

## AI in healthcare: Cyber Security, Governance and Risk Management

### Context

Earlier this year the CSBR launched a policy programme around AI and healthcare. We held a roundtable in the House of Lords in February which considered some key questions and helped identify the six key topics for the policy programme which have been identified as:

- Regulation and governance
- Data security, access, and management
- International benchmarking and lessons learnt.
- Breaking down barriers to innovation
- Cyber security, infrastructure, and resilience risks
- Education and training for more effective adoption and deployment

This document is part of a series of discussion briefs from this policy programme.

This policy brief was produced in collaboration with Mike Gillespie CEO of Advent IM who is leading on this policy work in partnership with the CSBR.



### Executive Summary

In this briefing document we look at some of the key Cybersecurity, Governance and Risk Management issues associated with introducing Artificial Intelligence (AI) Technology into the National Health Service (NHS) and consider the implications for use within wider healthcare settings.

### **Risks and Issues Needing Consideration**

#### **1. Key Cybersecurity Risks**

- **Increased Attack Surface:** AI systems introduce new endpoints and data flows, making NHS infrastructure more vulnerable to cyberattacks. The resultant aggregation of data will also make systems more attractive to attackers.
- **Sophisticated Threats:** AI can be both a tool for defence and a vector for advanced threats (e.g. adversarial attacks on diagnostic models).

- **Poor Cybersecurity Hygiene and Legacy Systems:** Many NHS systems still rely on outdated infrastructure and software, increasing susceptibility to breaches. Additionally, the NHS has, in the past, failed to adequately maintain its systems resulting in cybersecurity vulnerabilities being exploited. The WannaCry ransomware attack was successful due to a combination of factors, including the exploitation of unpatched vulnerabilities in outdated Windows operating systems, a lack of comprehensive IT security practices, and the ransomware's ability to spread rapidly.
- **Need for Continuous Monitoring and Response:** NHS Digital's Cyber Security Operations Centre provides 24/7 threat monitoring, but local organisations must also maintain robust defences, supported by effective Protective Monitoring and a robust Incident Management mechanism. Not all NHS Trusts adequately invest in these areas.
- **Supply Chain Vulnerabilities:** On several occasions over the years, the vulnerability has not always lain within core NHS Systems, but rather within those of their suppliers. To note, a ransomware attack on Synnovis, a pathology provider for the NHS in London, caused significant disruptions to services, including delays in blood tests and postponed appointments.

## 2. Key Data Management Risks

- **Data Privacy and Consent:** AI requires large datasets, raising concerns about lawful processing, patient consent, and secondary data use under UK GDPR. There are also concerns about Data Sovereignty.
- **Bias and Fairness:** Poorly curated datasets can lead to biased AI outcomes, disproportionately affecting vulnerable populations. A failure to adequately invest in the Integrity of Data before deploying any AI based analysis system could have fatal results. The Public Sector has a particularly poor track record in dealing with Data Integrity.
- **Data Quality and Interoperability:** Inconsistent or incomplete data across NHS systems can degrade AI performance and reliability. There are significant differences between various NHS Trusts, their diagnostic and analytics technologies and the systems utilised within Trusts. The sharing of data between Trusts has been a long-standing challenge.

## 3. Key Infrastructure Risks

- **Digital Maturity Gaps:** Not all NHS trusts are equally equipped with modern digital infrastructure, creating disparities in AI readiness. NHS Trusts have different technologies and there is no national consensus on how technology should be adopted. The What Good Looks Like (WGLL) framework has seven success measures which are applicable to all care settings for Digital Transformation, however, is not uniformly adopted.

- **Scalability and Integration:** AI solutions must integrate with existing systems like Electronic Patient Records (EPRs) and NHS Spine, which can be technically complex.
- **Resilience and Continuity:** AI systems must be designed with fail-safes to ensure continuity of care during outages or system failures.

#### 4. Data Privacy Issues

- **Lawful Basis for Processing:** Under UK GDPR, healthcare organisations must establish a clear legal basis for processing personal data with AI.
- **Secondary Use of Data:** AI systems often require large datasets for training and validation. Secondary use must be transparent, justified, and proportionate.
- **Transparency and Explainability:** Patients have the right to know how their data is used and how AI influences decisions about their care.
- **Special Category Data:** Health data is classified as special category data under UK GDPR, requiring enhanced protections and DPIAs.
- **Data Minimisation and Retention:** Only the minimum necessary data should be used, and it must not be retained longer than needed.
- **Data Sharing and Third Parties:** Robust data sharing agreements are required when NHS data is shared with AI developers or vendors.

