Written Evidence submitted by The AI Centre

(GAI0037)

The AI Centre was established in 2019 and is a consortium of hospitals, universities and companies researching and deploying healthcare AI. Our governance is overseen by a committee of public representatives, researchers, clinicians, and information governance representatives including Caldicott Guardians and data protection leads. Our responses will regard healthcare AI only.

A summary of our response:

1. **How effective is current governance of AI in the UK?**
   Response: current governance of AI (artificial intelligence) in healthcare is inadequate across the AI lifecycle. Healthcare AI development needs best practice in the form of guidance; best practice built into the infrastructure, as we have done with FLIP (Federated Learning and Interoperability Platform); and more resources to put the governance into practice. Furthermore, deployment and monitoring of this AI needs a clinical governance strategy.
2. **What measures could make the use of AI more transparent and explainable to the public?**
   Response: explainability and transparency can be supported in three ways. First, by having a national dialogue to clarify the definition of AI, because the public need to understand what AI is if they are to understand what AI does. Second, by defining explainability, and third by enforcing public involvement.
3. **How should decisions involving AI be reviewed and scrutinised in both public and private sectors?**
   Response: reviewing healthcare AI involves evaluating evidence generated on its use. However, generating evidence to evaluate AI in the NHS currently requires expensive, ad-hoc infrastructure, which is a problem that can be addressed by AIDE (AI Deployment Engine).
4. **How should the use of AI be regulated, and which body or bodies should provide regulatory oversight?**
   Response: regulation and regulatory bodies already exist, but the ecosystem that supports them needs to be more comprehensive and mature before regulation is efficiently enforced. For example, the backlog of applications to the notified bodies must be addressed.
5. **To what extent is the legal framework for the use of AI, especially in making decisions, fit for purpose?**
   Response: the legal framework for data protection is fit for purpose, but Article 22 of UKGDPR (2021) needs more guidance. Article 22 addresses automated processing and profiling. Also, the Medical Device Regulation needs updating.
6. **What lessons, if any, can the UK learn from other countries on AI governance?**
   Response: As each regulator was not specifically set-up for AI, there may be utility in a UK AI Board that regulators can submit AI questions to.

How effective is current governance of AI in the UK?

(What are the current strengths and weaknesses of current arrangements, including for research?)

The healthcare AI lifecycle can be partitioned into 4 stages: *prototyping, evaluation, deployment and monitoring, and decommissioning*. Existing governance frameworks in healthcare set a starting point for AI governance, but they inevitably leave gaps because they were not specifically designed for AI.

Best practice

AI governance frameworks for the first two stages (*prototyping and evaluation*) can borrow from established healthcare research practice. Research ethics enforces principles of good practice such as research integrity, beneficence, and privacy. Some of these principles take form in laws and other legal basis such as the Common Law of Duty of Confidentiality, Data Protection Legislation (including EU (2016) and UK (2021) General Data Protection Regulations and UK Data Protection Act (2018), Medical Device Regulation (2002), the Equality Act (2010), and the Human Rights Act (1998); and other principles are enforced in guidance and best practice. Additionally, the UK Policy Framework for Health and Social Care Research establishes oversight of health research using NHS data. These frameworks enable good, ethical research but they do not support best practice in AI.

For example, take the principle of integrity. Standard practice is for a healthcare research project to archive their raw data and documents at the end of a project in case the research is questioned in the future. If the results are questioned, the data and analysis method can be retrieved, inspected, and used to repeat the research. If the data, method and results are valid then the research maintains its integrity. However, this approach does not work so neatly for healthcare AI training research.

The current data archiving approach does not work for healthcare AI because:

- To replicate the research you need a copy of the AI, which could be protected from archiving with Intellectual Property terms.
- AI tends to be trained on multiple versions of large datasets, which dramatically increases the cost of archiving, making it potentially unfeasible.
- AI training is impacted by the system it is trained in, so you need to archive a snapshot of the system.
- Unstable AI can produce different results when exposed to the same data in two different tests.

For this reason, the AI Centre have been consulting with regulators, AI scientists, clinicians, data protection experts and the public to set standards for healthcare AI governance framework gaps, such as archiving, which we will be publishing on our website.

Innovative platforms

Best practice can also be built into the platforms supporting the AI lifecycle. For example, training AI often requires large quantities of data which presents several governance challenges. First, data needs to be sourced from several institutions, with idiosyncratic governance structures. Second, generally, the larger the dataset, the greater the privacy risk. Third, curating a dataset requires considerable amounts of time from NHS clinicians, analysts, IT and information governance teams. If you cannot curate a dataset then you cannot create acceptable AI.

