisor  
Dr Alexandra Smyth, Senior Policy Advisor  
Dr Natasha McCarthy, Associate Director, Policy  
Brittany Hsieh, Senior Policy Advisor

# References

- 1 *Safety and ethics of autonomous systems project overview*, Royal Academy of Engineering (2020). <https://raeng.org.uk/media/nqnhktgq/nepc-safety-and-ethics-of-autonomous-systems.pdf>
- 2 <https://nhsx.github.io/ai-dictionary?term=ai>
- 3 *Machine learning: the power and promise of computers that learn by example*, Royal Society (2017). <https://royalsociety.org/-/media/policy/projects/machine-learning/publications/machine-learning-report.pdf>
- 4 *How machine learning is embedded to support clinician decision making: an analysis of FDA-approved medical devices*, BMJ Health Care Inform (2021), <https://pubmed.ncbi.nlm.nih.gov/33853863/>
- 5 *The Moral Responsibility Gap and the Increasing Autonomy of Systems*, Zoe Porter et al. (2018). [https://eprints.whiterose.ac.uk/133488/1/Resp\\_Gaps\\_paper\\_21.pdf](https://eprints.whiterose.ac.uk/133488/1/Resp_Gaps_paper_21.pdf)
- 6 *Trusted autonomous systems in healthcare*, Rachel Hesketh, Kings College London (2021). [www.kcl.ac.uk/policy-institute/assets/trusted-autonomous-systems-in-healthcare.pdf](http://www.kcl.ac.uk/policy-institute/assets/trusted-autonomous-systems-in-healthcare.pdf)
- 7 Ibid.
- 8 *NHS Key Statistics: England 2022*, Carl Baker and Esme Kirk-Wade, House of Commons Library (2022). <https://researchbriefings.files.parliament.uk/documents/CBP-7281/CBP-7281.pdf>
- 9 *Housing an ageing population: a reading list*, Anastasia Lewis, House of Commons Library (2021). <https://researchbriefings.files.parliament.uk/documents/CBP-9239/CBP-9239.pdf>
- 10 *Robotics in social care: a connected care ecosystem for independent living*, UK-RAS Network Robotics and Autonomous Systems (2017). [www.housinglin.org.uk/\\_assets/Resources/Housing/OtherOrganisation/UK\\_RAS\\_robotics-in-care-report.pdf](http://www.housinglin.org.uk/_assets/Resources/Housing/OtherOrganisation/UK_RAS_robotics-in-care-report.pdf)
- 11 *The rise of artificial intelligence in healthcare applications*, Adam Bohr and Kaveh Memarzadeh (2020). [www.ncbi.nlm.nih.gov/pmc/articles/PMC7325854/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC7325854/)
- 12 *Artificial intelligence-enabled healthcare delivery*, Reddy et al. (2019). <https://journals.sagepub.com/doi/pdf/10.1177/0141076818815510>
- 13 *Artificial Intelligence: How to get it right*, NHSX (2019). [https://reform.uk/sites/default/files/2019-10/NHSX\\_AI\\_report.pdf](https://reform.uk/sites/default/files/2019-10/NHSX_AI_report.pdf)
- 14 *Artificial intelligence-enabled healthcare delivery*, Reddy et al. (2019).
- 15 *Machine learning: the power and promise of computers that learn by example*, (2017).
- 16 *Artificial Intelligence: Real Public Engagement*, Royal Society for the encouragement of Arts, Manufactures and Commerce (2018). [https://www.thersa.org/globalassets/pdfs/reports/rsa\\_artificial-intelligence---real-public-engagement.pdf](https://www.thersa.org/globalassets/pdfs/reports/rsa_artificial-intelligence---real-public-engagement.pdf)
- 17 *Exploring stakeholder attitudes towards AI in clinical practice*, Scott IA et al., BMJ Health & Care Informatics (2021). <https://informatics.bmj.com/content/bmjhci/28/1/e100450.full.pdf>
- 18 *Artificial intelligence-enabled healthcare delivery*, Reddy et al. (2019).
- 19 *The rise of artificial intelligence in healthcare applications*, Adam Bohr and Kaveh Memarzadeh (2020). [www.ncbi.nlm.nih.gov/pmc/articles/PMC7325854/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC7325854/)
- 20 *Artificial Intelligence: How to get it right*, NHSX (2019).
- 21 *Human factors challenges for the safe use of artificial intelligence in patient care*, Sujan et al. (2019). <https://informatics.bmj.com/content/bmjhci/26/1/e100081.full.pdf>
- 22 *Artificial Intelligence: Real Public Engagement*, Royal Society for the encouragement of Arts, Manufactures and Commerce (2018). [www.thersa.org/globalassets/pdfs/reports/rsa\\_artificial-intelligence---real-public-engagement.pdf](http://www.thersa.org/globalassets/pdfs/reports/rsa_artificial-intelligence---real-public-engagement.pdf)
- 23 [www.humanreliability.com/casestudies/sam\\_project/](http://www.humanreliability.com/casestudies/sam_project/)
- 24 As stated in a recent report published by the Royal Society, synthetic data is “data that has been generated using a purpose-built mathematical model or algorithm with the aim of solving a (set of) data science task(s).” Synthetic data – what, why and how? Royal Society (2022). <https://doi.org/10.48550/arXiv.2205.03257>
- 25 *Artificial intelligence-enabled healthcare delivery*, Reddy et al. (2019).
- 26 *Uptake and accessibility of surgical robotics in England*, Kyle Lam et al. (2020). <https://doi.org/10.1002/rcs.2174>