#### 5. Implications of Data Retention

- **Legal and Regulatory Compliance:** NHS organisations must follow the Records Retention and Disposal Schedule to avoid GDPR violations.
- **Risk of Re-identification:** Long-term retention of anonymised data can increase re-identification risks.
- **Purpose Limitation:** Retained data must not be repurposed without a lawful basis or consent.
- **Security Risks:** Older data may be stored in less secure systems, increasing breach risk.
- **Ethical Concerns:** Excessive retention can erode public trust in AI and digital health.
- **Operational Burden:** Managing retained data requires significant infrastructure and governance.

## Key Recommendations:

To facilitate the safe and effective introduction of artificial intelligence (AI) into the NHS and broader UK healthcare settings, the government ought to take several strategic actions to maximise its influence and accelerate safe AI adoption.

### Strengthen Regulatory Infrastructure

- The MHRA's AI Airlock sandbox.
  - The MHRA's AI Airlock sandbox is a controlled testing environment for AI tools in healthcare. It enables developers to trial AI systems under regulatory supervision before full deployment in the NHS.
  - The capability of this sandbox should be expanded and the use of the AI Airlock as a permanent regulatory tool for testing high-risk AI systems in controlled NHS environments be institutionalised.
  - Sandbox results should then be used to inform regulatory decisions and refine safety standards.
  - To expand the sandbox the Government will need to:
    - Increase funding and infrastructure to support more projects.
    - Broaden eligibility to include SMEs, researchers, and NHS trusts.
    - Embed regulators and multi-disciplinary panels for real-time feedback.
    - Integrate real-world NHS data and environments.
    - Develop modular, risk-based testing protocols.
    - Promote transparency through public reporting and patient engagement.
    - Align with national innovation strategies and the Regulatory Innovation Office.
  - Learning from International Examples.  
Several countries have launched similar initiatives:
    - EU: Regulatory sandboxes under the EU AI Act, with examples in Denmark and Spain.
    - USA: FDA's Digital Health Software Precertification Program for streamlined AI approval.
    - Singapore: AI Governance Testing Framework with risk assessment tools.
    - Canada: Health Canada's sandbox for digital health technologies.
    - Japan: AI Hospital Project supporting clinical validation and interoperability.
- Update medical device regulations to address adaptive and generative AI.
  - The government, through the MHRA and DHSC, is already undertaking a comprehensive reform of the UK Medical Devices Regulations 2002 (UK MDR 2002). As part of these reforms, the regulation can be further strengthened to meet the unique challenges of AI in healthcare.

- **Define Adaptive and Generative AI in Regulation**
  - Clearly distinguish between static AI, adaptive AI (which evolves post-deployment), and generative AI (which creates new content or decisions) in all AI related regulations and future legislation.
  - Introduce specific regulatory pathways for each category, reflecting their different risk profiles and oversight needs.
  
- **Introduce Lifecycle-Based Regulation**
  - Move from a one-time approval model to a continuous oversight framework. There is already precedence for this in the development of Secure by Design, the NSCS newly developed Principle Based Assurance Scheme. This framework should be designed to ensure that AI systems remain ethical, safe, and aligned with societal values throughout their lifecycle.
  - Require real-time monitoring, performance updates, and re-certification for adaptive AI systems that change over time.
  
- **Publish Clear Guidance and Best Practices**
  - Continue releasing MHRA guidance documents for developers of AI as a Medical Device (AIaMD), including templates for risk assessment, validation, and transparency.
  - Provide regulatory toolkits for SMEs and NHS innovators to navigate the evolving framework.

### **Mandate the adoption of ISO42001, supported by ISO23894 and ISO38507**

The adoption of these standards would significantly enhance the safe and responsible introduction of AI into UK healthcare by providing a structured, internationally recognised framework for governance, risk management, and oversight. They can be used to support the UK's evolving AI regulatory framework, including the MHRA reforms and the upcoming AI Bill and will demonstrate to patients, clinicians, and regulators that AI systems are safe, ethical, and well-managed.

