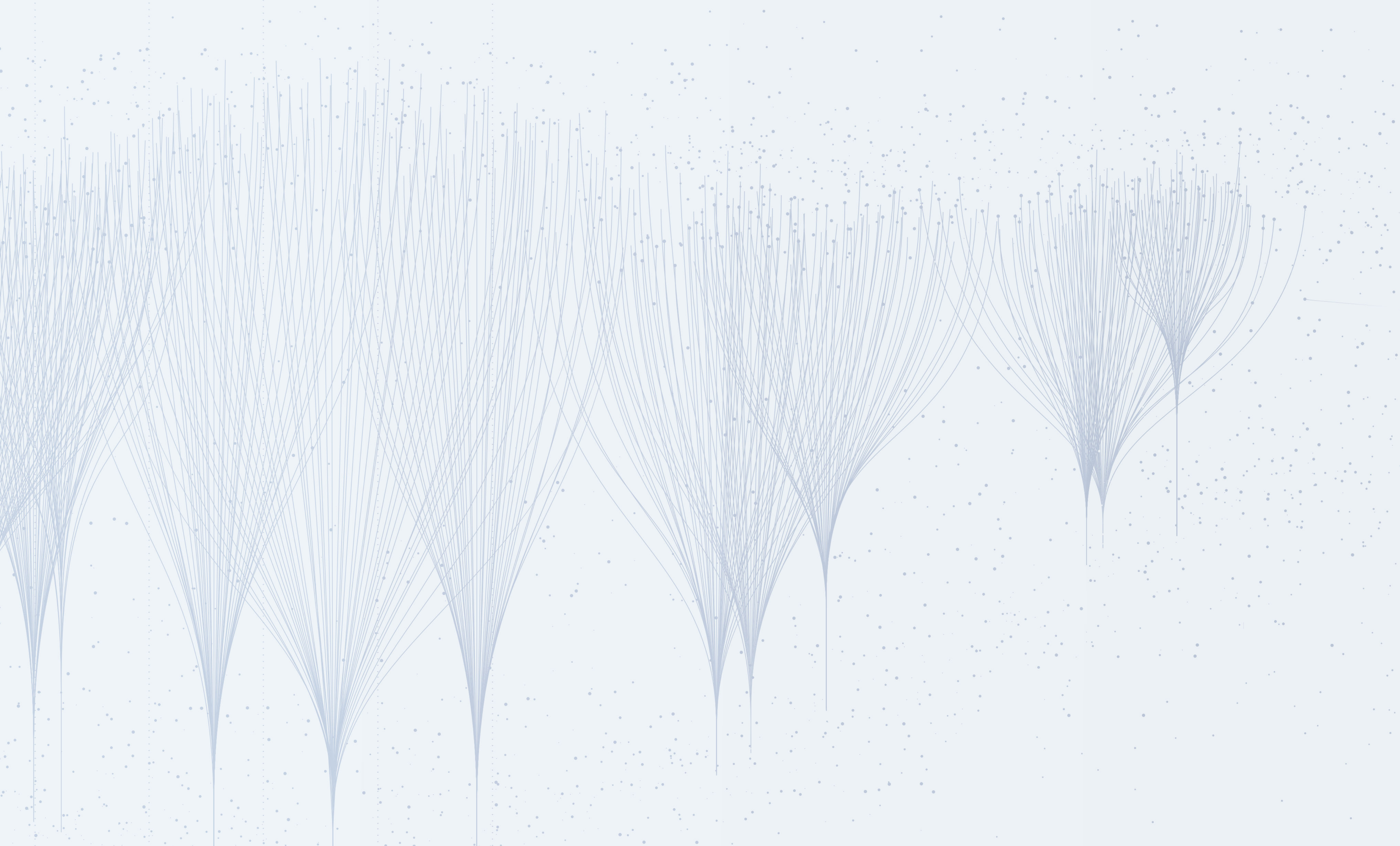
A Buyer's Guide to AI in Health and Care

10 questions for making well-informed procurement decisions about products that use AI

NOVEMBER 2020

ACCELERATING THE SAFE AND EFFECTIVE ADOPTION OF AI IN HEALTH AND CARE



CONTENTS

AUTHORS

Dr. Indra Joshi and Dominic Cushnan

Foreword

Artificial intelligence (AI) is already impacting positively on our health and care system - for example, supporting diagnostic decisions, predicting care needs, informing resource planning, and generating game-changing research. We need to harness its promise to reap tangible benefits for patients, service users and staff. That's why the Government announced a £250m investment last year to set up an Artificial Intelligence Laboratory (AI Lab).

We're clear on what we want the Lab to achieve: accelerating the safe and effective adoption of AI across health and care. Whilst that mission is straightforward to write down, it's a big undertaking to deliver on. We need to support innovation in targeted ways. We're committed to developing rigorous standards of ethics and safety - and creating technical infrastructure to help embed these standards. We will be developing methods for auditing and evaluation, in collaboration with other organisations doing great work. And we're clear about the importance of equipping people and organisations to embrace new ways of working - within teams, within organisations, across regions, and with commercial partners.

Amidst the heroic response of our health and care services to these very difficult times, we have seen how COVID-19 has been the catalyst for an extraordinary exploration and uptake of digital solutions, at pace and at scale. These are solutions aimed at improving people's lives and assisting the work of our front-line colleagues. And these solutions will be central to our recovery - short-term and long-term - as we rebuild in a world that looks very different to before. For example, AI technologies may be able to help optimise scheduling of appointments, forecast demand for healthcare from residents in care settings, speed up cancer screening tests, and advance our understanding of COVID-19 itself.

The possibilities presented by these technologies are hugely exciting. At the same time, we need to be assured that any products our organisations buy meet the highest standards of safety and effectiveness. Getting this assurance is no easy task in a nascent and fast-moving market.

This is where the Buyer's Guide to AI in Health and Care comes in. It is aimed at anyone likely to be involved in weighing up the pros and cons of buying a product that uses AI. It is recommended reading for clinical and practitioner leads, chief officers, senior managers, transformation experts and procurement leads. It offers practical guidance on the questions to be asking before and during any AI procurement exercise in health and care. What is the problem you are trying to solve? Why AI? Can you validate the product's performance claims? How well will the product work in your organisation? What do you need to have in place to give your project the best chance of success?

If we want to realise the transformational potential of AI for health and care, these questions matter. We need people who are equipped to ask these questions confidently, who can probe their suppliers appropriately, and who know when to seek expert input - from the AI Lab or elsewhere. We hope that this guide helps to support this ambitious, optimistic and discerning cadre of buyers that our health and care services need.

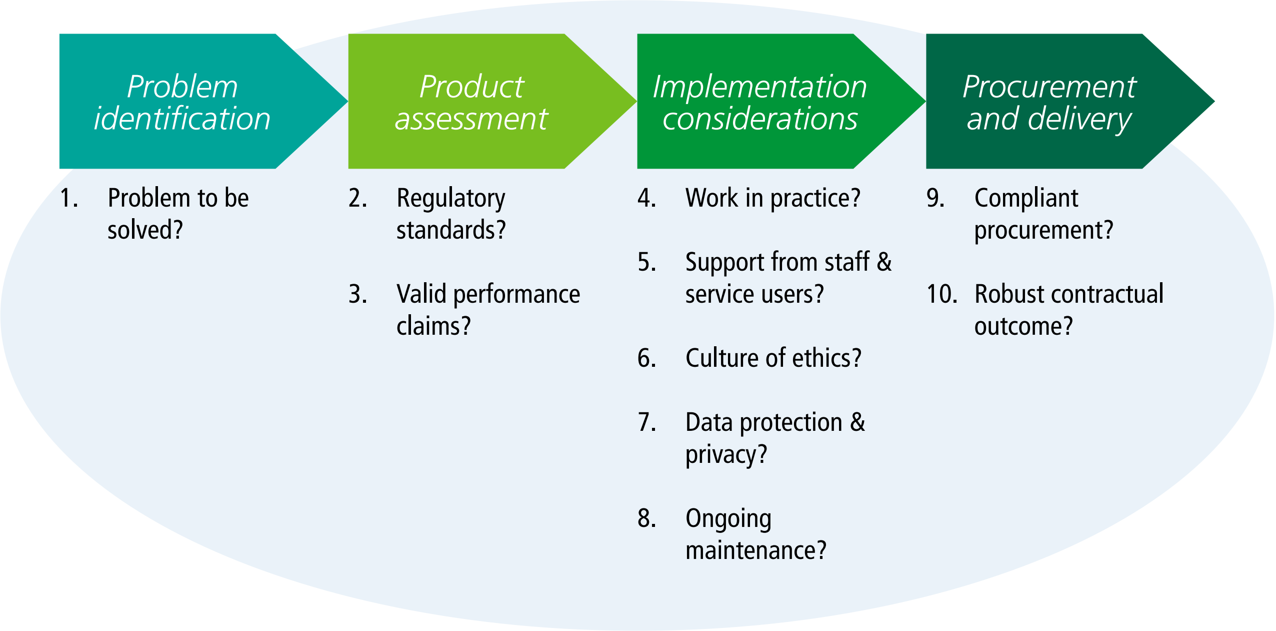
Matthew Gould,
CEO of NHSX

Julian Kelly,
NHS Chief Financial Officer

Executive summary

The exciting promise of artificial intelligence (AI) to transform areas of health and care is becoming a reality. However, deciding on whether a specific AI product is right for your organisation is no easy thing. Whilst prospective vendors are proliferating, discerning between their technologies is difficult.

This Buyer's Guide to AI in Health and Care sets out the important questions you need to consider in order to undertake robust procurement exercises and make well-informed buying decisions about AI products. The guide's ten questions are grouped under four categories: problem identification, product assessment, implementation considerations, and procurement and delivery.


Problem identification

1. Problem to be solved?

Product assessment

2. Regulatory standards?
3. Valid performance claims?

Implementation considerations

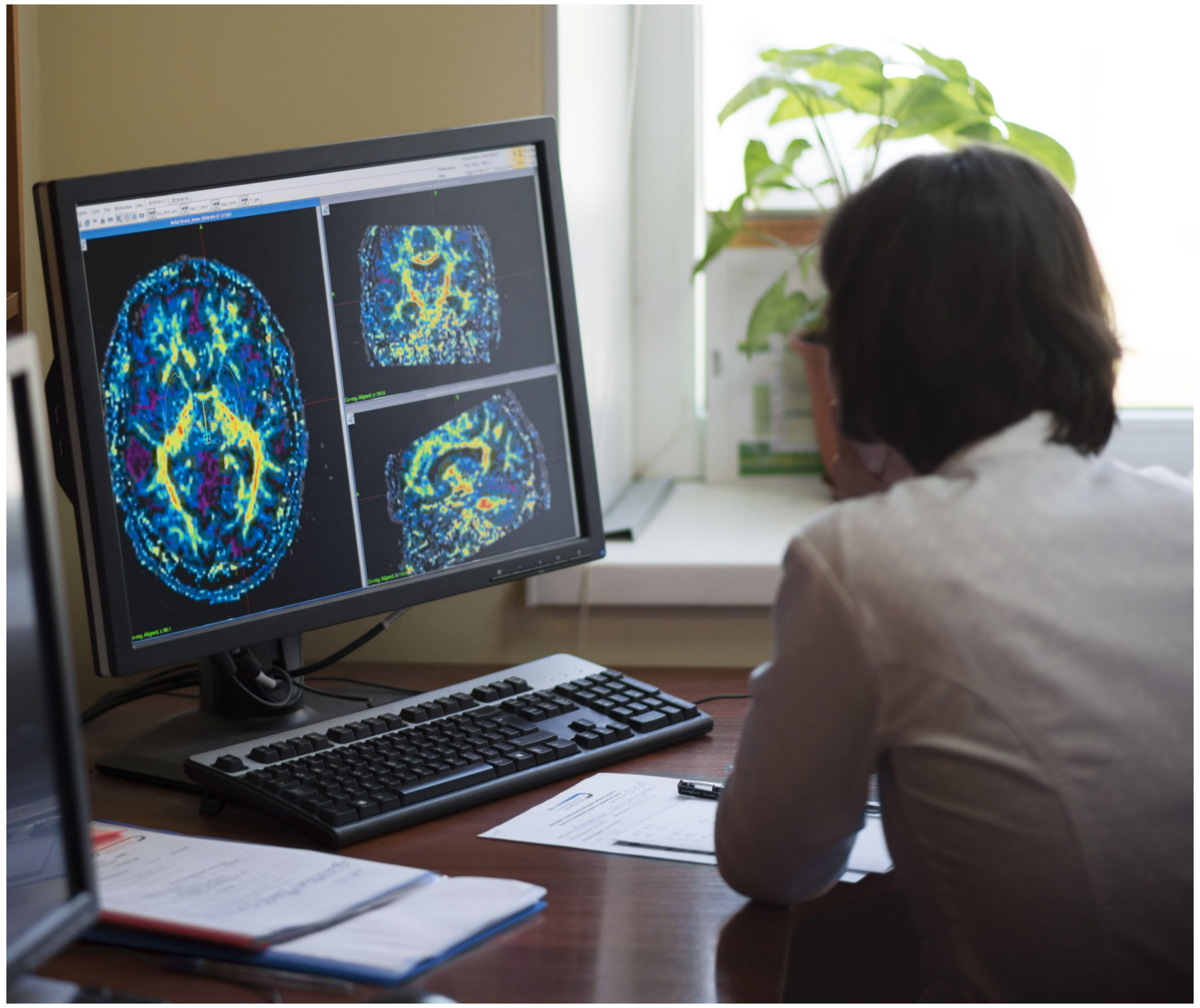
4. Work in practice?
5. Support from staff & service users?
6. Culture of ethics?
7. Data protection & privacy?
8. Ongoing maintenance?

Procurement and delivery

9. Compliant procurement?
10. Robust contractual outcome?

This executive summary sets out the key points under each of the questions. The main body of the guide addresses each question in detail, and signposts to other helpful guidance where relevant. The guide also comes with an assessment template, which provides a structured format for answering the questions posed in the guide.

If you would like to discuss anything in this guide, share your experience of buying AI, or make suggestions about how we can improve the guide, please get in touch with the AI Lab at ailab@nhsx.nhs.uk. Please also get in touch if your organisation has procured, or decides to procure, an AI product. We are collating all of these details, so that the Lab can be a central point for sharing use cases, successes and challenges.



