Putting the United Kingdom at the forefront of using artificial intelligence for the benefit of health and care: a mixed methods formative evaluation of the NHS Artificial Intelligence Lab

April 2025

Kathrin Cresswell, Professor of Digital Innovations in Health and Care, Usher Institute, University of Edinburgh

Robin Williams, Professor of Social Research on Technology, Institute for the Study of Science, Technology and Innovation, University of Edinburgh

Sheena Dungey, Senior Health Economist, NHS Arden GEM Commissioning Support Unit

Stuart Anderson, Professor of Dependable Systems, School of Informatics, University of Edinburgh

Miguel O. Bernabeu, Professor of Computational Medicine, Usher Institute, University of Edinburgh

Hajar Mozaffar, Senior Lecturer in Innovation, Business School, University of Edinburgh

Xiao Yang, PhD Student, School of Informatics & Institute for the Study of Science, Technology and Innovation, University of Edinburgh

Varun Sai, PhD Student, Usher Institute, University of Edinburgh

Sara Bea, Research Fellow, Usher Institute, University of Edinburgh

Sally Eason, Associate Director, NHS Arden GEM Commissioning Support Unit

Ray Foley, Project Coordinator, NHS Arden GEM Commissioning Support Unit

Table of Contents

How to read this document

This document reports on the background, methods, findings, discussion and conclusions of the independent evaluation of the NHS AI Lab. It is structured as an academic publication to help readers understand the scientific rigour of the work and describe what it adds to the existing literature. Drawing on the helpful example of the Ada Lovelace Institute on guiding readers to get the most out of this work,[1] we recommend the following:

Everyone, including members of the public

The executive summary (approximately 10 minutes reading time) presents the background, methods, and key lessons emerging from this work. We have produced two versions of this: one in lay language and one for a professional audience. If you are just interested in the key lessons, then please scroll to the last paragraph of the executive summary labelled "Summary of lessons identified".

Policy makers, strategic decision makers, health and care staff, system developers and suppliers

These audiences are likely to be most interested in the results (approximately 20 minutes reading time) and the discussion section (approximately 20 minutes reading time) detailing implications for policy and practice. If you have limited time, then we recommend focusing on the discussion section, as this also includes a summary of key findings.

[1] We have adapted this guide from: https://www.adalovelaceinstitute.org/report/new-rules-ai-regulation/ (last accessed: 01/01/2025).

Foreword

There is considerable global interest in developing and implementing Artificial Intelligence-based systems in health and care settings. However, we lack published evidence on how governments and healthcare systems can effectively intervene in this space to drive outcomes.

A pioneering initiative recently concluded in the United Kingdom (UK) where the National Health Service (NHS) Artificial Intelligence Lab (AI Lab) was set up as a five-year programme with £250 million of national funding to stimulate the development and adoption of AI-based systems in health and care settings.

We had the privilege to conduct a substantial, independent national evaluation of this programme exploring processes, impact, and value as well as extracting lessons identified to inform future efforts in AI and in other governmental programmes.

This work would not have been possible without the thoughtful guidance of Dominic Cushnan and Dr Alison Tweed who have patiently answered all our queries and requests for information. They also made every effort to allow us to maintain our independence whilst opening doors and connecting us with a range of stakeholders.

We also thank our Independent Advisory Board who have provided invaluable guidance: Professor David Bates, Professor Elske Ammenwerth and Professor Farah Magrabi.

All stakeholders we have interviewed and asked for information have, despite many pressures, willingly given their time and provided constructive and honest feedback for which we are incredibly grateful.

This evaluation presents an effort to curate evidence arising from specific projects investigating the impact of AI in health and care and surrounding the set-up and challenges of major governmental programmes. Learning from these initiatives is likely to be invaluable for the UK and beyond. These findings are being fed back to the new unit developing NHS AI strategy. We are deeply grateful for this opportunity to help shape evidence-based policy making.

Nevertheless, despite the significant effort and resources involved, this work only presents the beginning of our health and care system's transformation journey through AI. Key opportunities are still to be realised, and key challenges are still to be overcome.

We are looking forward to an exciting future – shaping it effectively will require evidence-based decision making and learning lessons from history.

Kathrin Cresswell (on behalf of the evaluation team)

Accessible executive summary (lay language)

Background

- In 2019, National Health Service (NHS) England jointly with the Department of Health and Social Care (DHSC) started a programme called the NHS AI Lab to help bring Artificial Intelligence (AI) into health and social care safely and effectively. It was the first programme like this in the world. The government also wanted to use AI to help grow the economy.
- At first, the programme was given £250 million, but in 2022 this was reduced to £143.5 million after a budget review. The programme aimed to improve health and care by focusing on four main areas:
    1. **Building national infrastructure**: Creating systems like a database of chest scans for COVID-19 and a platform to help use AI in healthcare.
    2. **Bringing people together**: Connecting experts, organisations, and other countries to share ideas and work on AI projects.
    3. **Testing and expanding AI**: Testing AI technology in NHS settings to find out what works and helping spread successful tools.
    4. **Rules and ethics**: Making guidelines to ensure AI is used safely, fairly, and responsibly without putting excessive burdens on companies producing AI tools.
- The goal was to find ways AI could make long-lasting improvements to health and care and make sure it is used in the right way.

The evaluation

- Between March and December 2024, we carried out an independent review of the NHS AI Lab to see how it worked, what impact it had, and whether it was valuable. Most of the work involved looking back at what had already been done although we are feeding into discussions about future AI strategy.
- Here is what we did:
    a. **Reviewing documents**: We went through 1021 documents, including plans, reports about lessons learned, meeting notes, evaluations of technologies, and records of benefits. Additionally, we also reviewed the different sections of the FutureNHS Website including case studies, forum, show and tell videos.
    b. **Interviews**: We spoke to 85 people involved in the AI Lab, such as current and former staff, decision-makers, technology developers, NHS workers, and others who evaluated aspects of the AI Lab's work.
    c. **Observations**: We observed 12 meetings and workshops.
    d. **Analysing information**:

   i. For written and interview data, we used a framework called TPOM, which looks at how technology, people, organisations, and the wider environment interact.
   ii. To understand the value and impact, we used the Triple E Framework, which focuses on economy (cost), efficiency (how well resources are used), and effectiveness (how well a programme achieves its goals).

e. **Looking at results**: We paid close attention to the AI in Health and Care Awards, which received the most funding and demonstrated the clearest results. We focused on technologies that were advanced and could show clear health, economic and other benefits for the NHS.

Our findings

- The AI Lab worked on identifying what helps and what gets in the way of using AI in health and social care. The work of the AI Lab has helped to place the United Kingdom (UK) as a leader in this area. If managed well, the foundations set by the AI Lab could continue to bring benefits in the future.
- **Progress and learning**: The AI Lab made good progress in experimenting with AI, and the biggest benefit has been the lessons identified. Support and coordination have been particularly useful, even though it is hard to measure their full impact.
- **Returns on investment**: Research in advanced AI has shown good results. But not all results are available yet, and it is hard to predict how valuable these new technologies will be in the future. Outcomes are also still being reported.
- **Challenges from a changing environment**: Progress has been slowed by big changes, such as the COVID-19 pandemic, changes in leadership, government spending reviews, and restructuring within government. Many groups were involved in coordinating the AI Lab, often with conflicting goals, which made things more complicated.
- **Scaling and adoption issues**: Expanding the use of AI has been tough, partly because it is unclear how to buy and use these technologies within the NHS. National plans sometimes did not match what the health and care system specifically needed. For example, many AI projects focused on popular areas like diagnostics but did not always address bigger system needs, such as helping with resource planning.
- **Long-term benefits**: Many benefits from the AI Lab will take time to show and are hard to measure, such as the impact of regulatory changes. However, there are also some near-term benefits which are easy to measure (e.g. those associated with doing things faster). Some of the most advanced projects funded by the AI Lab have already delivered substantial benefits – evidenced through the targeted funding for in-depth evaluation. One successful diagnostic AI Award provided £44 million in benefits (resulting from

increased rate and timeliness of critical treatment associated with improved patient outcomes and reduced ongoing health and care need). This was much more than the £1.9 million it cost and more than the £25.6m total awarded funding to the subset of eleven AI Awards of the highest maturity level which had completed in the timeframe of evaluation. Some projects also helped improve NHS clinical processes in line with the 2019 Long Term Plan (e.g. increasing the rate of mechanical thrombectomy for stroke patients to 10%).

Lessons identified

- Here are seven key lessons we identified about using AI in healthcare:
    1. **National support is key**: To successfully use AI in health and care, we need support from the entire system. This includes linking research with real-world use, building skills, sharing knowledge, creating the right infrastructure, and setting clear rules for using and managing AI. A coordinated effort is crucial to make AI solutions scalable and long-lasting. This also needs to involve helping to free up local resources, so organisations have capacity to implement and adopt technology.
    2. **Strong leadership and long-term planning**: A clear vision for the future, supported by stable leadership, is essential. Leadership changes should not disrupt progress. Long-term plans and strategies are needed to keep AI projects on track and ensure steady learning and improvement.
    3. **Focus on real needs**: AI should be designed to meet the needs of health and care providers and patients. Frontline staff, service providers, and the public must be involved in shaping AI solutions. If AI tools do not align with real-world needs, they will not deliver best outcomes and may not be used.
    4. **Transform, do not just automate**: Most of the AI Awards delivered automated discrete tasks - and AI can make some tasks faster and more accurate. Only a few AI Awards transformed existing care pathways - these offered much more substantial improvements in quality and efficiency of care. For this to work, AI tools need to be linked to changes in how care is provided, and this requires thoughtful planning and regulation.
    5. **Evaluate AI**: Every AI project and programme needs to be carefully assessed to understand its use and impact in the context in which it is to be deployed. Long-term studies using different methods and following the project in real-time are important to track progress, monitor risk and make evidence-based decisions about which technologies to use and how.
    6. **Balance risks and benefits**: AI is advancing quickly, so we need to carefully weigh its potential benefits against the risks and the evidence available. Developing effective

ways to gather evidence will help make informed decisions and increase the chances of safe and effective use.

7. **Build on what we have learned**: The work of the AI Lab has created a valuable resource of knowledge and experience. It is important to use what we have learned to guide future AI strategies and keep the momentum going.

- **The Risk**: If we do not learn from the AI Lab's experiences, disseminate them, act on them, and build on them, we risk losing the opportunity to make the most of what has been achieved so far. To do this, we need to continue analysing the data generated through the AI Lab's activities and see this initiative as the beginning of the journey towards the long-term transformation of health and care. Enabling and tracking scaling is a key next step if benefits are to be realised.

Executive summary

Background

- In 2019, National Health Service (NHS) England jointly with the Department of Health and Social Care (DHSC) launched a national Artificial Intelligence (AI) programme with the aim to accelerate the safe adoption of AI in health and social care settings (known as the NHS AI Lab).
- The AI Lab was the first of its kind in the world, with the goal of increasing adoption of AI in the NHS buttressed by a political ambition for the United Kingdom (UK) to stimulate the economy through AI.
- The programme was initially allocated £250 million national funding, although this was cut to £143.5 million, following a spending review in 2022.
- It sought to explore the potential role of AI through a systemic intervention involving several interlinked strands of work: **infrastructures** (National COVID-19 Chest Imaging Database, AI Deployment Platform); **community building** (communities of practice, international work, bringing together sectors, Skunkworks focusing on communities surrounding early technology development); **research and evidence** (AI in Health and Care Awards, a £120 million investment with the aim to build a real-world testing environment and accelerate spread and adoption of proven AI in the NHS); and **governance and ethics** (AI policy and regulation, AI ethics initiative).

The independent evaluation of the AI Lab

- We sought to assess how the AI Lab was implemented; its outcomes, impacts and value; and lessons learned.
- We conducted a retrospective independent formative and summative evaluation of the AI Lab between March and December 2024.
- We analysed 1021 project documents, conducted 85 semi-structured interviews with key stakeholders and observed 12 meetings and workshops. Documents included business cases, lessons learned reports, meeting minutes, evaluation reports of technologies, and benefits registers. We interviewed current and past AI Lab staff, local and national strategic decision makers, technology developers, NHS providers, and evaluators who were involved in the AI Lab.
- We analysed qualitative data with the help of the Technology, People, Organizations, and Macroenvironmental factors (TPOM) framework and within this assessed impact and value with the Triple E Framework (economy, efficiency, and effectiveness).

- The quantitative work focused on exploring the AI in Health and Care Awards, which received the majority of funding and generated evidence of outcomes. We focused on technologies with high Technology Readiness Levels (TRLs) where financial and non-financial benefits could be reliably estimated at a systemic level as part of the Award.

Key findings

- The work of the AI Lab has ensured that the UK remains a key player in this field and, if appropriately managed, the foundations established by the AI Lab will continue to deliver benefits in the future.
- The AI Lab has made considerable progress in an experimental area, with the primary benefit being the learning opportunities gained through these efforts. Support and coordination functions have proven to be highly valuable, but the impacts of these long-term systemic interventions are challenging to quantify.
- The AI Lab highlighted barriers to, and enablers for, the development, implementation and adoption of AI technologies in health and social care.
- Promising and reliable estimates of returns on investment (RoIs) have been identified, particularly in relation to research investments in relatively mature technologies and where technological system implementation was accompanied by process and pathway changes. However, the outcomes of experimental investments are difficult to demonstrate in the short-to medium term, and at the time of writing, not all outcomes have been reported and estimating the long-term value of potentially transformational technologies is highly uncertain.
- The turbulent macro-environment has presented notable challenges to progress, resulting in changes in leadership, objectives, personnel, and components of the AI Lab over time. Factors such as the COVID-19 pandemic, ministerial changes, organisational restructuring, and spending reviews have all contributed to the turbulence of this landscape. Additionally, the involvement of multiple stakeholders with conflicting needs and agendas necessarily complicated routes to progress.
- Scaling AI technologies has proved difficult, partly due to uncertainty surrounding procurement pathways, making it difficult to establish clear processes for adoption. Enabling and tracking scaling is a key next step if benefits are to be realised. In addition, aligning national strategy with the specific needs of the NHS remains a critical but complex task, that is crucial for future adoption. The targets for applying AI were driven by clinical and commercial interests (e.g. diagnostic AI), which were not always aligned with biggest potential benefits and system needs (e.g. AI facilitating resource allocation). However,

there is now a much better understanding of where and how AI may be applied to deliver improved efficiency and quality of care.

- Many benefits are likely to materialise over the long term and are hard to measure (such as the benefits of regulatory reform and advancing early innovations). However, there are also some benefits in the near term which are easy to measure associated with automation. Some of the most advanced projects funded by the AI Lab have evidenced significant benefits enabled by targeted funding for in-depth evaluation. One successful diagnostic AI Award provided £44 million in benefits (resulting from increased rate and timeliness of critical treatment associated with improved patient outcomes and reduced ongoing health and care need). This was much more than the £1.9 million it cost and more than the £25.6m total awarded funding to the subset of eleven AI Awards of the highest maturity level which had completed in the timeframe of evaluation. Some projects also helped improve NHS clinical processes in line with the 2019 Long Term Plan (e.g. increasing the rate of mechanical thrombectomy for stroke patients to 10%). Many of the most successful projects involved clinicians with an intimate knowledge of the pathway the AI device was deployed into. In these cases, the change tackled issues in the existing pathway without radical redesign. This led to beneficial change that could be quantified and evaluated within short- to medium-term timeframes. Issues, however, remain over how to effectively evaluate transformational change.