For this reason, the AI Centre has set-up several hospitals with FLIP (Federated Learning and Interoperability Platform). FLIP supports curating data to the same standard across participating hospitals without the data ever leaving its originating hospital. Researchers can then view high-level, anonymous statistics about the available data in a user interface. If a research project finds appropriate data and is approved to use it, then it can securely send AI to be trained in participating hospitals. It unlocks the potential of data whilst decreasing privacy risks and respecting the varied governance frameworks of participating hospitals. However, such a platform only works if appropriate staff time is available, and appropriate oversight exists.

The challenge of the former will continue as long as NHS staff are under-resourced and lack relevant training. Regarding the latter, a project can only use FLIP if it has approval from an oversight committee. The oversight committee comprises of data protection leads from the participating hospitals, scientists, clinicians and patients. The committee will review the project contract and Data Protection Impact Assessment to assess the project's value. Areas of review will include the quality of the projects input from the public, independent scientific review, purpose, and risk mitigations. After the necessary project updates the committee will vote on whether it should commence. The project will only commence in hospitals with their explicit approval.

Clinical governance

The AI governance frameworks of the last two lifecycle stages (*deployment and monitoring, and decommissioning*) can borrow from established healthcare frameworks, previously mentioned, and clinical governance. Clinical governance is the framework that holds NHS institutions accountable for the continual improvement of service quality and the safeguarding of high standards of care, following the 7 pillars of clinical governance:

1. Clinical effectiveness and research
2. Audit
3. Risk management
4. Education and training
5. Staff management
6. Patient and public involvement
7. Information and IT

Each Trust interprets the clinical governance requirements locally, but it is enforced during Care Quality Commission (CQC) inspections - minimum requirements include an annual audit and representation at the board level. The impact of cutting-edge clinical AI software brings complex new challenges that cannot be effectively managed by the current clinical governance structures in place across most NHS organisations.

1 - Clinical effectiveness

The gaps in current clinical governance arise immediately when directly comparing AI application management to traditional, rules-based clinical software. There is no guarantee of achieving the marketed performance of AI applications once clinically deployed, and review of clinical effectiveness during commissioning may require comparison of AI training data set against local populations to anticipate its suitability.

2 - Audit

The scale of this workload is often underestimated when resources are initially allocated, which can result in ongoing challenges in performing annual audits effectively. The naivety of the "AI premium" in deployment and maintenance activities means that AI initiatives can end up under-resourced and therefore prone to failure. Formal guidelines to evaluate healthcare AI technologies are yet to be published, and compliance will further lag behind until experience is obtained and shared. Audit frequency is determined by a risk assessment of the severity of incorrect outcomes, where higher audit frequencies directly after deployment can be reduced as confidence improve. AI performance can be sensitive to changes in input data so maintaining oversight is important to stop missed errors affecting large numbers of patients. It is analogous to radiologists reporting on X-rays of diminishing quality: where the culture of conversation/gossip/complaints between staff identifies errors quickly, AI applications cannot communicate unforeseen errors without being explicitly programmed to do so, highlighting the importance of audit.

3, 4 and 5 – Risk management, education and training, and staff management

As with any part of good clinical governance, the workforce should be appropriately educated and trained. However, AI training opportunities are minimal for most Allied Health Professionals and non-clinical staff who will interact with, and feel the effects of, AI applications, such as patient triage, diagnoses, and automation. The current digitalisation of the NHS aims to set the landscape for automation and AI, but real benefit will be revealed when the staff can identify how these technologies can improve work in their area. Initiatives like the Topol Digital Fellowship and the Fellowship in Clinical AI look to empower clinical staff and improve access to technology training.

AI applications pose new clinical risks above normal software, but these considerations are not incorporated into general clinical risk management training. Risks such as missed diagnoses due to training data bias, poor algorithm performance, overreliance on AI, changing demographics, and changing imaging techniques, need to be explored and worked examples provided.

6 – Patient and public involvement

Patient and public involvement (PPI) in introducing AI to hospitals is important for prioritising patient care, preventing the erosion of trust between patients and aligning healthcare AI with the goals of equality, diversity and inclusion. However, asking the public to make informed decisions on such complex software is difficult considering the layperson's current AI literacy. This literacy gap must be filled in order for the public to fully engage with PPI.