1. WHAT PROBLEM ARE YOU TRYING TO SOLVE, AND IS ARTIFICIAL INTELLIGENCE THE RIGHT SOLUTION?

You should start with the problem you're trying to solve. Once you've identified the challenge, can you explain why you are choosing AI? What additional 'intelligence' do you need and why is AI the solution? You should consider whether:

- The problem you're trying to solve is associated with a large quantity of data which an AI model could learn from
- Analysis of that data would be on a scale so large and repetitive that humans would struggle to carry it out effectively
- You could test the outputs of a model for accuracy against empirical evidence
- Model outputs would lead to problem-solving in the real world
- The data in question is available - even if disguised or buried - and can be used ethically and safely

If you can't satisfy these points, a simpler solution may be more appropriate.

You should also consider the appropriate scale for addressing your challenge - i.e. organisational, system, regional or even national. Organisations may experience specific challenges in common, making collaboration valuable. Key to this decision is the data required for the AI solution and at what scale the dataset is sufficiently large to ensure a minimum level of viability. It may be that data needs to be pooled across several organisations to achieve this.

Like any investment, you will need to produce a business case to justify expenditure. But owing to the experimental nature of many AI projects, this is not straightforward. Testing hypotheses on historical data, and then through a pilot project, may help you through some of the uncertainty.

2. DOES THIS PRODUCT MEET REGULATORY STANDARDS?

The Medicines and Healthcare product Regulatory Agency (MHRA) has overall responsibility for ensuring that regulatory standards for medical devices are met. A manufacturer must ensure that any medical device placed on the market or put into service has the necessary CE marking. CE marking should be viewed as a minimum threshold for certifying that the device is safe, performs as intended and that the benefits outweigh the risks. From 1 January 2021, because of the UK leaving the EU, there will be several changes to how medical devices are placed on the market. These changes include introducing the UKCA - UK Conformity Assessed - as a new route to product marking.

You should be clear about the 'intended use' of the product - i.e. what exactly it can be used for and under what specific conditions. This should enable you, in turn, to be clear about the product's risk classification. Medical devices can be classed as I, IIa, IIb or III in order of increasing risk. The requirements for obtaining a CE mark are similar across all the classes but the complexity and amount of effort increases as the risk class increases.

Where an AI product is not categorised as a medical device and is designed, for example, to improve operational efficiency rather than support specific clinical or care-related decision-making, manufacturers should develop their technology in line with ISO 82304-1 for healthcare software.

3. DOES THIS PRODUCT PERFORM IN LINE WITH THE MANUFACTURER'S CLAIMS?

As part of your procurement exercise, you will need to scrutinise the performance claims made by the manufacturer about the product.

The performance expected of a supervised learning model will vary in line with its intended use. For example, in the case of classification models used in diagnostic settings, there is an important trade-off between sensitivity - the proportion of actual positive cases correctly identified - and specificity - the proportion of actual negative cases correctly identified. The trade-off depends on weighing up the healthcare consequences and health economic implications of missing a diagnosis versus over-diagnosing. This trade-off may vary at different stages of a care pathway. Different metrics will shed light on different aspects of model performance and all have limitations. Different metrics will be more or less appropriate to different use cases.

Classification models often provide a probability between 0 and 1 that a case is positive, as opposed to a binary result. Discrete classification into positive and negative is obtained by setting a threshold on the probability. If the value exceeds the threshold, the model classifies the case as positive. If the value does not exceed the threshold, the case is classified as negative.

You should pay attention to the chosen threshold, particularly as performance metrics will change according to where the threshold has been set. Given the complexity of these metrics, the Area Under the Curve (AUC) is a helpful measure. It is a single metric which evaluates model performance without taking into account the chosen threshold.

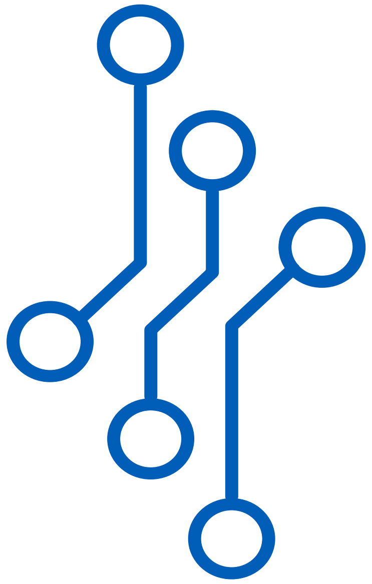
Whilst classification models predict discrete quantities (i.e. classes), regression models predict continuous quantities - e.g. how many residential care beds will be needed for patients discharged from hospital next week.

Performance metrics for regression models are affected differently by the presence of outliers in the data. You should consider how much of an issue outliers are for your use case, and then prioritise metrics accordingly.

You should request details of validation tests that have been performed, and should expect to see a form of retrospective validation. This entails testing the model on previously collected data that the model has not seen. The purpose of this is to test whether the model can generalise - i.e. carry across - its predictive performance from the data it was trained on to new data.

Investigating the model's safety credentials is key - you need to be confident about the model's robustness, fairness, explainability and privacy.

For any reported performance to be meaningful, it must be compared to the current state of play. For example, how does model performance compare with the product it will replace, or with the human decision-making it will augment?



4. WILL THIS PRODUCT WORK IN PRACTICE?

The key issue here is if the performance claims made in theory will translate to practical benefits for your organisation.

In terms of the evidence base for a product's effectiveness, the NICE Evidence Standards Framework for Digital Health Technologies sets out the evidence standards you should expect to see. These standards are stratified according to a product's function and potential risk. In the case of medical devices, clinical evaluation reports (CERs) are the primary source of clinical evidence to support the claims of a manufacturer. During procurement, you could ask the vendor about the product's effectiveness in other health and care organisations, and for contact details of people in those organisations involved in the product's implementation.

Practically speaking for your organisation, you should think about questions affecting project delivery, such as:

- How 'plug-in' ready is the product and will your organisation need to make significant changes in order to realise the promised benefits?
- How smoothly does the product operate for users, and how well does it meet user needs?
- How will your existing systems integrate with the new technology to ensure clear and reliable workflows?
- What are the product's data requirements, and does your organisation have the necessary data - labelled, formatted and stored in the right way?
- What data storage and computing power does the product need, and how will you ensure this is in place?
- How will you monitor whether your product is achieving in practice the benefits anticipated in theory?

5. CAN YOU SECURE THE SUPPORT YOU NEED FROM STAFF AND SERVICE USERS?

The breadth and depth of multidisciplinary backing needed for AI procurement is greater than for the procurement of a more traditional technical product. You should ensure that staff who will be end-users of the AI solution have been fully involved in every stage of the procurement process, to ensure that the product will meet their needs. Then, you should bear in mind that at the implementation stage, influencing and supporting people to change the way they do their work is difficult. Make sure you factor this into your plans and that your prospective vendor will supply any induction or training that you need.

You should assess how persuasive a story you can tell about the expected improvement in health and care outcomes. This will be important for getting buy-in from patients and service users. In addition, you should design a 'no surprises' communication approach about how the AI product is being used, how patients' and service users' data is being processed, and, where relevant, how an AI model is supporting decisions which impact on them.

6. CAN YOU BUILD AND MAINTAIN A CULTURE OF ETHICAL RESPONSIBILITY AROUND THIS PROJECT?

Ethics, fairness and transparency need to be at the front and centre of AI in health and care. Building a culture of ethical responsibility is key to successful implementation. Health and care staff who do not experience a technology being in keeping with their professional ethical standards are less likely to adopt it.

You should be confident that your AI project is ethically permissible, fair and non-discriminatory, worthy of public trust, and justifiable. To help you make an informed judgement, you should carry out a stakeholder impact assessment before making any procurement decisions.

7. WHAT DATA PROTECTION PROTOCOLS DO YOU NEED TO SAFEGUARD PRIVACY AND COMPLY WITH THE LAW?

Data protection must be embedded into every aspect of your project. You will need to create a data flow map that identifies the data assets and data flows - i.e. the exchanges of data - related to your AI project. Where the data flow map identifies instances of data being passed to and processed by a data Processor (i.e. the vendor) on behalf of a data Controller (i.e. your organisation), a legally binding written data processing contract - otherwise known as an information sharing agreement - is needed.

Further information governance measures depend on the purpose of the data processing and whether the data being processed could identify individuals. If individuals could be identified, this is considered sensitive personal data and you must complete a Data Protection Impact Assessment.

Where identifiable data is being processed, individuals have the right to:

- Be informed about how their personal data is collected and used
- Give consent to the use of their data
- Access their data

You will need to ensure that use of data for this AI project is covered under your organisation’s data privacy notice. You will also need to document what is in place to mitigate the risk of a patient or service user being re-identified - in an unauthorised way - from the data held about them.

8. CAN YOU MANAGE AND MAINTAIN THIS PRODUCT AFTER YOU ADOPT IT?

During procurement, you should investigate what support the vendor is offering for ongoing management and maintenance of the product, and ask about their post-market surveillance plans - i.e. plans to monitor the safety of their device after it has been released on the market. This includes asking about:

- Features of their managed service
- Their approach to product and data pipeline updates
- Their plan for mitigating product failure
- Their plan for addressing performance drift of the model outside of a margin that is acceptable to you

You should be clear about your organisation’s responsibilities and capabilities in relation to operation and maintenance.

You should also clarify any expectations the vendor has of your organisation sending back data to support their iteration of the model or development of other products. Depending on the product, this could be categorised as a clinical investigation and require separate approval. You will also need to address this in your information governance arrangements.

Decommissioning is a key final stage of the management and maintenance cycle. You should ensure that suitable plans are in place at the outset.

9. IS YOUR PROCUREMENT PROCESS FAIR, TRANSPARENT AND COMPETITIVE?

Like any technology, AI products need to be purchased on the basis of recognised public procurement principles. Early engagement with the market may identify new potential vendors, level the playing field and help vendors understand what buyers need. At the same time, you should be clear about and document your justification for talking to and inviting specific vendors to bid.

10. CAN YOU ENSURE A COMMERCIALLY AND LEGALLY ROBUST CONTRACTUAL OUTCOME FOR YOUR ORGANISATION, AND THE HEALTH AND CARE SECTOR?

This guide does not provide a comprehensive treatment of commercial contracting, but there are key questions to consider:

- Are you clear about exactly what you are procuring?
- Is it a lifetime product?
- Is it a licence?
- What is the accompanying support package?

You should set out a service level agreement as part of your contracting process. You should also ensure that the financial arrangements you are establishing are sustainable in the long-term.

In principle, your contracts should be as open as possible. Whilst confidentiality clauses are often invoked to prevent disclosure of commercially sensitive information, this can be detrimental to public trust.

Your contracts should recognise and safeguard the value of the data that you are sharing, and the resources which are generated as a result. Where your organisation sends back data to the vendor - whether for the purpose of auditing the product, re-training the model, or potentially developing a new product - you may be contributing to the creation of intellectual property. You should take advice early to address this appropriately. NHSX’s Centre for Improving Data Collaboration can offer tailored guidance - please contact the Centre via the AI Lab at improvingdatacollaboration@nhsx.nhs.uk.

Liability issues should not be a barrier to adoption of effective technology. However, it is important to be clear on who has responsibility should anything go wrong. Product liability and indemnity is therefore an important issue to address at contracting stage. In the case of data protection, your organisation - as a data Controller - is primarily responsible for its own compliance but also for ensuring the compliance of its data Processors. A Controller is expected to have measures in place to reduce the likelihood of a data breach, and will be held accountable if they have not done this.

Introduction

THE AI LAB

The mission of the NHSX AI Lab is to accelerate the safe and effective adoption of artificial intelligence in health and care. NHSX's report on Artificial Intelligence: How to get it right provides the policy context for the AI Lab, together with an overview of the opportunities and challenges of AI for the sector.

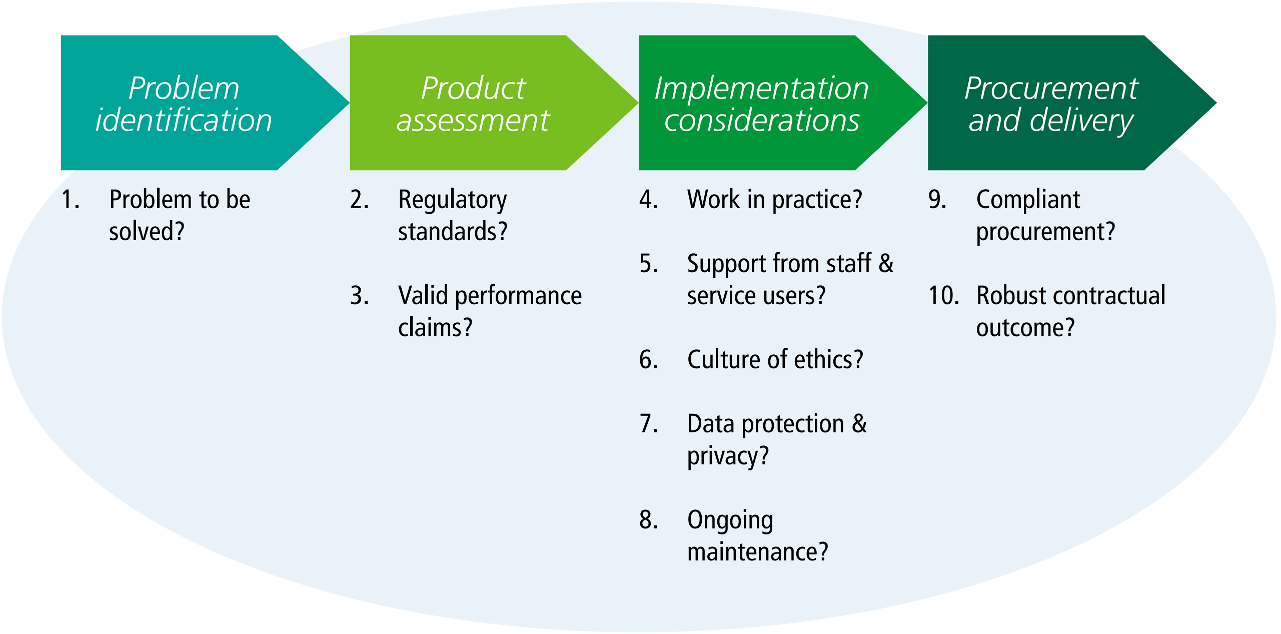
PURPOSE OF THE GUIDE

The exciting promise of AI to transform areas of health and care is becoming a reality. However, deciding on whether a specific AI product is right for your organisation is no easy thing. Whilst prospective vendors are proliferating, discerning between their technologies is difficult.