Summary of lessons identified

- Our insights highlight seven key lessons:
    1. **National support** and **system-wide intervention** are crucial for linking research and deployment, ensuring scalability and sustainability of AI initiatives. There is a clear need for a coordinated approach led by Government to foster innovation that is informed by the realities of practical implementation, including developing capabilities, building knowledge-sharing and innovation communities, establishing infrastructures, supporting implementation, and developing regulation and market management strategies. This also needs to involve understanding and resourcing local capacity and capability to implement technology.
    2. A **long-term vision of transformation and stability of leadership** to work towards that vision is a critical enabler for AI in health and care. This involves creating enduring strategic frameworks and visions that can weather leadership changes, ensuring consistent progress and learning over time.
    3. AI development must be **firmly rooted in system and service user needs**, with active involvement from service delivery organisations, frontline staff, and the

public. Technological strategies and system developments should align with the identified needs of health and care providers. Otherwise, there is a risk that developed systems will not be adopted. Hence work must identify common and cross-organisational needs, at task and at service level.

4. The focus should be on **transformation**. AI may replace some existing tasks to achieve better outcomes (and it can make some tasks much faster and more accurate), but more importantly, AI must be seen as a key tool in the redesign and improvement of health and care pathways rather than merely speeding up existing processes. In turn, AI technologies should be able to adapt to new requirements arising from a redesigned pathway. As such, AI needs to be viewed as part of a sociotechnical intervention, involving education and process re-design. This transformation will require input from service designers and careful regulatory guidance considering domain shift, approaches to retraining, and post-market surveillance. It will also require efforts to promote the creation of a learning ecosystem, improving knowledge flows and engagement/ownership by frontline care providers.
5. **Evaluation** must be a key component of this process. Both individual projects and larger programmes must be rigorously assessed, ideally through longitudinal mixed-methods studies and in real-time. This includes establishing robust baselines and incorporating formative evaluation to mitigate emerging risks. Evidence from evaluations should become central to the process of selecting and scaling technologies.
6. There must be a balance between **risk, benefits, and evidence**. Rapid technological improvements mean that stakeholders must weigh the potential advantages of AI against the need for a comprehensive and pragmatic evidence base. Developing effective methods to generate evidence and real-time post-market surveillance is essential for informed decision-making and successful implementation and adoption.
7. The AI Lab has evidenced the **benefits of a concerted approach** to the safe and effective adoption of AI in the NHS, providing strategic guidance, shared learning, development of communities of practice and a favourable culture.
8. The results of the AI Lab work are a unique asset that is rapidly depreciating as technology continues to advance. It is important to **retain the momentum** of the AI Lab work and **build lessons identified** into future strategy.

- Many of these lessons apply not only to AI but also to other digitalisation initiatives, illustrating the importance of socio-organisational factors in technology development, deployment and adoption. However, AI-enabled medical devices also have some

fundamental features that differentiate them from conventional digital health devices. These include the rapid development, the dependence on data and learning features changes over time.

- A key risk now is the failure to learn from these experiences, disseminate them, and act on them. Outcomes are still being reported and need to be integrated in future work. Going forward there is an imperative to build on the wealth of evidence generated. The AI Lab's efforts are the beginning of a long-term journey towards the future of an AI-enabled health and social care system.

Abbreviations

AI	Artificial Intelligence
DHSC	Department of Health and Social Care
GMPP	Government Major Projects Portfolio
HIT	Health Information Technology
IP	Intellectual Property
NCCID	National COVID-19 Chest Imaging Database
NHS	National Health Service
NHSE	NHS England
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health and Care Research
MHRA	Medicines and Healthcare products Regulatory Agency
PPIE	Patient and Public Involvement and Engagement
QALY	Quality Adjusted Life Year
RoI	Return on Investment
TRL	Technology Readiness Level
TSETs	Technology Specific Evaluation Teams
UK	United Kingdom

Abstract

Background: Internationally there is a drive to implement Artificial Intelligence (AI)-based systems to improve safety, quality and efficiency of health and care. However, many efforts remain localised and do not progress beyond early-stage technology development. National Health Service England (NHSE) jointly with the Department of Health and Social Care (DHSC) launched a national programme in 2019 with the aim to accelerate the safe adoption of AI in health and social care settings (known as the NHS AI Lab). We sought to assess how the AI Lab was implemented; its outcomes, impacts and value; and lessons learned.

Methods: We conducted a mixed-methods, retrospective, formative evaluation between March and December 2024 consisting of analysis of 1021 project documents and 85 semi-structured interviews with key stakeholders. Documents included business cases, lessons learned reports, meeting minutes, and benefits registers. We interviewed current and past AI Lab staff, local and national strategic decision makers, system developers, NHS providers, and evaluators who were involved in the AI Lab. We analysed qualitative data with the help of the Technology, People, Organizations, and Macroenvironmental factors (TPOM) framework and within this assessed impact and value with the Triple E framework (economy, efficiency, and effectiveness). Many programme benefits are challenging to quantify (such as the benefits of regulatory reform, training, and advancing early innovations). Therefore, the quantitative work focused on review of evidence generated from the AI in Health and Care Awards, a £120 million investment in research and development aiming to build a real-world testing environment and accelerate spread and adoption of proven AI in the NHS. We analysed project evaluations commissioned as part of the AI Awards to assess efficiency and effectiveness of implementations in realising impact and financial benefit demonstrated during the award funding period.

Results: The AI Lab was an ambitious experimental initiative in an emerging area. Learning happened on route as this was a first-of-type, and some failures were inevitable. However, these provided invaluable learning. The AI Lab has made considerable progress in contributing to national regulatory guidance and policy; helping to understand existing gaps and identified needs in the system; and in contributing to learning how to develop, implement, deploy, and evaluate AI technologies. It has also helped to place the United Kingdom (UK) in the vanguard of discussions surrounding the deployment of AI models at scale and pace. If appropriately managed, these foundations will continue to deliver benefits in the future. However, progress was inhibited by a turbulent macro-environment leading to changing leadership, objectives, scope, and resourcing. Consequent changes in components of the AI Lab over time, staff turnover and limited capacity. The AI Lab also faced challenges surrounding the implementation and scaling of technologies, including difficulties aligning systems with service needs and issues

surrounding pathways to procurement. These challenges were much harder than originally envisaged and led to delays with some projects not fulfilling anticipated outcomes. Reviewing benefits as reported, one successful diagnostic AI Award provided £44 million in benefits (resulting from increased rate and timeliness of critical treatment associated with improved patient outcomes and reduced ongoing health and care need). This was much more than the £1.9 million it cost and more than the £25.6m total awarded funding to the subset of eleven AI Awards of the highest maturity level which had completed in the timeframe of evaluation. Some projects also helped improve NHS clinical processes in line with the 2019 Long Term Plan (e.g. increasing the rate of mechanical thrombectomy for stroke patients to 10%). These findings demonstrate both the challenges to implementation but also the potential for impact.

Conclusions: Many of the lessons identified apply not only to AI but also to other digitalisation initiatives, illustrating the importance of socio-organisational factors in technology development, deployment and adoption. National programmes can stimulate development and implementation of innovative technology in health service settings but there is limited understanding of how to support scaling and the achievement of sustainable long-term benefits. Local-level procurement and deployment is time-consuming and fails to utilise acquired experience. There is therefore a requirement to better understand the existing needs of the ecosystem, and for coordination of learning and strategic alignment. However, it is not yet clear how best to organise this technically and institutionally. The AI Lab experience puts the UK at the forefront of systemic attempts to implement AI to support health and social care. There is now a need to build upon this momentum and sustain coordination, viewing the AI Lab's efforts as the beginning of a long-term journey towards a future of an AI-enabled health and social care system.

Introduction

The NHS (National Health Service) Artificial Intelligence Laboratory (hereafter AI Lab) was established in 2019 as a major government initiative with an initial funding allocation of £250 million.[2] Its creation came at a time when there were international calls to leverage health information technology (HIT) and AI to improve the safety, quality, and efficiency of health and care delivery.[3] Building on the vision of data-driven transformation, supported by the increasing volumes of available data, there were high expectations around the deployment of AI tools, particularly in radiology and other diagnostics. However, the uptake of AI had been slow, and the health and care sector had limited experience in the effective use of AI.[4] [5]

More generally, progress in digital transformation of health and care has been slow and uneven.[6] There is an ongoing tension around how best to digitalise health services — oscillating between top-down and bottom-up strategies —which have been played out in trial-and-error experimentation over time.[7] It was within this context that the NHS AI Lab was envisioned to support the wider development and deployment of AI technologies at scale across the health and care system.[8]

The AI Lab was intended to position the NHS as a pioneering centre for AI in health and care, with dual goals of enhancing AI adoption and promoting the growth of the United Kingdom's (UK)

[2] Health Secretary announces £250 million investment in artificial intelligence. Available from: (last accessed: 27/09/2024).https://www.gov.uk/government/news/health-secretary-announces-250-million-investment-in-artificial-intelligence (last accessed: 27/09/2024).

[3] Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. Journal of global health. 2018 Dec;8(2).

[4] Sheikh A, Anderson M, Albala S, Casadei B, Franklin BD, Richards M, Taylor D, Tibble H, Mossialos E. Health information technology and digital innovation for national learning health and care systems. The Lancet Digital Health. 2021 Jun 1;3(6):e383-96.

[5] Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM. Artificial intelligence applications in health care practice: scoping review. Journal of medical Internet research. 2022 Oct 5;24(10):e40238.

[6] Cresswell K, Sheikh A, Williams R. Accelerating health information technology capabilities across England's National Health Service. The Lancet Digital Health. 2021 Dec 1;3(12):e758-9.

[7] Fennelly O, Cunningham C, Grogan L, Cronin H, O'Shea C, Roche M, Lawlor F, O'Hare N. Successfully implementing a national electronic health record: a rapid umbrella review. International Journal of Medical Informatics. 2020 Dec 1;144:104281.

[8] About the NHS AI Lab. Available from: (last accessed: 27/09/2024).https://transform.england.nhs.uk/ai-lab/about-the-nhs-ai-lab/ (last accessed: 27/09/2024).

health technology and AI sectors.[9] Alongside its healthcare objectives, the AI Lab thus had broader economic goals. Its significance was underscored by its inclusion in the Government Major Projects Portfolio (GMPP), and it received the backing of the Cabinet Office, including the incumbent Prime Minister Boris Johnson and his Chief Advisor Dominic Cummings.[10] The AI Lab was jointly delivered by NHSE and the Department of Health and Social Care (DHSC). NHSE leads the NHS in England, and DHSC is responsible for overall health policy.

The AI Lab sought to change the health and care ecosystem at various levels including NHS readiness and skills for AI deployment, supplier-user relationships and markets, and regulatory contexts. Its primary objectives were to develop evidence-based interventions, support strategic AI alignment across sectors, and generate evidence to create clear adoption pathways for AI in health and care. The work encompassed several linked components, including the development of critical infrastructures such as the National COVID-19 Chest Imaging Database (NCCID) and a national AI Deployment Platform (AIDP).[11] It also sought to foster collaboration and learning through community-building activities and early-stage prototype development. Here, it supported 86 early-stage and "market-ready" AI projects through its AI in Health and Care Awards (AI Awards).[12] Additionally, the AI Lab focused on governance, helping to shape AI policy, regulation, and ethics within the health and care sector. Box 1 illustrates the different components of the AI Lab.

Box 1: Different components of the AI Lab

Infrastructures

- National COVID-19 Chest Imaging Database
- AI Deployment Platform: pilot deployment in two radiology networks

Knowledge sharing and community building

- Training and upskilling
- Cultural change and building confidence in developing and implementing AI
- Foster a community of practice of AI practitioners in health and care by providing opportunities for co-production

[9] Arora A, Wright A, Cheng TK, Khwaja Z, Seah M. Innovation pathways in the NHS: an introductory review. Therapeutic Innovation & Regulatory Science. 2021 Sep;55(5):1045-58.

[10] Boris Johnson pledges £250m for NHS artificial intelligence. Available from: (last accessed: 27/09/2024).https://www.theguardian.com/society/2019/aug/08/boris-johnson-pledges-250m-for-nhs-artificial-intelligence (last accessed: 27/09/2024).

[11] The NHS AI Lab. Available from: https://transform.england.nhs.uk/ai-lab/ (last accessed: 27/09/2024).

[12] AI in Health and Care Award winners. Available from: https://transform.england.nhs.uk/ai-lab/ai-lab-programmes/ai-health-and-care-award/ai-health-and-care-award-winners/ (last accessed: 05/01/2025).

Innovation, research, and evidence

AI in Health and Care Awards

- Aimed to build a real-world testing environment and accelerate spread and adoption of proven AI in the NHS
- Delivery targeted four phases in the projected innovation/uptake journey
    - Phase 1 to show technical and clinical feasibility of the concept
    - Phase 2 to evaluate and develop prototypes and generate early safety and efficiency data
    - Phase 3 to support first real-world testing in health and social care settings including evidence for routes to implementation
    - Phase 4 to support technologies with market authorisation but insufficient evidence for large scale commissioning or deployment. Phase 4 projects had independent evaluations through Technology Specific Evaluation Teams (TSETs). Evaluation teams were matched to companies through a competitive process.
    - Phase 5 was initially planned to support wider roll-out, (through other funding national or local commissioning based on the specific clinical pathway the technology related to, once the Phase 4 projects completed) but no Phase 5 projects were funded.
- The Awards were managed on a day to day outside of the AI Lab until late 2023 when they were managed in-house
- Biggest and most expensive component of the AI Lab

Skunkworks projects

- Examined proof-of-concept projects using agile development
- Injected AI capability into organisations for a short time and helped create their own capability
- Demand identification and connecting suppliers to problems

Governance and ethics

AI policy and regulation ecosystem

- AI Futures Programme: policy development on futures, meaningful human control, and liability
- Multi-Agency Advisory Service: aimed at easing the navigation of the regulatory environment both for research and for service transformation

AI ethics initiative

- Striving for health equity: researched issues arising from bias and discrimination in algorithms and how to safeguard against this
- Governing the use of data: looked at different stewardship models, patient participation, and Algorithmic Impact Assessments

- Building confidence in AI: oriented mainly to workforce training to build capability and confidence

The application of AI in medicine has developed rapidly since the AI lab was established in 2019. At that time Generative AI for instance was in early stages of development, which is why these applications were not considered by the NHS AI Lab. At the time of writing there are no approved medical devices that deploy generative AI.

Our evaluation of the NHS AI Lab aimed to assess processes surrounding development and delivery, impact on care processes and patient outcomes, and value for money. It sought to understand the short-term and long-term effects of the programme, capture key lessons, and provide accountability to stakeholders through external independent review.

