Medicines & Healthcare products Regulatory Agency

National Commission into the Regulation of AI in Healthcare

Report on the research and engagement programme for the National Commission into the Regulation of AI in Healthcare

Contents

Foreword 4
Executive Summary 6
Introduction 8
Cross-cutting insights from the evidence 12
Next Steps 16
1. Call for Evidence 17
2. Deliberative Research 21
2.1 National Voices Online Focus Groups 21
2.2 Public deliberation on the regulation of AI in healthcare 25
3. Stakeholder Engagement 27
3.1 'Ask Me Anything' Open Public Meeting 27
3.2 Professional Bodies Roundtable Discussions 30
3.3 Post-Market Surveillance and Liability Roundtable with techUK 34
3.4 Liability Roundtable with Ada Lovelace Institute 38
4. Survey and Polling 41
4.1 Public attitudes to technology and AI 41
4.2 Ipsos' UK Knowledge Panel 43
5. Evidence Synthesis 46
5.1 Public attitudes to AI in healthcare in the UK 46
5.2 Categorising harms on AI in healthcare 48
6. AI Airlock Sandbox Programme 51
Ongoing work on liability in healthcare AI 55
Acknowledgements 56
Glossary of terms 57

© Crown copyright 2026

Open Government Licence

Produced by the Medicines and Healthcare products Regulatory Agency.
www.gov.uk/mhra

You may re-use this information (excluding logos) free of charge in any format or medium, under the terms of the Open Government Licence. To view this licence, visit http://www.nationalarchives.gov.uk/doc/open-government-licence or email: psi@nationalarchives.gsi.gov.uk.

Where we have identified any third-party copyright material you will need to obtain permission from the copyright holders concerned.

The names, images and logos identifying the Medicines and Healthcare products Regulatory Agency are proprietary marks. All agency's logos are registered trademarks and cannot be used without the agency's explicit permission.

Foreword

Healthcare has always evolved alongside advances in science and technology. From the discovery of antibiotics to the development of modern imaging, each generation has faced the challenge of understanding how new innovations can be used to improve care while maintaining the trust and confidence of the people they are intended to serve.

AI is proving to be one of those defining moments. Already, these technologies are beginning to support earlier diagnosis, more personalised treatment and new ways of delivering healthcare. Their potential is significant. Yet alongside that opportunity come important questions. How should these technologies be used? What safeguards are needed? How can innovation be encouraged while maintaining public trust? And how can regulation keep pace with change while continuing to protect patients?

As Chair and Deputy Chair of the National Commission into the Regulation of AI in Healthcare, we believe these are questions that deserve broad and thoughtful discussion. AI is not simply another technological development. It has the potential to affect many aspects of healthcare and, ultimately, many of the people who use and deliver it.

The Commission was established to help inform the future regulation of AI in healthcare. It brings together expertise from across healthcare, science, technology and regulation. From the outset, we recognised that expertise alone cannot answer every question that AI raises. Understanding what matters to patients, healthcare professionals and the wider public is equally important.

That is why listening has been such an important part of the Commission's work. Decisions about the future regulation of AI in healthcare must reflect both what is technically possible and the experiences, expectations and values of the people who will ultimately be affected by them.

As AI continues to develop at pace, there is also a broader question about how we respond to that change. The future of AI in healthcare should not simply happen to patients, healthcare professionals and the public. It should be shaped with them. Listening has therefore been about more than gathering evidence. It has been about helping to ensure that the priorities and expectations of the people who use and deliver healthcare help inform the direction of travel and how we get there.

The work reflected in this report has helped ensure that the conversation extends to as many people as possible, bringing together experiences from all parts of society and across the UK. It has provided an opportunity to hear from a broad range of perspectives and lived experiences, all of which are informing the Commission's deliberations.

As we consider the future regulation of AI in healthcare, it is striking how successful innovation is built on trust. The innovations that have transformed healthcare most successfully are those that have earned the confidence of patients, healthcare professionals and the public. The challenge ahead is to ensure that innovation continues to develop in ways that are safe, proportionate and worthy of that trust, while also enabling patients and the public to benefit from timely access to valuable new technologies.

The insights gathered through this work will play an important role in informing the Commission's recommendations to the MHRA. They remind us that questions about the future regulation of AI are ultimately questions about people: how we maintain trust, protect patients and ensure that innovation serves the public good.

We are grateful to everyone who has contributed to this work. Publishing this report marks an important milestone, but it is not the end of the conversation. As the Commission continues its work, engagement and evidence will remain central to our deliberations, helping us to test, challenge and refine our thinking as we develop recommendations for the MHRA, health system, and wider government on the future regulation of AI in healthcare.

Professor Alastair Denniston
Chair, National Commission into the Regulation of AI in Healthcare
Professor of Regulatory Science and Innovation at the University of Birmingham
Honorary Consultant Ophthalmologist, University Hospitals Birmingham NHS Foundation Trust
Executive Director, Centre of Excellence for Regulatory Science in AI & Digital HealthTech (CERSI-AI)

Professor Henrietta Hughes
Deputy Chair, National Commission into the Regulation of AI in Healthcare
Patient Safety Commissioner for England
General Practitioner and Visiting Professor, Institute of Medicine, University of Greater Manchester

Executive Summary

Artificial intelligence (AI) is changing how healthcare is delivered, with potential to improve patient outcomes, support healthcare professionals and increase the efficiency of the NHS. The MHRA established the National Commission into the Regulation of AI in Healthcare in September 2025 as an independent advisory body to bring together expertise from across sectors and inform future regulatory approaches, with recommendations due in late summer 2026.

The UK has a clear ambition to lead in the safe and effective use of AI in healthcare. Delivering this ambition will require a regulatory framework that supports timely access to innovation while maintaining public trust, professional confidence and effective oversight once technologies are in use.

This report draws together a wide range of research and engagement commissioned by the National AI Commission, bringing together perspectives from patients and the public, healthcare professionals, industry and system partners. It combines deliberative research, stakeholder engagement, surveys, evidence reviews and regulatory sandbox activity to provide both in-depth insight and broader system perspectives. Together, these approaches provide a comprehensive evidence base on how AI is used in practice and the challenges it presents for regulation.

Across the evidence gathered through the research and engagement activities, there is broad recognition of the potential for AI to improve outcomes, support healthcare delivery and, in some cases, widen access to services. However, support for its use is conditional and varies depending on how it is used in practice and the level of risk involved. Safety, effectiveness and meaningful human oversight emerge as the foundation of trust, with confidence shaped by whether AI can be shown to improve outcomes in real-world settings.

A consistent finding across the evidence is that existing regulatory approaches are not well matched to systems whose behaviour may vary across contexts or change over time, creating a need for lifecycle-based oversight that extends beyond initial approval. Risks and outcomes are shaped not only by technical performance, but by how AI interacts with people, workflows and organisations in real-world settings.

Trust is closely linked to transparency, accountability and the ability to understand and challenge AI-supported decisions. Equity considerations, including bias and uneven access, are central to safety and legitimacy, while uncertainty around responsibility and liability remains a significant barrier to safe and effective adoption.

Taken together, the evidence suggests the central question is not whether AI should be used in healthcare, but what conditions are necessary for it to be used safely, effectively and legitimately. These conditions include proportionate scrutiny before deployment, clear standards for safety and effectiveness, meaningful human oversight, ongoing monitoring in use, and safeguards to support fairness and inclusion. Across stakeholders, there is a clear expectation that regulation should strike an appropriate balance, providing the safeguards needed to build trust and ensure safety, while remaining proportionate, enabling innovation and avoiding unnecessary barriers to beneficial adoption.

This report does not set out policy or final recommendations. Instead, it provides a transparent account of the evidence gathered through the National AI Commission’s work to date, forming the evidence base that will inform the Commission’s later recommendations. These will focus on how regulation can support both public confidence and timely access to beneficial innovation, in a way that remains credible and trusted, reflecting the dynamic nature of AI and the real-world environments in which it is used.

Introduction

Background

The UK Government has committed, through the 10 Year Health Plan for England and the Life Sciences Sector Plan, to establishing a modern regulatory framework for medical devices, including artificial intelligence (AI). AI presents significant opportunities to improve patient outcomes and health system productivity, but also introduces new challenges for safety, performance and oversight.

If an AI system is intended by the manufacturer to perform a medical purpose, it is likely to qualify as a medical device under the Medical Devices Regulations 2002 (MDR 2002) and be regulated accordingly[1]. Systems that do not meet this definition may still fall under other regulatory frameworks, such as data protection or consumer protection law. The Medicines and Healthcare products Regulatory Agency (MHRA) is responsible for regulating medical devices in the UK, ensuring they meet appropriate standards of safety, quality and performance, and overseeing ongoing post-market monitoring to identify risks as they emerge in real-world use.

AI is used across a wide range of healthcare applications, from clinical decision support to administrative tools and direct-to-consumer products, and not all are regulated in the same way. For those AI technologies that are regulated as medical devices, the level of oversight depends on the risk they pose. As adoption increases, challenges have emerged around qualification and classification of AI as a medical device (AIaMD), visibility of use across the health system, ongoing monitoring of performance and safety in diverse real-world environments, and the application of existing frameworks to AI systems that evolve over time, highlighting the need for approaches that better reflect how AI operates in practice.

The National Commission into the Regulation of AI in Healthcare was established in September 2025 as an independent, non-statutory advisory body to support the MHRA in ensuring that patients in the UK have timely access to safe and effective AI-enabled medical technologies.

The role of the National AI Commission is to bring together expertise from across the health system, industry, devolved nations and wider society to inform the development of a future regulatory approach that enables innovation while protecting patients. In doing so, the National AI Commission seeks to address the unique challenges posed by AI, while supporting its safe and effective adoption through proportionate and adaptable regulation.

[1] The MDR 2002 applies in Great Britain (England, Scotland and Wales). Medical devices in Northern Ireland are regulated under the relevant EU medical device regulations.

The National AI Commission is expected to publish its recommendations in late summer 2026.

To inform its work, a wide range of research and engagement activities was commissioned, bringing together evidence from patients and the public, healthcare professionals, industry and wider health system partners. This helps ensure the recommendations are grounded in both lived experience and system-level insight.