### **Mandate the use of the Government's Secure by Design methodology.**

The UK Government's Secure by Design methodology (Secure by Design) ensures that cybersecurity is not an afterthought, but a foundational element of AI system architecture. This is critical in healthcare, where AI systems often handle sensitive patient data and influence clinical decisions.

The framework aligns with UK GDPR, NHS data governance policies, and ethical standards, supports transparency and accountability, key to maintaining public trust in AI-driven healthcare, encourages threat modelling and risk assessments early in the AI lifecycle (helping, amongst other things, to identify and mitigate risks such as model manipulation, data poisoning, bias amplification and unintended clinical consequences). The methodology also includes guidance on incident detection, response, and recovery, which is essential for AI systems that operate in high-stakes environments like hospitals and other healthcare environments.

Secure by Design also promotes a “security is everyone’s responsibility” culture, encouraging collaboration between Developers, Clinicians, Data protection officers and Cybersecurity teams.

It complements the UK’s AI regulatory reforms, ISO/IEC 42001, and the MHRA’s AI Airlock sandbox. This alignment ensures that AI systems are interoperable, certifiable, and export ready.

Mandating the use of Secure by Design would therefore significantly support the safe, lawful, secure, and ethical introduction of AI into UK healthcare by embedding cybersecurity and risk management principles into every stage of AI system development and deployment.

## **Shape International Standards**

The Government should use the UK’s leadership position in the HealthAI Global Regulatory Network to co-develop global safety, efficacy, and ethical standards for AI in healthcare to maximize its influence and accelerate safe AI adoption. This needs to include the alignment of UK regulations with international frameworks to streamline cross-border AI approvals and foster global trust.

- **Share Early Safety Signals and Best Practices**

Establish real-time data sharing protocols with other member countries to detect and respond to emerging risks.

Share insights from the AI Airlock sandbox and NHS deployments to help other regulators refine their own testing environments.

- **Promote UK Innovation Globally**

Showcase UK-developed AI tools (e.g. for lung diagnostics or cancer care) as case studies for safe and effective implementation.

Support UK health tech companies in navigating international regulatory pathways, boosting exports and global partnerships.

- **Foster Collaborative Research and Trials**

Launch joint clinical trials and validation studies with other pioneer countries to build robust, diverse evidence bases.

Encourage academic and NHS collaboration with international institutions on AI safety, ethics, and performance monitoring.

- **Build Capacity and Knowledge Exchange**

Host international workshops, training programs, and conferences to share UK expertise in AI regulation and deployment.

Create fellowship or exchange programs for regulators, clinicians, and developers across member countries.

---

- **Align with the UK's 10-Year Health Plan**

Integrate global learnings into the UK's long-term strategy for AI in healthcare, ensuring safe, lawful, ethical, and responsible innovation and patient-centred care.

### **Ensure Clinical Validation and Real-World Testing**

Mandate shorter and robust clinical trials and real-world evidence.

Support ongoing monitoring and post-deployment surveillance.

### **Build Workforce Confidence and Capability**

Invest in digital literacy and AI training for NHS staff.

Engage clinicians in AI development and testing.

### **Address Ethical, Legal and Governance Issues**

Implement clear AI governance policies.

Ensure patient data protection and public trust.

## **Appendix 1: Ensuring current Best Practice is adopted.**

Within the healthcare environment, any development, adoption or deployment of AI systems must always follow best practice. The following is by way of example and not intended to be an exhaustive list. Best practice is well documented in many of the standards listed in the section titled **Existing Legislation, Regulation and Standards**.

### **Ensure Robust Security and Privacy by Default**

- **Technical Safeguards:**
  - Use strong encryption.
  - Implement Role Based Access Controls and Multi-Factor Authentication.
  - Maintain audit trails and secure development practices.
  
- **Organisational Policies:**
  - Conduct DPIAs for all AI systems.
  - Provide regular staff training.
  - Establish and test incident response plans.
  