This document reports on overall findings. Separate in-depth papers are being prepared reporting on specific sub-programmes.

Methods

We received ethical approval from the Research Ethics Committee at the University of Edinburgh School of Social and Political Science.

As the programme was extensive and complex, several challenges emerged in the evaluation. Our methods have evolved over time as we have engaged with these challenges (Box 2).

Box 2: Overview of key challenges of the evaluation

The AI Lab programme involved such an extensive range of activity that direct assessment of all outcomes was not feasible within the limited timescale and resources available for this evaluation. However, an extensive programme of evaluation activity had already been commissioned and undertaken by the AI Lab. We drew on and extended this. Our work thus had aspects of a "meta-evaluation", synthesising some results of individual evaluations already conducted.

Most of the AI Lab's evaluation effort focused on specific projects — especially the AI Awards where work centred primarily on project-oriented Key Performance Indicators (KPIs). Some KPIs were clearly defined and were easy to gather (e.g. the number of patients involved in an AI trial). Outcome-oriented data which assesses the real-world impacts of AI projects proved more difficult to collect within the programme timeframes. Process-oriented evidence was

collected for lessons-learned reviews of individual components of the AI Lab but was not conducted at programme level.

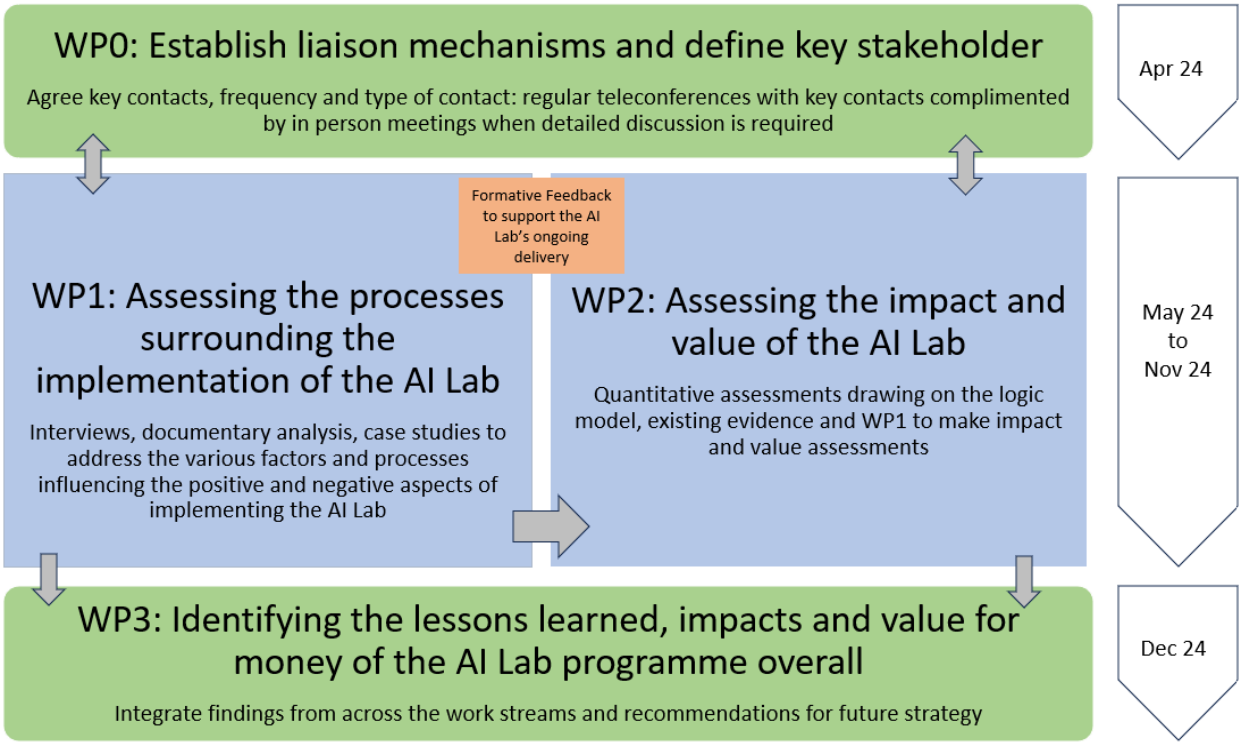
The availability of quantitative baseline data was patchy and varied considerably depending on the maturity of the vendor, deployment contexts, and stakeholders. This posed challenges in the extent to which we could generalise outcomes, attempt to estimate the benefits of scaling and aggregate benefits across the entire programme. There was also a lack of quantitative baseline data which accentuated the challenges of attributing benefits to the programme and drawing a counterfactual in such an active space. Moreover, COVID-19 and other confounding factors limited our ability to attribute changes to the AI Lab's work, making the quantitative aspects of the evaluation particularly challenging.

The evaluation of most components of the AI Lab depended therefore mainly on qualitative methods, predominantly relying on evidence from interviews. The quantitative work focused on review of evidence generated through AI Awards.

Many of the benefits of the AI Lab are yet to materialise and may take years to fully emerge. This evaluation is therefore unlikely to capture many of the long-term outcomes of the AI Lab quantitatively. These broader outcomes are also often difficult to assess as much of the value of the programme lies not just in immediate service improvements but in learning how to effectively develop, implement, and deploy AI technologies. Quantifying the value of this learning process is difficult.

We divided our evaluation into four main work packages (WPs, Figure 1).

Figure 1: Overview of the methods



The initial, start-up phase (WP0) comprised engaging with key stakeholders to map their interests and insights, obtaining ethical approvals, and developing a detailed evaluation strategy. This phase also included a series of scoping interviews with key decision-makers and review of overarching documents.

WP1 focused on exploring factors that have influenced the implementation of the AI Lab's components, and examining how these factors have led to outputs, outcomes, and impacts. This involved an analysis of activities, resources, and contexts, alongside exploring planned and actual activities. WPs2 and 3 built on the findings of WP1 by making quantitative assessments surrounding the outcomes and benefits of the AI Lab.

Sampling

Sampling of respondents was purposeful and encompassed a wide range of stakeholder viewpoints including tool developers, managers, implementers, strategic decision makers, regulators, evaluators, academic experts, and professional bodies. Gatekeepers provided us with an initial list of key stakeholders in the AI Lab covering all components (AI Awards, infrastructures, ethics and regulation, communities). This included current stakeholders associated with the AI

Lab, its delivery partners, AI Award holders and evaluators, NHSE, and DHSC. In addition, we drew on our networks and snowball sampling to widen the range of individuals, focusing on those who had left the AI Lab to mitigate against potential "survivor bias" in our account.

We also sampled AI Award case studies seeking to achieve insights into a range of experiences. These were selected in line with emerging findings as exemplars of the trajectory enabled by AI Award-related activities. They included typical, positive and negative experiences to illustrate the range of processes involved over time. We initially planned to conduct a series of in-depth case studies but struggled with recruitment, so we had to adapt our sampling strategy to include a wider range of settings and examining these in less depth than originally planned.

Documents were sampled purposefully to represent different components of the AI Lab (AI Awards, infrastructures, ethics and regulation, communities), overall strategic and contextual documents, benefits registers and other documents developed by the AI Lab, and meeting minutes of key strategic groups.

Data collection

Data collection took place between April and December 2024, when the AI Lab was in its final year of delivery, so most of the activities were retrospective in nature.

We conducted documentary analysis and in-depth semi-structured interviews across all components of the AI Lab including both ongoing and completed projects. Gatekeepers provided an extensive list of documents which we reviewed in full, although some documents were reviewed in more depth than others depending on the relevance of the information to the aims of the evaluation. Documents included business cases, lessons learned reports, monthly reports, meeting minutes, and evaluation data (including existing analyses, benefits maps, project reports, and existing evidence of outcomes and impact documented in logic models).

Interviews explored experiences over time (including changes in the AI Lab's focus), perceived outcomes (anticipated/unanticipated, positive/negative), challenges and ways to address these, ways to maximise impact, and potential lessons for future programmes. We fed back emerging findings iteratively to commissioners and helped to refine ongoing strategy. Sample topic guides are provided in Box 3.

Box 3: Sample topic guide

Topics were tailored to participant roles and in line with emerging themes.
Key areas covered included:

What has been your involvement in the AI Lab?
What are your opinions of the AI Lab as an initiative?

- Initial vision and to what extend this has been realised
- Nature of programme activities and innovations
- Expected and realised benefits
  - Impact on safety, quality, efficiency of care;
  - Impact on development processes of AI
  - Impact on implementation of technology into practice
  - Impact on regulations and policies
  - Impact on market development
  - Value for money
  - Any other types of impact
- Any unintended/unanticipated consequences (positive and negative)?
- Challenges and potential ways to address these
- Differences between AI Lab components

Lessons learned

- Contribution/attribution (e.g. what do you feel has been the impact or change as a direct result of the AI Lab's work? What has the AI Lab influenced/changed in the wider AI/healthcare landscape? What would not have happened if the AI Lab had not existed?)
- How could impacts and value be maximised in the future?
- Lessons learned and implications for future initiatives

We extracted quantitative data from AI Lab Phase 4 awardees and Technology Specific Evaluation Team (TSET) reports designed to address an evidence gap within the context of the funding call. In addition, for all award phases, we analysed closure reports that followed a semi-structured template provided by the AI Lab to include details of adoption, impact, and company commercial metrics. We also examined contract documentation and related journal publications.

Participants provided written informed consent to be interviewed. Interviews were conducted online via Microsoft Teams. They were audio-recorded and transcribed verbatim with the inbuilt transcription software.

Given that the evidence base was most robust in the AI Awards, we conducted a range of further analyses, including data extraction of Award reports and a series of case studies of individual

projects to supplement the breadth of data collected through interviews and documentary analysis. Within each case study we conducted internet searches of published supplier documentation and conducted targeted interviews with suppliers and evaluators.

We developed a data extraction sheet to guide AI Award and case study data extraction and allow comparisons across projects in relation to trajectories and impact. Extraction sheets are shown in Appendix 1.

Analysis

Transcripts and documents were uploaded into NVivo 12 software to facilitate coding. We used TPOM as a coding framework to ensure that various technological, social, organisational, and macro-environmental dimensions were represented in the analysis.[38] We extracted different experiences (positive and negative), perceived outcomes (anticipated and unanticipated), and key opportunities and challenges (including skills, resources, relationships, and strategies) associated with different AI Lab components and from various stakeholder perspectives. In doing so, we extracted barriers and enablers to the AI Lab's outputs, outcomes, impacts and benefits; assessed to what extent the vision of the programme had been realised and how; and extracted how adaptations to interventions could deliver future improvements. This helped us to identify challenges and understand what factors fostered positive progress. Additionally, we highlighted other relevant factors that were likely to lead to a positive impact and to contribute to lessons learned.

The quantitative work focused on a review of evidence generated through AI Awards. We developed a benefits register capturing "benefits realised" - defined as a measurable and evidenced improvement attained during the award funding period. We focused on establishing the success of projects at generating evidence to support future large-scale commissioning or deployment of market authorised products when this was absent at the beginning of the project.

For Phase 4 projects, we analysed TSET evaluations. This provided assessments of the efficiency and effectiveness of the projects in demonstrating impacts and benefits during the award funding period. We also reviewed markers of maturity and appropriateness of the evaluation programme. This examined whether and to what extent key aspects required for a successful implementation design had been identified including clinical context, clinical pathways, bottlenecks and current optimised clinical guidelines.

Whilst TSET evaluations were not commissioned for the Phase 3 Awards, we reviewed available project documentation, and developed a narrative, building on findings from the review of Phase 4 evaluations, to set out early indications of impact. Our analysis focused on ascertaining evidence of achieving real-world testing in health and social care settings including evidence for routes to implementation.

We extracted the overall reported financial benefits of the projects and the corresponding average patient-level savings (compared to the baseline current best practice pathway) using the study patient cohort size. Whilst the approach to, and focus of, measuring economic benefit for Phase 4 was variable across all projects, those which reported pointed to common categories including:

1) efficiency at point of use (e.g. for a clinician carrying out a diagnostic procedure);
2) short- and long-term efficiencies in care provision (NHS and social care); and
3) patient benefit measured as gain in quality-adjusted life years (QALYs) resulting from the intervention.

Reflecting these categories, we constructed a "benefits register" to summarise the findings and allow reporting at programme level. Despite being highly heterogeneous in the way that they were derived, we totalled these monetary benefits for reporting. We elaborate on the limitations of this approach in the Discussion section.

We initially analysed data within each AI Lab component, including AI Awards, infrastructures, ethics and regulation, communities, and benefits/impact. We presented emerging findings at analysis workshops including the whole team. These included exploring TPOM dimensions and links between these, similarities and differences, synergies, emerging tensions and trade-offs and developments over time.

Results

We interviewed 85 individuals, reviewed 1021 documents, and conducted 12 observations.

A detailed description of the dataset is provided in Table 1.

Table 1: Overview of the dataset

<table>
<tr><td colspan="2">Data Collected</td></tr>
<tr><td>85 interviews</td><td>We interviewed 34 females and 55 males, some interviews had more than one interviewee:

<table><tr><th>Number of interviewees in the interview</th><th>Number of interviews</th></tr><tr><td>1</td><td>76</td></tr><tr><td>2</td><td>8</td></tr><tr><td>3</td><td>1</td></tr></table>
Main roles included: 22 NHSE staff, 15 academics, 14 suppliers, 8 DHSC staff, 6 previous AI Lab staff, 3 regulators, and 17 others

5 people were interviewed twice

146 individuals were approached for interview (resulting in a response rate of 58%)</td></tr>
<tr><td>12 observations</td><td>Two project progress meetings, four AI Lab board meetings, six workshops</td></tr>
<tr><td>1021 documents</td><td>173 project final reports, 156 contractual papers, 154 progress reports, 200+ management & governance papers, 29 presentations and 18 annual reports.

Broad subject areas by document count are given below:</td></tr>
</table>

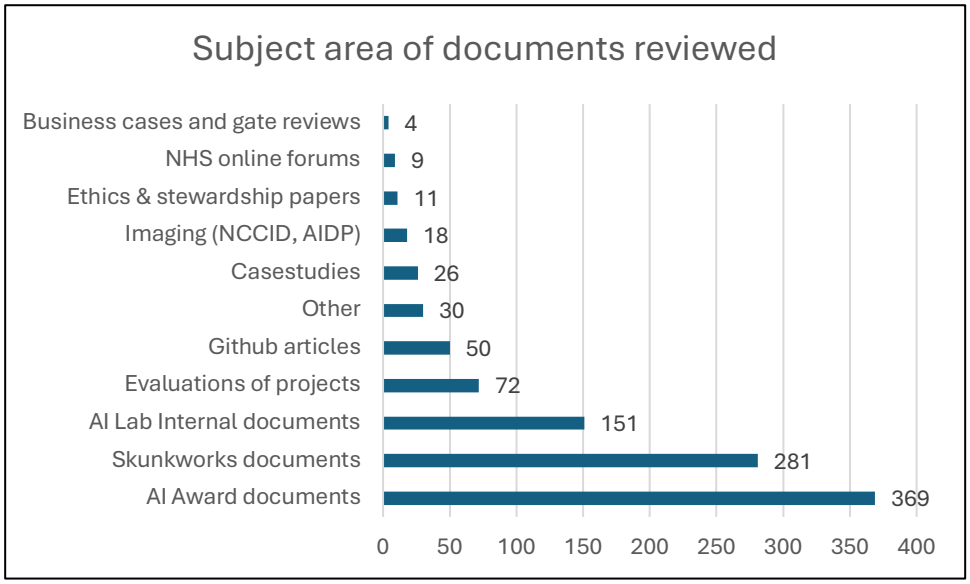


Table 2 provides an overview of emerging findings, which we will explore in more detail in the subsequent paragraphs.