Activities have been delivered with a range of partners, including the Health Foundation as the research partner, National Voices through commissioned engagement with underrepresented perspectives, techUK and the Ada Lovelace Institute as partners delivering stakeholder roundtables, and SurveyOptic in supporting the delivery and analysis of the Call for Evidence.

Together, these activities provide a comprehensive evidence base on the opportunities, risks and practical challenges associated with AI in healthcare, and how these should inform future regulatory design, including how regulation can support both safety and timely access.

The MHRA has supported the National AI Commission's work by facilitating its research and engagement activities, alongside providing governance, secretariat and operational support. As the sponsoring organisation, the MHRA has overseen delivery of this work, including commissioning research activities and managing associated governance and data requirements.

The National AI Commission has been delivered within existing UK Government funding. Commissioners and working group members are unpaid, reflecting the Commission's advisory role.

Purpose of this report

This report presents a summary of the evidence gathered through the research and engagement undertaken. It provides an account of the methods used, the perspectives provided by stakeholders, and the key findings emerging from this work.

It also sets out the full range of issues identified through the Commission's research and engagement activities, providing a transparent and objective view of the current landscape. Not all of these issues are regulatory in nature or require regulatory solutions; some reflect wider system or research challenges. They are nonetheless included in full both for transparency and as a resource for researchers, policymakers and others considering where further work or collaboration may be needed.

This evidence represents one component of the National AI Commission's wider work and will be considered alongside additional analysis and broader policy considerations in

developing its final recommendations, including how to balance public trust, patient safety, innovation and access in a proportionate way. It does not represent government policy or set out final recommendations, nor does it seek to fully reproduce the evidence gathered through this work, although many of the activities described in this report will be published in full elsewhere.

Structure of this report

This work comprises a set of interrelated workstreams, grouped into six categories reflecting the different types of evidence and perspectives gathered. These categories also provide the structure for how findings are presented in this report.

Call for Evidence provides stakeholder perspective on the effectiveness of the existing regulatory framework, and the distribution of responsibility across AI system development and deployment.

Deliberative Research involves structured discussions with members of the public to understand their views and expectations around AI in healthcare.

Stakeholder Engagement brings together patients, professionals, industry and experts through targeted discussions, workshops and roundtables to explore practical challenges surrounding AI in healthcare.

Survey and Polling provide nationally representative quantitative data on public and workforce attitudes, awareness and preference relating to AI, offering a broad view of perceptions across groups over time.

Evidence Synthesis reviews existing research and literature to identify established knowledge, gaps and emerging themes relevant to AI in healthcare.

MHRA's AI Airlock Programme uses a regulatory sandbox to test AI as a medical device, generating insights into how regulatory frameworks operate in practice and where they need to adapt.

Each section in this report follows a consistent structure to make it easier to compare different types of evidence. Sections begin with key findings, followed by an overview of the

work and a description of the approach and scope. This is then supported by a summary of the discussion or findings, and a final section highlighting the key insights from the work.

Research and engagement timeline

The timeline shows the sequence of workstreams within the research and engagement activity, which began in December 2025 and is ongoing. Further activity is underway to develop the evidence base, including dedicated work on liability (see “Ongoing work on liability in healthcare AI”).

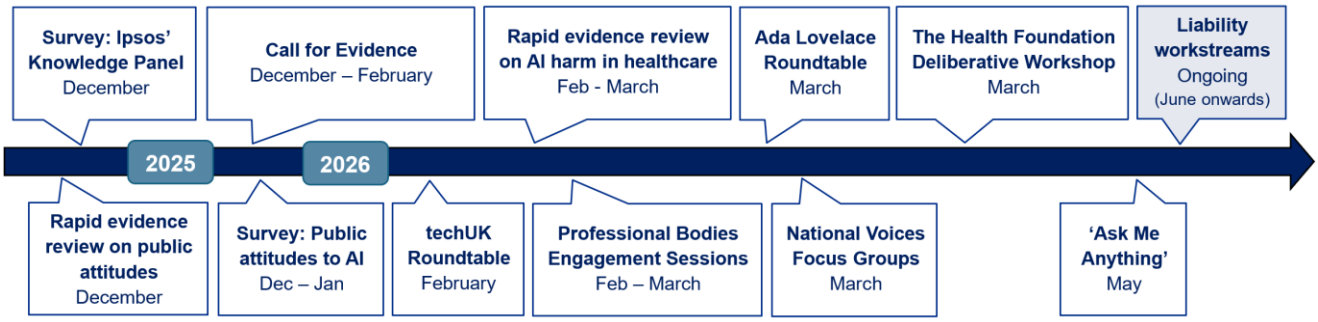

Figure 1. Research and engagement timeline

The figure excludes MHRA’s AI Airlock Programme, as the pilot phase was completed prior to the establishment of the National AI Commission, while phase 2 programme was delivered in parallel with the National AI Commission’s work. Insights from both phases are included in this report as part of the wider evidence.

Cross-cutting insights from the evidence

Across the research and engagement undertaken, a consistent set of system-level insights emerged. While perspectives differed across public, professional, industry and policy settings, there was broad alignment on the factors shaping whether AI can be used safely, effectively and with public confidence. Together, these highlight that the core challenge is not whether AI should be more tightly regulated, but how to create a regulatory environment that supports safety, trust and transparency while enabling beneficial innovation and timely access.

The right regulatory environment can enable trust and innovation

Across the evidence, stakeholders emphasised the importance of a regulatory environment that supports both patient safety and innovation. Regulation is seen not only as a mechanism for managing risk, but also as a means of building public confidence and enabling the responsible adoption of new technologies.

There is recognition that innovation in AI has the potential to improve patient outcomes, increase efficiency, and expand access to care. Realising these benefits depends in part on ensuring that regulatory approaches are proportionate, predictable and supportive of development and deployment in real-world settings.

At the same time, some evidence highlights the risks of over-regulation, including creating barriers to innovation, slowing access to beneficial technologies and services, and placing undue burden on developers and healthcare providers. This underscores the importance of striking an appropriate balance between enabling innovation and ensuring safety.

Support for AI is conditional, not automatic

Support for AI in healthcare is consistently conditional and closely linked to trust in how systems are used and governed in practice. Public and professional confidence varies depending on the use case, level of risk and degree of human involvement.

Lower-risk and administrative uses are generally viewed as more acceptable, while applications that influence diagnosis, treatment or access to care attract significantly greater caution. Across the evidence, confidence depends less on abstract support for innovation and more on whether AI can be shown to deliver safe, effective and beneficial outcomes in real-world settings.

Current regulatory frameworks are not well matched to AI

Across the research and engagement, stakeholders from across the system – including patients and the public, healthcare professionals, industry and policymakers – highlighted challenges in how existing regulatory frameworks fit the characteristics of AI systems.

Existing regulatory approaches were largely designed for fixed, well-defined products assessed at a single point in time. By contrast, AI systems are often adaptive, context-dependent, and shaped by how they are deployed, used and integrated into clinical workflows.

The evidence therefore highlights tensions between current regulatory assumptions and the realities of AI in practice, suggesting a need to consider how approaches might better reflect system behaviour over time.

Lifecycle oversight is important

Given these challenges, there is consistent support across stakeholder groups for complementing pre-market assessment with more sustained oversight. Evidence across workstreams points to the importance of ongoing monitoring, post-market surveillance, and the ability to respond to changing performance, emerging risks and unintended consequences.

This suggests the value of a regulatory approach that operates across the full lifecycle of an AI system, with mechanisms to identify, assess and act on risks once systems are in use.

AI risks are shaped by context and real-world performance

The evidence indicates that risks associated with AI cannot be understood solely through technical performance metrics. Outcomes are shaped by how systems interact with users, organisational processes and wider healthcare environments.

Factors such as human oversight, professional behaviour, workflow integration and data quality all influence how AI performs in practice. This highlights the importance of assessing AI not only in controlled settings, but also in the context in which it is used.

Trust depends on transparency, accountability and routes for challenge

Trust in AI is closely linked to visible and meaningful safeguards in practice. Across the evidence, there is a clear expectation that people should understand when AI is used, what role it plays, and how decisions can be questioned or challenged.

Transparency, accountability and routes for challenge are consistently identified as important parts of trustworthy deployment, including clear communication, defined responsibilities and mechanisms for raising concerns.

Equity and inclusion are core considerations

Concerns relating to bias, uneven performance, digital exclusion and differential access recur throughout the evidence base. These issues are consistently framed not only as ethical considerations, but as relevant to safety, quality and public confidence.

The evidence suggests that equity should be considered throughout the lifecycle of AI systems, including how they are designed, tested, evaluated and implemented, to help ensure that benefits are realised across different groups.

Responsibility and liability remain areas of uncertainty

Across the workstreams, responsibility for AI in healthcare is described as complex and not always clearly defined. There is no consistent view on how liability should be allocated across manufacturers, healthcare providers and professionals, particularly where systems evolve over time and involve multiple actors.

This lack of clarity is seen to create challenges for accountability, introduces uncertainty for professionals, and potentially affects public confidence. Greater clarity in approaches to responsibility and liability is therefore likely to support safer and more effective adoption.

Overall implication

Taken together, these insights point to the importance of a regulatory approach that is risk-based, proportionate and adaptive, supporting both safety and innovation in the use of AI in healthcare.

The evidence suggests that the central challenge is not whether AI should be used in healthcare, but under what conditions its many different applications can be used safely, effectively, fairly and with public confidence. These conditions include appropriate scrutiny before deployment, clear standards for safety and effectiveness, meaningful human oversight, ongoing monitoring in use, and safeguards to support fairness and inclusion.

Across stakeholders, there is a clear expectation that regulation should strike an appropriate balance, providing the safeguards needed to build trust and ensure safety, while remaining proportionate and enabling innovation, adoption and timely access to beneficial technologies. This reflects the importance of ensuring that patients and the public can benefit from advances in AI, including improvements in care, access and system efficiency, without creating unnecessary barriers to development or adoption.

Taken together, this points towards a regulatory approach that is more agile, lifecycle-based and responsive to real-world use, while remaining risk-based and proportionate to the varying levels and types of risk presented by different AI-enabled technologies, and avoiding unnecessary barriers where risks are lower.

Next Steps

This report marks an important stage in the National AI Commission's work, providing a transparent account of the evidence gathered to date and setting the foundation for its next phase.