In May 2020, the AI Lab published a short, initial version of a Buyer's Checklist for AI in Health and Care. This was in the immediate context of COVID-19, and the unprecedented opportunity this presented for adopting innovative digital solutions at pace and scale. Health and care organisations continue to receive multiple approaches in respect of AI applications that may improve the quality and ease the burden of their work. At the same time, organisations need to be assured that any AI technology they do buy meets the highest standards of safety and effectiveness. This assurance is not straightforward: specific characteristics of AI generate important considerations for buyers, over and above those related to digital products in general.

This guide is a revised and more comprehensive version of the checklist, which it now replaces. The guide comes together with an assessment template, which provides a structured format for answering the questions posed in the guide. These resources are designed to equip organisations with the background knowledge and important questions they need to consider in order to undertake robust procurement exercises and make well-informed AI buying decisions. The diagram summarises the ten questions addressed by the guide - categorised under problem identification, product assessment, implementation considerations, and procurement and delivery.

In highlighting how much there is to consider, we do not want to put off potential buyers from making AI purchases and reaping their benefits. At the same time, we recommend caution and healthy scepticism. Whilst some organisations may be well-placed to address all of the questions posed here, we recognise that others will not have all of the in-house capacity for this. The guide references several additional resources that are available. You should also feel free to contact the AI Lab at ailab@nhsx.nhs.uk - we will be able to signpost you to further support.



1. Problem to be solved?
2. Regulatory standards?
3. Valid performance claims?
4. Work in practice?
5. Support from staff & service users?
6. Culture of ethics?
7. Data protection & privacy?
8. Ongoing maintenance?
9. Compliant procurement?
10. Robust contractual outcome?

SCOPE OF THE GUIDE

This guide is concerned with procuring ‘off-the-shelf’ AI applications - i.e. products packaged by vendors as ready for deployment. It does not focus on bespoke projects - i.e. research or build collaborations between heath/care organisations and developers. The distinction is more blurred than might appear though: even products labelled as off-the-shelf will need customising - to varying degrees - to the operational environments of specific organisations.

Another point on scope: this guide does not offer detailed guidance on implementation and delivery. Where it does address these issues, it is insofar as they form important considerations prior to procurement. A procurement decision should be aimed at setting up successful implementation and delivery.

For further support, you can find a list of materials at the back of this guide that have been published by public bodies to help guide you through this process.

CONTACT US

If you would like to discuss anything in this guide, share your experience of buying AI, or make suggestions about how we can improve the guide, please get in touch with the AI Lab at ailab@nhsx.nhs.uk.

Please also get in touch if your organisation has procured, or decides to procure, an AI product. We are collating all of these details, so that the Lab can be a central point for sharing use cases, successes and challenges. Comparing notes between organisations is critical to identifying opportunities and avoiding common pitfalls.

Problem identification

1. WHAT PROBLEM ARE YOU TRYING TO SOLVE, AND IS ARTIFICIAL INTELLIGENCE THE RIGHT SOLUTION?

Challenge-driven, not solution-led

Starting with the problem you are trying to solve should be the first step of any purchasing decision. Mapping the pathway in question, identifying pain points and speaking to end-users are all important aspects of this. Starting with the problem is particularly important for AI: the novelty and ingenuity of AI technologies can make it easy to overlook this critical step.

Once you have identified the problem, can you articulate the rationale for choosing AI? What is it about AI that makes it a powerful choice for addressing your challenge? What additional "intelligence" is it that you require? Guidance to help you assess if AI is the right technology for your challenge, produced by the Government Digital Service and the Office for AI, suggests that organisations consider whether or not:

- The problem to be solved is associated with a large quantity of data which an AI model could feasibly learn from
- Analysis of that data would be on a scale so large and repetitive that humans would struggle to carry it out effectively
- The outputs of a model could be tested for accuracy against a ground truth - i.e. empirical evidence
- Model outputs would lead to problem-solving that achieves outcomes in the real world
- The data in question is available - even if disguised or buried - and can be used ethically and safely

You should constantly review your rationale though the procurement process. If your rationale does not satisfy the points above, it may be that a simpler solution is more appropriate.

You should also consider the appropriate scale for addressing your challenge - i.e. organisational, system, regional or even national. This is an important decision; it may be that organisations experience specific challenges in common, making collaboration more valuable. From a transactional perspective, working at a system-level, for example, may offer economies of scale through 'doing things once' and increased buying power. Key to this decision is the data required for the AI solution and at what scale the dataset is sufficiently large to ensure a minimum level of viability. It may be that data needs to be pooled across several organisations to achieve this.

Credible business case

You will need to produce a business case to justify investment in an AI solution. Because of the experimental nature of many AI projects, this is not straightforward. What is the baseline you are looking to improve, and what metrics matter in measuring this improvement? Can you test the AI product on historical data in your organisation to evaluate its potential impact? Can you then implement a pilot project to ascertain whether this impact is achieved in a live, operational setting? This approach will help to test hypotheses in the business case and navigate some of the uncertainty.

As far as possible, you should articulate and quantify the quality improvements and/or savings and efficiencies which would be delivered for your organisation. Certain AI solutions may result in small advances to the functionality of a service, or minor improvements to an end user's experience. These might be 'nice-to-haves' but will not be significant enough to develop a credible business case.

Section B of the NICE Evidence Standards Framework for Digital Health Technologies sets out evidence standards for making judgements on the potential economic impact of digital health technologies. It specifies:

- Key economic information inputs for making these judgements
- Appropriate levels of economic analysis based on the type of commissioning decision and level of risk to the buyer - you could potentially ask for health economic analyses as part of your procurement exercise

The Framework also provides a template for undertaking budget impact analyses. These analyses can become complicated given that an AI product may deliver a proportion of cost savings for a different service, or organisation, to the one using the product. In such cases, this will raise questions about allocating costs and might require co-developing a business case.

Product assessment

2. DOES THIS PRODUCT MEET REGULATORY STANDARDS?

Medical devices

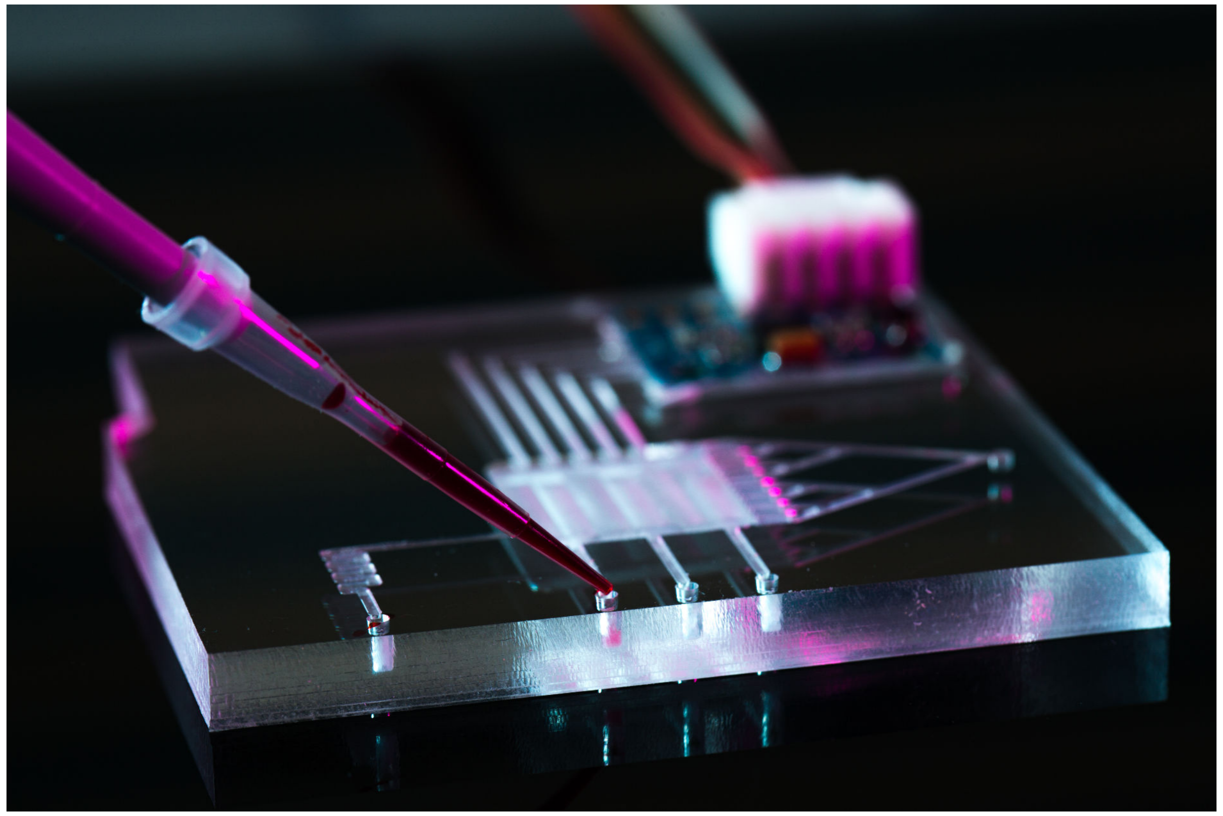
The Medicines and Healthcare product Regulatory Agency (MHRA) has overall responsible for ensuring that regulatory standards for medical devices are met, with certification activities being undertaken by organisations known as Notified Bodies. A manufacturer is required to conform to the CE marking requirements for any product defined as a medical device that is placed on the market or put into service. The MHRA has published guidance on when software applications are considered to be a medical device and how they are regulated.

From 1 January 2021, because of the UK leaving the EU, there will be several changes to how medical devices are placed on the market. These changes are summarised at the end of this section.

Note that manufacturers can apply to make an investigational device available to a specific group as part of a clinical investigation for CE marking, i.e. before the device has been certified. Note also that in-house produced devices are currently outside medical device regulations, provided that they cannot be found on the market, are developed within an organisation and are used only for patients within that organisation.

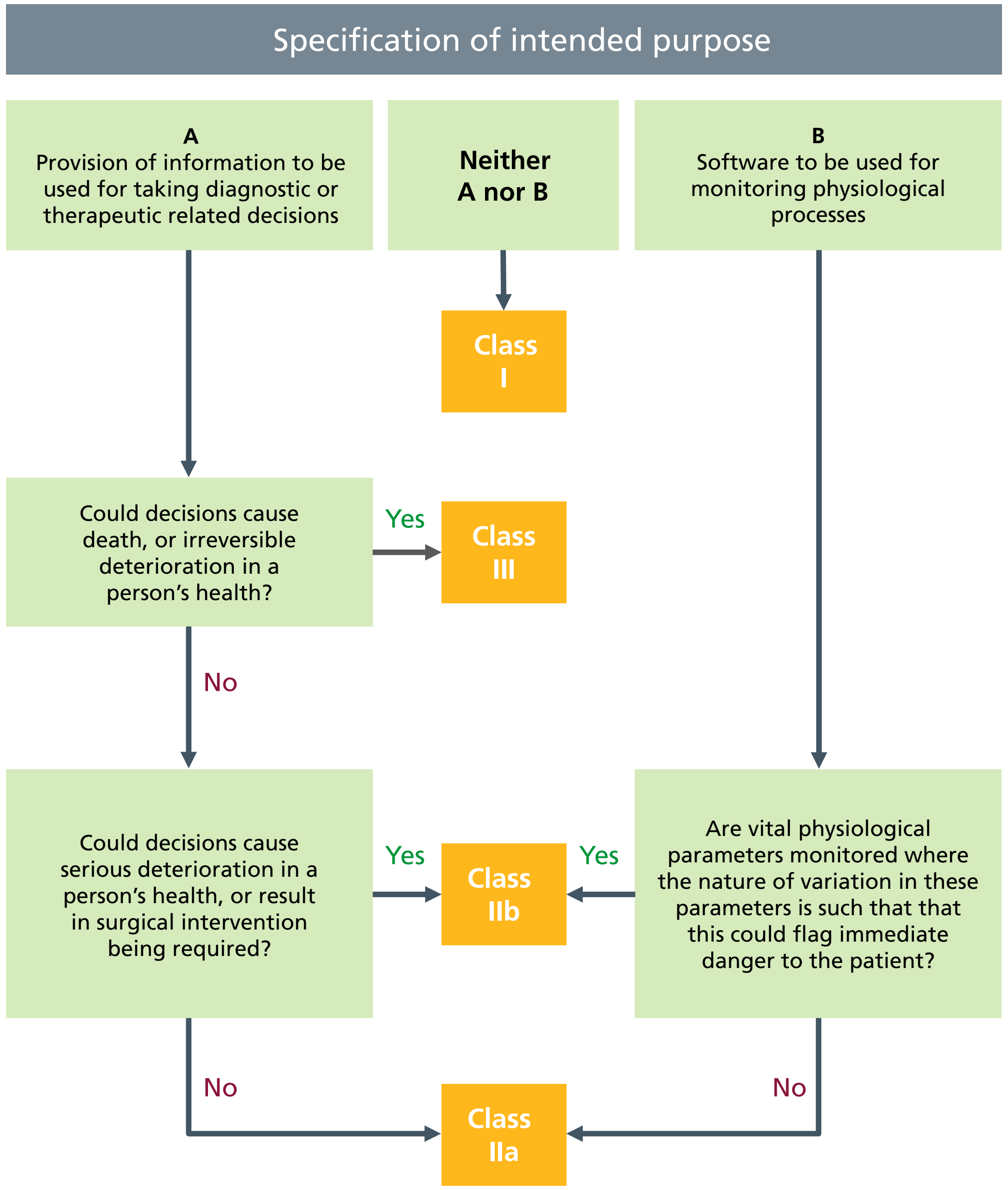
Intended use

You should be clear about the ‘intended use’ of the product - i.e. what exactly it can be used for and the exact conditions under which it can be used. Conversely, you should also be clear about what counts as misuse - i.e. what falls outside of the product’s remit. The intended use statement can be found on the device’s declaration of conformity, CE marking certificate and instructions for use. Being clear on the product’s intended use is also important for judging the product’s performance.



Device risk classification

You should be clear about the device's risk classification. Medical devices can be classed as I, IIa, IIb or III in order of increasing risk - this classification follows from the intended use statement. The requirements for obtaining a CE mark are similar across all the classes, however, the complexity and amount of effort increases as the risk class increases. It is the manufacturer's responsibility to determine the risk class of their device. However, it is worth checking that you agree with this by taking the intended use statement and following the flow chart below (originally sourced from an article by VDE). The European Commission's Guidance on Qualification and Classification of Software in Regulation provides further detail.