Table 2: Overview of findings

An ambitious experimental initiative in a complex ecosystem

- Learning associated with the experiences of the AI Lab is likely to be the biggest benefit
- The AI Lab made significant progress in areas such as generation of evidence, data stewardship, ethics, navigating the regulatory landscape and enabling the workforce
- Deployment challenges were underestimated by vendors, adopters, evaluators, programme managers but the AI Awards (and AIDP Pilot) started to map and quantify these challenges

Turbulent macro-environmental influences and political drivers

- Unstable governance, objectives and components of the AI Lab over time (COVID-19, changing ministers, re-structuring, spending review)
- Multiple stakeholders with conflicting agendas driven by broad economic objectives and a grand ambition to change whole ecosystem

Generating real-world evidence to support market authorisation and large-scale adoption

- Promising significant returns on investment in relation to research investment in a few mature technologies with a short journey to market but long-term value uncertain

Scaling of technologies and pathways to procurement, reimbursement, deployment and operation

- Uncertainty surrounding procurement pathways and commercial models of AI technology
- Aligning technology strategy with system needs to promote integration into practice and use
- Limited organisational deployment and operational models and associated infrastructure inhibiting adoption progress

An ambitious experimental initiative in a complex ecosystem

The NHS AI Lab represents a unique and ambitious initiative, intervening across several overlapping ecosystems, including vendors (and the associated anticipated development of an innovation ecosystem stimulating the formation and growth of start-ups), regulators, the NHS, and policy. As such it faced challenges in balancing multiple, sometimes conflicting, objectives, such as focusing on both the development and commercialisation of early-stage AI innovation and the large-scale deployment of established AI technologies. This required compromises that in some instances meant neither objective was fully met.

> *The question of how to accelerate adoption and scaling within the NHS is a recurring topic among Number 10 and the Secretary of State. Leveraging the NHS's spending power and streamlining the process for companies and innovative ideas to scale across the system are major challenges. The NHS's structure is not optimised for innovation, which can be especially difficult for start-ups or ideas that lack evidence or change frequently.*
> AI Lab Board 24, Minutes, May 2023

> *…the original high level objective is about testing and accelerating the use of AI in in health and care, but I don't know from my perspective it felt like well surely then we should be looking at the system and looking at where the problems are and wondering where could AI feasibly help, whereas it didn't feel like that's what we were being driven by. I was never really clear what it was being driven by.*
> Interview 5, DHSC

The AI Lab was also experimental by nature, operating in a rapidly evolving area with little precedent to guide its activities. There were for instance different priorities at different times. At the beginning there was a focus on proof-of-concept projects through the Skunkworks which was closed mid-way through the programme. Responding to the COVID-19 pandemic the AI Lab developed an image repository to develop and test diagnostic tools: the National COVID-19 Chest Imaging Database (NCCID). Experiences from this fed into a vision to develop a national AI Deployment Platform to support the validation and deployment of AI tools at scale. A pilot

implementation of a commercial product was launched in two imaging networks towards the end of the programme. There were also various attempts to create a knowledge ecosystem such as setting up the AI Lab Virtual Hub community, but these dissipated over time.

Without prior knowledge, and given the perception that urgent action was needed, it was necessary to adopt an accelerated experimental learning approach: learning-by-doing and learning across different contexts. One perhaps inevitable consequence was uneven success across different programme components and different projects within those components.

> *…think about it like it ran a series of experiments looking at you know, can we stand up a national imaging platform which is centralising health data which we know how bad that went the first time. So now we're going to try it again, but in a different way. You know, we're going to stimulate industry through targeted investments in new technologies. So haven't been tried before and we're going to run trials with them, which is fantastic.*
> Interview 43, Previous AI Lab Member

The AI Lab's accelerated launch limited opportunities to establish robust processes, project plans, and baselines, while the COVID-19 pandemic presented serious disruption. The pandemic shifted strategic priorities and diverted resources, but it also created new opportunities for AI, such as the development of the NCCID. More widely COVID-19 also focused the NHS on data, automated interpretation of data and the benefits of sharing infrastructure in enabling the response of the health system. This shifted attitudes towards digitalisation of health and care delivery, coupled with rapid product development by vendors, and opened the minds of many to the potential benefits of the work of the AI Lab.

> *And then COVID came. And COVID caused a multitude of problems...people were pivoted to work on COVID programmes, it was the COVID kind of data vaccine database... All sorts of things have been set up where people were called from teams. And so, the team in the AI lab shrunk hugely, … the director at time… was moved to work on other urgent programmes. And we pivoted some of the work of the Lab because it felt wrong to continue on the same track. We continued with some of the programmes we've already set up, continued with the regulatory aspects, but pivoted to set up a national COVID chest imaging database.*
> Interview 54, Previous AI Lab Member

The AI Lab has made considerable progress in coordinating and facilitating conversations across stakeholder groups which have become more mature and better informed about the

opportunities and challenges of deploying AI in health and care. The overall structure of the AI Lab was rational and coherent but, in operation, too often individual projects within the AI Lab were siloed and not integrated effectively with wider activity.

The AI Lab has also contributed to national regulatory guidance, helped identify existing gaps and needs in the healthcare AI ecosystem, began to understand AI markets, and facilitated learning on how to develop, implement, and evaluate AI technologies.

> *Every minute we spend talking about something shines a little bit of light on the subject… And even actually you learn a lot more from your failures than these successes. You know, from my perspective, that's a tremendous value. And doing a lot of things under if not controlled conditions at least you know being able to you know have the opportunity to look most NHS hospitals wouldn't talk to you about their failures at all. So having a link into…lots of similar projects in lots of different and relatively similar organisations and understanding why they fail for me is a tremendous opportunity. It's a huge opportunity to be able to work out a really powerful and evidence-based logical model for complex technical change in complex organisations.*
> Interview 18, DHSC

As such, the AI Lab has begun to establish the foundations for safe and effective adoption of AI in UK health and care. Initial over-hyped optimism about the readiness and profitability of AI has given way to a more realistic understanding amongst some (but not all) stakeholders of the contexts and challenges associated with development, regulatory approval, and procurement. The extensive exploration of data stewardship models highlighted the tensions between various approaches to stewardship and the importance of Patient and Public Involvement and Engagement (PPIE). This work deepened understanding of the challenges facing a data-driven NHS.

> *...we need to have public and patient involvement...we work with the health research authority as well, and obviously through their ethics lens, that's a big part of things. One of the big things that came out of the AI Lab is definitely the digital and AI regulation service...that's genuinely good guidance on how to go about generating an AI product and that does advocate for clinical and public engagement. And we sponsored a piece of work looking at patient stewardship, data stewardship.*
> Interviewee 85, Previous AI Lab Member

However, there remains a risk that valuable lessons, especially those emerging from the AI Lab's challenges, may be overlooked in favour of focusing selectively on successes revealed in immediate outcomes. There is also a risk that lessons identified are not taken forward to inform future strategy and that longer-term developments are not sufficiently followed up.

> *…in terms of like the AI deployment platform for instance, the way that that's been procured with a view to you know this would be a national roll out after the pilot if there's all, all is successful, what we're seeing is that actually a national rollout might not be the most appropriate route, that's a huge benefit because it saves what could have been. …you know, a disaster; we don't know yet. The pilot's still in full swing at the moment, but you know. … Early conversations are saying potentially this might not be the way to go moving forward. …So, although it's a bit of a, you know, a loss from our side, overall, it's a really big win because it gives them, it gives you an opportunity to actually see, right, that wasn't the right way to do it. This is another way that we could be doing this from the lessons learned that we're taking from this particular project. So that's a that's a massive benefit.*
> Interview9, DHSC

Turbulent macro-environmental influences and political drivers

The AI Lab strategy and delivery was heavily influenced by a turbulent macro-environment and political drivers. The programme was established in part to stimulate the UK economy and position the UK as an AI superpower driven by ministerial enthusiasm for innovative technologies (particularly around diagnostics). However, despite these ambitious goals, the budget was modest compared to other large government programmes.

> *The work was done and ultimately with Boris Johnson and Matt Hancock's sponsorship, the 240 million was put together to run the AI Award and the AI Lab to move all of that forward with the launch in...early 2019 with a view to…putting England, at least at the forefront of using artificial intelligence for the benefit of health and care, so that just reflecting back to you, that's kind of the value idea. Wow, that's hard, really hard. I mean, really, really hard for all sorts of different reasons. I look at that and I, the cynic in me can understand….it was really clear that it was - this Halo of...trying to position England at the forefront of using AI for healthcare.*
> Interview 7, NHSE

The turbulent political landscape has significantly shaped the AI Lab's activities. Since its inception, there have been six Health Ministers and four Prime Ministers, each with different

priorities and changing governance arrangements. In addition, regions have acquired greater responsibility in planning and funding local services since 2022 which further contributed to changes in stakeholder priorities. The dissolution of NHSX in 2022, which was the initial delivery unit for the AI Lab, led to staff changes, uncertainty, and restructuring within the programme. Departures and associated disruption meant that painfully acquired expertise and knowledge dissipated.

> *You've had all of these, sort of, mergers and demergers at the centre. You've had a political system, and you've had how many health secretaries have you had? I mean, honestly, working in the NHS right now is so hard. It's so hard. And part of it is you don't know if you're… guaranteed that you won't have consistent leadership, consistent funding. Consistent strategies and things that would have gone...one month get blocked the next.*
> Interview 21, ASHN

> *...you have a hiatus with political change, you also have periods where people are aware, you know, often weeks, if not months in advance about when a government is reaching the end of its period and there's going to be an election. So that means that decisions are slowed even then, so it's like, "well, the elections coming, let's not make any buying decisions until after that point". So, you you're kind of floating along through those layers.*
> Interview 56, supplier

Additionally, a DHSC spending review in 2022 resulted in a ~£107 million reduction in the AI Lab's funding and has driven a shift away from experimentation toward a focus on evidencing the delivery of tangible benefits.

> *The AI Lab receiving a challenging spending review settlement, which removed its ring-fenced funding for AI related activities and the departmental reprioritisation reducing the AI Lab's future years remaining Capital budget down from £180m down to £75m for 22/23 - 24/25. This required pausing and/or stopping a number of programmes.*
> AI Lab Board 21, Minutes

We observed that, associated with this, AI Lab staff spent significant efforts on progress reporting, trying to track and account for programme impacts and benefits. Reporting activity has been escalated by requests from multiple disparate stakeholders with different reporting requirements and timeframes across departmental boundaries, driven by the need to demonstrate value of investment.

> *I mean again for the wider team probably did affect quite a lot because there's a lot of governance that's attached to being a government major programme. There's a lot of reporting, rightly so. There are gateway reviews, there's the whole kit and caboodle of things that you have to do if you're a government major programme because you are given some money to go and do stuff and you have to see how you're delivering against that stuff.*
> Interview 5, DHSC

The AI Lab was unique in its ability to connect policy, technology development, evidence generation, and healthcare services. This position created significant opportunities but also posed challenges in aligning the diverse needs and priorities of stakeholders. For example, participants noted a widespread overestimation of the maturity and capabilities of AI and its supply chain, leading to inflated expectations about progress. Additionally, the lengthy timelines required to achieve and demonstrate outcomes often clashed with the shorter-term nature of funding cycles.

> *It's a problem with like how government funding models work, because it's between each financial year and then you know it creates the wrong behaviours of like how we're trying to deliver and do things. And it's just it doesn't work for innovation and needs to be changed. … There needs to be more flexibility because you don't want to have a funding constraint in terms of you need to spend, you know your budget is only per financial year driving like how quickly you stand things up because otherwise you do things incorrectly to kind of meet a timeline, whereas actually it's just you know it's extremely difficult to plan these things anyway. Then you add on all these constraints of like you know outcome based milestones that you can only get paid once you've delivered and then it's just it just creates the wrong, or just an unnecessary set of challenges to try and deal with on top of the actual challenge of trying to deliver. An innovation project anyway, which is already difficult.*
> Interview 32, Previous AI Lab Member

The AI Lab's distributed structure and governance further complicated delivery. Rather than being a single, established organisational entity, the AI Lab was an evolving network of actors and collaborations that changed over time. Its joint governance between NHSE and DHSC added complexity, sometimes resulting in misaligned objectives, strategic directions, and blurred boundaries between the two organisations.

> *The policy to just give you an insight into how complex this space is … all of those organisations, all of whom are constantly competing for their own right to exist and their*

> *own money, and none of them have the same objectives, and yet they are supposed to somehow bring about. …the same end goal now when you're thinking about AI, that's even more complicated*
> Interview 13, Previous AI Lab Member

While bringing together organisations from across the ecosystem—including policy makers, providers, suppliers, and regulators—was seen as a strength, there was a noticeable lack of well-established communities of practice at the frontline of service delivery. Early efforts to build these communities were made, but they were not sustained due to limited resources and lack of forward planning for a sustainable community. The strain on the NHS, particularly during the COVID-19 pandemic, further diverted capacity away from these initiatives.

> *There was in the beginning. Definitely. They had these communities of practice, but it petered out cause I didn't hear much about it…I think [name] did kind of bring in a bit of that communities of practice, but I feel like as the Lab dwindled that sort of fell away and that that's actually quite helpful because. I think it needs to be a bit more it. There's needs to be a balance of organic development and because people within pockets of energy just know what the needs are. But then if you match them up with those that have more experience in the technology, then you can probably get that supply and demand you can get that intersection right.*
> Interview 53, Previous AI Lab Member

Generating real-world evidence to support market authorisation and large-scale adoption

Phase 4 Awards aimed at supporting technologies with market authorisation but insufficient evidence for large-scale commissioning or deployment. Table A2.1 (Appendix 2) and Table 3 summarise data extraction for the Phase 4 Award projects.