Over the summer, the National AI Commission will continue to engage with stakeholders to test and refine its emerging recommendations, including on complex issues such as liability, accountability and implementation in practice. This phase will focus on how a future regulatory approach can provide the safeguards needed to build trust and ensure safety, while remaining proportionate, enabling innovation and supporting timely access to beneficial technologies.

The National AI Commission expects to publish its final recommendations in late summer 2026. These will build on the evidence in this report, alongside continued engagement and analysis of the work undertaken by Commissioners and working groups since September 2025, to translate the findings into recommendations on how the future regulation of AI in healthcare can be safe, effective, proportionate to risk and trusted, while supporting both public confidence and timely access to beneficial innovation.

Following publication of the National AI Commission's recommendations, the Government - led by the MHRA - will consider the National AI Commission's advice and outline its response and next steps, helping to shape the future direction of AI regulation in healthcare.

1. Call for Evidence

A Call for Evidence is a formal process used by the UK government or Parliament to gather information, data and views from stakeholders. It invites responses to a set of targeted questions on a specific issue, with the aim of building a robust evidence base. This evidence is then used to inform and shape future policy development and decision-making.

Delivery partner	MHRA; SurveyOptic
Format	Online Call for Evidence
Date(s)	19 December 2025 – 02 February 2026
Output(s)	Summary of findings published 11 June 2026

Key Findings

- All stakeholder groups support significant reform, but not a complete redesign
- Views on impact on innovation diverge across groups,
- Strong support for risk-based, lifecycle regulation and improved processes
- Responsibility and liability remain unclear, with no clear consensus

Overview

This Call for Evidence gathered views from a wide range of stakeholders on how AI in healthcare should be regulated in future. It used a structured survey combining quantitative and qualitative questions to capture both broad perspectives and detailed insights.

Approach and Scope

The Call for Evidence was conducted as an online survey, combining fixed-response and free-text questions. It ran for approximately six weeks (between 19 December 2025 and 2 February 2026) and received 761 unique responses. The launch and promotion were supported by the government, the MHRA, and members of the National AI Commission.

The Call for Evidence invited views on the adequacy of the current regulation of AI in healthcare, how it could be improved, how safety should be monitored in use and how responsibility and liability should be allocated.

Summary of findings

Appetite for reform

Across healthcare professional, healthcare provider and industry respondent groups, a majority supported significant reform of the regulatory framework, though not a complete overhaul. In contrast, public views were more mixed, with similar proportions supporting significant reform and a complete overhaul.

Diverse opinions on the current regulatory framework

There was a range of views on the current regulatory framework, including its impact on both patient safety and enabling innovation. Most patients and public respondents (77%) described the current framework as 'somewhat loose' or 'too loose', while most industry respondents (65%) described it as 'somewhat restrictive' or 'too restrictive' and creating barriers to innovation. The views of healthcare providers and healthcare professionals, responses were more evenly distributed, with 38% health providers and 52% health professionals stating that the current regulatory framework is “somewhat loose” or “too loose”, and 48% health providers and 37% health professionals stating that the current regulatory framework is “somewhat restrictive” or “too restrictive”.

Priorities for improving the framework

Respondents proposed a range of improvements, including:

- clearer risk-based classification of AI systems
- stronger clinical evidence requirements
- enhanced post-market surveillance
- greater coordination across regulatory bodies

Some respondents called for improvements to be made to regulatory processes, including:

- facilitating manufacturers’ access to regulatory documentation
- further guidance on version control and maintaining regulatory documentation
- streamlined regulatory approval processes

Some stakeholders discussed the need for better coordination between regulators and improvements to be made to NHS procurement processes, noting that fragmentation can slow adoption and make it harder to comply with regulation.

A group of respondents commented on the need for clearer guidance on classification and safety standards for AI systems. This would help health systems better understand what types of AI systems are medical devices and how risks should be managed.

Responsibility and governance

There was broad agreement by respondents that current arrangements for responsibility are unclear. Many respondents called for clearer, shared responsibilities to support coordination and safer deployment of AI systems.

Some stakeholders emphasised the importance of stronger governance structures to oversee how AI systems are developed, deployed and monitored across their lifecycle. Others pointed to the need for more clearly defined clinical safety responsibilities, including responsibilities for monitoring performance, managing risks and ensuring appropriate safeguards are in place.

Many responses also discussed the need to ensure that access to, and use of, healthcare data is trustworthy. These respondents noted that undue access to healthcare data could affect patient confidence in the use of AI in healthcare and reduce their willingness to engage with AI-enabled healthcare settings.

Liability and accountability

There was a consistent view that clearer legal and regulatory frameworks are needed to better define liability and establish more effective accountability mechanisms in practice. This was also commonly linked to a recognition that the safe use of AI health technologies depends on many different individuals and institutions appropriately fulfilling their roles and responsibilities, and that considerations of liability and accountability should reflect this.

When respondents considered the question of liability, they stressed the different responsibilities of the key roles and institutions. For example, some respondents pointed to the healthcare provider who deploys the AI system, and their responsibility to ensure that the system is being deployed safely in their institution. Others highlighted the role of the professional and stating that they should be responsible for the final decision regarding an individual patient's care. Others emphasised the role of the manufacturer, and that they have responsibility to ensure that their product is not only safe at the point of sale, but also that it continues to operate safely in the health setting it is deployed into.

What the Call for Evidence highlighted

The Call for Evidence shows strong stakeholder appetite for reform, with broad agreement that current arrangements do not yet provide a sufficiently future-ready framework for AI in healthcare. While views differ on the balance between innovation and regulatory checks, there is consistent emphasis on the need for clearer roles and responsibilities for different actors, stronger lifecycle governance and more workable approaches to accountability and liability in practice.

2. Deliberative Research

Deliberative research is a structured approach to engaging members of the public to explore their views and expectations on a particular issue, such as AI in healthcare. It typically involves informed discussion over time, allowing participants to consider evidence, reflect, and reach considered views. This provides deeper insight into public perspectives and how opinions evolve when people are given time and information to engage with the topic.

2.1 National Voices Online Focus Groups

Delivery partner	National Voices
Format	Online workshops
Date(s)	March 2026
Output(s)	Report publication by National Voices expected on 15 June 2026

Key Findings

- Trust depends on clear information, human oversight and safeguards in practice
- Regulation is expected to ensure safety, transparency and accountability
- AI should support – not replace – clinical judgement
- Strong concerns remain about fairness, exclusion and uneven access

Overview

This targeted engagement explored how underserved groups view the regulation of AI in healthcare. It used facilitated online workshops and real-world examples to examine what would be needed to build trust and support inclusive governance. The full report will be published by National Voices in June 2026.

Approach and Scope

Commissioned by the MHRA and delivered by National Voices, this work focuses on engaging underserved groups. A total of 26 participants were recruited via the Voluntary,

Community and Social Enterprise (VCSE) partners: 11 young people (16-25); 9 unpaid carers; and 6 people living with a learning disability.

Three online workshops (two hours each) were delivered to reduce participation barriers and to allow engagements in familiar environments. Sessions were co-facilitated by National Voices and VCSE partners, combining research expertise with community insights. Accessibility support was tailored to each group and participants received an honorarium.

Discussions were structured around three use cases (AI scribe, AI triage and booking system and diagnostic AI), exploring key trade-offs, regulatory expectations and potential safeguarding mechanisms.

Insights from the engagement

Participants brought varied familiarity and experience with AI, but a shared uncertainty about its reliability and use in healthcare. While some participants described everyday use and enthusiasm, others expressed scepticism or anxiety, particularly where AI was introduced without clear explanation.

Across all groups, participants emphasised the need for clear, plain-language information, confidence underpinned by evidence that AI is used to improve care, and the ability to speak to a person and opt out where needed.

Regulation was viewed as the mechanism for making AI in healthcare trustworthy in practice. Participants expected it to be meaningful to patients, enabling people to understand what is happening and what safeguards apply.

Across groups, there was a shared expectation that AI should remain under human control and support, rather than replace, professional judgement.

Concerns were raised about commercial incentives, use of health data, and the potential erosion of the human aspects of care, particularly in high-stakes or sensitive contexts. For some groups, particularly people with learning disabilities or those with complex care needs, this was closely linked to expectations around dignity, autonomy and accessible information.

Participants emphasised that governance should be shared, independent and accountable, involving multiple stakeholders, including clinical, regulatory and voluntary sector perspectives, alongside clear and navigable accountability.

Across groups, participants highlighted the importance of ensuring that responsibility is clear in practice, and that governance arrangements support transparency, oversight and trust in the use of AI in healthcare.

Trade-offs

Analysis identified key themes of consensus and divergence across four core trade-offs:

Safety vs Innovation

Participants supported innovation in principle and acknowledged the need to drive efficiencies, but consistently prioritised safety. Acceptability of AI depended on the consequences of error, the ability to detect and correct mistakes, and who is affected. Lower risk administrative uses were generally viewed as more acceptable than tools that influence diagnosis, triage, or treatment.

- Young people were concerned about “shortcuts” in the pressured system, and cautious about tools influencing clinical decisions.
- People with a learning disability set clear ‘red lines’ around high-stakes decisions and expected much higher threshold for acceptable error and clear evidence of safety.
- Unpaid carers emphasised limits of AI in capturing emotional context and nuance, highlighting human judgement as essential.

Speed vs Scrutiny

Participants acknowledged potential efficiency gains but emphasised that regulation and scrutiny must keep pace with deployment. There was consistent support for rigorous validation before use, active post-market surveillance, and clear mechanisms to address or halt systems where harms emerge. Participants expected scrutiny to increase with risk level.

- Young people expected governance to be in place before widespread adoption and expressed concerns about accountability. They also supported learning from errors to improve the system without increasing access to identifiable data.
- People with a learning disability emphasised the need for clear proof of safety to establish trust and highlighted that trust can be easily undermined by serious mistakes.
- Unpaid carers focused on practical approaches to scrutiny in real-world use, including clear benchmarks, oversight by healthcare professionals, and stop mechanisms. Views differed on the acceptable balance between speed and reassurance.

Equity vs Efficiency

Equity and inclusion were seen as essential, with the risk of exclusion seen as a patient safety risk. Key concerns included digital exclusion, communication barriers, and uneven rollout.