Specification of intended purpose
A
Provision of information to be used for taking diagnostic or therapeutic related decisions
Neither
A nor B
B
Software to be used for monitoring physiological processes
Class
I
Could decisions cause death, or irreversible deterioration in a person's health?
Yes
Class
III
No
Could decisions cause serious deterioration in a person's health, or result in surgical intervention being required?
Yes
Class
IIb
Yes
Are vital physiological parameters monitored where the nature of variation in these parameters is such that that this could flag immediate danger to the patient?
No
Class
IIa
No

CE marking requirements - as per the EU Medical Devices Regulation (MDR)

All devices require a set of technical files including a clinical evaluation/performance report. All devices above Class I require a certified quality management system (e.g. ISO 13485 accreditation) and their conformity to the medical device requirements to be certified by a Notified Body. The technical files for a device should contain all the evidence for meeting the requirements. This includes a requirements checklist, product development documentation, usability file, validation testing, clinical evaluation/performance report, security and risk assessments, labelling and post market surveillance. You can ask to see all of this documentation. An article by Psephos Biomedica sets out in full the requirements for CE marking of medical devices.

Current regulatory position in the UK

Because the UK has left the EU, there will be a number of changes to how medical devices are placed on the market in Great Britain, and separately in Northern Ireland after the transition period, which come into force from 1 January 2021. These changes are set out in guidance from the MHRA. In summary:

- All medical devices placed on the UK market will need to be registered by their manufacturers with the MHRA. There will be a grace period for registering. The length of this grace period depends on the risk class of the device.

- A new route to product marking - UKCA (UK Conformity Assessed) - will be available for manufacturers wishing to place a device on the Great Britain market (England, Scotland and Wales). The technical requirements, together with the processes and standards to demonstrate conformity, will be largely the same as they are now for CE marking.

- CE marking will continue to be used and recognised until 30 June 2023, after which only UKCA marking will be recognised for the Great Britain market.

- The UKCA mark will not be recognised in the EU, EEA or Northern Ireland markets. Products that currently require CE marking will still need a CE mark for sale in these markets.

Given that CE marking will continue to be used and recognised until June 2023, the current position of AI medical devices in relation to EU regulation remains significant for the UK. The current position is set out in this blog article from Hardian Health:

- Under the current EU Medical Devices Directive (MDD), AI decision support systems can be classed as low risk (Class I). Manufacturers of class I devices can apply a CE mark following self-certification, which means the manufacturer decides that they have met the requirements of the legislation and signs the declaration of conformity without external audit.

- This is due to be redressed by the new EU Medical Devices Regulation (MDR), where almost all AI systems will be upgraded to at least medium risk (Class IIa).

- This means that a formal external audit on quality systems, together with technical documentation by an independent regulatory body, is required for CE marking.

- The MDR was due to come into operation in May 2020. However, the EU Parliament voted to delay the MDR by one year to May 2021 as a result of COVID-19.

- Even when the new directive does come into force, devices previously self-certified under their Class I designation through MDD can keep that certification until 2024, as long as there is no significant change to the device (e.g. no change to the model) - note this guidance from the European Commission.

In summary, CE - and UKCA - marking should be viewed as a minimum threshold for certifying that the device is safe, performs as intended and that benefits outweigh risks. You should ensure that you agree with the manufacturer's determination of risk class by reviewing the intended use of the product against the flow chart above. It is also wise, especially in the case of Class I devices, to review enough of the clinical and other evidence about the device's safety and performance for you to feel satisfied with the claims being made - refer to questions 3 and 4.

For certain types of AI devices that carry out regulated clinical activity independently of a healthcare professional - such as analysis and reporting of X-Ray, CT or MRI images - registration as a service through the Care Quality Commission (CQC) is required in addition to CE/UKCA marking. Section 4.2 of CQC's report on Using machine learning in diagnostic services covers this in more detail.

Operational software

In this context, operational software refers to technology which improves operational efficiency, as opposed to supporting specific clinical or care-related decision-making. An AI example might be a virtual flow assistant which optimises the matching of discharged hospital patients to onward care settings.

Healthcare software technologies should be developed in line with ISO 82304-1.

There are further clinical safety standards that apply to non-medical devices (technology that does not fall within the scope of the medical device regulations), as well as to medical devices. These are:

- DCB 0129 - risk management standards for developers of health software
- DCB 0160 - risk management for the deployment of health software

The two standards are companions. To comply with DCB 0129, the supplier must carry out risk management activities throughout the lifecycle of the product, log these in a hazard log with details of mitigating actions and, prior to release of a particular version, create a safety case report that summarises the risks, mitigations and overall safety position. You can ask to see documents covering:

- The clinical safety management system (a description of the processes followed by the company to review and respond to clinical risks
- The clinical safety management plan (a description of risk management processes specific to a particular product)
- The hazard log (the risk register for the product)
- The clinical safety case report (a summary of the safety position of the particular release)

It is then the responsibility of the deploying organisation to perform a parallel risk assessment for the deployment of the software (DCB 0160). This looks at how the product will fit into real world systems.

The standards mandate that risk management activities involve and are signed off by appropriately qualified personnel, namely a Clinical Safety Officer. This is a person with professional clinical accreditation, experience with software and who has completed the NHS Digital Clinical Safety Officer training. Refer to this this article from Safehand for a summary of the standards.

In cases where your organisation is not considering an 'off-the-shelf' product but rather a collaborative research partnership, usual research governance requirements apply. Approval for health and social care / community care research in the UK is facilitated through the Integrated Research Application System (IRAS). For clinical research projects, the relevant Health Research Authority (HRA) Research Ethics Committee for medical devices research must grant ethical approval. In addition, the MHRA should be notified about a clinical investigation.

3. DOES THIS PRODUCT PERFORM IN LINE WITH THE MANUFACTURER'S CLAIMS?

To undertake full due diligence as part of your procurement exercise, you will need to scrutinise the performance claims made by manufacturer about their products. Question 2 explains where to find these performance claims in the product's documentation. This due diligence requires relatively detailed technical knowledge of AI. If your organisation does not have this knowledge in-house, you can contact the AI Lab at ailab@nhsx.nhs.uk and we will signpost you to available support.

Actionable output

You should be confident that the product's target prediction - i.e. the variable predicted by the model - results in an output that supports practical action. If not, the product's utility will be limited, irrespective of how impressive the model's technical performance.

Model performance metrics

Models can be categorised according to whether they operate through supervised or unsupervised learning. If the training dataset includes the variable to be predicted, a supervised learning model can learn the relationship between that variable and the input variables. On the other hand, if the training dataset does not include the variable to be predicted, an unsupervised learning model generates predictions on the basis of the most apparent pattern in the data. This section on performance metrics focuses on supervised learning models, as these are the most commonly used in real-life applications. The performance expected of a model will vary in line with its intended use and whether it carries out classification or regression.

1. Classification models

Classification models predict the class labels, or categories, of new data. In health and care, this type of model is most commonly associated with diagnostics, with the class labels being positive and negative. Here, the trade-off between metrics of sensitivity and specificity is key. This trade-off depends on weighing up the healthcare consequences and health economic implications of missing a diagnosis versus over-diagnosing.

For example, if the product is to be used in early breast cancer screening tests, then its ability to capture positive cases may be particularly important - for example, you don't want to 'send home' someone who actually has a tumour. It follows that the model requires a high level of sensitivity - i.e. the proportion of actual positive cases that are correctly identified as such. At the same time, if sensitivity is prioritised excessively, the model's specificity - i.e. the proportion of actual negative cases that are correctly identified as such - becomes unacceptably low. In practice, this means that the model raises too high a number of false alarms to be useful as a triaging tool.

On the other hand, if the product is to be used for testing further along a breast cancer care pathway, its ability to correctly detect negative cases may become more important. For example, false alarms may cause unnecessary medical intervention and anxiety. It follows that the model requires a higher level of specificity. Reliably capturing positive cases is clearly still important, so the trade-off with sensitivity must be considered. The greater risk in terms of litigation claims actually comes from false negatives at any stage of the process - i.e. adverse outcomes resulting from failing to identify someone who should have been diagnosed and prioritised for treatment.

Sensitivity and specificity are one possible pair of metrics that quantify the performance of a model. An alternative pair is Positive Predictive Value (PPV) and Negative Predictive Value (NPV). PPV answers the question "out of the cases that were predicted to be positive, what fraction of them was classified correctly?" NPV answers the question "out of the cases that were predicted to be negative, what fraction of them was classified correctly?" For each of these metrics, the closer to 1 the result is, the better the performance of the model judged against that metric.

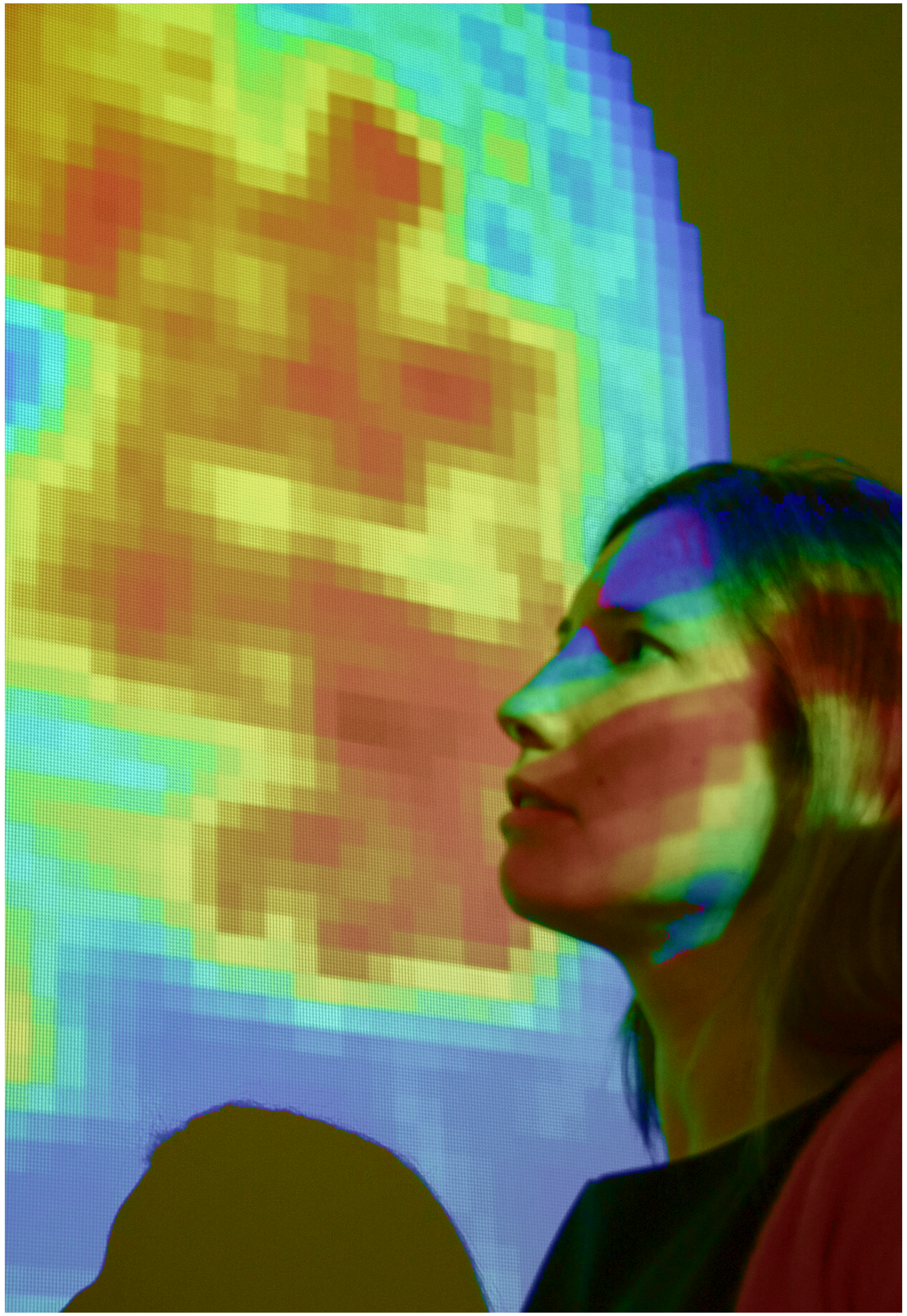
Different metrics shed light on different aspects of model performance, and have different limitations. For example, when the classes are imbalanced - i.e. the number of negative cases far exceeds the number of positive ones - certain metrics become misleading. Model accuracy - defined as the proportion of correctly identified positive and negative cases - is prone to this. A model that classifies all cases as negative would score very highly on accuracy in spite of its inability to recognise positive cases. This is because the metric is skewed by the large number of negative cases identified correctly. The lesson here is that you should select performance metrics that are appropriate to the use case.

It is also worth noting that the metrics themselves have multiple names used by different communities of experts.

You should expect to see confusion matrices - sometimes known as contingency tables - as part of a product's reported performance. Confusion matrices display numbers of actual cases in the columns and numbers of predicted cases in the rows. As shown in the example - covering 10,000 patients - each case is allocated to the relevant cell. Relevant metrics can then be calculated from the entries in the table. Confusion matrices are a useful way of describing model performance because they are intuitive to read, and cover all possible combinations of correct and incorrect predictions.

		Actual outcome		
		Positive case	Negative case	Metrics
Predicted outcome	Positive case	97	210	Positive Predicted Value (PPV) = 97/(97+210) = **31.60%**
	Negative case	3	9,690	Negative Predicted Value (NPV) = 9960/(3+9690) = **99.97%**
Metrics		Sensitivity = 97/(97+3) = **97.00%**	Specificity = 9,690/(210+9690) = **97.88%**	

Example confusion matrix



Models often provide a probability between 0 and 1 that a case is positive, as opposed to a binary result. Discrete classification into positive and negative is obtained by setting a threshold on the probability. If the value exceeds the threshold, the model classifies the case as positive; if the value does not exceed the threshold, the case is classified as negative.

You should therefore pay attention to the chosen threshold, and be sure that this is reported. The sensitivity and specificity change according to where the threshold has been set. For example, prioritising sensitivity involves setting a low threshold, so that even merely suspected positive cases are flagged.

Given the complexity of these performance metrics, the Area Under the Curve (AUC) is a helpful measure. It evaluates model performance without taking into account the chosen threshold. The AUC is a single metric with values between 0 and 1, where a perfect model would score 1. A downside of the AUC is its theoretical nature, which makes it difficult to link scores to practical conclusions. Moreover, when the AUC is defined in terms of the Receiver Operating Characteristic (ROC) curve, it can be misleading in a similar way as accuracy is for use cases with imbalanced classes.

2. Regression models

Whilst classification models predict discrete quantities (i.e. classes), regression models predict continuous quantities. For example, regression models would be needed to predict how many people will attend A&E on a given day, or how many residential care beds will be needed for patients discharged from hospital next week.