Table 3: Benefits realised for completed Phase 4 projects with sufficient evaluation

Award	Benefit category	Benefit	Latest patient count	Average benefit per patient	Extrapolated benefit for project patient reach
Project 1, Phase 4	b. Efficiency – care short term	Increased surgical procedures associated with enhanced triaging and diagnosis	150,019	-£110	-£16,477,087
Project 1, Phase 4	c. Efficiency – care longer term	Savings in social care over 5 years associated with enhanced triaging, earlier diagnosis and subsequent appropriate care	150,019	£208	£31,128,943
Project 1, Phase 4	d. Patient health and wellbeing	Increased patient quality of life over 5 years (measured as QALYs) associated with enhanced triaging, earlier diagnosis and subsequent appropriate care	150,019	£195	£29,303,711
Project 7, Phase 4	b. Efficiency – care short term	Reduced 90-day histopathology tests associated with more accurate imaging reading reducing need for further tests	1,471	£36	£53,489
Project 7, Phase 4	c. Efficiency – care longer term	Subsequent lifetime management savings to NHS associated with more accurate imaging reading reducing activity in pathways of care	1,471	£32	£46,516
Project 7, Phase 4	d. Patient health and wellbeing	Insufficient evidence	1,471		
Project 6, Phase 4	a. Efficiency – point of delivery	Clinician reporting time saved associated with more automated triaging in diagnostic pathway	597	£7	£4,293
Project 6, Phase 4	b. Efficiency – care short term	Increase in management referrals with enhanced triaging in diagnostic pathway increasing patients receiving appropriate care	597	-£4	-£2,178

Project 6, Phase 4	c. Efficiency – care longer term	Reduced NHS care costs associated with earlier stage diagnosis as a result of enhanced triaging in diagnostic pathway	597	£4	£2,249
Project 6, Phase 4	d. Patient health and wellbeing	Insufficient evidence			

Whilst the potential for impact was not limited to Phase 4 projects, the maturity of the projects and focus on evidence generation meant that a systematic analysis of impact and benefit was possible. Ten Phase 4 projects were completed with final reports available at the time of writing which were included in the analysis. One project stopped without delivering a final evaluation report. The overall funding provided by the AI Awards for the 11 projects included in the analysis amounted to £25.6 million.

We found that while all included projects correctly identified clinical contexts and pathways, as would be expected from technologies with market authorisation, some failed to identify important bottlenecks/pain-points that could be targeted to deliver improvements in efficiency or effectiveness (3/11). There were also instances (2/11) where technologies supported by the award were successfully deployed in care activities but did not demonstrate a clear improvement in health or economic terms when compared to existing best-practice care pathways. For example, some compared less favourably with other technologies or varying staff skill mix. Two projects (2/11) delivered a retrospective study only, limited by timeline and budget constraints which impacted on the opportunity to complete the originally planned prospective study. However, the latter is required to provide evidence suitable for regulatory purposes. This highlights challenges with aligning multiple stakeholders to deliver prospective evaluations within the AI Award timeframes in the absence of extensive prior engagement efforts.

Two projects (2/11) were able to measure significant improvements to effectiveness and efficiency. For example, in Project 1 *"treatment rates rose to 5.7% at [technology name] hospitals compared to the national average of 3.6% […] highest performing hospitals taking part in the evaluation reached mechanical thrombectomy rates over 10%, the target set out in the NHS Long Term Plan baselined at 1% in 2019."* This supports the potential of the technology to introduce efficiencies that support the delivery of best practice treatment targets. Furthermore, in Project 6: *"The use of [technology name] impacted management recommendations in […] 7.9% of cases where there were nodules detected [of which 69%] would otherwise be dismissed by the reader and instigated a follow-up recommendation, constituting a major change in management."* This

highlights the potential of the evaluated technology to improve effectiveness in radiology by demonstrating a superior adherence to radiology guidelines by less experienced clinicians.

Phase 4 Awards provided the greatest scope for delivering measurable RoI but we found significant heterogeneity in the level of evidence and approach to measurement. We could only value the RoI and the associated economic benefits for three completed projects (Table 3).

Where we could evaluate RoI, we observed heterogeneity in approach and scope of evaluation. In some cases, evaluations were not designed to capture long-term impacts. Many technologies did not fulfil some stakeholders' prior expectations of large-scale adoption but nevertheless reported some substantial impacts. These are discussed below.

Project 1 implemented a diagnostic tool in a non-elective care setting across a range of regional networks within the NHS. The technology provided a set of decision support tools that aided frontline clinicians to make time critical treatment decisions. Increased rate of optimal treatment was hypothesised to improve patient outcomes and reduce associated costs to health and social care.

Considering short term care efficiency, there was a reported increase in optimal treatment, leading to a modelled average of £110 increase in cost of care per patient. This was offset by efficiencies to longer term care and improved patient outcomes modelled as a five year (discounted[13]) saving to social care and increase in QALYs. The project valued this close to £400 per patient leading to a significant cost saving estimate of nearly £44 million across the approximately 150,000 patient cohort.

The evaluation was not able to conclude whether the technology provided point of delivery efficiency. Whilst this is not likely to be a significant value compared to the five-year projected benefit, understanding of implications to delivery of care is critical for implementation and adoption.

Project 6 deployed a diagnostic AI platform, integrated into existing software systems, to optimise oncology pathways in secondary care (elective and non-elective). The evaluation examined the impact on the quality of patient management recommendations, including decisions such as

[13] The Green Book (2022). Available from: https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-government/the-green-book-2020#a6-discounting (last accessed: 02/01/2025).

discharge, scheduling a re-scan at a later date, or pursuing further diagnostic workup or review. This technology, unlike Project 1, was reported to save clinician time at the point of use. Patient quality of life benefit from earlier diagnosis was not modelled due to a lack of external evidence for the impact of early diagnosis on disease progression and patient outcomes.

Project 7 deployed diagnostic AI-software to support pathologists in identifying and diagnosing tumours in secondary care. It anticipated that lab-based testing activity would be reduced by increasing accuracy of initial imaging review. The evaluation calculated short-term care efficiencies of £36 per patient. However, the analysis focused on 90 days histology impact and not clinician time at the point of delivery. The project could not draw conclusions about long-term care efficiency and patient outcomes.

Only one evaluation conclusively addressed the impact of the use of the AI on patient quality of life. Here, the associated benefit was reported to be substantial. This highlights a core challenge in assessing the impact of AI technologies: the most significant benefits or disbenefits often emerge over extended timeframes, exceeding the typical duration of procurement and evaluation periods. Similarly, only one evaluation addressed the immediate impact on service delivery - understanding of this is crucial for service planning to support adoption of new technology.

Projects 1, 6 and 7 illustrate some challenges. These arose from omissions in data collection, which resulted in key impacts and benefits not being measured, and gaps in the evaluation design, where the connection between the data gathered and the expected outcomes was not clearly articulated. Additionally, some impacts and benefits were inherently difficult to quantify, making their measurement more complex. These issues were further compounded by ambitious evaluation goals that could often not be realised within the constraints of the available project budget.

These projects did, however, demonstrate broadly appropriate evaluation strategies and coupled with their maturity and alignment within their deployment setting, have provided promising indications of the capability of AI technologies to realise a RoI.

For the Phase 4 projects examined, those demonstrating potential RoI tested a diagnostic or screening tool in a clinical setting. Administrative tools in operational settings were included in Phase 3 projects (discussed below). In clinical contexts, the most significant return was based on early diagnosis to avoid downstream patient care costs and impacts on quality of life. An overarching factor for “success” was therefore that projects aligned to identified national health

service and whole system priorities and targets for improvement, underpinned by level of burden and unmet need with significant monetary implications and patient outcomes.

Ultimately, our analysis indicates that a projects' success in generating evidence for large-scale commissioning or deployment depended both on technology maturity/penetration and appropriateness of the evaluation programme. Table A2.1 summarises the dimensions of appropriateness of evaluation programme, technology maturity and alignment for all Phase 4 projects under study.

Whilst benefits were reported in relation to AI Award funding, which we assumed covered all costs of technology and implementation, reports did not always robustly analyse RoIs based on the cost of the technology to the healthcare provider when deployed as part of normal service delivery. Extrapolating both cost and patient benefits is challenging due to various local factors, including digital infrastructure, a healthcare provider's capacity to act on the availability of earlier diagnosis, and their ability to deliver preventative care. We discuss this further in the Limitations section.

Having reviewed the outputs from Phase 4 AI Awards, we will now describe findings from Phase 3 Awards. Supplementary Table A2.2 (Appendix 2) summarises data extraction for the Phase 3 Award projects that had completed at the time of writing. Again, we observed a great deal of heterogeneity in evaluation approaches, likely influenced by technology maturity, innovation stage, and degree of NHS penetration. Approaches ranged from assessing feasibility of technology integration with existing clinical pathways (Project 5) to randomised controlled trials aimed at establishing effectiveness of interventions compared to standard care (Project 3). The latter matched the scope of a Phase 4 AI Award, which underscores the difficulty in clearly distinguishing between the two phases.

A considerable proportion of Phase 3 AI Award projects (7/10) reported sufficient evidence of having achieved the main objective of this phase, which was to support first real-world testing in health and social care settings including evidence for routes to implementation. This suggests that they could be advanced to the next AI Award phase, irrespective of whether the innovators pursued this option or not. Interestingly, most technologies (4/7) were already generating revenue, some (3/4) in an international context, despite great variability in the level of self-reported maturity (ranging TRL 3 to being CE marked and having Food and Drug Administration clearance). A higher maturity level was not necessarily associated with success in Phase 3, which highlights that mature technologies can also encounter implementation challenges. The three

projects (3/10) that did not make enough progress to generate significant outcome evidence pointed to the following confounders:

a) changes in clinical context (e.g. diagnostic test guidelines) that invalidated the main use case for the innovation;
b) focus on effectiveness of the technology without addressing integration in existing (or new) clinical pathways;
c) innovators and/or project managers lacking expertise in integrating systems with local information technology infrastructures.

Nine out of 10 technologies provided evidence of successful integration into an existing clinical pathway, eight out of 10 addressed efficiencies in the pathway, and five out of ten addressed cost-effectiveness. Notably, project 10 was able to quantify significant savings in administrative tasks arising from the implementation of the technology as part of the evaluation. Finally, half of the technologies (5/10) demonstrated some degree of adoption in new settings arising from the AI Award. However, this varied (from sites involved in the AI Award continuing to use the technology after the project to double digit new site deployments contracted during the project, see Supplementary Table A2.2). We cannot ascertain medium-to-long term sustainability. There were also challenges to gaining buy-in to support procurement. For example, one project found that the specialty team deploying the technology and achieving a reduction in emergency admissions, did not directly benefit as budgets remained unchanged. This disconnect reflects the fragmented NHS budget system, where savings in one area (e.g., emergency admissions) do not translate to budget reallocations for responsible teams. More generally Phase 3 projects do not appear to have followed any standard approach to procurement.

Scaling of technologies and pathways to procurement, reimbursement, deployment, and operation

The AI Lab was a research and development programme and not a delivery programme. Nevertheless, we observed significant uncertainty surrounding the procurement pathways of technologies developed and implemented as part of the AI Lab. For example, we found that many sites participating in AI Awards had no transition plans to procure the implemented technologies after completion of the AI Awards as they lacked evidence to support clinical or economic cases at the time that the AI Award funding was due to run out.

This uncertainty was seen as a threat to the sustainability of these technologies, as participating sites might face financial challenges in maintaining the systems. The situation was further complicated by the coordination difficulties stemming from national procurement guidance and

the inherent complexity of NHS procurement, which is managed independently by hospitals with diverse needs.

> *…it seems ludicrous if you've got 12 people doing a procurement for a specific type of thing that you don't turn around to them in the centre and go here's a template that you can all use for doing it. … Because to me, why would I have 12 people go off and write their own template? Makes no sense whatsoever. It's the same thing for all of the stuff that's kind of going on here. We're saying, well, if...they've got the approval to deploy their AI, why don't we get their clinical safety approval and share it with everyone else? … Because otherwise...I'm going to get 12 of them or all of the Trusts are going to do their own thing, because the best role that NHS England in the centre can play is to connect the people together and go: "You're doing what this person over here did it six months ago? Here's all the stuff that they used to do it".*
> Interview 29, NHSE

> *And the trouble is you just fall off the cliff edge at the end of it. So you do, you know, you implement something, you evaluate it, you show that it's, you know, you could, if you're lucky, within the small time that you've got and you're able to show clinical effectiveness and you can cost effectiveness, which is hard to do within a time scale. You know, even with all of that, you build a business case, and you've got the evidence, there is no route to procure procurement.*
> Interview 23, Supplier

Many suppliers were primarily focused on the United States market, and some reported that the complexity of UK procurement pathways discouraged companies from scaling their technologies within the country. The existing NHS procurement processes were viewed to be ill-equipped to accommodate the dynamic nature of innovative technologies, prompting some companies to withdraw from the UK market altogether.

> *The procurement within the NHS in general is lagging very far behind the technology that exists and could be available. So, if we get a tender specification from NHS supply chain it is…For older technology and it is actually, and I'm talking about technology that has been around for 60 years and hasn't changed…and it doesn't even give you the option of explaining how your technology could benefit this hospital. Because it doesn't ask the right questions in the tender specification.*
> Interview 57, Supplier

In addition, although the AI Lab worked with National Institute for Health and Care Excellence (NICE) on adapting the Evidence Standards Framework, in the UK, regulatory hurdles were reported to be more complex than in other countries (including Medicines and Healthcare products Regulatory Agency (MHRA) and NICE requirements).

Significant public support for mature product development, implementation, and rollout mobilised expectations about RoI for the taxpayer. This presented novel challenges for NHSE/DHSC, who had to develop benefit sharing agreements, based on licensing or fee discounts or share ownership, on an ad-hoc, case-by-case basis. Efforts to establish a stable framework encountered intractable and not fully resolved issues. For example, some stakeholders surfaced concerns that fee discounts or shareholding might distort the market and commit the NHS to products that would become obsolete.

> *… it wasn't clear what the [commercialisation] process was, and they were kind of building that process and filling that process on the fly*
> Interview 56, Supplier

AI Awards for less mature products presented fewer challenges than more mature products as investments were lower and concrete benefits were further from being realised. The National Institute for Health and Care Research (NIHR) which managed these projects had mechanisms for research funding that involved fluid arrangements in relation to measuring and recovering benefits. Early stage approaches to benefit sharing avoided applying commercial terms, as the value of these technologies was not yet known.

Intellectual property (IP) considerations added another layer of complexity. Established companies, with strong background IP, were better able to get their products adopted by the NHS, while those developing foreground IP offered better commercial value. This tension underscored the challenge of aligning commercialisation strategies with both market dynamics and public-sector goals. Careful balance is needed to avoid market distortions that could stifle competition or unfairly advantage specific suppliers.

> *So, we usually blend the model, so we'll say, you know, if you do that, we'll do, we will take X percentage, but if you go for the more commercial route, we'll take Y and that sort of gives the company a bit of flexibility. But yeah, I think the principle is one that is sound, but you have to be sort of careful about how you apply it to ensure that you're not adversely influencing the market or making the funding you know unattractive because that's not what anyone's here to do.*

Interview 52, External Contractor

The AI Lab had many stakeholder groups, but some were more actively involved in strategy and delivery than others. Participants mentioned that frontline health and care service delivery staff, in particular, had limited involvement, resulting in activities in some instances not sufficiently being aligned with system needs. For example, some observed that the range of applications supported by the AI Awards was influenced by supplier interest, rather than emerging from a bottom-up process of service delivery staff to establish health system requirements. Similarly, some suppliers identified an opportunity to secure funding for AI products without ensuring their products were tailored to service delivery and patient priorities. We also observed limited development of operational models and associated infrastructure (i.e. application of AI to improve efficiency of service delivery), with most projects focusing on clinical applications.