- Young people emphasised the importance of inclusive testing prior to deployment, particularly across different demographics.
- People with learning disabilities highlighted accessibility barriers in digital-first systems.
- Unpaid carers cautioned against a two-tier system, stressing the need for alternative pathways so efficiency gains don't disadvantage those with complex needs or limited digital access.

Transparency vs Commercial interests

Transparency and informed consent were key to trust, with participants seeking clarity on AI use, its role in care, and available choices. Accountability for harm was prioritised over technical explainability.

- Young people raised concerns about undisclosed AI use and private sector involvement, highlighting strong expectations around data protection and clear consent.
- People with learning disabilities emphasised the need for plain-language explanations and simple, accessible consent processes, linking transparency to autonomy and choice.
- Unpaid carers focused on risks around the use of data beyond care and third-party access, emphasising the need for greater visibility of how information is handled.

What this work highlighted

Across participant groups, trust in AI was shaped not only by technical performance, but by how systems are explained, governed and experienced in practice. Participants consistently expected clear, accessible information, meaningful choice and the ability to speak to a person, with these issues especially important for groups who may otherwise face barriers to participation or care.

Participants also emphasised that AI should support, not replace, professional judgement, with safety, scrutiny and accountability prioritised over speed or efficiency, particularly in higher-risk contexts.

Equity, digital exclusion and uneven rollout were recurring concerns across groups, alongside expectations that governance should be shared, transparent and clearly accountable. Together, these findings highlight that inclusive deployment is central to public trust, not an optional add-on.

2.2 Public deliberation on the regulation of AI in healthcare

Delivery partner	The Health Foundation
Format	Workshops
Date(s)	Research conducted between March and April 2026
Output(s)	Report publication expected late Summer 2026

Key messages heard by the Commission

From the perspective of participants in this research, an ideal regulatory framework for AI in healthcare would prioritise accuracy, effectiveness and human oversight. It should be risk proportionate and not rely solely on pre-market authorisation.

An advance copy of the report sharing the findings of this research was shared with the AI commission under embargo. The final report is expected to be published in September 2026.

Overview

As part of its role as research partner to the Commission, the Health Foundation conducted a programme of deliberative research to explore public views on the regulation of AI in healthcare. The research was designed to go beyond what survey data can tell us – providing in-depth, considered public perspectives on the principles and trade-offs that should underpin a new regulatory approach.

Approach and Scope

The deliberation engaged 75 members of the public across three locations in England and Wales, using a structured three-workshop process. Participants were informed by expert input, explored real-world AI use cases and regulatory approaches, and deliberated on the conditions needed to build public trust in AI regulation. The research was conducted in partnership with Ipsos UK.

The deliberation focused on four core questions: what the public believe the primary aims of AI regulation in healthcare should be; how they would prioritise and trade off outcomes such as accuracy, speed and equity; what makes regulatory approaches feel trustworthy; and

what conditions and safeguards are needed to build public confidence in the use of AI in healthcare.

Summary of findings

The core output was the development of a series of principles, which participants believed should underpin regulation of AI in healthcare. These principles and the underlying context for each will be set out in full in a report to be published by the Health Foundation, expected in September 2026.

Two important messages heard by the Commission were: 1) that participants would support regulation which is robust but risk proportionate, with greater scrutiny for higher-risk AI tools and more streamlined routes for those which are lower risk; and 2) that an effective regulatory framework would include continuous monitoring of tools to ensure ongoing safety and performance over time.

How the Commission integrated the findings of this research

Ensuring that the voices and perspectives of patients and the public were reflected during policy development was a key priority for the National AI Commission into the Regulation of AI in Healthcare.

Representatives of the Commission were involved in designing this research and were invited to attend each in-person deliberative discussion as observers.

An advance copy of the report sharing the findings of this research was shared with the AI commission under embargo, and members of the core commission were briefed on the findings by the Health Foundation and Ipsos.

3. Stakeholder Engagement

Stakeholder engagement is a structured process to gather insights from across the health and care system, drawing on the experience and expertise of those involved in developing, regulating and using AI. It brings together a range of perspectives to explore challenges and opportunities, and to identify practical considerations for implementation.

This provides system-level insight into how AI operates in practice, including operational, regulatory and professional issues that may not be captured through research or public engagement alone.

3.1 'Ask Me Anything' Open Public Meeting

Delivery partner	MHRA
Format	Webinar
Date(s)	20 May 2026 at 9.30 AM
Output(s)	Recording of the session published 09 June 2026

Key Findings

- Participants focused on safety, transparency, accountability and human involvement in AI-enabled healthcare
- There was significant interest in how AI systems will be regulated, monitored and governed in practice, particularly as systems evolve over time
- Participants highlighted the need to balance innovation with robust safeguards, particularly in higher risk clinical settings
- There was a strong emphasis on clear communication, transparency and ongoing oversight to maintain trust

Overview

This open public meeting provided a moderated forum for patients, carers and the wider public to engage directly with the National AI Commission's leadership.

Approach and Scope

This open public meeting was held on 20 May 2026 as a one-hour, publicly accessible webinar hosted by the MHRA. The ‘Ask Me Anything’ (AMA) session provided patients, the public and stakeholders with an opportunity to engage with the National AI Commission’s leadership, supporting openness and transparency around the Commission’s work.

The event brought together participants from patient groups, charities, and wider public and stakeholder networks with the Commission’s Chair and Deputy Chair, and the MHRA’s Chief Executive, to discuss the Commission’s priorities, its approach to AI regulation in healthcare, and key themes emerging from the wider research and engagement work. The event was open to all, with outreach designed to encourage broad and diverse participation.

A total of 928 individuals registered for the event, with 363 participants joining live. The session combined pre-submitted questions with moderated live Q&A to enable real-time engagement while ensuring responses remained accurate and accessible. In total, 180 questions were received, including 79 submitted in advance and a further 101 submitted during the webinar. Several questions were addressed during the discussion, with the remaining questions helping to inform the Commission’s wider understanding of public concerns and priorities, alongside the ongoing development of its work and recommendations.

Insights from the meeting

The AMA discussion reinforced themes emerging across the Commission’s wider research and engagement activities. Questions submitted both in advance of, and during, the session highlighted significant interest in how AI will be used, governed and regulated within healthcare systems.

Across the questions submitted, recurring themes included patient safety, transparency, accountability, fairness, implementation and the role of human oversight.

Across the session, there was broad recognition of the potential of AI in healthcare, but confidence was consistently described as conditional on how it is used in practice and the safeguards in place.

Safety, accuracy and human oversight

Safety, accuracy and meaningful human oversight were central priorities. Questions frequently focused on how risks would be identified, monitored and managed over time, including issues relating to accountability, post-market surveillance, auditability and liability where harm or errors occur.

Many questions reflected support for AI where it supports clinicians and improves efficiency, but with a consistent expectation that AI should support, rather than replace, professional judgement, particularly in higher risk clinical contexts such as diagnosis and treatment decisions.

Support varies depending on how AI is used

Questions submitted during the session reflected variation in support for AI in healthcare depending on use cases. Questions relating to administrative processes, operational efficiency and support for healthcare professionals often focused on how AI could improve services and reduce burden.

In contrast, questions relating to diagnosis, treatment decisions and direct patient care reflected greater concern about safety, accountability, oversight and the potential consequences of errors.

This suggests that expectations for safeguards, oversight and accountability are strongest where AI may have more direct impact on patient care and outcomes.

Transparency, communication and choice

Transparency, communication and choice were recurring themes. Questions emphasised the importance of patients being informed when AI is used, understanding the role it plays in care and having access to human support where needed. Expectations focused on clear, proportionate communication rather than technical detail.

Questions also raised issues relating to patient choice (including opt-in/opt-out), data protection and wider concerns relating to data governance, privacy, commercial involvement and the use of patient information in the development and deployment of AI systems.

Ongoing oversight and accountability

There was strong emphasis on the need for ongoing oversight and clear accountability once AI systems are in use. There was a strong expectation that monitoring extends beyond initial approval, with mechanisms to identify and respond to risks over time, alongside clearer arrangements for responsibility and support if issues arise.

Questions reflected significant interest in how AI systems would continue to be monitored after deployment, including post-market surveillance, auditability, explainability, liability and the governance of adaptive AI systems.

Overall, there was an expectation that regulatory approaches should remain responsive to technological change while ensuring AI systems remain safe, transparent, and challengeable in practice.

What this work highlighted

The open public meeting reinforced wider findings emerging across the Commission's research and engagement, showing that public confidence in the use of AI in healthcare is conditional on strong safeguards, meaningful human involvement and clear accountability in practice.

Questions submitted by patients, the public and stakeholders highlighted significant interest in how AI systems will be regulated, monitored and governed over time, particularly in higher risk clinical contexts.

The session also reinforced the importance of transparency, communication and ongoing public engagement in supporting trust in the future use and regulation of AI in healthcare.

3.2 Professional Bodies Roundtable Discussions

Delivery partner	MHRA
Format	Roundtable discussions
Date(s)	February – March 2026

Key Findings

- AI must demonstrate real-world patient benefit, not just technical performance
- AI should support – not replace – clinical judgement
- Accountability and liability remain unclear across the system
- Existing governance and monitoring are not suited to evolving AI systems

Overview

These sessions brought together healthcare professionals and organisations to gather views on AI in healthcare, using structured, facilitated focus groups organised around predefined themes to capture perspectives across a range of professional roles and care settings.

Approach and Scope

These engagement sessions gathered views from healthcare professionals and organisations on current and emerging uses of AI, including both regulated medical devices and wider applications. A series of two-hour online focus groups were conducted with provider organisations and professional networks. In total, 117 participants from 31 organisations contributed to the discussions.

Discussions were structured around six *a priori* themes, meaning thematic areas identified at the outset to organise the discussion. These were: quality, safety and effectiveness; clinical judgement and professional autonomy; the relationship between patients and healthcare professionals; workforce impact and liability; governance and safe deployment; and ongoing safe use and post-market surveillance.

Insights from the discussion

Contributions were analysed against the six *priori* themes used to structure the discussions. Emergent cross-cutting themes were identified where issues did not map cleanly onto this framework and were retained where they demonstrated significance within sessions or recurrence across contributions. This enabled the analysis to capture both the depth of discussion and the consistency of concerns raised by participants, ensuring that findings reflect both the structured thematic framework and broader system-level insights.