7 - Information and IT

The clinical governance of IT systems and infrastructure within a health institution will need to extend to include a range of independent AI systems that interact with it in differing capacities. The extent to which the operational scope of existing IT systems and infrastructure faces an "AI premium" is dependent on the digital maturity of the health institution, as data producers and controllers, and how well-equipped their systems are to meet the data demand for these AI systems

(as opposed to standard reporting). This may include varying levels of digital transformation, such as implementing improved electronic health records (EHR), increased physical and virtual storage, and data cleaning and enrichment capabilities, and may require significant upfront human and capital investment before the AI systems are fully deployed and operational. Connecting on-premises and cloud-based AI applications requires oversight from Information governance and Information safety, and additional risk management for data transfer outside the Health Institutions IT domain. There are additional considerations for vendors that act as data processors/data controllers; the types of data protection agreements required are similar to large enterprise level systems, and management of these agreements will become a growing burden to IT departments. Processing data through AI models can require large computational power which is often provided by dedicated hardware installed onsite, which further adds to the equipment that an IT service needs to manage.

Clinical governance summary

These issues are all summarised by the challenges of performing effective audits. Having the technical skills and knowledge to analyse an AI application is hard enough, but doing this routinely on all AI applications deployed across a health institution requires a clear strategy in terms of equipment, staff, training, Trust IT capacity, clinical benefit and workforce planning. Top management must be able to interpret audit results effectively in terms of workforce planning, having an informed risk appetite, and awareness of AI limitations - only then can true accountability be demonstrated.

What measures could make the use of AI more transparent and explainable to the public?

There are three measures that could help with transparency and explainability: defining AI, defining explainability, and enforcing public involvement for effective transparency.

Defining AI

A large problem for explainability is the public narrative surrounding AI. The AI reference point for many people is popular media: films like the Terminator and news stories about international technology companies. These reference points highlight some valid concerns, such as how is data protected, but they also distract from other valid concerns by focusing on science fiction. The public need to understand what AI is if they are to understand what AI does. Then there can be a conversation regarding the limits of an AI product, where it will be inserted in the patient pathway, and how the patient can share concerns about the AI's outputs. At the AI Centre we have invested resources into understanding how to discuss healthcare AI with the public, but this only impacts the select public that we engage. A larger Government funded program would be needed to impact the nation.

Defining explainability

A standard for explainability must be set. For example, most people think explaining AI means explaining how it works, but this is not always possible for clinical products. For example, we do not know how all drugs work, which is why we undertake drug clinical trials and have post-market

reporting - to gather evidence on who can take a drug and what adverse effects they can expect. We need the same standard for healthcare AI: a focus on safety over an understanding of how it works. For this reason, at the AI Centre, when we deploy AI with AIDE, we deploy it with an information packet, called a model card, that describes uses and limitations of the product.

Enforcing public involvement for effective transparency

Representative public involvement in healthcare AI development is best practice, but not mandatory. It is best practice because it builds a relationship with the public which in turn cultivates trust, and in the long run improves the design of the AI. It achieves this by discovering information about the patients' experience of the NHS that is not captured in the dataset, because not everything is recorded or quantifiable. Understanding *what* the public values and *how* to discuss it creates effective transparency. Transparency must not be a tick-box exercise achieved by a large block of text, because effective and meaningful transparency ensures the information is relevant and understood.

How should decisions involving AI be reviewed and scrutinised in both public and private sectors?

(Are current options for challenging the use of AI adequate and, if not, how can they be improved?)

Testing of the current options for reviewing and scrutinizing healthcare AI are in their early stages. When there is enough evidence on a healthcare product it can be reviewed by hospitals using frameworks like the NICE Evidence Standards Framework. However, few products have enough evidence for an effective evaluation, and the evidence produced for other products is often not to the standards expected in the clinical community. There is a real concern that the evidence standards for healthcare AI will remain unacceptably low.

One reason for the issue with evidence standards is the cost of deployment. Currently, when you want to deploy AI into a hospital in order to evaluate it, you need ad-hoc technology to install the algorithm. This is expensive in both time and money. For this reason the AI Centre has developed AIDE and installed it in several hospitals to standardize AI deployment and make it accessible.

AIDE is an open-source, enterprise-grade software platform designed to seamlessly integrate with existing hospital clinical information systems. The entire AIDE ecosystem was designed around transparency of deployed AI applications to engender trust in users of the platform. AI applications can be reviewed and scrutinised on multiple levels within AIDE.

1) All applications on AIDE are featured within the AIDE App Store, where they each have a store page which provides all manner of information about the application, for example, how the underlying AI works, what training data was used to create the AI algorithm, and what are the limitations of the AI.

2) AI applications run in hospitals can be passed through AIDE Clinical Review. Clinicians use this graphical user interface to review the performance of any AI applications installed at their hospital. They can choose to accept or reject the result of the AI application based on its performance. These decisions can be collated and used as evidence generation.