> *… you start with the problem with the clinical need and for whom. So rather than you start with a solution which is AI and who's regulating it... So, I think that it's basic we just start there. Unfortunately, what they did was to...discover that they were in some cases solving the wrong problems.*
> Interview 14, Academic

Delivery of AI Award projects was hindered by conflicting service priorities and a lack of active service delivery involvement. As alluded to above, we observed cases where after successful evaluation of projects (both in AI Awards, and proof-of-concept work), adopter organisations were unable to sustain the system beyond its initial pilot. This was due to a lack of evidence needed to establish a clear business and safety case for adopting the application into routine organisational practice.

> *For certain products, in certain situations, the stumbling block came that the evidence, clinical safety and value people weren't sure what to do with it outside of its initial context, people didn't know if it was translatable fundamentally [...] both on the financial perspective and on the safety net accuracy perspectives.*
> Interview 48, Previous AI Lab Member

In addition to these challenges, adopter organisations lacked the capacity, IT skills, and resources to support implementation and evaluation. This was further exacerbated by the pandemic, which shifted attention to COVID-19-related priorities and delayed non-COVID-19 projects.

> *We are still very much reeling from the COVID epidemic, we're really feeling the effects of a reduced workforce and huge waiting lists and so actually whereas you know, when we designed the programme. … It was pretty fair to imagine a world in which workforce had a bit of capacity to engage with this. The reality that we're faced with now is that. …You know even. … Staff that want to see improvements that have a real interest in research and AI simply do not have the capacity to engage in what is, you know, exploratory research based*
> Interview 12, DHSC

Discussion

Summary of main findings

The NHS AI Lab was a pioneering initiative designed to address barriers to AI development, deployment and adoption in health and care by intervening across an ecosystem of vendors, regulators, service delivery, and policy. It helped to foster innovation through product development and piloting, provided regulatory guidance, identified gaps in the AI health and care ecosystem, and facilitated learning surrounding AI development, implementation, and evaluation. The AI Lab also contributed to building AI literacy and confidence among stakeholders, providing a space for collaboration and co-creation between technology developers and adopters. It was in the vanguard of efforts to develop new technical and institutional arrangements for the safe and cost-effective validation, procurement, deployment, and post-market surveillance of multiple AI models across different settings. Here it has contributed to a shift away from a focus on specific AI models towards an evolving flow of improving models. The NHS now also has a much better understanding of where and how benefits of AI adoption can be realised. The experimental nature of the AI Lab allowed for flexibility and agility in the face of changing demands, dynamics, and the pace of technological development[14][15]

However, the AI Lab also faced challenges in reconciling at times conflicting objectives associated with the wide range of stakeholders involved. This included for instance supporting early-stage innovation while promoting large-scale deployment of mature technologies. Over time, changes in leadership, funding models, and other factors associated with a turbulent macro-environment

[15] Eilers K, Peters C, Leimeister JM. Why the agile mindset matters. Technological Forecasting and Social Change. 2022 Jun 1;179:121650.

narrowed its focus to measurable and quantifiable outcomes. Assessment of these was often constrained by a lack of baseline data and limited strength of quantitative evidence. While efforts were made to engage diverse stakeholders, supported technologies did not always align with system needs. There was in addition limited integration of different components or sub-programmes and difficulty moving from development and validation to scaling.

Strengths and limitations of this evaluation

We have provided important insights into a world-first national programme seeking to stimulate the adoption and scaling of AI in health and care settings. However, the retrospective nature and tight delivery timeframe of the evaluation posed some challenges. For example, key personnel had moved on, and historical data was likely partial and variable in quality (e.g. many AI Award reports included self-reported benefits). We have attempted to mitigate this risk through triangulation of prospective and retrospective data sources, and through actively approaching stakeholders who had left their work in the AI Lab at the time of data collection and others less directly involved. Consulting service users proved to be difficult given the timelines of the evaluation and the heterogeneous nature of experiences, use cases and technologies.

Our formative findings were incorporated real-time into ongoing decision making, and we became active actors in shaping concurrent efforts of the AI Lab. This position and intervention in the field through our formative feedback strategy may have influenced the information participants provided to us.

We further relied on gatekeepers to facilitate access to documents and participants. While this approach opened many doors and ensured we had all necessary documents at our disposal, it may have inadvertently limited our exposure to certain perspectives and insights. Nevertheless, we were granted access to relevant documents and interviewees whenever requested. To maintain independence, we also kept a separate list of interviewees that was not accessible to the commissioners.

Assessing value and impact was hampered by the lack of baselines. Existing evaluations did not always identify and quantify best practice pathways prior to the intervention, and therefore economic modelling of the efficiencies introduced versus costs incurred was limited. We also found evidence gaps in relation to patient outcomes and indirect, long-term outcomes which are likely to emerge over extended timeframes and may be hard to attribute. Therefore, potential variances in aggregated financial benefits and impacts are substantial and our findings must be interpreted with caution. Notwithstanding these challenges, even if appropriate efforts to measure baselines were made, the changing landscape may have thwarted attempts at

quantification for those activities that experienced turbulence (e.g. changes in markets, technological developments, and strategic direction).

At the time of writing, the market and maturity of AI products was limited, and tools were not routinely integrated in service delivery. There had been little opportunity to enhance benefits of AI models by optimising let alone transforming care processes. Our outcome modelling is based on a relatively small cohort of projects that had completed and fully reported when we left the field. Outcomes are still being reported and consent to commercialise agreements are being negotiated at the time of writing.

Even with a robust assessment of cost effectiveness and impact on longer term patient outcomes, there are limitations to how far this evidence can be used to inform:

1) the benefits of using a technology when it has scaled; and
2) the ongoing benefits and long-term RoI. An assessment of scalability is often based on linear extrapolation to large scale populations with limited attention to variations in performance across different demographic, operational and technological settings.

Furthermore, costs associated with various parts of the AI lifecycle have not been fully explored through the course of the AI Lab. There are therefore uncertainties associated with projecting future RoI due to difficulties in calculating long-term costs, for example, arising from post-deployment recalibration.

Nevertheless, we have identified some initial quantitative evidence of benefits and impact, as well as important contextual process factors that can now act as a foundation for ongoing work in this area.

Integration of the findings with the existing literature

The AI Lab encountered many challenges that are typical of large governmental digitalisation programmes in the NHS and elsewhere. These include the role of central bodies, setting up contracts, benefits realisation, evaluation, funding, strategy and vision, nurturing a learning

ecosystem, and staff involvement and engagement. They are summarised in Box 4.[16] [17] [18] [19] [20] [21] [22] [23]

Box 4: Challenges facing large governmental digitalisation programmes in the NHS

Role of central bodies

Changing policies and priorities in the NHS impede digitalisation

Commercial interests inhibit experience sharing

Mismatch between political timescales and the time necessary to bring about organisational change

National support can aid innovation but issues in coordination of strategic initiatives

Uncertainty surrounding sustainability of developed and implemented systems

Need for a shift in focus from short-term gains to sustainable long-term digitalisation strategies

[16] National Audit Office. Available from: https://www.nao.org.uk/wp-content/uploads/2024/02/digital-transformation-in-government.pdf (last accessed: 23/09/2024).

[17] Beginning a joint digitally enabled transformation and learning journey in the English National Health Service. Full Report of the Independent Evaluation of the Global Digital Exemplar (GDE) Programme. Available from: https://www.ed.ac.uk/sites/default/files/atoms/files/final_report_gde_evaluation_programme.pdf (last accessed: 01/09/2023).

[18] The Long and Winding Road: An Independent Evaluation of the Implementation and Adoption of the National Health Service Care Records Service (NHS CRS) in Secondary Care in England. Kathrin Cresswell, Maryam Ali, Anthony Avery, Nicholas Barber, Tony Cornford, Sarah Crowe, Bernard Fernando, Ann Jacklin, Yogini Jani, Ela Klecun, Valentina Lichtner, Kate Marsden, Zoe Morrison, James Paton, Dimitra Petrakaki, Robin Prescott, Casey Quinn, Ann Robertson, Amirhossein Takian, Katerina Voutsina, Justin Waring and Aziz Sheikh (2011). Available from: http://www.cphs.mvm.ed.ac.uk/grantdocs/526%20%20Final%20report%20v31st%20Mar%20FINAL.pdf (last accessed: 01/09/2023).

[19] National evaluation of the Vanguard new care models programme: Interim report: understanding the national support programme. Available from: https://www.research.manchester.ac.uk/portal/files/103375904/Interim_report_of_the_NCM_external_evaluation_final_v1.pdf (last accessed: 01/09/2023).

[20] Independent Review of NHS and Social Care IT, commissioned by Stephen O'Brien MP, Chaired by Dr Glyn Hayes, August 2009. Available from: https://ntouk.files.wordpress.com/2020/11/nhs-and-social-care-it-review-2009.pdf (last accessed: 01/09/2023).

[21] Making IT work: Harnessing the Power of Health Information Technology to Improve Care in England: Report of the National Advisory Group on Health Information Technology in England: Chair Robert Wachter. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/550866/Wachter_Review_Accessible.pdf (last accessed: 01/09/2023).

[22] National Audit Office, Digital Transformation in the NHS, 15 May 2020. Available from: https://www.nao.org.uk/wp-content/uploads/2019/05/Digital-transformation-in-the-NHS.pdf (last accessed: 01/09/2023).

[23] National Audit Office. Available from: https://www.nao.org.uk/wp-content/uploads/2021/12/Evaluating-government-spending.pdf (last accessed: 01/09/2023).

Contracts
Contractual re-negotiations over time leading to changing objectives and relationships
Limited contact between developers and provider organisations
Commercial interests may result in lack of sharing lessons
Issues around supplier responsibilities to address deployment issues

Benefits realisation
Unrealistic assumptions of achieving cost savings or returns on investment
Lack of clarity on where biggest benefits of deployment can be expected
Quantifiable benefits difficult to measure in the early stages of implementation
Lack of capability and capacity in benefits realisation
Projects with clinical leadership and robust benefits management are most successful in delivery (but they may be too narrow in scope)
Level of support required by provider organisations is often underestimated

Evaluation
Evaluations of complex transformation initiatives are challenging
Evaluations are often not systematic, independent, and on-going with formative/summative and qualitative/quantitative components
Views of end users are not always considered, including experiences of clinicians and views of patients and members of the public
Evaluation is often an afterthought and does not start before the beginning of the project, baselines are often not well established
Alignment of various stakeholder needs and interests is difficult

Funding
Funding issues surrounding capital and revenue are hard to reconcile
Reduced funding due to economic circumstances presents challenges to development and implementation
Limited flexibility of carrying forward funding across financial years

Vision and strategy
An overall continued commitment to digitalisation is not always present within deployment contexts
Uncertainty surrounding scaling and sustainability
The importance of retaining and applying lessons is not always understood by all stakeholders

Learning ecosystem
Challenges of creating a balance between formal and informal knowledge sharing
Issues with investing in a sustainable learning ecosystem

Staff involvement and engagement

Issues with buy-in at local level
Issues with clinical engagement
Lack of commitment from the organisation's leaders and end-users

Although these issues are well-known, unfortunately lessons are often not learned as macro-environmental contexts are not easily changed and strategic decision makers operate within certain constraints. However, our work underscores the importance of these contexts in influencing progress. When compared to other programmes, the AI Lab's governance, and staffing environment (with high turnover and reductions in head count) was particularly turbulent, which amplified some of these issues.

What makes this work unique is the focus on AI, an emerging and fast evolving area, with a yet limited empirical evidence base. Although specific targeted applications have shown promise in particular settings and we now know where benefits can be anticipated, [24] [25] [26] [27] Strategic decision makers therefore must balance the enthusiasm and potential benefits of emerging AI innovations, which is often associated with a degree of risk, with the need to make evidence-based decisions on safe and ethical development, implementation, and procurement. This also needs to involve managing expectations among stakeholder groups balancing enthusiasm for innovation with the slow emergence of robust evidence.[28]

Our work has shown that national policy can and should play a key role in helping to frame individual deployments through leading on regulatory frameworks, infrastructures, and pathways to procurement. However, the notion of success and failure of AI deployments may need to be replaced with a more nuanced understanding of deployment processes and pathways.

[24] Alami H, Lehoux P, Papoutsi C, Shaw SE, Fleet R, Fortin JP. Understanding the integration of artificial intelligence in healthcare organisations and systems through the NASSS framework: a qualitative study in a leading Canadian academic centre. BMC Health Services Research. 2024 Jun 3;24(1):701.

[25] He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nature medicine. 2019 Jan;25(1):30-6.

[26] Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. A review of challenges and opportunities in machine learning for health. AMIA Summits on Translational Science Proceedings. 2020;2020:191.

[27] Yeo K, Li Z, Lin W (2021). Barriers and facilitators to adoption of AI in healthcare: A systematic review. BMC Medical Informatics and Decision Making, 21(1), 54.

[28] Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: Mapping the debate. Big Data & Society. 2016 Nov;3(2):2053951716679679.

The literature shows that sustained integration of AI into health and care practices is likely to be facilitated by close attention to the needs of adopter organisations. Critical factors include advancing digital leadership;[29] building necessary capacity, capability, and culture; maintaining a system thinking perspective; clearly articulating evidence surrounding benefits; financial resources; [30] harnessing the power of adopter communities; and managing supplier relationships.[31] [32] The AI Lab began to address some of these factors, such as promoting evidence generation, connecting stakeholders, and experimentation with building capacity and capability.

The integration of AI requires a fundamentally novel approach to managing risks and benefits within provider organisations. Unlike traditional interventions, AI's inherent dynamism means that it is not possible to fully evaluate all risks prior to implementation. Instead, organisations must accept the reality of implementing AI with a degree of risk, ideally accompanied by continuous evaluation as algorithm performance evolves over time and across diverse contexts.[33]

Post-market surveillance of AI is a key element in controlling risk but the costs and requirements on infrastructures are to date uncertain. Such a commitment can only be secured if the innovation aligns with existing organisational agendas, needs, and priorities.

The most significant benefits of AI will arise not from automating isolated tasks and processes but from pathway redesign, population health, prevention, patient flow, and administration. [34] However, much work (including that of the AI Lab) to date has focused on building and deploying AI models rather than optimising their integration into real-world settings.[35] To address this,

[29] Tagscherer F, Carbon CC. Leadership for successful digitalization: A literature review on companies' internal and external aspects of digitalization. Sustainable Technology and Entrepreneurship. 2023 May 1;2(2):100039.

[30] Bevan H, Ketley D, Cawthorne R, Stavropoulou C, Scarbrough H. Spreading and scaling innovation and improvement: understanding why the differences matter. BMJ Innovations. 2024 Jul 1;10(3).

[31] Hinder S, Cresswell K, Sheikh A, Franklin BD, Krasuska M, The Nguyen H, Lane W, Mozaffar H, Mason K, Eason S, Potts HW. Promoting inter-organisational knowledge sharing: a qualitative evaluation of England's Global Digital Exemplar and Fast Follower Programme. Plos one. 2021 Aug 2;16(8):e0255220.