Priori Themes

Quality, Safety and Effectiveness

There was strong scepticism of approaches focused solely on accuracy metrics or efficiency gains in controlled settings, with a clear preference for evidence demonstrating how AI performs in real-world clinical contexts, including complex and variable patient populations.

Concerns about bias and representativeness were consistently framed as patient safety issues, reflecting risks that AI systems trained on unrepresentative data may produce unequal or unsafe outcomes for different groups.

Clinical Judgement and Professional Autonomy

There was a recurrent theme that AI should function as decision support, not replace clinical judgement. Participants highlighted the risk of uneven reliance across the workforce, potential for gradual de-skilling and the importance of ensuring clinicians can understand how AI systems generate their outputs.

Professional accountability was expected to remain with healthcare professionals, reinforcing the need for tools that are understandable and challengeable in practice.

Patient-Clinician Relationship

Participants noted both opportunities and risks for the patient-clinician relationship. While AI may support communication and shared decision-making, concerns focused on erosion of human interaction, impacts on trust and risks of misunderstanding AI outputs.

Participants also raised concerns relating to consent and patient understanding, particularly where AI outputs may be misinterpreted or presented without sufficient context. This included the need for clear, accessible explanations of AI-supported decisions, and awareness of how AI-generated information could influence patient expectations or decision-making in ways that may not fully reflect clinical reality.

Workforce Impact and Liability

Workforce considerations were closely linked to liability. Participants described uncertainty about responsibility when AI influences decisions, distinguishing between accountability and legal liability, with participants noting that healthcare professionals may remain accountable for decisions in practice, even where legal responsibility for harms arising from AI is unclear or shared across multiple actors.

Concerns also focused on the increased cognitive and emotional burden placed on staff, and expectations that clinicians retain oversight of AI outputs despite limited clarity on responsibility, appropriate use and how to challenge AI recommendations in practice.

Governance and Safe Deployment

Governance was seen as a foundational issue. Participants highlighted that existing approaches are often insufficient for systems that evolve over time and are too frequently treated as one-off activities rather than continuous processes. This was seen as misaligned with AI systems that may update, adapt or perform differently across contexts, creating risks that only emerge in real-world use.

Clearer roles, responsibilities and coordination across organisations were consistently identified as priorities. Participants highlighted the importance of governance approaches that operate across care settings, without which AI systems risk fragmented oversight and limited capacity to monitor or respond to emerging issues.

Ongoing Safe Use and Post Market Surveillance

Participants identified significant gaps in monitoring AI in real-world use, including the limited ability to capture incidents and near misses, difficulty tracking performance over time, and resource constraints for continuous oversight. This was seen as a critical barrier to safe and effective deployment.

Emergent Cross-Cutting Themes

Participants identified a set of cross-cutting themes that sit outside the six a priori areas, reflecting systemic, behavioural and longitudinal dynamics in the use of AI in healthcare. These cross-cutting themes may not all require regulatory responses.

- **Automation bias and over-reliance:** Participants highlighted the risk of over-trusting AI outputs, particularly under time pressure. This risk is amplified where systems are perceived as authoritative or endorsed.

- **Psychological safety to challenge AI:** The ability to question or override AI depends on organisational culture. Lack of psychological safety may discourage challenge – particularly among less experienced staff – reducing decision quality and learning.

- **De-skilling and erosion of expertise:** Long-term reliance on AI may reduce independent clinical reasoning and limit opportunities for experiential learning, particularly for trainees. This also has implications for system resilience if AI systems fail or underperform.

- **Equity, data representativeness and structural inequality:** Participants emphasised risks of reinforcing existing inequalities through biased data and uneven adoption across care settings. Equity concerns were linked to both system design and patterns of implementation.

- **Regulatory boundary confusion and accountability gaps:** Uncertainty over whether certain AI tools fall within existing frameworks – particularly general-purpose and generative AI – was seen to create gaps in accountability and oversight.

- **Fragmentation across care settings:** Participants raised concerns that AI tools introduced in one part of the system may not be visible or interoperable across others, risking duplication, inconsistency and challenges to continuity of care.

What the engagement sessions highlighted

The engagement sessions highlighted a consistent mismatch between existing product-based regulatory models and the realities of how AI behaves in clinical practice. Participants emphasised that safe and effective use depends on understanding AI across its lifecycle, including how it interacts with professional judgement, organisational context and human factors in real-world settings.

They also pointed to the need for clearer accountability across manufacturers, deployers, healthcare professionals and organisations, alongside stronger

approaches to monitoring, incident reporting and organisational learning after deployment.

The findings further show that trust and safety are shaped not only by technical performance, but by workforce capability, psychological safety to challenge AI, transparency to patients, and the wider system conditions in which AI is deployed.

Taken together, these discussions suggest that AI regulation in healthcare must operate at system level as well as product level if it is to support safe, effective and equitable adoption.

3.3 Post-Market Surveillance and Liability Roundtable with techUK

Delivery partner	techUK; MHRA
Format	Roundtable discussion
Date(s)	02 February 2026

Key Findings

- Current regulation does not reflect AI's evolving, non-static nature
- Risk and assurance require more flexible and context-sensitive approaches
- Greater emphasis is needed on real-world monitoring and post-market oversight
- Clearer liability and responsibility across the lifecycle are needed

Overview

This roundtable brought together businesses working in AI for healthcare to discuss regulatory challenges and complement wider industry input to the National AI Commission. It used a facilitated in-person discussion to capture practical insights from organisations working directly in the field.

Approach and Scope

The discussion focused on post-market surveillance and liability, examining how monitoring could be adapted for iterative AI models and how responsibility should be shared across the supply chain. Discussions were structured around practical, scenario-based questions, and

also drew on questions from the Call for Evidence and was designed to capture diverse industry perspectives. The session was held in-person at techUK's London office, with participants primarily comprising companies operating within the AI healthcare sector. A total of eight organisations took part.

Insights from the discussion

Mismatch between static regulation frameworks and dynamic AI systems

Participants highlighted a fundamental misalignment between existing regulatory approaches and the lifecycle of AI systems.

Current frameworks assume fixed products, whereas AI systems are iterative, adaptive, and may evolve post-deployment. Point-in-time assessments and binary classification do not adequately capture key risk drivers such as autonomy, generalisability, and context of use. This creates challenges in assessing systems where real-world performance diverges from their initially approved state.

This reflects a broader shift from product-based to lifecycle-based regulation, where oversight must extend beyond initial approval and continue throughout deployment and use. There was also recognition that defining "meaningful ongoing oversight" for non-deterministic systems – where outputs may vary for the same input – remain difficult to oversee using traditional assurance approaches that assume predictable behaviour.

Risk-based and context-dependent regulatory approaches

There was broad support for risk-based approaches that account for system characteristics, such as autonomy, architecture and deployment context.

Participants noted that the classification of risk may extend beyond the potential impact of the tool, and should also include the system architecture, levels of automation, and the specific deployment environment. In particular, the same AI system may present different risk profiles depending on local data, workflows, and patient populations.

A more risk-proportionate approach to post-market surveillance would therefore reflect these granular characteristics, recognising that risk is shaped dynamically through real-world use rather than defined at a single point in time.

These challenges extend to how systems are classified and regulated, particularly where functionality evolves over time or spans multiple categories.

Evolving approaches to assurance

Participants highlighted limitations in traditional, document-based assurance approaches, which struggle to reflect evolving systems.

There was interest in more dynamic models of assurance, such as model-based artefacts that track safety assumptions and evidence over time. This reflects a broader shift toward continuous, system-level assurance, where post-market surveillance is embedded throughout the lifecycle.

Suggested approaches included designing systems with built-in logging and monitoring capabilities, incorporating human-AI interaction signals (e.g. patterns of clinician reliance), and moving from retrospective reporting to proactive, integrated monitoring architectures

System-level enablers and limitations of existing frameworks

Effective regulation was seen as dependent on system-level infrastructure. Key enablers identified included, real-world testing environments, phased or iterative authorisation pathways, and mechanisms to capture and share incident data across organisations.

Participants highlighted challenges related to fragmentation across systems and limited interoperability, which can reduce visibility of system performance and constrain learning across organisations.

Decentralised approaches to incident reporting were discussed as a way to balance local usability with system-wide visibility. Existing reporting frameworks (such as the Yellow Card Scheme) were seen as poorly suited to AI, reflecting product-based assumptions.

Proposed improvements included integrating reporting into clinical systems (e.g. Electronic Health Records), requiring built-in logging and structured outputs, and updating definitions to reflect software lifecycles. Participants emphasised that clinician-level reporting alone is insufficient, underscoring the need for deeper system-level telemetry.

Liability considerations in system design and use

Participants raised strong concerns about over-burdening healthcare professionals, or that they may become “liability sinks”, where responsibility for errors is effectively transferred onto healthcare professionals despite the influence of AI systems and wider system design. There was a preference for clearer guidance on the boundary between product liability and clinical responsibility over legislative reform.

What this work highlighted

The roundtable highlighted the need for regulatory approaches that reflect the dynamic, context-dependent and system-level nature of AI technologies.

Participants consistently pointed towards lifecycle-based oversight, more granular risk assessment, and assurance models that can respond to changing performance in real-world use.

They also emphasised the importance of system-level enablers, including real-world testing, data sharing, integrated monitoring and clearer boundaries for responsibility across the AI lifecycle.

3.4 Liability Roundtable with Ada Lovelace Institute

Delivery partner	Ada Lovelace Institute; MHRA
Format	Roundtable discussion
Date(s)	18 March 2026

Key Findings

- Existing legal frameworks are not designed for AI systems
- Responsibility is fragmented and insufficiently defined across actors
- Healthcare professionals face uncertainty in relying on AI outputs
- Patients may lack clarity on AI use and routes to redress

Overview

This roundtable brought together stakeholders to examine how liability for AI in healthcare should be allocated across the lifecycle of AI systems, highlighting the complexity created by multiple actors and system components. Participants were invited to discuss these issues through structured, scenario-based analysis, exploring how liability challenges arise in practice.