The table sets out the most common metrics used to evaluate the performance of regression models.

Abbreviation	Name	Description	Unit	How to judge performance against this metric
RMSE	Root Mean Square Error	Square root of the average square of the prediction error.	The same unit as the variable being predicted (e.g. number of residential care beds needed).	The closer the value to zero, the better.
MAE	Mean Absolute Error	Average of the magnitude of the prediction error.		
R^2	R-Squared	Quantity of variability in the data that is explained by the model.	This metric has no unit - it generates a number less than 1 that may take negative values as well as positive ones.	The closer the value to 1, but above zero, the better. Values below zero mean that model performance is worse than simply predicting a mean value.

Faults with data collection or unusual events can result in outliers, which are extreme values of the quantity being predicted. Outliers do not follow the overall trend in the data, and can compromise how well the regression model fits the data by skewing it towards extreme values. Because of this, you should be mindful of how outliers affect different metrics, and whether this is an issue for your use case. Because the RMSE contains a square power, it is sensitive to outliers, and its value is changed dramatically by extreme deviations. The MAE is less affected by outliers, but as a result it may underestimate the model's prediction error.

Retrospective validation

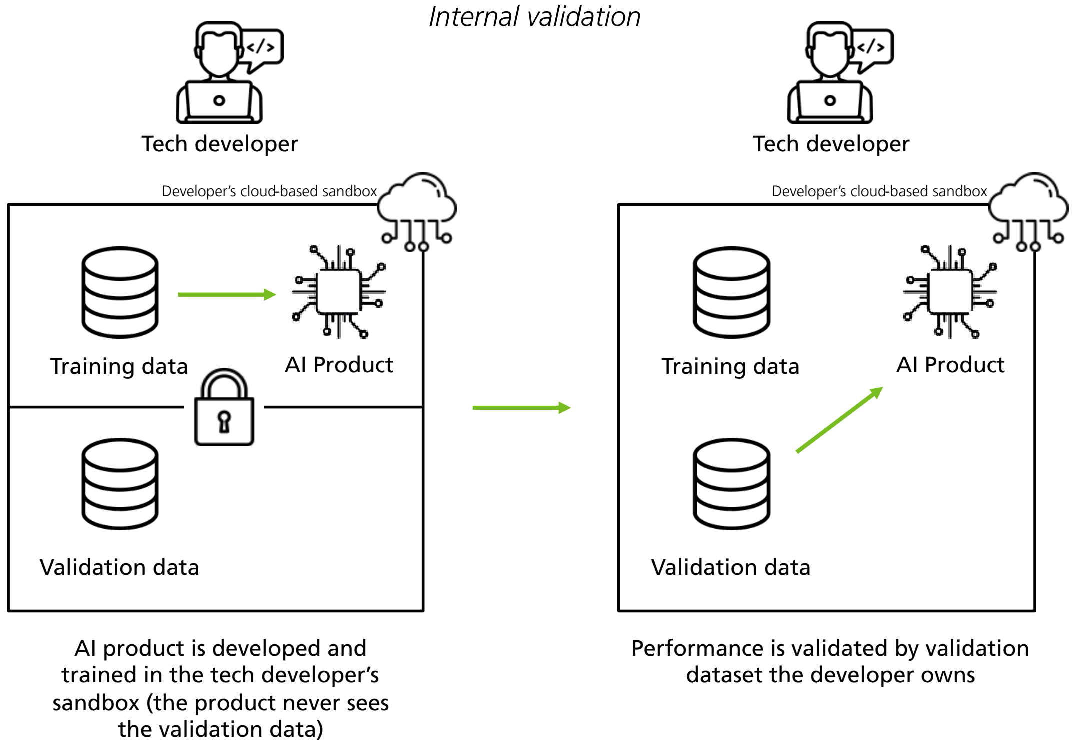
As part of your procurement exercise, you should request details of validation tests that have been performed, and should expect to see a form of retrospective validation.

Retrospective validation entails testing the model on previously collected data that the model has not seen. The purpose of this is to test whether the model can generalise - i.e. carry across - its predictive performance from the data it was trained on to new data. You would usually expect to see a deterioration in performance, when compared against performance on data the model has been trained on.

As highlighted in A Clinician's Guide to Artificial Intelligence: How to Critically Appraise Machine Learning Studies, there should be no overlap or leakage between the seen and unseen data. Data used for validation must be kept separate from the data used for training the model initially. AI models learn from the data in a way that is similar to how students learn from past exam questions. The validation data, however, corresponds to this year's exam. Validation data that overlaps with training data is equivalent to testing the knowledge of students on exam questions they have already seen. The separation of the training and validation data should be clearly documented. There is some nuance between "internal" and "external" validation:

- Internal validation - where the unseen data is generated by holding back and separating data from the same dataset used for training
- External validation - where the unseen data comes from an independent dataset - which the developer has not had access to - and where the testing is done by a third-party investigator

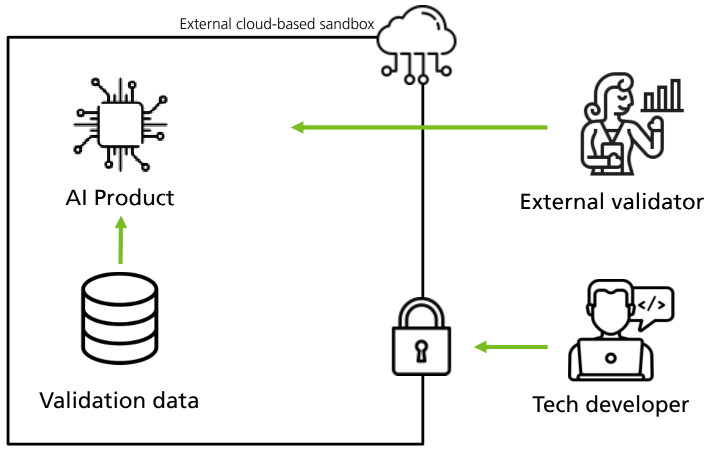
Internal validation



AI product is developed and trained in the tech developer's sandbox (the product never sees the validation data)

Performance is validated by validation dataset the developer owns

External validation



The trained model is deployed on a cloud-based sandbox owned by the external validator

The tech developer is locked out, and the external validator uses their validation data to validate model performance

As set out in A Clinician's Guide to Artificial Intelligence, external validation may include using a dataset that is:

- Independent of the original dataset but similar in terms of its population and setting
- Independent, but different in either the population (for example, ethnicity, socioeconomic status) or the setting (for example, screening, primary care, secondary care, geographical location)
- Representative of the same or new populations over time - this would help to assess degradation of the model performance as the population evolves
- Different for technical reasons (for example, on account of images taken by different scanners)

It is important that the validation dataset has been sampled fairly and representatively, and ideally that it incorporates edge cases - for example, data points from demographic groups under-represented in the data. To assess this, you will need to see descriptive details of the dataset.

Retrospective validation may be a precursor to prospective evaluation, where the model is further tested on site - this time on live data being fed to it in real-time. Whilst retrospective validation emphasises the performance of a model, prospective evaluation emphasises the usability and impact of the product as part of a care pathway.

Model safety credentials

Investigating the model's safety credentials is a key aspect of due diligence on the product, and integral to minimising risks associated with its use. This is critical in health and care, where the consequences of incorrect information and decision-making are so weighty. The rigour of these safety credentials will impact buy-in from health and care professionals, who will be held accountable for decisions made with AI support, and will be expected to explain how they arrived at these decisions.

Safety credentials for a model should incorporate the criteria set out below. (Note that safety considerations in relation to the data itself that is processed by the model are set out under question 7.)

Robustness

Can the model make reliable predictions, given that data is subject to uncertainty and errors? Does the model remain effective even in extreme or unexpected situations?

Fairness

What measures are in place to prevent the model from discovering hidden patterns of discrimination in its training data, reproducing these patterns and making biased predictions as a result? This article from the Information Commissioner's Office (ICO) on Human bias and discrimination in AI systems elaborates on how bias can feature in AI models, and how this can be mitigated.

Explainability

Can predictions made by the model be explained in terms that both a trained user of the product and a patient/service user would understand?

Good explainability means explaining the reasoning behind a particular model-supported decision or outcome in plain, easily understandable, and everyday language. A report on explaining decisions made with AI, produced by the ICO and Alan Turing Institute, breaks this down into six explainability tasks which will help to build trust about the safety and equity of an AI model without having to dive deeply into the model's complexity.

- Select priority explanations by considering the domain, use case and impact on the individual
- Collect and pre-process data in an explanation-aware manner
- Build your system to ensure you are able to extract relevant information for a range of explanation types

- Translate the rationale of your system's results into usable and easily understandable reasons
- Prepare implementers to deploy your AI system responsibly and fairly
- Consider how to build and present your explanation to different audiences

You will need to ask prospective vendors about how the model they are selling performs in relation to these tasks. Annex 1 of the same report sets out how these tasks could be applied in the case of providing an explanation for a cancer diagnosis.

Existing technical tools - summarised in Annex 3 - for quantifying the explainability of "black-box" models are experimental, and so you should assess their suitability with the help of an expert.

In general, AI technologies used in health and care do not currently provide solely-automated decision-making about individuals. However, you should be aware that if this were to be the case, the General Data Protection Regulation (GDPR) would grant individuals the right to be informed about and have access to meaningful information about the model logic involved, and the significance and envisaged consequences for the individual. Article 22 of the GDPR actually gives individuals the right not to be subject to a solely automated decision that produces legal or similarly significant effects.

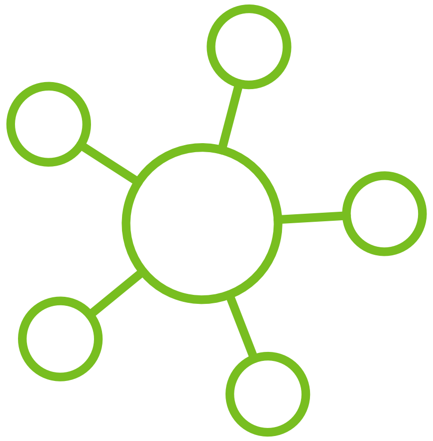
Privacy

Is the model resilient against unauthorised attempts to re-identify individuals whose data was used in the model's training set? This resilience is important because the parameters contained within a model inherit information from the dataset used for training. This information may be used to re-identify individuals, or information about individuals, through adversarial attacks such as:

- Membership inference attack - given access to the model and a data point, the attacker tries to determine if this known data point was in the model's training set. Knowing that an individual was part of a training dataset would imply additional information about them - for example, based on where and when the training data was collected.
- Model inversion - given access to the model and some publicly available information about an individual the model was trained on, the attacker tries to predict other private information about that individual.

AI models used in health and care settings will typically be accessed by a restricted number of authorised users. This reduces the risk of adversarial attack: a non-authorised person would need to have a very high level of technological sophistication to gain access to the model. Nonetheless, the possibility of a model's vulnerability being exploited can never be entirely discounted. Even though the individuals whose data is contained in the training dataset of an off-the-shelf product will most likely be unconnected with your organisation, you still have a responsibility to ensure the model's privacy credentials.

Two techniques used to bolster model privacy are differential privacy and federated learning. Differential privacy, in the context of model training, refers to the addition of random noise to the data in order to limit the amount the model can learn about specific individuals. Federated learning involves generating a remote model by aggregating copies of it that are trained at multiple local sites. This makes it possible to bring 'the model to the data', as opposed to the conventional arrangement of bringing 'the data to the model'. As a result, federated learning enables data to remain with its local controller. Safe implementation of federated learning relies on differential privacy or encryption solutions.



Comparative performance

For reported model performance to be meaningful, it must be compared to the current state. For example, how does model performance compare with a system that it might replace, such as in the case of a clinical coding process? You should be sure that the comparator being considered matches closely enough to the use case of your product.

Where a model is designed to augment human decision-making, identifying the additional value it will offer requires nuance. For example, one hospital trust found that a model which detected lung nodules to support clinical decision-making provided value in a scenario they had not initially envisaged (this could qualify as off-label use). They had originally intended to use the model to support decision-making in senior multidisciplinary team (MDT) meetings about cancer patients. But because cases are vetted and reviewed by several specialists before they get discussed at an MDT, the intelligence provided by the model was of limited value. However, the model did provide significant additional value where it augmented a screening exercise in a broader patient population. In this scenario - where images were being reviewed by non-thoracic-specialist clinicians - the model was able to 'opportunistically' identify lung abnormalities, even though this was not the primary purpose of the screening. The lesson here is that the seemingly obvious comparator current state may not be the best place to look for potential value.

If you need support with assessing comparative performance, please contact the AI Lab at ailab@nhsx.nhs.uk - we can signpost you to further help.

Implementation considerations

4. WILL THIS PRODUCT WORK IN PRACTICE?

The key point in this section is to consider the operational utility of the product. Will the performance claims made in a theoretical context translate to practical benefits in your specific organisation?

Operational utility is a particularly important issue in the case of AI. To illustrate: this article from the MIT Technology Review - based on a study by Google Health of the impact in clinical settings of its deep-learning system for spotting signs of eye disease in diabetic patients - makes the point that Google’s medical AI was super accurate in a lab. Real life was a different story.

Evidence base for effectiveness

The NICE Evidence Standards Framework for Digital Health Technologies establishes the evidence base required of a technology product’s effectiveness before it is deemed to be appropriate for adoption. (Note that this framework is applicable to AI technologies that are based on fixed algorithms.)

In Section A, the Framework classifies technologies by function and stratifies them into evidence tiers based on their potential risk to users. For example, an operational function improving system efficiency is considered less risky than a diagnostic function. The framework then sets out evidence categories - with associated standards - expected of each tier. These categories include:

- Relevance to current health and care pathways
- Achievement of demonstrative health and care outcomes
- Credibility with practitioners
- Acceptability with users
- Contribution to challenging health inequalities

Using this framework, you can ascertain which function and evidence tier your prospective product comes under, and then identify the standards you should expect of the evidence provided by the manufacturer. Clinical evaluation reports (CERs) are the primary source of clinical evidence for medical devices supporting the claims of a manufacturer. These reports are a prerequisite for CE marking and need to be updated throughout the life cycle of the device. This means they become a source of real world evidence once the device is on the market.

Insight from other organisations

As part of your procurement exercise, you could ask the vendor about the product’s effectiveness in other health and care organisations, and for contact details of people in those organisations involved in the product’s implementation. Speaking to such people will give you a flavour of how widely and successfully the product has been deployed to date, and crucial insight into some of the human factors at play that are fleshed out in questions below.