- 27 *The rise of artificial intelligence in healthcare applications*, Bohr and Memarzadeh (2020).
- 28 *Patient Willingness to Undergo Robotic Surgery: Identification and Validation of a Predictive Model*, Emily C. Anania, Dissertations and Theses, Embry-Riddle Aeronautical University (2019). <https://commons.erau.edu/cgi/viewcontent.cgi?article=1434&context=edt>
- 29 *Uptake and accessibility of surgical robotics in England*, Kyle Lam et al. (2020).
- 30 *Why is it so difficult to govern mobile apps in healthcare?* Magrabi et al. (2019). <https://doi.org/10.1136/bmjhci-2019-100006>
- 32 Ibid.
- 33 Ibid.
- 34 *Safety and ethics of autonomous systems project overview*, (2020).
- 35 *Artificial Intelligence: How to get it right*, NHSX (2019).
- 36 *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)*, U.S. Food and Drug Administration (2019). [www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf](http://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf)
- 37 Ibid.
- 38 *Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan*, U.S. Food and Drug Administration (2021). [www.fda.gov/media/145022/download](http://www.fda.gov/media/145022/download)
- 39 *Recommendations for the safe, effective use of adaptive CDS in the US healthcare system: an AMIA position paper*, Petersen et al. (2020). <https://doi.org/10.1093/jamia/ocaa319>
- 40 *Cyber Safety and Resilience: strengthening the digital systems that support the modern economy*, Royal Academy of Engineering (2018). [www.raeng.org.uk/publications/reports/cyber-safety-and-resilience](http://www.raeng.org.uk/publications/reports/cyber-safety-and-resilience)
- 41 *The Moral Responsibility Gap and the Increasing Autonomy of Systems*, Zoe Porter et al. (2018). [https://eprints.whiterose.ac.uk/133488/1/Resp\\_Gaps\\_paper\\_21.pdf](https://eprints.whiterose.ac.uk/133488/1/Resp_Gaps_paper_21.pdf)
- 42 [www.cs.york.ac.uk/research/trusted-autonomous-systems/](http://www.cs.york.ac.uk/research/trusted-autonomous-systems/)
- 43 *Explainable AI: the basics*, Royal Society (2019). <https://royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf>
- 44 *Artificial intelligence in health care: accountability and safety*, Habli et al. (2020). <https://www.who.int/bulletin/volumes/98/4/19-237487.pdf>
- 45 *Artificial Intelligence: How to get it right*, NHSX (2019).
- 46 *Towards trusted data sharing: guidance and case studies*, Royal Academy of Engineering (2019). <http://reports.raeng.org.uk/datasharing/cover/>
- 47 *The Topol Review: Preparing the healthcare workforce to deliver the digital future*, NHS (2019). <https://topol.hee.nhs.uk/>
- 48 *Public attitudes towards AI: Tracker survey*, CDEI (2021). [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1064525/Public\\_attitudes\\_to\\_data\\_and\\_AI\\_-\\_Tracker\\_survey.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1064525/Public_attitudes_to_data_and_AI_-_Tracker_survey.pdf)
- 49 *The Topol Review: Preparing the healthcare workforce to deliver the digital future*, (2019).
- 50 [www.bradfordhospitals.nhs.uk/tag/dr-tom-lawton/](http://www.bradfordhospitals.nhs.uk/tag/dr-tom-lawton/)
- 51 *One year On: Progress on the recommendations from the Topol Review*, NHS (2020). [www.hee.nhs.uk/sites/default/files/Topol%20One%20Year%20On.pdf](http://www.hee.nhs.uk/sites/default/files/Topol%20One%20Year%20On.pdf)
- 52 *Transforming health through innovation: Integrating the NHS and academia*, The Academy of Medical Sciences (2020). <https://acmedsci.ac.uk/file-download/23932583>
- 53 *Human factors challenges for the safe use of artificial intelligence in patient care*, Sujan et al. (2019).
- 54 *Exploring stakeholder attitudes towards AI in clinical practice*, Scott IA et al., BMJ Health & Care Informatics (2021). <https://informatics.bmj.com/content/bmjhci/28/1/e100450.full.pdf>
- 55 Ibid.
- 56 *Human factors challenges for the safe use of artificial intelligence in patient care*, Sujan et al. (2019).



## THE ROYAL ACADEMY OF ENGINEERING

The Royal Academy of Engineering is harnessing the power of engineering to build a sustainable society and an inclusive economy that works for everyone.

In collaboration with our Fellows and partners, we're growing talent and developing skills for the future, driving innovation and building global partnerships, and influencing policy and engaging the public. Together we're working to tackle the greatest challenges of our age.

## NATIONAL ENGINEERING POLICY CENTRE

We are a unified voice for 43 professional engineering organisations, representing 450,000 engineers, a partnership led by the Royal Academy of Engineering.

We give policymakers a single route to advice from across the engineering profession.

We inform and respond to policy issues of national importance, for the benefit of society.