Approach and Scope

This roundtable session, held at the Ada Lovelace Institute, was delivered through a structured, facilitated roundtable with legal, clinical and regulatory stakeholders across the health and care system. A total of 18 organisations participated, including legal firms, medical defence unions, universities, regulators and NHS representative bodies.

The session was organised around five core questions and an ‘anchor scenario’ to explore how liability uncertainty may arise in practice, including through system updates, performance drift, and organisational deployment pressures.

Discussions covered the current liability landscape, approaches to allocating responsibility across the AI lifecycle, expectations for manufacturers and developers, distinctions between different actors in the supply chain (including foundation model developers), and considerations for general-purpose and direct-to-consumer AI systems.

Insights from the discussion

Limitations of current liability frameworks

Participants highlighted that existing legal and regulatory frameworks are designed for stable products and clear, human-led causation, whereas AI systems are more dynamic, context-dependent and embedded within complex value chains.

The performance of an AI system may change during its lifetime through updates, retraining, and environmental changes, meaning performance assessed at approval cannot be relied upon over time. At the same time, opacity of reasoning can obscure the link between harms and their causes, which together could challenge traditional liability approaches to negligence and product liability. Limited case law and cross-jurisdictional deployment further increase uncertainty.

Fragmented responsibilities across the AI value chain

AI-enabled systems involve a wider and more interdependent set of actors than traditional devices, yet distribution of responsibilities across this chain remains unclear. This leads to fragmented accountability in key areas, such as post-market performance monitoring, oversight of model updates and escalation and resolution of issues.

In practice, liability often falls to healthcare professionals, despite their limited influence over the design or deployment conditions. This creates a situation in which clinicians, and, by extension, the NHS may end up bearing most of the financial risk when AI products in healthcare go wrong, while other upstream actors are shielded from that risk. Variable contracting practices and the framing of AI as “decision support tools” can reinforce this reality.

Clinical uncertainty and reliance on AI

Healthcare professionals remain the primary decision-makers under current frameworks but face uncertainty about when reliance on AI is appropriate or defensible. Limited access to information about performance, updates and system behaviour constrains their ability to evaluate outputs and respond to risk. This is compounded by variations in AI literacy, difference in risk appetite, organisational context and pressures to adopt AI.

Existing post-market surveillance systems were viewed as poorly adapted to frequently updated or context-sensitive systems, contributing to diffuse accountability.

Implications for patients and pathways to redress

Participants noted that AI introduces challenges for transparency, consent and access to compensation, as patients may not know when AI has influenced their care and complex

systems make causation difficult to establish. These issues are amplified when harms emerge over time or at scale.

Concerns were also raised about direct-to-consumer and agentic AI, which may influence decisions with limited professional oversight and operate outside established frameworks, creating uncertainty over responsibility when harm occurs.

What the roundtable discussion highlighted

Across discussions, participants emphasised that existing legal and regulatory approaches are not well aligned to adaptive, multi-actor AI systems.

A recurring theme was the need for clearer allocation of responsibility across the lifecycle, alongside more effective approaches to transparency, accountability and redress.

Participants also identified practical areas for development, including stronger manufacturer responsibilities, clearer contractual and professional expectations, better support for healthcare professionals and improved public visibility of how AI influences care.

4. Survey and Polling

Survey and polling work provides quantitative evidence on public and workforce attitudes to AI in healthcare. This work captures nationally representative data to understand awareness, perceptions, and trends, and provides a broad view of public attitudes at scale, identifying patterns across demographic groups, and tracking how perceptions are changing over time.

4.1 Public attitudes to technology and AI

Delivery partner	The Health Foundation
Format	Survey analysis
Date(s)	Annual surveys conducted 2023, 2024 and 2025; data in this summary drawn primarily from Wave 3 (July – October 2025)
Output(s)	Report publication expected late Summer 2026

Key Findings

- Awareness of AI in healthcare is low with some increase in familiarity over time
- Support is stronger for administrative uses than clinical decision-making
- Acceptance depends on human oversight, transparency and safeguards
- Safety and evidence are prioritised over speed or convenience
- Attitudes vary significantly by age, gender and socio-economic group

Overview

This survey summarises annual data on public and NHS staff attitudes to AI in healthcare, providing a representative view and allowing comparisons over time. It draws on nationally representative survey data and quantitative analysis to track awareness, attitudes and preferences across different groups. The full report will be published by the Health Foundation in late Summer 2026.

Approach and Scope

Conducted by the Health Foundation as a research partner to the National AI Commission, this analysis draws on nationally representative survey data collected since 2023, including an 8,000-person public sample and a 2,027-person NHS staff sample in 2025.

The analysis covers public awareness of AI in the NHS; attitudes to administrative and clinical uses of AI; the conditions shaping public support, including human oversight, transparency and evidence of effectiveness; concerns around accuracy, loss of human interaction, bias and data privacy; public priorities when needing to balance factors such as safety and speed; and variation in attitudes across demographic and socioeconomic groups.

Annual surveys were conducted online using Censuswide's panel, including booster sampling of people at risk of digital exclusion. Data were double opt-in, weighted to UK Census demographics, and tested for statistical significance at the 95% confidence level. Findings were triangulated with wider polling and evidence review outputs.

Summary of findings

Public awareness of AI in healthcare remains low, particularly among older and lower-income groups. Despite this, overall sentiment is cautiously positive. There has been a modest increase in positive sentiment since 2024, though this is concentrated among those in higher-income groups and driven primarily by increased support for administrative uses of AI rather than clinical applications.

Public attitudes towards AI are shaped by expectations of human oversight, transparency, and safeguards. Support is significantly higher where healthcare professionals remain involved and declines sharply for fully autonomous systems – especially in clinical decision-making. There is also a strong preference for being informed when AI is used, especially in higher-risk contexts.

Perceptions of AI's impact are mixed. The most recognised benefit is reducing administrative burden, while key concerns include accuracy, loss of human interaction, bias and data privacy risks. The public consistently prioritises safety, evidence and human checks over speed or convenience, favouring stronger safeguards over rapid deployment.

Attitudes towards AI vary across demographic and socioeconomic groups. Women, younger people (aged 16-24), and those with lower or less secure incomes tend to be less supportive and more concerned about safety and quality.

While willingness to share data for AI development is relatively high overall, it varies depending on the sensitivity of the data and concerns about how it will be used.

What this work highlighted

The findings reinforce that public attitudes to AI in healthcare are cautiously positive but strongly conditional. Support depends on confidence in safety, transparency, human oversight and demonstrated benefits in practice.

Trust is also shaped by wider system factors, including data use, fairness and public understanding. Variation across groups highlights the importance of clear communication and equitable implementation.

Overall, the evidence underlines that building and sustaining trust will require visible safeguards and consistent delivery in practice – not just technical performance.

4.2 Ipsos' UK Knowledge Panel

Delivery partner	The Health Foundation
Format	Survey analysis
Date(s)	December 2025
Output	Report publication expected late Summer 2026

Key Findings

- Awareness of AI use in the NHS remains low and uneven
- Acceptability varies by use case and presence of human oversight
- Support is highest for administrative tasks and lowest for clinical decision-making
- Public expectations centre on high safety standards, accountability and transparency

Overview

This survey analysed new nationally representative data on public attitudes to AI in healthcare to provide up-to-date insight for the National AI Commission. It used a probability-based survey and quantitative analysis to examine responses to questions developed with the MHRA. The full report can be found on The Health Foundation website.

Approach and Scope

This analysis, conducted by the Health Foundation as a research partner to the National AI Commission, draws on a survey of 2,214 UK adults carried out in December 2025 using Ipsos' KnowledgePanel, which comprises over 25,000 participants recruited through random probability address-based sampling, with weighting applied to ensure representativeness.

The survey examines perceptions of AI use in the NHS, including acceptability across different use cases (with and without human oversight), safety expectations, demographic variation in attitudes, public use of consumer large language models, and priorities for AI regulation. It also includes a new module on public attitudes to AI, developed with the MHRA and Commission leadership to inform the National AI Commission's work.

Data were analysed using subgroup comparisons where relevant, with statistical significance tested at the 95% confidence level. Findings were triangulated with other survey data and evidence review outputs.

Summary of findings

Public perceptions of AI use in the NHS are mixed and uncertain. While some identify uses such as appointment scheduling, scan interpretation and chatbots, over a quarter report not knowing where AI is used. Awareness is socially patterned, with higher understanding among graduates and greater uncertainty among those with lower levels of education.

Acceptability is strongly shaped by use case and human oversight. Support is highest for administrative tasks and declines as AI becomes more involved in clinical decision-making – particularly for triage and prioritisation, which are consistently least acceptable. Removing human oversight reduces acceptability across all use cases, with the largest declines for higher-risk clinical applications.

Public expectations for safety are high and, in some cases, polarised. Many believe AI must be as safe as or safer than the best healthcare professionals, while a consistent minority believe it will never be sufficiently safe. Across groups, safety, evidence and oversight are prioritised over speed or convenience.

Use of consumer large language models for health advice is already widespread, particularly among younger groups, but trust remains relatively low compared to traditional sources. Trust increases significantly when AI is delivered through the NHS or healthcare professionals.

When considering regulation, the public prioritises patient safety, accountability, and explainability. Less emphasis is placed on innovation, speed or economic benefit.

What the polling highlighted

The findings reinforce that public confidence in AI is strongly shaped by use case, setting and oversight. Support is higher for lower-risk applications and for AI delivered within trusted healthcare contexts, particularly the NHS.

Across all findings, there is consistent emphasis on high safety standards, meaningful human oversight, transparency and fair implementation.

Overall, the evidence underlines that trust depends not only on the technology itself, but on how it is governed and deployed in practice.

5. Evidence Synthesis

Evidence synthesis brings together existing research and literature on AI in healthcare. It reviews and analyses the wider evidence base to identify established knowledge, gaps and emerging themes. This provides a structured understanding of what is already known, supporting consistency and context for findings from primary research and engagement.

5.1 Public attitudes to AI in healthcare in the UK

Delivery partner	The Health Foundation
Format	Rapid evidence review
Date(s)	December 2025
Output	Report publication expected later Summer 2026

Key Findings

- Public attitudes are cautiously positive but conditional on context and use case
- Human oversight is the most consistently cited condition of public acceptability
- Acceptability is shaped by interrelated factors including effectiveness, safety, equity, transparency and data governance
- Concerns focus on accuracy, loss of human contact, data use, fairness and governance

Overview

This rapid evidence review brought together existing UK research on public attitudes to AI in healthcare. Using a structured literature search and thematic analysis, it identified common patterns in perceptions and experiences across study types and use cases. The full report will be published later this Summer.