It is also worth asking the AI Lab at ailab@nhsx.nhs.uk to connect you with other organisations using, or thinking of using, the same or a similar AI product. This will give you further insight on performance, or the opportunity to share thoughts on procurement.

Deliverability

Be cautious of ambitious promises about how ‘plug-in’ ready the product is for your organisation specifically, even if it is being sold as off-the-shelf. If significant changes to your organisation’s ways of working are needed to realise the benefits promised by the product, is this possible? If implementation will cause short-term disruption, how will you manage this? If appropriate, you might initially consider running the new technology and existing system in tandem. If you are replacing an older system with the new technology, have you factored in time, costs and potential complications of dealing with a legacy system?

Getting to the point of achieving a fully operational AI implementation in your organisation is a significant task. Therefore, you should consider starting off with a pilot project - and even then one with a tightly defined scope and set of success metrics. This will enable you to test your judgement about the AI product’s performance in practice. At the same time, if your intention is to scale-up, you should check at the outset how feasible this is for the technology in question - against each of the sub-sections, as a minimum. Conversely, you should also consider the operational implications of an exit strategy, in case it transpires that the product does not meet your organisation’s needs.

Usability and integration

How smoothly the product operates from a user perspective, and how well it meets user needs, is key to practical implementation. One of the most important considerations here is how your existing systems will integrate with the new technology to ensure clear and reliable workflows for staff. Back-end integration with key systems is often necessary to establish reliable workflows. As part of this, it may be helpful to ask for the product's software architecture diagram. In addition, the product should make use of open standards to promote interoperability - such as FHIR, HL7v2, DICOM. Manufacturers are required by regulation to provide all the information on how to connect their device to other devices.

You should consult widely to ascertain that integration will be possible for your prospective AI product, or, as a minimum, acknowledge the uncertainty and scope your project accordingly.

To illustrate how important this is, consider the unsatisfactory result of a product that is designed to augment the diagnostic decision-making of a radiologist, but is only integrated with the imaging storage system and not the radiology information system. It will not be integrated into the radiologist's clinical workflow, which is managed through the radiology information system. As a result, the radiologist would need to manually open a patient's imaging set to view the product's predictive output. Without automatically flagging to the radiologist those images that are worthy of concern, the product's usability is compromised and therefore its utility severely limited. The Royal College of Radiologists has issued detailed guidance on integrating AI with the radiology reporting flow.

Data compatibility

You should be clear about the product's data requirements - as set out in the product's instructions for use - and whether your organisation is able to fulfil these. Do you have the data needed, in the right format? Can the data be labelled and stored in the right way? How reliable is the quality of this data? This relates to both data accuracy and completeness. Data quality might be a particular issue in cases of unstructured data that is inputted by humans - for example, free text.

Data storage and computing power

What are the data storage and computing power requirements of the product? If this infrastructure is not provided by the vendor, can your organisation cover the associated costs? Cloud-based computing can offer scalable and accessible solutions, but transitioning away from local servers may not be a straightforward move for many organisations. If your project will use cloud-based servers, you should be clear about where these are based and whether this raises any issues of compliance with GDPR. You should also consider the scalability of storage and computing costs: as data processing requirements increase, do these costs increase in a linear or exponential way?

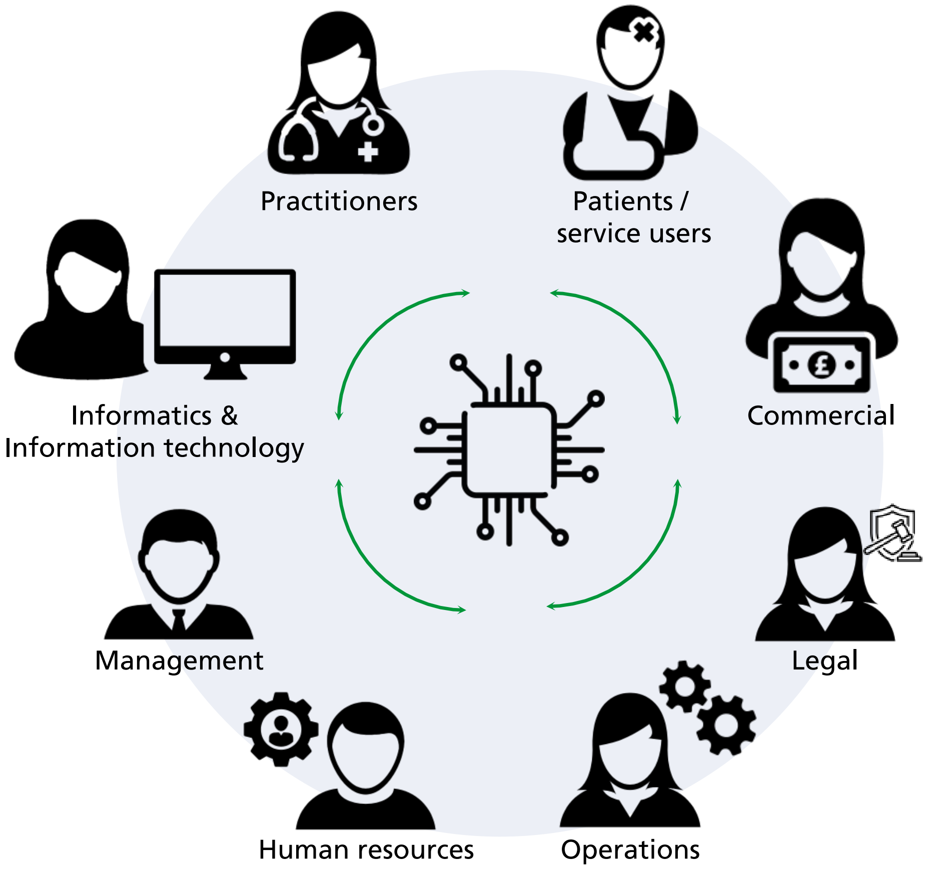
Auditing and evaluation

Once your implementation is underway, you will need to monitor whether your product is achieving in practice the benefits anticipated in theory. Therefore, you should build auditing and evaluation into your project design early on. You should consider the data, resourcing and cost implications of this early on, prior to procurement. The AI Lab can offer guidance about this - you can get in touch with us at ailab@nhsx.nhs.uk.

5. CAN YOU SECURE THE SUPPORT YOU NEED FROM STAFF AND SERVICE USERS?

Buy-in from staff

First, you should ensure that staff who will be end-users of the AI solution have been fully involved in every stage of the procurement exercise, to ensure that the product will meet their needs. Then, you should bear in mind that the breadth and depth of multidisciplinary backing needed for an AI procurement is greater than for the procurement of a more traditional technical product. The complexities are greater, as are the number of people who need to be involved. A successful AI procurement will need input from a range of clinicians/practitioners, together with colleagues from strategy and management, information governance, commercial, legal, information technology, informatics, operations and human resources.



Successful AI procurement and adoption needs buy-in from a diverse range of stakeholders

A common stumbling block of technology adoption - in health and care, and beyond - is failing to take account of changes to end users' workflows. Influencing and supporting people to change how they perform their day-to-day work is difficult and time-consuming - make sure this is factored into your plans. You should ensure that your prospective vendor will supply any induction or training that is needed in your organisation - not only about the product's functionality but also about its limitations. This should equip staff not only to use the product effectively, but also - where applicable - to provide their patients and service users with clear information about how the product will be used, and to secure their informed consent.

Buy-in from patients and service users

You should assess how persuasive a story you can tell about the expected improvement in health and care outcomes. This is important for any intervention which will change the way people receive and experience their services. It is even more important when their personal data is being used and processed. Understanding Patient Data's Foundations of Fairness report makes clear that the public recognises the potential of data to improve outcomes, but are also sceptical that the benefits promised will always be achieved.

In addition, you should design a 'no surprises' communication approach with patients and service users about how the AI product is being used, how their data is being processed, and, where relevant, how an AI model is supporting decisions which impact on them. Context here is key. As argued in Explaining decisions made with AI, the importance of providing explanations to individuals, the reasons why individuals want them, and the level of detail you should expect to go into, all depend on what the decision is about. For example, if a decision might result in an individual potentially losing out or identifying a need for redress, they would attach more importance to explainability.

This transparency is a fundamental requirement of the GDPR, which requires that any information addressed to a data subject or to the public is easily accessible, concise and easily understood - through the use of clear and plain language.

6. CAN YOU BUILD AND MAINTAIN A CULTURE OF ETHICAL RESPONSIBILITY AROUND THIS PROJECT?

The Department of Health and Social Care's (DHSC) Code of conduct for data-driven health and care technology advocates for ethics, fairness and transparency to be at the front and centre of AI in health and care. Aside from the moral and legal arguments, building a culture of ethical responsibility is key to successful implementation. If health and care staff do not experience a technology as consonant with their professional ethical standards, this is a recipe for non-adoption and abandonment - as set out in Beyond Adoption, there in the context of clinicians' resistance to new healthcare technology.

Guidance on how to use AI ethically and safely, produced by the Government Digital Service and the Office for AI, sets out four aspects of responsibility which should be embedded into decision-making and design. You should be confident that your AI project, and the product in question, is:

- Ethically permissible - the impact on the wellbeing of affected stakeholders and communities has been considered
- Fair and non-discriminatory - potential discriminatory effects on individuals and social groups have been considered, and potential biases which may influence the AI model's outcome have been mitigated
- Worthy of public trust - the public interest motivation for using the product, together with accountability for and robust governance of its use, has been guaranteed as much as possible
- Justifiable - transparency in project design and implementation, together with interpretability of the model's decisions and behaviours, has been prioritised

You should also assess your project against the seven principles of the Data Ethics Framework, which guides the design of appropriate data use in government and the wider public sector. Other sections of this guide address many of these principles.

- Start with clear user need and public benefit
- Be aware of relevant legislation and codes of practice
- Use data that is proportionate to the user need
- Understand the limitations of the data
- Ensure robust practices and work within your skillset
- Make your work transparent and be accountable
- Embed data use responsibly

The Data Ethics Workbook sets out specific questions under each of these principles, to support you in assessing your project against the framework.

In assessing these ethical issues, you will likely develop an intuitive sense of how much importance prospective vendors attach to them. Does this give you a view of the vendors' values and commitment to the public good?

Before making any procurement decision, you should carry out a stakeholder impact assessment - as part of your engagement with staff, patients, service users and the broader public. This is a practice recommended by The Alan Turing Institute's report on Understanding artificial intelligence ethics and safety (which comes with an impact assessment template). Stakeholder impact assessments serve several purposes, which include:

- Helping to build public confidence that the design and deployment of your AI project has been done responsibly
- Facilitating and strengthening your accountability framework
- Bringing to light unseen risks that may impact on individuals and the public
- Underwriting well-informed decision-making and transparent innovation practices
- Demonstrating forethought and due diligence, not only within your organisation but also in relation to the wider public

7. WHAT DATA PROTECTION PROTOCOLS DO YOU NEED TO SAFEGUARD PRIVACY AND COMPLY WITH THE LAW?

The Data Protection Act (DPA) 2018 requires integration of data protection into every aspect of data processing activities. This approach is identified as 'data protection by design and default' in the GDPR. In practice, this means that organisations need to consider data privacy at the initial stages and throughout the complete development process of new products or services that involve processing data.

In the context of health and care, the importance of data protection and privacy for safeguarding cannot be underestimated. You should therefore consider your organisation's responsibilities - and ability to comply with them - well in advance of making any AI procurement decision, guided by your Data Protection Officer.

You will need to create a data flow map that identifies the data assets and data flows - i.e. the exchanges of data - pertaining to your AI project. You should be clear on where the data sits at all times, and to what extent it is encrypted - both in transit and at rest.

Where the data flow map identifies instances of data being passed to and processed by a data Processor (i.e. the vendor) on behalf of a data Controller (i.e. your organisation), a legally binding written data processing contract - otherwise known as an information sharing agreement - is needed. This is a requirement under both the DPA and GDPR. The agreement must stipulate what the Processor can and cannot do with the data - in relation to data retention, deletion and secondary uses of the data (i.e. uses beyond the direct care of the patient/service user, for example, training new models). The first page of this Data Sharing Checklist from the ICO sets out in more detail the key points you should include.

Further information governance measures you will need to take depend on the purpose of the data processing and whether identifiable data is being processed:

What is the purpose of the data processing?	Does this purpose justify the processing of identifiable data?	What is the expectation of you as a service provider?
Direct care of patients / service users.	Yes.	Develop a Data Protection Impact Assessment - this sets out how proposed flows of personal identifiable data will be governed, together with the controls you have in place to ensure lawful processing. If you have carried out a DPIA that identifies a high risk which cannot be mitigated, prior consultation with the ICO is required under GDPR. Verify that a technology supplier is suitably accredited or qualified to process sensitive data (e.g. ISO/IEC 27001 certified).
Logging issues to improve your organisation's safe use of the model.		
Developing a new AI product - including training of new models.	No - the identity of patients must be protected through the data privacy techniques described above.	As above, plus consult your organisation's Data Protection Officer. Ensure that the consent process offers people real choice and control over whether their data is used for these purposes. Be mindful that any terms and conditions in a contract which permit the vendor to further process data for these purposes would make the vendor a data Controller in their own right, regardless of any contractual statements to the contrary.
Monitoring the overall performance of a product, enabling software developers to improve the performance and safety of the product they offer.		

The six lawful bases for processing personal data are set out in Article 6 of the GDPR - consent being the first and most obvious of these. In addition, a key transparency requirement set out in the GDPR is that individuals have the right to be informed about how their personal data is collected and used, and be able to access this data. These legal imperatives of consent and transparency are also enshrined under the common law duty of confidentiality. You will need to ensure that use of data for this AI project is covered under your organisation's data privacy notice, which identifies:

- Contact details of the Data Protection Officer in your organisation
- The purposes - in relation to both direct care and any reasonable secondary purposes - for which personal data is collected and used
- Your legal basis for processing personal data
- How long data is kept
- How data is used and disclosed
- How you monitor the compliance of your data Processors, and what contractual sanctions are in place for misuse and breaches

You will also need to document what is in place to mitigate the risk of a patient or service user being re-identified - in an unauthorised way - from the data held about them. As explained in this Nature article, re-identified patient records may be a lucrative target for health insurance companies wishing to reduce their financial risk by discriminating against individuals with certain illnesses. As a result, the sale of re-identified medical records has become a business model for unscrupulous data-mining companies. One basic way to reduce the risk of re-identification is to collect and process as little personal data as possible in the first place. Known as the minimisation principle, this means ensuring that any personal data is adequate, relevant and limited to what is necessary for the purpose it is being processed. Common data privacy techniques to mitigate the risk of re-identification - which you can ask prospective vendors about - are set out here.