[32] Johnson M, Mozaffar H, Campagnolo GM, Hyysalo S, Pollock N, Williams R. The managed prosumer: Evolving knowledge strategies in the design of information infrastructures. Information, Communication & Society. 2014 Aug 9;17(7):795-813.

[33] He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nature medicine. 2019 Jan;25(1):30-6.

[34] Cresswell K, Anderson S, Montgomery C, Weir CJ, Atter M, Williams R. Evaluation of Digitalisation in Healthcare and the Quantification of the "Unmeasurable". Journal of General Internal Medicine. 2023 Dec;38(16):3610-5.

[35] He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nature medicine. 2019 Jan;25(1):30-6.

service redesign focusing on prevention, population health and integrated care is crucial in the development of AI tools.

AI-specific markets present additional challenges. They are currently characterised by small, quickly evolving startups and fragile operating models. National coordination of public sector AI has been argued to be important in this context, but there is limited understanding how this may be operationalised in health and care.[36] Our work has shown that ongoing and coordinated governance and regulation work is required to tackle emerging challenges and liaise with national and international bodies. In addition, ongoing national market management will be needed to promote diversity of supply. One approach could involve focusing on generic AI capabilities and target AI application areas that are informed by practice but not tied to a particular product. [37]

Strategic shifts, such as the move toward platform deployment models trialled by the AI Lab, reflect efforts to scale AI nationally and streamline implementation and procurement processes.[38] However, although promising, the AI Lab experience has shown the difficulties associated with concerted approaches to the procurement of platforms.[39] There are parallels to other national procurements where it has been shown that local input in decision making is crucial to ensure that systems align with existing needs and priorities of provider organisations.[40]

Recommendations for policy and practice

Developing, deploying, and maintaining bespoke AI tools is a significant undertaking. Ideally, this should only be pursued when a clear need has been identified across multiple contexts with

[36] Wirtz BW, Weyerer JC, Geyer C. Artificial intelligence and the public sector—applications and challenges. International Journal of Public Administration. 2019 May 19;42(7):596-615.

[37] Mikhaylov SJ, Esteve M, Campion A. Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration. Philosophical transactions of the royal society a: mathematical, physical and engineering sciences. 2018 Sep 13;376(2128):20170357.

[38] Bounfour A. Platforms and Artificial Intelligence. Springer International Publishing; 2022.

[39] Our empirical work has shown that the attempt to pilot the AI Deployment Platform in two regional radiology imaging networks encountered an array of unanticipated challenges rooted in differences in information governance and in local technology infrastructure (arising with both different Picture Archiving and Communications Systems (PACS) and different implementations of 'the same' PACS) between the 12 provider organisations involved. A decision was taken to close the Pilot when it became clear that the delays implementing the platform and AI tools selected were such that there was no prospect being able to collect adequate data about real world performance to assess the specific radiology models being trialled. This experience provided a wealth of learning opportunities.

[40] Sheikh A, Cornford T, Barber N, Avery A, Takian A, Lichtner V, Petrakaki D, Crowe S, Marsden K, Robertson A, Morrison Z. Implementation and adoption of nationwide electronic health records in secondary care in England: final qualitative results from prospective national evaluation in "early adopter" hospitals. Bmj. 2011 Oct 17;343.

potential for scale and impact, where market failure exists, and where it aligns well with national and local transformation strategy.

Most transformation is likely to be achieved through deployment of AI acquired through procurement. The AI Awards have demonstrated that appropriate involvement of frontline NHS staff can inform change that leads to significant health benefits and cost savings and these changes can be achieved in relatively short timescales. Involvement needs to continue and should be linked to strategic optimisation and transformation that draws on evidence from local deployment experience. This may also require a shift in how evidence is assessed and used to inform decision making in diverse and changing environments and technologies more broadly. For example, local deployment of tools is simpler to assess than systemic effects, which require longer timeframes and are associated with more complex implementation and optimisation processes.

Resources need to focus on strengthening and adapting existing information infrastructures and procurement/implementation capabilities and skills within implementing organisations, as the basis for re-designing pathways enabled by new AI-based technologies.[41] This will require a nuanced approach to community engagement that may be difficult to plan and will need to involve building trust and relationships. It is likely to be best achieved through working with community leaders to promote active local involvement. There is also a need for greater knowledge-sharing across the health and care system. Establishing and sustaining communities of practice will help to facilitate learning from the AI Lab's experiences.

Further sustained work is needed at national and regional levels to guide the commissioning and adoption of AI technologies. There has been a strategic shift in approach from promoting stand-alone AI tools to the adoption, validation, operation, and review of multiple models at scale. Partly because of the AI Lab's activities, the UK is at the forefront of discussions about the potential contribution of AI validation and deployment platforms, including real-time post-market surveillance of tools. However, it is not clear how these emerging platforms will eventually be configured (for example between devolved, centralised, or federated approaches to scaled AI).

[41] Priorities for an AI in health care strategy. Available from: https://www.health.org.uk/reports-and-analysis/briefings/priorities-for-an-ai-in-health-care-strategy#:~:text=An%20AI%20in%20health%20care%20strategy%20should%20ensure%20the%20NHS's,the%20development%20of%20AI%20systems (last accessed: 01/01/2025).

A centrally- led approach to market and vendor management as well as information governance will be critical in guiding this work. It will require a delicate balance between concerted adoption of nationally endorsed systems and the stimulation of aspects of an innovation ecosystem to allow new entrants in the market.[42] There may be scope for government to be responsible for overall planning and commercial companies to focus on delivery. [43]

While a one-size-fits-all approach to identifying needs and evaluating AI solutions is unlikely to be appropriate, early baselining and budget impact modelling will be necessary to justify the operational and economic rationale for adoption of developed technologies as well as long-term benefit realisation. We summarise key baselining and budget impact modelling considerations in Box 5. This applies the principles set out in the recently published HM Treasury guidance on impact evaluation of AI interventions (as part of the Magenta book).[44] Integral to this approach is understanding technology costs, for implementation and scale-up, and future costs of maintenance, monitoring and changing digital environments. The benefits register we have developed as part of this work is an important tool that now needs to be built on.

Box 5: Baselining and budget impact modelling considerations

A. Identify relevant clinical context, use case, and relevant clinical/operational pathway.
B. Identify most important bottlenecks/pain-points to address and research opportunities for introducing efficiency and effectiveness benefits with the help of a new technology and develop a theory of change.
C. Evaluate and (if possible and appropriate) iteratively refine the technology. Benefits are hypothetical at this stage.
D. Evaluate with a focus on substantiating hypothesised efficiency/effectiveness benefits and demonstrating financial savings via budget impact modelling:
    1. Model current best practice or 'optimised' clinical pathway in terms of financial costs (the baseline).
    2. Evaluate technology hypothesised to introduce benefits in terms of the efficiency/effectiveness introduced (i.e. shifting patients between different pathways, increasing throughputs) and associated costs incurred.
    3. Model updated clinical context (with the technology adopted) in terms of financial costs.

[42] Cresswell K, Sullivan C, Theal J, Mozaffar H, Williams R. Concerted adoption as an emerging strategy for digital transformation of healthcare—lessons from Australia, Canada, and England. Journal of the American Medical Informatics Association. 2024 May 1;31(5):1211-5.

[43] Mazzucato M. Mission economy: A moonshot guide to changing capitalism. Penguin UK; 2021 Jan 28.

[44] The Magenta Book. Available from: https://www.gov.uk/government/publications/the-magenta-book (last accessed: 06/01/2025).

4. Compare 1 and 3 to validate or refute hypothesised benefits. In the latter case, innovators may need to go back to Stage B.
E. Set out a model to project longer term benefits realisation and benefits realisation across contexts: this should include identifying relevant cohorts, contextual variations, evidence around disease progression and associated NHS and social care costs. To support this objective there is a need for an increased focus on preventative care both in policy design but also how this is weighted in measurement of health and care productivity. Identifying cohort and contextual variations is essential.

There is also a need to continue to support evaluation activity to understand processes and impacts, inform procurement and regulatory decisions, and facilitate learning from experiences at project and at programme level. The fast evolving and dynamic nature of AI is likely to require a phased approach, balancing progress with continuous evaluation and adjustments, ensuring no critical factor is overlooked while maintaining momentum. [45] Part of this will involve communicating that AI development and deployment is a long complex journey that should not be rushed but requires a systematic approach to assessing safety and regulatory implications.[46]

It will need to involve establishing and maintaining open lines of communication along the whole AI supply chain involving developers, vendors, configuration, deployment, and post-market operation with effective PPIE. This will enable products to be iteratively developed and adapted to contextual specificities, while maintaining appropriate security and privacy.

The success of AI in health and care in the NHS will require building on the experiences of the AI Lab. The establishment of a permanent unit to guide, coordinate and facilitate strategy, deployment, scaling and continuing the evaluation work initiated under the AI Lab is a welcome development in this respect. Sustained funding, central leadership and support, and working towards a common vision (whilst retaining a degree of flexibility) will be crucial going forward. A key risk is the failure to exploit and apply lessons learnt and build on the experience gained from the AI Lab work. We summarise lessons for various stakeholder groups including government, programme managers, healthcare digitalisation programmes and AI-specific programmes in Table 4.

[45] Mozaffar H, Candi M. Extending the process frontier of digital transformation: A flow-oriented perspective. Information Systems Journal. 2024 Aug 10.

[46] McKinsey. The state of AI in 2022—and a half decade in review. Available from: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review (last accessed: 23/09/2024).

Table 4: Lessons for various audiences

Lessons for major digitalisation programmes in other areas – government	• Funding of work through discrete programmes - lack of strategic integration between them - failure to build cumulatively on programmes
	• From projects to programmes and from local to system innovation – longer timeframes needed
	• A level of risk must be expected (and planned for e.g. through flexible timeframes/contracts in programmes that involve innovation)
	• Explore links and potential alignment of programmes to avoid silos and duplication. For example, several publicly funded initiatives independently supported the development of platform infrastructures to validate and implement AI tools - including the now defunct UKRI funded iCAIRD (the Industrial Centre for Artificial Intelligence Research in Digital Diagnostics) and the AI Lab's Deployment Platform - but these experiences were not systematically integrated.
	• Accept that progress is unlikely to be linear and outcomes are not predictable as technologies, needs, and environments change
	  • Nevertheless, need monitoring that captures progress towards the desired outcome
	• Seek to mitigate effects of volatile macro-environments (e.g. changing ministers with their agendas, budgets and objectives, COVID-19) e.g. by establishing long-term priorities and vision
	  • Seek stability in objectives, staff and allocation of resources of programmes
	  • Give programme managers autonomy and flexibility to deploy resources in line with emerging need
	• Establish clear lines of accountability, especially in programmes that span organisational boundaries and involve various organisational departments
	• Streamline reporting requirements as efforts associated with reporting can distract from delivery

	• Understand and resource local capacity and capability to implement technology

Lessons for major digitalisation programmes in other areas – Programme leaders/ managers	Programme leaders
	• Big transformation programmes have medium and long-term benefits that need to be captured
	• Many benefits result from learning (especially in experimental initiatives); lessons and capabilities are sometimes dissipated on programme closure, need to think how to capture these
	• Delivery team must possess (or have access to) the right skills and resources, and a degree of autonomy
	• Understand what incentivises different stakeholders and align these with programme needs
	• Independent real-time formative evaluation to proactively identify and mitigate risks
	  • Evaluation of GMPs now mandated by Cabinet Office which is encouraging
	  • Evaluate "failure" as carefully as "success" Programme managers
	• Robust project management processes to ensure oversight, streamline processes and handovers if team members change
	• Establish baselines to track progress and evidence benefits
	• Evaluate impact and value

Lessons for healthcare digitalisation programmes	• Healthcare digitalisation is a long and difficult journey - more difficult than many who are attracted to the health domain realise
	  • Introducing innovation and scaling: The AI Lab hypothesis that these were difficult was amply confirmed – indeed they were more difficult than anticipated
	  • Stakeholders struggled to implement, use and evaluate AI models in the timeframes of the AI Lab
	  • Longer-term and more flexible arrangements are needed to support real-world deployments

	• Healthcare is a high-risk environment and hard for suppliers to get into market, cannot generalise across all sectors (specifics of health and AI)
	• An ongoing learning journey
	• Need to focus on whole technology lifecycle (address flow of AI products through multiple lifecycles rather than single tools) and systemic changes enabled - strategic policy importance of focusing on scaling and sustainability (issue that nobody knows how to do this in an organisation of the size and complexity of the NHS so experimentation essential)
	• Tension between managing uncertain processes and risk around increasing flow of innovations and DHSC/NHSE risk-averse approach; lack of established processes for managing experimental developments like AI (project management/benefits realisation frameworks adopted were inhibiting), question of how to enhance rigour without inhibiting rigidity
	• Digital tools add to the process of service (re)design but should not be seen as the primary driver
	  • Systemic effects from redesigning pathways around AI opportunities (but difficult as they emerge gradually and are hard to predict and control)
	• Need to attend to innovation lifecycle and continued evolution of techs over multiple cycles – with ever-shortening development cycles - if we can get regulation right
	• Active involvement of NHS and needs-based approaches are crucial
	• Embedding patient and public engagement approaches from the beginning will improve technology development
	• Need to acknowledge that some issues will never be resolved (“wicked” problems) but require ongoing management
	  • Ongoing issues with top-down versus bottom-up strategies, national guidance and local involvement

	• Role of government in orchestrating and in sending signals to the market

Lessons for AI projects and programmes	• Need to balance potential benefits of emerging models with desire for comprehensive evidence base on existing models
	• New approach to managing risks and benefits is needed
	  • Cannot fully evaluate all the risks of AI in one go, before it is implemented. Each setting must implement it at risk and continuously evaluate as algorithm performance changes over time and across contexts.
	• Benefits are clearer/larger in situations where a service/pathway is redesigned rather than just automating a specific task
	  • Focus to date has been on building models and not on optimising their exploitation
	  • Need to either integrate service redesign in tool design or look at how organisations learn to make effective use of these tools in subsequent optimisation
	• Pay attention to variations in context – need for ongoing scrutiny of performance between settings and over time
	• AI presented unique challenges that went beyond regular project management, such as handling data governance, legal implications, and the rapid pace of AI technological development
	• Key is to manage expectations of various stakeholders (that are likely to be conflicting) and ensure evidence-based decision-making (balancing enthusiasm for innovations with slowly emerging evidence base)
	  • Build incentive structures that align with these goals
	• Understanding AI-specific markets (small/start-ups, quickly evolving, fragile operating models)
	  • There was a strategic shift towards platform deployment models – the AI Lab was in the vanguard of development – but

	future pathways are still not fully resolved – this is a key area for further investigation • Governance and regulation are key – maintain active links with regulatory bodies • Need to develop new agile but evidence-based models of evaluation (independent to a degree) and link these to procurement to align early • More granularity is required in guidelines for evaluating AI, to support identification of specific metrics relevant to clinical pathway transformation, and to enable consistency in evaluation of impact and value • Need to consider processes for national procurement of well-evidenced technologies • The NHS needs to confront long term uncertainty about which aspects can best be managed centrally and which locally (e.g. scope to standardise elements of information governance frameworks nationally whilst many elements of infrastructure must be installed locally)

Conclusions

The AI Lab has helped to place the UK at the forefront of efforts to implement AI safely in health and care settings. It bridged the gap between research programmes and deployment initiatives, creating a targeted space where insights could inform practical implementation. It also created a wealth of evidence and experience that now needs to be built upon.