3) AIDE provides continuous longitudinal oversight of all applications deployed within the host clinical environment via the AIDE Admin Console. This feature enables local AIDE administrators to review every run of every single AI application. For longitudinal monitoring, this information can be gathered for statistical analysis, which may help with identifying changes in usage patterns – perhaps if the clinicians no longer favour the AI application – or changes in performance of the AI application over time.

How should the use of AI be regulated, and which body or bodies should provide regulatory oversight?

A framework to regulate taking healthcare AI to market, and monitoring its data use, already exists. Data use is overseen by the Information Commissioner's Office (ICO), and higher-risk products are evaluated by the Medicines and Healthcare products Regulatory Agency (MHRA) or an MHRA monitored body, before they enter the market. However, there are two issues. The first issue is resources: there are currently only three notifying bodies, who have a tremendous backlog, preventing innovative products from entering the market. The second issue is the maturity of the current ecosystem.

AI presents new regulatory issues when bringing software to market. The current regulation, UK Medical Device Regulation (2002), lacks significant clarity on the additional requirements for medical software, as opposed to the physical products. Taking a medical software device to market requires building the device within a quality management system (QMS), having a qualified independent Clinical Safety Officer to perform clinical risks management activities through the software development lifecycle, and conduct a clinical investigation to prove the advertised performance of the product. AI applications have additional considerations at each of these key stages.

AI products often start their lifecycle in academic institutions without a formalised quality management system which asserts best design and development principles and generates the required technical files. Retrospectively putting mature AI software through a QMS has significant cost, workload burden, and requires regulatory expertise. Similarly, involving a Clinical Safety Officer is often an oversight when creating a hazard log during development, so the gaps in the intended clinical use, benefit, and safety may not be realised until much later in the development process. The clinical investigation to prove intended performance involves engaging a healthcare institution through an R&D contract which can take a long time to get approvals from the departments for information governance, information safety, and Trust IT. Additionally, data protection agreements and data sharing contracts must be signed and curation within a health institution can become costly.

Where academic institutions wish to deploy AI software clinically within an NHS Trust, there can be a misunderstanding on the application of the "in-house exemption", where non-CE/UKCA marked products can be used clinically if developed by the Host healthcare institution. This entails the healthcare institution becoming the manufacturer of the device, therefore taking on all the clinical risk, responsibility for deployment, ongoing maintenance, clinical investigations, validation tests, frequent audit and IT infrastructure costs. This is a significant burden which will likely require a new department dedicated to medical software development in hospitals to replicate the benefits and safety provided by a company supplying a CE/UKCA marked product.

Once on the market with a UKCA/CE mark, proof of expected performance in routine clinical settings will often still be required by health institutions, especially on their local patient demographics. They may require this to be completed even if an independent clinical investigation has been performed on the product, if it cannot be convinced that the training data is congruent with institution's patient population and use case. The manufacturer should be prepared to wait for a considerable time, in some cases years, for this satisfactory evidence before signing contracts. There is a responsibility for AI software developers to perform post market surveillance on all deployments at regular intervals to ensure the product performs as intended. Where an AI application fails and causes patient harm, the incident will be reported to the MHRA which will trigger an investigation into the product.

To what extent is the legal framework for the use of AI, especially in making decisions, fit for purpose?

(Is more legislation or better guidance required?)

As previously mentioned, due to the maturity of the ecosystem it is hard to gauge how fit for purpose the legal framework is for healthcare AI. However, best practice for AI must be established as the current healthcare governance frameworks do not adequately address AI development and deployment. The AI Centre has been working closely with industry and the public sector to create best practice for each gap we find.

The current legal framework for data protection is fit for purpose, however more guidance should be provided for how UKGDPR Article 22 applies, for example, by defining the term "significant affect" in the healthcare context. Finally, the UK Medical Device Regulation (2002) needs updating – which is currently underway by the MHRA.

What lessons, if any, can the UK learn from other countries on AI governance?

The EU GDPR set the gold standard for trustworthy data use, and a lesson the UK must learn is that high data protection standards need to be maintained if data use by AI is to be supported by the public. High data privacy standards cultivate trust from the public, and changing data protection standards, as proposed with the recent Data Protection and Digital Information Bill, could damage the trust that has been earned.

The draft EU AI Act establishes an AI Board to provide advice. Given each regulator in the UK was set-up for a purpose other than AI, it may be helpful to establish a UK AI Board where AI questions can be escalated for guidance. The Board can have representation from each regulator's AI team, experts and observers, and should increase standardisation and reliability across regulators.

(November 2022)