Approach and Scope

The review was undertaken by the Health Foundation as a research partner to the National AI Commission, focusing on public attitudes to AI in healthcare in the UK.

A structured search identified original research published since 2020. In total, 503 papers were screened, with 32 studies included in the final review, spanning qualitative, quantitative and mixed methods designs. The studies covered a range of applications, including prevention, diagnosis, treatment and administrative uses. Most explored hypothetical scenarios, with fewer examining direct experience of AI in practice. Findings were extracted and analysed thematically to identify consistent patterns in perceptions and perceived impacts.

Summary of findings

The evidence base focuses on acceptability, perceived impacts, and the factors shaping public views. While few studies directly address regulation, those that do highlight concerns around accountability, access and potential harms, with wider findings informing expectations of trust, safety and governance.

Most studies found AI acceptable to a majority of participants, particularly when used to support healthcare professionals rather than replace them. However, acceptability was highly conditional. Comfort depended on healthcare professionals retaining oversight and responsibility for decisions, with significantly lower confidence where AI operates autonomously – especially in high-stakes contexts.

Views on the impact of AI were mixed. Perceived benefits included improved efficiency, quality of care, reliability, communication and reduced staff workload. Perceived risks focused on reduced accuracy, loss of human interaction, bias and discrimination, data privacy and security, and exclusion of certain groups. Across the evidence, perceptions were shaped by a mix of factors in four interrelated domains:

- AI implementation and oversight (human oversight and accountability)
- Patient and public characteristics (demographic factors and familiarity with AI)
- Perceived impacts of AI (evidence of effectiveness, safety, and equity)
- Trustworthiness and data (trustworthiness and data security and privacy).

What the research highlighted

The review found that public attitudes to AI in healthcare are broadly positive but strongly conditional. Acceptability is shaped by a range of interrelated factors, of which human oversight was the most consistently cited. Other important factors included explainability, evidence of effectiveness, data security, equity, and demographic characteristics — though findings were not always consistent across studies.

The evidence base has significant limitations: studies vary widely in quality, most explore hypothetical rather than real-world scenarios, and samples are often small and unrepresentative. A full peer-reviewed article based on this review is in preparation.

5.2 Categorising harms on AI in healthcare

Delivery partner	The Health Foundation
Format	Rapid evidence review
Date(s)	February – March 2026
Output(s)	Report publication expected late Summer 2026

Key Findings

- AI harms extend beyond direct patient harm to system-level impacts
- Existing frameworks are useful but incomplete for AI regulation
- Defining and assessing harm is complex and context-dependent
- There is limited real-world evidence on AI harms

Overview

This rapid evidence review examined how harms linked to AI in healthcare are identified and categorised. It used a targeted literature review, thematic analysis and expert input to identify common approaches, limitations and gaps in how harms are currently understood. The full report will be published by the Health Foundation in late Summer 2026.

Approach and Scope

This review, conducted by the Health Foundation as a research partner to the National AI Commission, examined how harms from AI in healthcare are defined and categorised.

A targeted search of academic and grey literature identified 267 records, with 16 studies included, supplemented by four expert interviews. The analysis considered how harms are conceptualised across individual, community and societal levels, with findings synthesised thematically to identify common patterns and gaps.

Summary of findings

The review identified a wide range of harms, extending beyond physical injury to include clinical, psychological and system-level impacts across individual, community and societal levels. Examples include:

- inaccurate advice and misdiagnosis
- bias and inequality
- privacy and data security breaches
- loss of trust in human interactions
- reduced patient agency
- resource misallocation
- accountability challenges

Alongside harms common to other health technologies, the review identified risks specific to AI, including hallucinations and unreliable outputs, misalignment with human values, and adaptive system behaviour. It also highlighted indirect harms arising through wider systems, such as the spread of health misinformation.

Existing frameworks (including the Health AI Risks Taxonomy and epistemic, normative and traceability approaches) provide useful but partial perspectives. No single existing framework fully meets the needs of an effective regulatory system, particularly in supporting the prioritisation and assessment of risks in a way that reflects the needs of patients, the public, and the healthcare system as a whole.

Defining harm is inherently complex. The review highlights tensions between the narrow clinical definitions, and broader concepts including wellbeing, systemic injustice, and societal impact. Harms are highly context-dependent, varying by technology, user, vulnerability and setting. This suggests that rigid, product-level regulatory approaches may be insufficient.

The review found that limited real-world evidence constrains understanding, prioritisation and categorisation of harms, indicating a need for further work and higher-quality evidence, including direct involvement of patients and the public.

What the research highlighted

This review shows that AI harms in healthcare extend beyond individual patient safety to include psychological, social and system-level effects. It points to the need for regulatory approaches that are broader and more adaptive than traditional product-focused models, particularly because harms are context-dependent, may emerge indirectly, and are not yet well evidenced in real-world settings.

6. AI Airlock Sandbox Programme

A regulatory sandbox is a tool used by governments and regulators to test new technologies, services or business models in a controlled and supervised environment. It allows organisations to trial innovations safely before wider rollout or formal regulation. This helps generate evidence to inform future policy and regulatory approaches.

Delivery partner	MHRA; Team AB (UK Approved Bodies); DHSC; NHS
Format	Regulatory sandbox
Date(s)	Pilot phase April 2024 – March 2025 Phase 2 April 2025 – March 2026
Output(s)	Pilot Programme Report published 16 October 2025 Phase 2 Programme Report published 09 June 2026

Key Findings

- AI exposes gaps in current medical device regulatory frameworks
- Existing approaches are not specifically designed to address some of the unique risks associated with AI
- Effective regulation requires continuous monitoring, real-world evidence and attention to human factors
- Accountability, explainability and lifecycle responsibilities are insufficiently defined

Overview

The MHRA AI Airlock was launched in Spring 2024 as the MHRA's first regulatory sandbox for artificial intelligence as a medical device (AIaMD). The programme is designed to enhance understanding of, and accelerate solutions to, novel regulatory challenges associated with AIaMD.

Using real-world products, the AI Airlock brings together expertise from across the MHRA and key partners, including UK Approved Bodies, Department of Health and Social Care (DHSC), the NHS, and other regulators. It generates practical evidence on how existing regulatory frameworks operate in practice, where challenges arise, and where future guidance or policy could adapt.

While the pilot phase pre-dated the National AI Commission, Phase 2 has run in parallel with the National AI Commission's work. This phase is testing a more complex set of regulatory challenges and has completed in May 2026, with reports expected in Summer 2026.

The MHRA announced on 8 April 2026 that funding has been secured to continue the AI Airlock programme. This new multi-year funding will enable the programme to scale beyond the constraints of annual financial cycles, supporting more ambitious long-term testing models and helping to create a more sustainable regulatory pathway for future AI medical technologies. Design of the Phase 3 sandbox programme is now in progress, with further information due to be announced later in the year.

Approach and Scope

The AI Airlock is intended to inform the MHRA's future regulatory approach by generating practical insight into how existing frameworks apply to AIaMD in practice and where gaps or uncertainties arise. It uses a regulatory sandbox model, enabling selected innovators to work directly with regulators and stakeholders in a controlled environment to explore challenging questions without affecting patient care.

The programme has included open calls for candidate products spanning a range of AI technologies, clinical contexts and levels of regulatory maturity. Over the past 2 years, work has combined product-specific testing with wider regulatory analysis, expert input and structured engagement with relevant partners.

Testing has typically followed a structured process covering orientation, planning, testing and review. Depending on the use case, this has involved simulation workshops, virtual testing and controlled real-world exploration. This approach enables the programme to examine how AI systems behave under different conditions, surface regulatory issues that may not be visible through pre-market assessment alone, and test how evidence could be generated to support future regulatory decisions.

Alongside product testing, the programme has included regulatory mapping, gap analysis and engagement with healthcare professionals, technical experts and other stakeholders to explore how existing requirements apply across the AI lifecycle. Its primary focus is AIaMD within the Great Britain regulatory framework, while also considering relevant international standards and cross-system learning.

Summary of findings

The AI Airlock has identified important regulatory gaps and challenges in applying existing frameworks to AIaMD. While the specific issues varied across products, a consistent set of themes emerged. These findings were developed through work with innovators and are being further tested and refined through Phase 2 of the programme. Taken together, they

show that many of the most important regulatory questions for AI do not arise only at the point of market entry, but through how systems are developed, validated, deployed and monitored in practice.

Synthetic data and LLMs

Current medical device regulations and guidance do not clearly define how synthetic data can be used in training and testing to demonstrate compliance, and there is limited direction on how such data should be generated, validated and assessed for regulatory purposes. This is particularly relevant for AI systems where access to suitable real-world data may be constrained.

Large language models (LLMs) may have the potential to support the generation and assessment of synthetic clinical data, but this also highlights important challenges around validation, quality assurance and the risk of circularity, where one AI system is used to evaluate another. These findings point to the need for clearer validation standards, independent scrutiny and a stronger evidence base to support the safe use of synthetic data in regulatory contexts.

Hallucinations, Non-Determinism and Risk Management

Existing risk management frameworks provide an important foundation for medical devices, but they do not fully address challenges associated with AI such as hallucinations, non-deterministic outputs, and model drift. These characteristics create uncertainty about how risks should be identified, controlled, and evidenced within current regulatory approaches.

Techniques such as retrieval-augmented generation can help reduce some of these risks by grounding outputs in trusted sources, but this also highlights the need for more comprehensive approaches to risk management that reflect how AI behaves in practice. This includes greater emphasis on real-world monitoring, clearer expectations for post-market oversight, and more explicit approaches to managing uncertainty over time.

Explainability and Transparency

Current regulatory frameworks provide limited guidance on explainability and transparency for AI systems, and do not yet set out clearly what different users, regulators or stakeholders may need in order to understand and safely use these technologies. This becomes more important where systems evolve over time or operate in clinical settings.