Aggregation

This entails pooling records of individual data to generate summary, aggregated results. How feasible this technique is will depend on the use case for the model. There is a residual risk of re-identification if the aggregated group is too small, and especially if it is possible to link this group with other datasets.

Pseudonymisation

This entails replacing potential individually identifiable data fields with pseudonymous data or blank data entries. As explained in the ICO's guidance on what is personal data, pseudonymisation may involve replacing names or other identifiers which are easily attributed to individuals with, for example, a reference number. Whilst you can tie that reference number back to the individual if you have access to the relevant information, technical and organisational measures are put in place to ensure that this additional information is held separately. There is still a residual risk of re-identification if pseudonymised data is linked in an unauthorised way with other datasets.

Synthetic data generation

This entails generating synthetic data that is representative of the real data without containing any 'real' individuals within it. Synthetic data has the advantage of maintaining the same level of granularity as the real data - for example, no aggregation has been performed - whilst reducing the risk of compromising individuals' privacy. In practice, a compromise has to be struck between the utility of the synthetic data and the degree of privacy achieved. The synthetic data must be similar enough to the real data so that statistical analysis and modelling is still possible, but not so similar that the privacy of individuals is compromised. This can be achieved using tools such as differential privacy that enable precise control of the utility-privacy tradeoff, and quantification of the residual privacy risk in the synthetic data.

The appropriate mitigations against patients or service users being re-identified are governed by the ICO's Anonymisation code of practice (note that this code is being revised in light of GDPR).

8. CAN YOU MANAGE AND MAINTAIN THIS PRODUCT AFTER YOU ADOPT IT?

Several features specific to AI products result in a greater emphasis on management and maintenance than is required for other technologies. You should consider these implications up-front, so that they don't emerge as a surprise once contracts have been signed.

Vendor's responsibilities

As part of your procurement exercise, you should investigate what the vendor is offering vis-a-vis ongoing management and maintenance, and ask about their post-market surveillance plans, i.e. plans to monitor the safety of their device after it has been released on the market:

- Are they providing a managed service?

- What is their approach to product and data pipeline updates, and who pays for these? The more these updates are automated, installed remotely and continuously "pushed", the more confidence this should give you in the vendor. At the same time, uniform updates may not suffice; you may need updates that are bespoke to your organisation's data and use of the product. You should also consider how the vendor's timetable of updates will align with your organisation's internal maintenance cycle.

- What is their plan for mitigating adverse events - i.e. if the AI product fails or is compromised? You should ask to see their risk management report, which will address this. In addition, are they clear on their reporting requirements via MORE, and - where applicable - under EU Medical Devices Regulation?

- What is their plan for addressing performance drift? Performance drift refers to the degradation of model performance over time, caused by shifts in the distribution of input variables in the live data - for example, owing to changing demographic patterns - compared with the model's training data. You should agree a margin of acceptable drift - dependent on the use case of the product - with the vendor at the start of a contract, and continually assess whether the model's performance remains within this margin.

Your organisation's responsibilities

You should also examine your own organisation's role and capabilities:

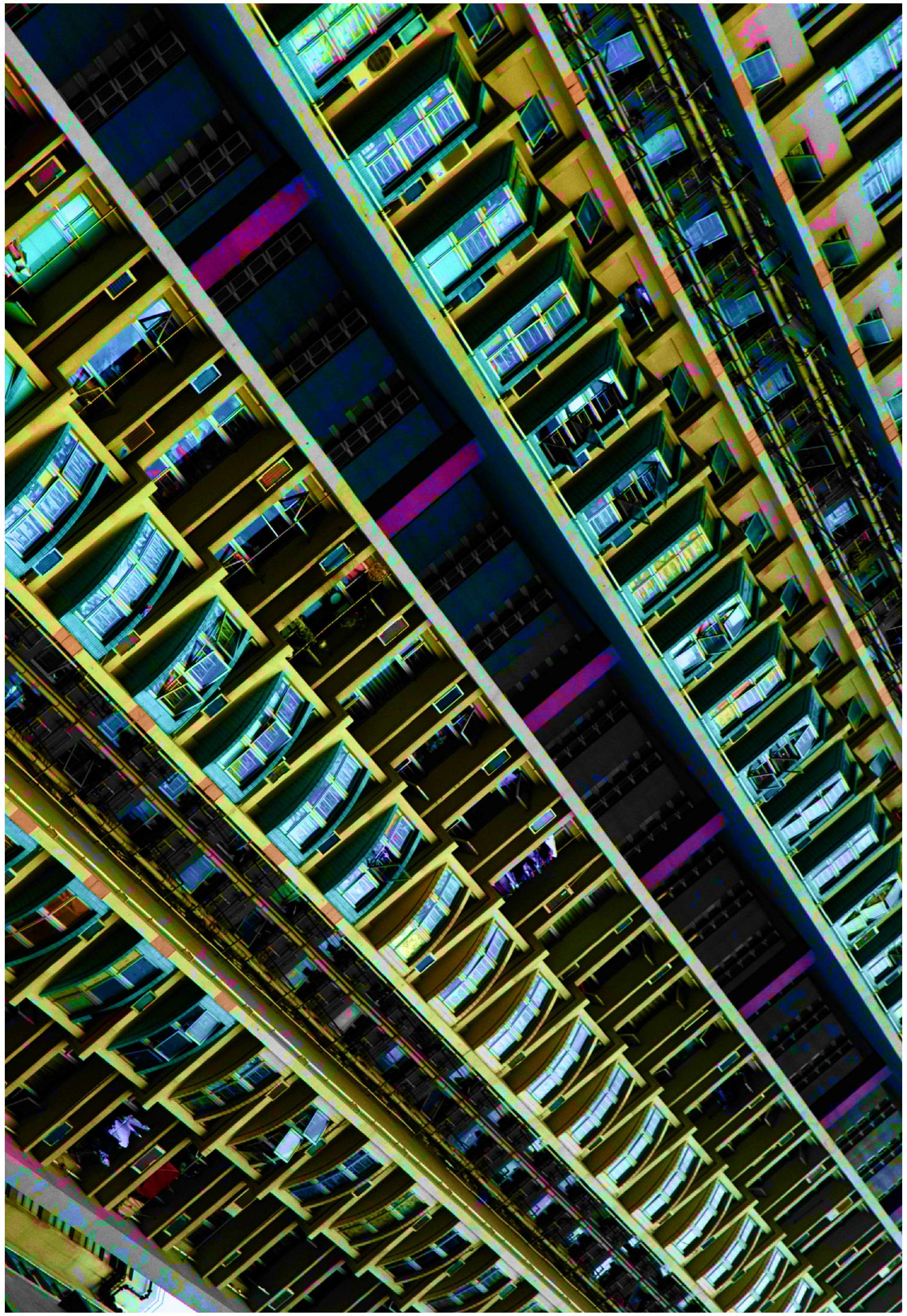
- If you are not buying into a managed service, do you have the IT capability in-house? You should consider expertise in system administration (for example, Windows, Linux, Unix), networking and firewalls, and single sign-on and access controls (for example, Okta).

- Can you develop a sufficiently robust understanding of relevant data feeds, flows and structures, such that if any changes occur to model data inputs, you can assess any potential impacts on model performance - or signpost questions to the vendor?

- Are you clear on your organisation's requirement to report adverse events via the Yellow Card scheme - if your prospective product is a medical device? Adverse events can include misdiagnosis or incorrect treatment, which has involved use of the device

- How resilient would your service be if the vendor were to cease operating?

You should also be clear about any expectations the vendor has of your organisation sending back data to support their iteration of the model or development of other products. Where the use of this data goes beyond immediate product safety monitoring, this could be categorised as a clinical investigation by the MHRA and require separate approval - dependent on the product. It will also likely require separate consent from the patient or service users whose data is being used. You should therefore clarify exactly what the vendor means by model iteration and development, and ensure that your information governance arrangements address this.



Decommissioning

Although it might seem strange to recommend thinking about this before you have even bought the product, decommissioning is a crucial final stage of the management and maintenance cycle. This is the case either when your use of the product has reached a natural end-point, or if you need an earlier exit strategy because the product is not meeting your organisation's needs. You should incorporate the questions below into your procurement exercise.

- What will happen to any data that is stored outside of your organisation's systems, at the point of decommissioning? Will it be deleted, or archived? If it is to be deleted, are there any restrictions on the period of time it needs to be kept before deletion? If it is to be archived, where will it be archived, who will have access to it, and who will pay for storage costs? How will the vendor evidence that they have deleted or archived the data appropriately?

- How will you ensure that you still have access to any data or analysis you require that is due to be deleted or archived? Will it be in a machine readable format? A machine readable format is better in case this data needs to be analysed or processed again.

- How will you ensure that the vendor's access to any part of your organisation's infrastructure - for example, a virtual private network (VPN) service - is revoked in full?

Procurement and delivery

9. IS YOUR PROCUREMENT PROCESS FAIR, TRANSPARENT AND COMPETITIVE?

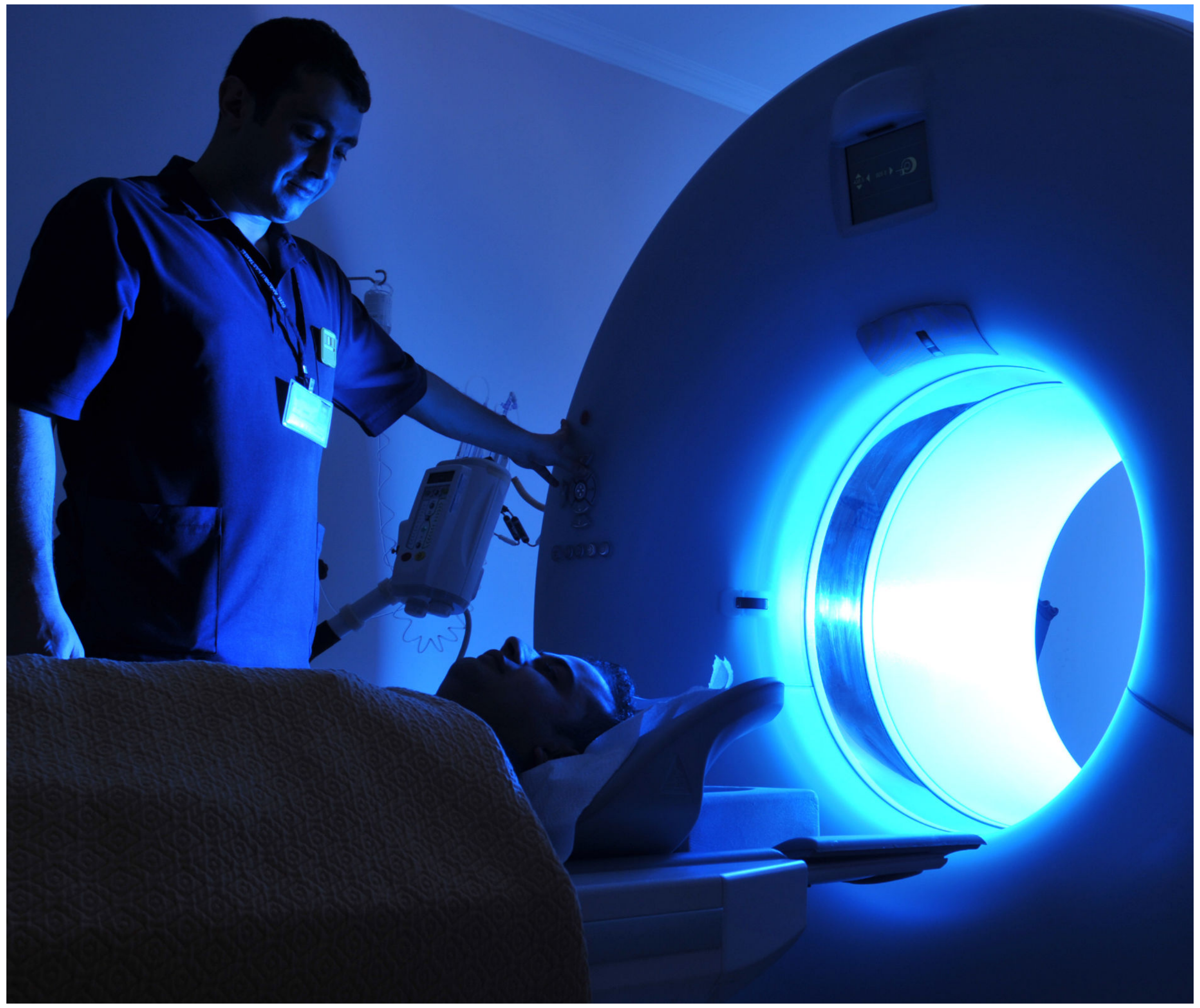
Like any technology, AI products need to be purchased on the basis of recognised public procurement principles that promote fairness, transparency and competition. (This applies to off-the-shelf products; early-stage research partnerships come with more flexibility).

Engaging the market early on may identify new potential vendors, level the playing field and help vendors understand what buyers need. At the same time, you should be clear about and document your justification for talking to and inviting specific vendors to bid. Establishing a competitive process may seem difficult in a nascent market such as AI. However, competition is a powerful tool for buyers, as well as being critical to promoting fairness and transparency.

Guidelines for AI procurement - published by the Office for AI - set out specific points of good practice, which include:

- Focusing on the challenge, rather than a specific solution
- Involving multidisciplinary teams in the procurement
- Determining in advance if and how you will share data with vendors for the purpose of the procurement exercise
- Highlighting known limitations of your organisation's data in the invitation to tender, and requiring tenderers to describe how they would address these shortcomings
- Making ethical considerations part of your evaluation criteria for proposals
- Making knowledge transfer and training a requirement of vendors in the invitation to tender

A vendor may approach your organisation offering their product for free - perhaps as an introductory sales technique, or perhaps as an opportunity for them to learn more about the product's performance in the real world. In such a situation, be careful to remain compliant with procurement requirements. Any instance you enter into a contractual agreement with a vendor - even where the product is offered for free - is considered to be a procurement. You should also consider whether or not you would be able to enter into an open competitive procurement at the end of the free offer period. If you were to be locked-in to the existing product, this would be problematic.



10. CAN YOU ENSURE A COMMERCIALLY AND LEGALLY ROBUST CONTRACTUAL OUTCOME FOR YOUR ORGANISATION, AND THE HEALTH AND CARE SECTOR?