- **Data Governance:**
  - Integrate the concept of Privacy by Design, Privacy by Default, into AI system design.
  - Apply data minimisation, purpose limitation and anonymisation principles to minimise data collection and sharing:
    - Use personally identifiable information only where necessary.
    - Understand the spectrum of identifiability and assess re-identification risks.
    - Use a combination of anonymisation techniques: masking, generalisation, suppression, noise addition, and aggregation.
    - Apply pseudonymisation where full anonymisation is not feasible.
    - Conduct regular risk assessments and reassess anonymisation effectiveness.
    - Follow NHS retention policies.
  - Ensure third-party compliance with GDPR through robust contracts and third-party assurance audits.
  
- **Continuous Monitoring:**
  - Use NHS Digital's CSOC for real-time threat detection.
  - Regularly update security policies.

## Appendix 2: Existing Legislation, Regulation and Standards:

### Legislation

- **UK General Data Protection Regulation (UK GDPR)**
  - Governs the processing of personal data, including health data.
  - Requires a lawful basis for processing, data minimisation, transparency, and accountability.
- **Data Protection Act 2018**
  - Supplements UK GDPR and includes specific provisions for health and social care data.
  - Covers automated decision-making and profiling.
- **Common Law Duty of Confidentiality**
  - Requires patient consent for the use of confidential health information unless there is a legal basis or overriding public interest.
- **Human Rights Act 1998**
  - Article 8: Right to respect for private and family life, which includes data privacy.

### Regulatory Guidance

- **Information Commissioner's Office (ICO) Guidance on AI and Data Protection**
  - Covers fairness, transparency, explainability, and accountability in AI systems.
  - Emphasises the need for Data Protection Impact Assessments (DPIAs) and safeguards for special category data.
- **NHS England & NHS Transformation Directorate Guidance**
  - Provides practical advice on integrating AI into care pathways while maintaining information governance standards.
- **National Data Guardian (NDG) Principles**
  - Advocates for transparency, patient choice, and data security in health and care data use.

### Standards and Frameworks

- **UK Government's AI Regulation White Paper (2023–2025)**
  - Proposes a context-specific, pro-innovation framework for AI regulation.
  - Encourages sector regulators (like the MHRA and CQC) to apply principles such as safety, transparency, and fairness.
- **Artificial Intelligence Playbook for the UK Government**
  - offers guidance on using AI safely, effectively, and securely for civil servants and people working in government organisations.

- supports in better understanding what AI can and cannot do, and how to mitigate the risks it brings.

- **NHS AI Ethics and Safety Framework**

- Aims to ensure AI tools are safe, effective, and aligned with NHS values.
- Includes requirements for clinical validation, bias mitigation, and human oversight.

### **ISO/IEC Standards**

- **ISO/IEC 27001:** Information security management – Provides a framework for establishing, implementing, maintaining, and continually improving an Information Security Management System (ISMS).
- **ISO/IEC 42001:** AI Management Systems - provides a framework for organisations to establish, implement, maintain, and continually improve an Artificial Intelligence Management System (AIMS).
- **ISO/IEC 23894:** Risk management for AI - provides guidance on how organizations that develop, produce, deploy, or use products, systems and services that utilize artificial intelligence.
- **ISO/IEC 38507:** Governance implications of the use of artificial intelligence by organisations - provides guidance to enable and govern the use of Artificial Intelligence, emphasises aligning AI initiatives with organisational objectives, promoting accountability, transparency, and ethical considerations.
- **ISO 13485:** Quality management for medical devices (relevant for AI as a medical device).

### **NIST AI Risk Management Framework**

- A structured approach developed by the U.S. National Institute of Standards and Technology to help organisations manage risks associated with artificial intelligence.
- It is especially relevant for sectors like healthcare where trust, safety, and accountability are critical.
- The framework aims to promote the trustworthy and responsible design, development, deployment, and use of AI systems by helping organisations identify, assess, and manage AI-related risks.

### **MHRA Guidance on AI as a Medical Device (AIaMD)**

- Regulates AI tools that meet the definition of a medical device.
- Requires conformity with UKCA marking and clinical safety standards.

### **NHS Digital Technology Assessment Criteria (DTAC)**

- A framework for assessing the safety, security, and suitability of digital health technologies used in the NHS and social care.

- It ensures that these technologies meet national standards and gives staff, patients, and citizens confidence in their use.

Ends