However, the AI Lab faced challenges around achieving a balance between leveraging the potential benefits of emerging technology and developing a strong evidence base to ensure safe and efficient deployment in a turbulent environment. Some of the challenges experienced are not unique to the AI Lab. Digital transformation programmes must develop resilience to withstand various forms of organisational and environmental turbulence, ensuring their goals and outcomes are sustained regardless of shifting political priorities and evolving technologies and needs. Establishing clear baselines at the beginning of each programme is essential for tracking progress and assessing impact.

Nevertheless, AI, as a rapidly evolving field, has specific needs that require careful consideration. To maximise its benefits for health and care while avoiding harm, national guidance and oversight are essential. There are challenges of orchestrating local and national efforts whilst avoiding

duplication, enabling knowledge exchange, and maximising benefits of economies of scale. Key is to establish a shared understanding across stakeholder groups as to how this ecosystem should evolve. Achieving such a shared understanding would be a significant achievement given the political significance, scale of the NHS and the highly distributed nature of decision making.

Although there is to date only limited evidence of impact and value, it is important to recognise that we are only at the beginning of the journey towards transformation of health and care through AI. Key in this journey towards a future, AI-enabled, health and social care system will require looking beyond the short-term benefits associated with task automation towards more far-reaching transformation opportunities that may emerge gradually with greater deployment experience.

Contributors: KC led the drafting of the manuscript and led on data collection and analysis for the AI Awards. RW led the infrastructures work, SA led the ethics and regulations work, HM led the knowledge sharing and communities of practice work, and SD and MB led the value and impact work. XY extracted and analysed AI Award data, SB supported data collection and analysis of Skunkworks and VS led on a case study. SE provided overall strategic guidance and RF led project management. All authors analysed the data and commented on drafts of the manuscript.

Conflict of interests: MB, RW and KC were commissioned as the TSET for one of the Phase 4 AI Award projects.

Funding: This article has drawn on a programme of independent research funded by NHS England. The views expressed in this publication are those of the authors and not necessarily those of the NHS, NHS England, or the Department of Health and Social Care.

Acknowledgements: We gratefully acknowledge the input of the participants, and the AI Lab team, in particular Dr Alison Tweed and Dominic Cushnan. We also thank the Independent Advisory Board of this project consisting of Professor David Bates, Professor Elske Ammenwerth and Professor Farah Magrabi.

Appendix 1

Case study data extraction template

Name of company and its product	Company Product
Date of case study completion	
Data sources	Interviews Documents Press releases Company website
Methods	Data collection and analysis
Summary	Brief overview of the company and the purpose of the case study Summary of the main findings or conclusions
Company background	Company overview: history, mission, vision, and core values Products/services: description of the main offerings Industry context: overview of the industry and market positioning Size and location: number of employees, headquarters, and geographical presence Implementation/adoption progress
Timeline and history	Brief history of the company, how and when did it become involved in the AI Award? How much funding did it get from the AI Awards and other sources? Where is the company now? Changes in ownership?
The innovation	Brief history of innovation Building on prior development? Who initiated, how developed, and matured towards market readiness (internal players plus external partners/settings)?

Validation and regulation	Performance trials completed; regulatory approvals secured
Process	Main achievements Challenges/tensions faced and the way these were addressed Sustainability and scalability (including commercial models going forward) Relationships with other stakeholders (NHSE, DHSC, evaluators) Lessons learned
Impact and value	Key metrics: revenue, market share, cost savings, efficiency improvements, or customer satisfaction How did the results compare to the original objectives?
Future plans	
Any other information	

AI Award reports data extraction

Supplier	
Project/Product	
Phase	1,2,3,4
Uses	Diagnostic, administrative, preventative/screening, remote patient monitoring, p4 medicine, generative AI
Modality	Imaging (X-rays, CT scans), wearables, electronic health records
Setting	e.g. hospital, primary care, telemedicine
Organisation	Type: university-affiliated, SME, hospital, large UK or multi-national (if multi-national then origin)

Reported timeframe	
Independent evaluation	Yes/no
Contract value	
Contract end date	
Consent to commercialise	Yes/no/to be confirmed
Proposed impact	
Surfaced impact	
Number of publications	
Feedback for AI award team	
Value keywords	
Process insights	
Limitations	Reported and not reported
Other Comments	

Appendix 2

Table A2.1: AI Award Phase 4 project impact and benefit review summary

This table provides the following information:
a) a summary of the realised benefits reported, whether they were short- or long-term and comparator group;
b) the broad medical category or whether operational in nature;
c) the use case for the technology;
d) whether the setting was primary or secondary care; and
e) whether the realised benefits were explicitly recorded in the awardee/TSET reports. The last two columns provide a red-amber-green assessment of our two main axes of evaluation: 1) the degree of technology maturity and healthcare penetration at the beginning of the project, and 2) the appropriateness of the evaluation programme design considering potential for realising benefits a priori.

Award	Benefits summary	Category	Use	Setting	Benefits Logged?	Technology maturity	Company/ TSET
Project 1, Phase 4	Benefits evidenced: early diagnosis and subsequent treatment adhering to national targets and leading to a positive RoI based on reduced 5-year care costs and improved quality of life.	Neurology	Diagnostic	Hospital	Yes		
Project 2, Phase 4	Evaluation completed. Benefits not clearly evidenced - potential for re-design of model.	Operational	Administrative	Hospital	No		

Project 3, Phase 4	Evidence suggests no benefit over optimised non-AI pathway (whilst a positive RoI is reported comparing to scenario in which General Practitioners carry out triaging, this is not the case if the triaging is carried out by administrative staff)	Operational	Administrative	Primary care	No		
Project 4, Phase 4	Retrospective study, so no opportunity for benefit realisation under current definition	Ophthalmology	Diagnostic	Primary care	No		
Project 5, Phase 4	Evidence suggests no benefit over optimised non-AI pathway (whilst a positive RoI is reported comparing to a face-to-face pathway, this is less that comparing non-AI digital pathway to face-to-face)	Oncology	Diagnostic	Secondary care	No		
Project 6, Phase 4	Short term benefits evidenced, however, longer term benefits unclear. Evidence supports early diagnosis - so high potential for longer term benefits	Pulmonary	Diagnostic	Hospital	Yes		
Project 7, Phase 4	Short term benefits evidenced, however, longer term benefits unclear. Evidence suggests potential reduction in patient quality of life	Oncology	Diagnostic	Hospital	Yes		

Project 8, Phase 4	Company change in direction to focus on new devices / diseases - no plans to commercialise in UK- value of award is in learning	Cardiology			No		
Project 9, Phase 4	Benefits not clearly evidenced, lack of evidence to generalise to real-world settings	Operational	Administrative	Secondary care	No		
Project 10, Phase 4	Delayed – funding increased	Oncology			N/A		
Project 11, Phase 4	Stopped	Oncology	Diagnostic	Hospital	N/A		
Project 12, Phase 4	Ongoing as delayed – no cost extension	Endocrinology	Preventative screening	Social care	N/A		
Project 13, Phase 4	Ongoing as delayed – no cost extension	Cardiology	Diagnostic	Hospital, primary care, social care	N/A		
Project 14, Phase 4	Retrospective study, so no opportunity for benefit realisation under current definition.	Orthopaedic	Diagnostic	Hospital	No		
Project 15, Phase 4	Ongoing as later stage (not delayed)	Oncology	Diagnostic	Hospital	N/A		

Project 16, Phase 4	Ongoing as later stage (not delayed)	Gastroenterology	Diagnostic	Hospital	N/A		

Table A2.2: AI Award Phase 3 project impact and benefit review summary

This table summarises the following information: a) a red-amber-green (RAG) assessment of whether the technology underpinning this Phase 3 project has potential for entering Phase 4; b) a RAG assessment of whether the project has generated learning around real-world testing and routes to implementation; c) assessment comments accompanying the previous two columns; d) considerations for scaling at the end of the award including barriers identified; e) the level of self-reported technology maturity (including regulatory evidence); f) knowledge of revenue generation; g) knowledge of the product being exported; h) private funding raised; i) public funding raised; j) NICE approval; k) whether the project has validated the clinical pathway in which the technology is integrated; l) whether the project has uncovered efficiencies in the clinical pathway under study; m) whether the project has generated evidence of cost effectiveness; n) whether the project has generated evidence in the scientific literature or other specialised outlets; o) regulatory clearance achieved; p) whether adoption increased as a result of the project; q) the broad medical category or operational nature; r) the use case for the technology; and s) the clinical setting (e.g. primary, secondary care).

Project ID	(a) RAG assessment - indicator for Phase 4 success	(b) RAG assessment - indicator of project learning value	(c) Assessment comment	(d) Considerations for scaling
1	red	green	Study was multi-centre Randomised Controlled Trial (RCT) aimed at showing non-inferiority with standard of care. Company changed focus during trial due to changes in clinical context and does not intend to commercialise technology.	No further development
2	green	amber	Economic evaluation not included, however, plan for assessment considers both pathway, longer	

			terms NHS efficiencies and patient benefits. Strong commercial and publication record.	
3	green	green	Retrospective/prospective/RCT studies all show effectiveness of intervention. Triages 33% of cases away from Multidisciplinary Team (MDT) meetings, potential for savings not quantified. Proof of concept IT integration and patient acceptability. Health economic analysis underway.	Data quality challenges noted stemming from electronic health record (EHR) - suspected barriers to scaling due to local EHR systems
4	amber	amber	Whilst positive commercial indicators, the project does not address integration to pathway. Project is framed around the impact of technology on patient mental wellbeing but evidence not statistically significant due to small sample size	Small target population - may be an indicator of need for optimising towards a more prevalent cohort
5	green	green	Study focused on assessing acceptability, technical feasibility and safety of providing live AI predictions in MDT meetings. No evaluation of efficiencies introduced or cost-effectiveness.	One provider organisation commissioned the supplier to undertake a retrospective analysis of winter pressures data to identify the patients missed by standard analysis during winter 2023 with a view to deploying the model suite in operation for winter 2024
6	amber	green	Open-source medical device - this non-commercial project explores and offers valuable learning around deployment cloud-based solutions in the NHS that can save other innovations time and money	Status of adoption – stalled. Project reported to have no mechanism to deploy a cloud based open-source medical device back to the NHS without the support of a third party. They have identified a third party willing to provide cloud as a service and act as legal manufacturer.
7	green	green	Four parts: retrospective accuracy/fairness evaluation, evaluation of reader+AI, integration feasibility study Patient and Public Involvement and Engagement (PPIE)	A business case is being developed under a health technology assessment grant to transform its role from a clinical research tool to a clinical utility tool
8	green	green	Significant commercial success / ongoing and growing adoption of technology through the project sites. Whilst not widely published (perhaps	

			due to commercial rather than academic nature of company) they have a growing presence in a workforce advisory capacity.	
9	green	green	Combination of removing administrative inefficiencies, optimising clinical tasks (e.g. medication reviews), and identifying actionable events that would have been otherwise missed leading to potential negative patient outcomes	Strong international foothold
10	green	green	Retrospective evaluation based on research-level EHR data. Smaller pilot of real-world testing. Standardised final report not available so cannot complete commercial review. Report presents simulated economic assessment on NHS resources upstream to diagnosis and pathway for diagnosis of diseases and acknowledges longer term impact although due to the rare nature of disease the impact is not modelled. Key partnerships with genomics networks have been established.	

(e) Level of maturity	(f) Is the product generating revenue?	(g) Product exported ?	(h) Private funding?	(i) Public funding?	(j) NICE approval?	(k) Validation of an existing pathway?	(l) Address efficiency in pathway?	(m) Address cost Effectiveness?
CE mark	no	no	no	yes	n/a	yes (but negative outcome)	yes	no
Efficacy demonstrated, effectiveness demonstrated, final prototype for regulatory approval, real world evidence generated, adopted in the NHS	yes	yes	yes	high profile award	no	yes	yes	yes
Proof of concept, feasibility study, efficacy demonstrated, effectiveness demonstrated, prototype, final prototype for regulatory approval, CE/UKCA marking or other obtained, Real-World Evidence generated, adopted in the NHS, IP licensed	no	no	yes	no (only this award)	no	yes	yes	yes
Technology Readiness Level (TRL) 9	yes	no	yes	yes	yes (for research)	no	no	yes

Acceptability demonstrated, implementation in NHS	yes	no	n/a	yes	no	yes	no	no
Clinical deployment awaiting regulatory approval by an identified manufacturer	no	no	no	no	included in NICE Medtech Innovation Briefing	yes (although not reported here)	yes (although not reported here)	yes (although not reported here)
Technology Readiness Level (TRL) 3	no	no	no	n (only this award)	no	yes	yes	no
Commercially available and being sold globally	yes	yes	yes	no	n/a	yes	yes	yes
Technology Readiness Levels (TRL) 8 - 9	yes	yes	yes	yes	n/a	yes	yes	yes (although details of evaluation not available)
Class 1 device						yes	yes	Not modelled But acknowledged

(n) Generating Evidence?	(o) Regulatory clearance	(p) Adoption from Award?	(q) Category	(r) Uses	(s) Setting
no	CE, FDA	no	cardiology	diagnostic	hospital
high impact journal publications	UKCA mark, CE mark, FDA approval	100 GP sites	operational	diagnostic	primary care
yes, publications under preparation	CE mark	1 hospital site / team	oncology	P4	hospital
one discussion article submitted at point of final report	Class 1 Medical Device license, Cyber Essentials, HIPPA, DTAC	During the project we also secured commercial agreements for 14 NHS Talking Therapy Services	psychiatry	generative	non-acute, outpatient care
yes, publications under preparation	ISO13485 – Quality Management ISO14155 – Clinical Investigation ISO27001 – Information Security DCB0129 DCB0160 DTAC DSPT WAI WCAG 2.1 Level AA	no	pulmonary	remote patient monitoring	non-acute, outpatient care

yes, high impact journal publication	ISO 9001 for Radiotherapy QMS ISO 16485 for development IEC 62304:2006 for lifecycle management	one site continued, one discontinued to use a commercial solution	oncology	diagnostic	hospital
yes, publications under preparation	None, plans for CE marking	no	oncology	diagnostic	hospital
no significant publications but presence in related field via advisories	Cyber Essentials Plus certified, NHS Data Security and Protection Toolkit, HIPAA and GPDR compliant		operational	administrative	hospital
yes, publications	DCB0169 / DCB0129 / DTAC	yes	operational	administrative	secondary care
publication planned	Class 1 medical device	Working with the NHS Genomics Medicine Service "Genomics Network of Excellence" programme to expand across the Central and South and Southeast regions, hoping to expand to up to 20 million patients in 2025	rare disease	diagnostic	primary care