Commercials

This guide is not intended to provide a comprehensive treatment of commercial contracting. However, there are some key points to consider. First and foremost, are you clear about exactly what you are procuring? Is it a lifetime product? Is it a licence? What is the accompanying support package? You should set out a clear specification in your invitation to tender, and establish a service level agreement as part of your contracting process, to secure the quality, availability, flexibility and performance that you need from the vendor.

You should make sure that the financial arrangements you are establishing are sustainable in the long-term. It is worth noting that there are a variety of payment mechanisms available - including stage payments and outcome-based specifications - which may transfer more of the risk from the buyer to the vendor where this is appropriate.

You should ensure that there are provisions in place for contract termination and handover to another supplier. You should also consider if you need provision for a change in ownership of the supplier - this is a plausible scenario for AI start-up companies.

In principle, your contracts should be as open as possible. Whilst confidentiality clauses are often invoked to prevent disclosure of commercially sensitive information, this can be detrimental to public trust. You can establish the importance you attach to commercial transparency by stating early on in the procurement exercise that you will need to publish contracts, or, at the very least, details of your commercial partnership.

Intellectual property

You should be 100% clear on plans for the rights of each party over personal data - including data generated by the partnership - at the outset.

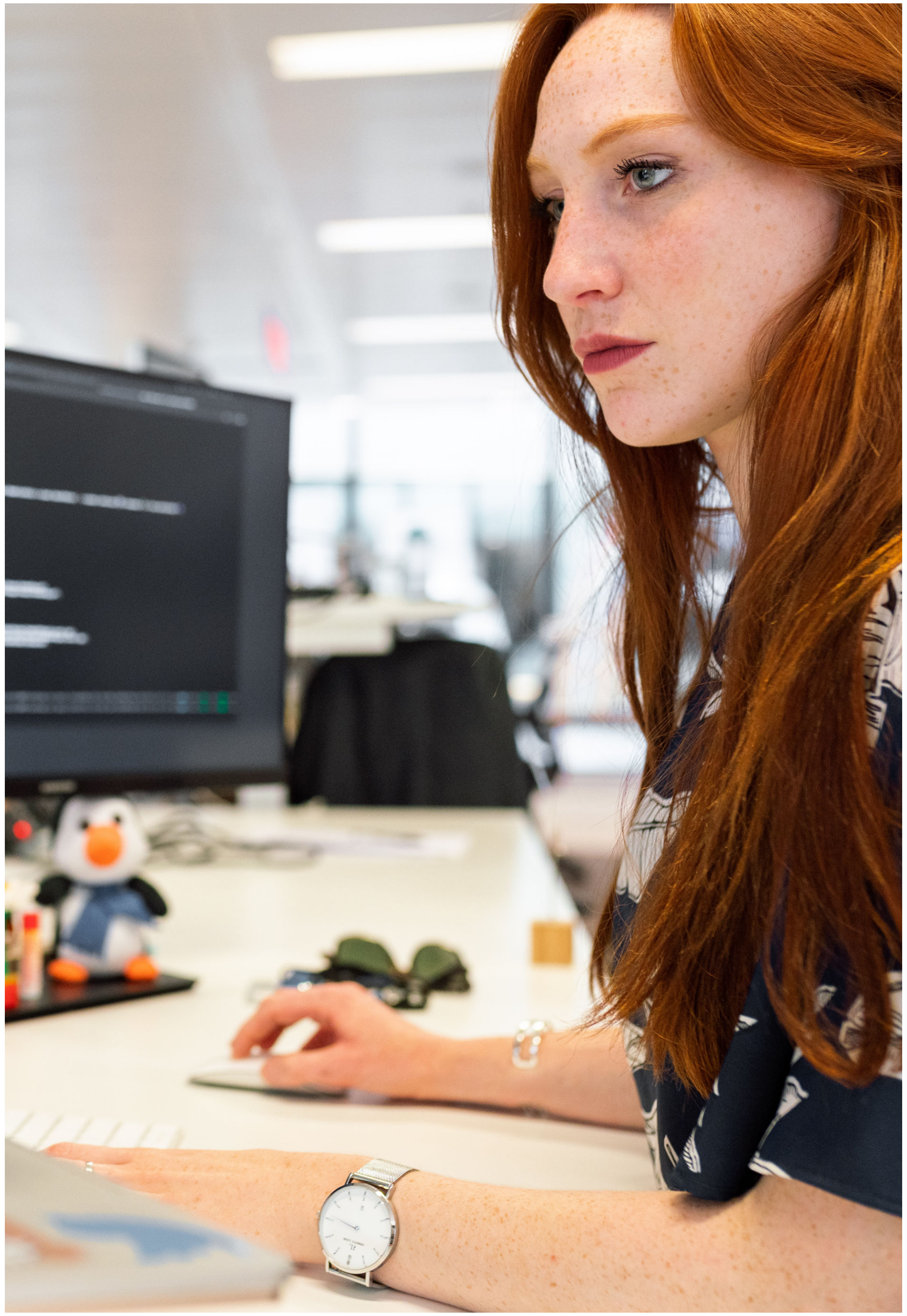
NHS and partner organisations entering into agreements involving data must consider the DHSC's five guiding principles for realising benefits to patients and the NHS where data underpins innovation.

These principles are designed for the NHS specifically. Whilst the underlying philosophy of these principles also offers good practice to the care sector, the less centralised structure of the care sector and the varied nature of data controlling organisations within it mean that the principles would need to be adapted.

Principle 2 requires that NHS organisations ensure agreements are 'fair', including recognising and safeguarding the value of the data that is shared and the resources which are generated as a result. You should note that in any situation where your organisation sends back data to the vendor - whether for the purpose of auditing the product, re-training the model, or potentially developing a new product - you may be contributing to the creation of intellectual property.

Principle 3 requires that data sharing agreements do not undermine the ability of the NHS to maximise the value or use of NHS data. This includes being careful not to sign up to conditions that limit any benefits which accrue from being applied at a national level. In addition, agreements should not undermine the wider NHS digital architecture, which in practice means emphasising the need for open standards and interoperability.

You should also be mindful about the implications of your relationship with a vendor where you might support them to create a more sophisticated version of an early product they offered to you at discount, which they go on to offer to other organisations at a much higher cost.



You should take advice early on to ensure your organisation is addressing these principles in its commercial negotiations. NHSX's Centre for Improving Data Collaboration can offer tailored guidance to NHS organisations - please contact the Centre via the AI Lab at improvingdatacollaboration@nhsx.nhs.uk. The Centre for Improving Data Collaboration will publish a detailed policy framework later in 2020, building on the five guiding principles.

Liability

Liability issues should not be a barrier to adoption of effective technology. However, it is important to be clear on who has responsibility should anything go wrong. Product liability is therefore an important issue to address at contracting stage:

- Does the contract provide any indemnities from the vendor and are they clearly set out?

- Is it clear what is considered as product failure versus human error in using the product? The concept of 'off-label use' – i.e. use of a product that is not in line with its intended use as described in the manufacturer's instructions – also needs to be clearly delineated in relation to medical devices.

- What is the extent of cover your own indemnifier or insurer can provide in the event of product failure or human error? Do you need to purchase additional cover or extend existing cover?

In the case of data protection, the ICO's guidance on Responsibilities and liabilities for controllers using a processor stipulates that a data Controller is primarily responsible for its own compliance but also for ensuring the compliance of its data Processors. Bear in mind that a Controller is expected to have technical and organisational measures in place to reduce the likelihood of a data breach - the Controller will be held accountable if they have not done this. Patients and service users are likely to hold accountable their health and care setting, rather than the technology vendor, and will frame their complaints accordingly.

Acknowledgements

The NHSX AI Lab is grateful to the following people who contributed to the development of this guide:

Natalie Banner, *Understanding Patient Data*

Nick Barlow, *University Hospitals Birmingham NHS Foundation Trust*

Andrew Boxall, *Western Sussex Hospitals NHS Foundation Trust*

Diarmaid Crean, *Sussex Community NHS Foundation Trust*

Rishi Das-Gupta, *Royal Brompton & Harefield NHS Foundation Trust*

Daniel de Rozarieux, *NHS England/Improvement*

Alastair Denniston, *University Hospitals Birmingham NHS Foundation Trust*

Natalia Domagala, *Department for Digital, Culture, Media and Sport*

Jacques Du Preez, *Psephos Biomedica*

Hacer Evans, *NHS Digital*

Greg Fletcher, *NHS Digital*

Nicola Forshaw, *Lifeways*

David Grainger, *Medicines & Healthcare products Regulatory Agency*

Maddy Griffiths *and wider team at the Information Commissioner's Office Innovation Hub*

David Gurusinghe, *NHS Resolution*

Hugh Harvey *and team, Hardian Health*

Shaid Hussain, *NHSX Information Governance team*

Stefan Janusz, *Office for Artificial Intelligence*

Hardeep Jhutty, *NHSX Centre for Improving Data Collaboration*

Chris Johnson, *Royal Papworth Hospital NHS Foundation Trust*

Monica Jones, *Health Data Research UK*

Tim Jones, *University Hospitals Birmingham NHS Foundation Trust*

Tom Jones, *General Medical Council*

Amrita Kumar, *Frimley Health NHS Foundation Trust & Shuri Network*

Xiao Liu, *University Hospitals Birmingham NHS Foundation Trust*

Rizwan Malik, *Bolton NHS Foundation Trust*

Richard Mitchell, *Sherwood Forest Hospitals NHS Foundation Trust*

Catherine Pollard, *NHSX Centre for Improving Data Collaboration*

Luke Readman, *NHS England/Improvement*

Emma Rush, *NHSX Data Policy team*

Danny Ruta, *Guy's and St Thomas' NHS Foundation Trust*

Haris Shuaib, *Guy's and St Thomas' NHS Foundation Trust*

Paul Simpson, *Surrey and Sussex Healthcare Trust*

Peter Smith, *Procurement Excellence*

Tom Stocker, *Care Quality Commission*

Peter Thomas, *Moorfields Eye Hospital NHS Foundation Trust*

Katie Thorn, *Digital Social Care*

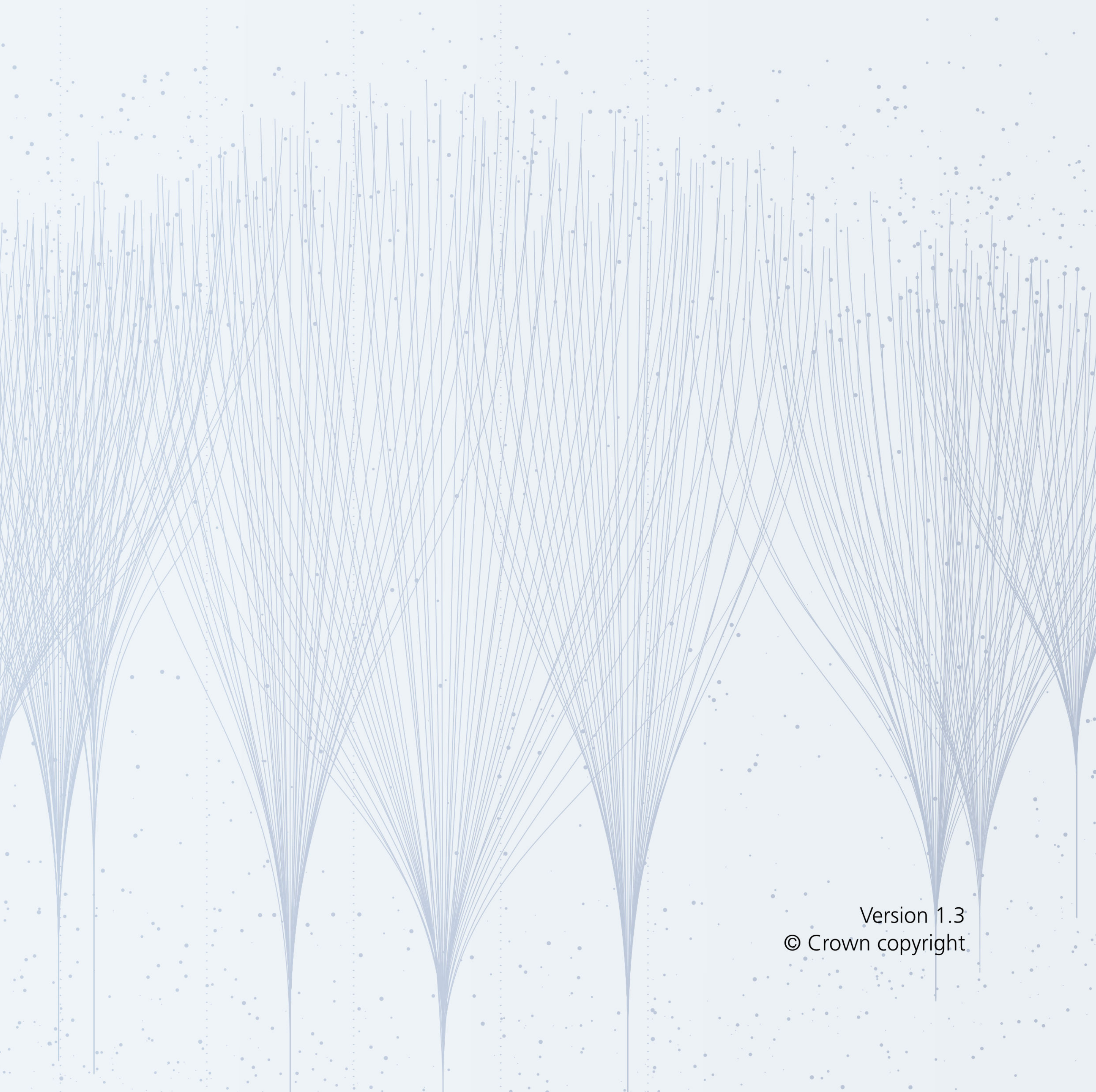
James Wilson, *Office of the National Data Guardian*

Nina Wilson, *NHSX Clinical Specialist*

FURTHER INFORMATION

There are many materials published by public bodies to help guide you through this process. Several key resources are listed below. The NHSX AI Lab can answer queries as well. You can contact the Lab at ailab@nhsx.nhs.uk.

- Artificial Intelligence: How to get it right (NHSX)
- A guide to using artificial intelligence in the public sector (GDS and OAI)
- Code of conduct for data-driven health and care technology (DHSC)
- Data Ethics Framework (DCMS)
- Data Protection Impact Assessments and AI (ICO)
- Evidence Standards Framework for Digital Health Technologies (NICE)
- Explaining Decisions made with AI (ICO and The Alan Turing Institute)
- Foundations of Fairness (Understanding Patient Data)
- Guidelines for AI Procurement (OAI and the World Economic Forum)
- Human Bias and Discrimination in AI Systems (ICO)



Version 1.3