A range of explainability approaches may support understanding and trust, but their usefulness depends on the context, the intended audience and the purpose they are meant to serve. This suggests a need for clearer and more consistent approaches to how

explainability is defined, evidenced and communicated to stakeholders across the AI lifecycle.

Post-Market Surveillance and Real-Time Monitoring

Existing post-market surveillance arrangements provide established reporting mechanisms, but they are not specifically designed to suit the way AI systems can change over time or perform differently across settings. They also provide limited support for issues such as drift, automation bias and the interaction between AI systems and human users in real-world environments.

Continuous and real-time monitoring may help identify emerging risks earlier, including changes in model behaviour and patterns of human-AI interaction. It also highlighted the importance of human factors, such as professional experience and working conditions, in shaping safety outcomes. Overall, this points to the need for stronger approaches to lifecycle oversight, clearer thresholds for action, and better-defined responsibilities for managing risk once AI systems are in use.

What the programme highlighted

The AI Airlock programme confirms that while the existing regulatory framework provides a foundation for regulating AIaMD, some revisions, targeted guidance and clarification are needed in specific areas to ensure it remains effective as AI technologies develop and deploy at pace. Key risks may emerge not only from technical properties, but from how systems perform over time and interact with users, workflows and organisations in practice, which are becoming ever more complex.

The programme reinforces the need for lifecycle oversight, continuous monitoring and clear approaches to explainability, human factors and shared accountability, particularly where systems are meant to change after deployment. It also demonstrates the value of sandbox-based testing in surfacing issues that may not be visible through pre-market assessment alone.

More broadly, the findings show how regulatory sandboxes can generate practical evidence, test emerging approaches and help identify where future frameworks can adapt.

Ongoing work on liability in healthcare AI

In addition to the research and engagement activities undertaken to date, the MHRA is progressing further engagements from June 2026 onwards to explore professional liability in the context of AI in healthcare, in collaboration with CERSI-AI and the Professional Standards Authority.

This engagement responds to persistent uncertainty identified in earlier work around how responsibility is allocated when AI-influenced decisions result in harm, particularly across healthcare professionals, provider organisations, and technology developers and manufacturers.

The work is being taken forward through a series of targeted engagement sessions with professional regulators, accredited bodies, defence organisations, system partners, and legal and ethical experts. These sessions are designed to test emerging thinking, capture sector-specific perspectives, and inform the development of shared approaches to liability in AI-enabled healthcare.

In parallel, the National AI Commission is also examining questions of liability through commissioned legal analysis, exploring how existing UK frameworks apply to the use of AI systems in healthcare practice.

This analysis considers how liability may arise across the AI lifecycle, including the roles of manufacturers, developers, healthcare providers and professionals. It highlights areas of uncertainty around responsibility and causation and examines how these challenges may affect patients seeking redress when harm occurs. This analysis is intended to support ongoing policy and regulatory development by providing a clearer understanding of how liability operates in practice.

Insights from both the ongoing engagement activity and commissioned legal analysis will contribute to the National AI Commission’s consideration of liability in the context of AI in healthcare and will inform its recommendations, to be published in late summer 2026.

Acknowledgements

The MHRA and National AI Commission would like to thank all those who have contributed to the research and engagement activities undertaken.

We extend particular thanks to the patients, members of the public and carers who shared their experiences, perspectives and insights. Their contributions have been central to shaping the evidence presented in this report. We also extend our thanks to the many individuals and organisations who contributed their time and insight, including healthcare professionals, industry representatives, regulators and system partners.

We are grateful to our delivery partners, including the Health Foundation as the National AI Commission's research partner and National Voices for their commissioned engagement work, for their expertise and collaboration in delivering key elements of the activities set out in this report. We extend thanks to SurveyOptic, who supported the delivery of the Call for Evidence and led on the subsequent analysis of responses that were received. We also thank techUK, the Ada Lovelace Institute, Ipsos UK and other organisations who supported specific workstreams.

We also acknowledge the contribution of the Commissioners and Working Group members, whose expertise and commitment supported both the consideration of insights and wider outreach in key public engagements.

Glossary of terms

This list contains a glossary of commonly used words, phrases, and abbreviations from this report. Please note that this glossary is a tool to assist readers with their understanding of this report, which includes technical language and terminology. While some of the definitions below are set out in medical device regulations, such as the UK MDR and EU MDR, in some cases these definitions are set out in guidance or are the National AI Commission's interpretation of these terms.

Term	Definition
Accountability (in AI systems)	The assignment of responsibility for the design, deployment, performance and outcomes of an AI system, including who is responsible when things go wrong across its lifecycle
Adaptive systems	Systems whose behaviour can change over time, for example through updates, retraining, or learning from new data after deployment
Agentic AI systems	AI systems capable of acting autonomously to pursue goals, often coordinating multiple tasks or tools with limited direct human oversight.
AI as a Medical Device (AIaMD)	AI software that performs a medical function and is regulated as a medical device under the UK medical device regulations as a subset of Software as a Medical Device (SaMD)
Artificial intelligence (AI)	Technologies that simulate human intelligence to perform complex tasks by learning from data, recognising patterns, making predictions, or generating outputs
Automation bias	The tendency to over rely on or favour AI system outputs, even when they may be incorrect
Clinical decision support	AI systems or tools that assist healthcare professionals in making decisions about patient care, rather than replacing human judgement
Context-dependent performance	Variation in how an AI system performs depending on factors such as local data, deployment environment, user capabilities, or organisational settings
Continuous oversight	Ongoing monitoring and review of an AI system's safety and performance after deployment, rather than relying solely on pre-market checks
EU Medical Device Regulations (EU MDR)	Regulation (EU) 2017/745
Explainability	The extent to which the reasoning or outputs of an AI system can be understood and interpreted by users, regulators or other stakeholders

Foundation models	General-purpose AI models trained on extensive datasets that can be adapted for a wide range of tasks and downstream applications
General-purpose AI systems	AI systems designed to perform a broad range of tasks across multiple contexts, rather than a single defined use
Ground truth	The reference standard used to determine the correctness of an AI system's output, often based on expert judgement or validated data, and used to train AI systems or evaluate performance
Hallucinations (in AI)	Incorrect or misleading outputs generated by AI systems, often presented as if accurate or factual
Human oversight	The involvement of a human in reviewing, validating or controlling AI outputs
Human-AI interaction	The ways in which people interpret, use and respond to AI system outputs in practice, including how decision-making is influenced and how users rely on, question or override those outputs
Intended purpose	The use to which the device is intended for according to the data supplied by the manufacturer on the labelling, the instructions for use and/or the promotional materials
Large Language Models (LLMs)	AI systems trained on large volumes of text data that can interpret and generate natural language (e.g. chatbots)
Liability	Legal responsibility for harm or adverse outcomes, including determining who is accountable under law when an AI system contributes to an error
Lifecycle regulation / oversight	An approach to regulation that considers an AI system across its full lifecycle, including design, training, development, deployment, use, ongoing monitoring, and decommissioning
Manufacturer / Developer	(a) the person with responsibility for the design, manufacture, packaging and labelling of a device before it is placed on the market under their own name, regardless of whether these operations are carried out by that person themselves or on their behalf by a third party; or (b) any other person who assembles, packages, processes, fully refurbishes or labels one or more ready-made products or assigns to them their intended purpose as a device with a view to their being placed on the market under their own name, apart from a person who assembles or adapts devices already on the market to their intended purpose for an individual patient
Medical device	Any instrument, apparatus, appliance, software, material or other article, whether used alone or in combination, together with any accessories, including the software intended by its manufacturer to be used specifically for diagnosis or therapeutic purposes or both and necessary for its proper application, which –

	(a) is intended by the manufacturer to be used for human beings for the purpose of- (i) diagnosis, prevention, monitoring, treatment, or alleviation of disease, (ii) diagnosis, monitoring, treatment, alleviation of or compensation for an injury or handicap, (iii) investigation, replacement, or modification of the anatomy or of a physiological process, or (iv) control of conception; and (b) does not achieve its principal intended action in or on the human body by pharmacological, immunological or metabolic means, even if it is assisted in its function by such means, and includes devices intended to administer a medicinal product or which incorporate as an integral part a substance which, if used separately, would be a medicinal product and which is liable to act upon the body with action ancillary to that of the device
Medical Device Regulations (MDR)	The Medical Devices Regulations 2002 (SI 2002 No 618, as amended)
Model drift	Changes in an AI system's performance over time due to shifts in data, environment or use, which may reduce accuracy or reliability
Non-determinism	The characteristic of some AI systems where the same input may produce different outputs, introducing variability and uncertainty
Post-market surveillance (PMS)	Activities carried out by manufacturers to proactively collect, and review experience gained from devices placed on the market or put into service for the purposes of identifying any need to apply corrective or preventive actions
Pre-market assessment / approval	Regulatory evaluation a medical device must undergo before it can be placed on the market to ensures the device meets the requirements for safety, performance, and design through conformity assessments, technical documentation review, and clinical investigations
Real-time monitoring	Continuous observation of an AI system's performance as it operates, enabling early detection of issues or risks
Real-world evidence	Data and insights generated from the use of technologies in real-world settings, as opposed to controlled testing or trials, used to assess performance and safety over time
Retrieval-Augmented Generation (RAG)	An AI approach where a model retrieves information from trusted data sources to support and ground its outputs, improving accuracy
Risk-based regulation	A regulatory approach in which the level of scrutiny applied depends on the potential risk of the system to patients or users
Sandbox	A collaborative, time-bound environments where innovators can test novel technologies, generate evidence, and refine practices within

	existing regulatory frameworks under regulator supervision, before wider deployment
Software	A set of instructions that processes input data and creates output data
Software as a Medical Device (SaMD)	Standalone software and apps that meet the definition of a medical device (including AI as a medical device (AIaMD))
Synthetic data	Artificially generated data used to train or test AI systems, designed to replicate characteristics of real data without using actual patient information
System-level approach	An approach that considers how AI interacts with people, organisations and processes, rather than treating it as a standalone product
Transparency (in AI systems)	Openness about how and when AI is used, including its role, limitations, and how decisions are made or supported.
Validation (in AI systems)	The process of assessing whether an AI system performs as intended and produces reliable and accurate outputs
Yellow Card Scheme	The UK system for collecting and monitoring information on safety concerns such as suspected side effects or adverse incidents involving medicines and medical devices