

<https://doi.org/10.1038/s41746-025-01805-w>

A mixed methods formative evaluation of the United Kingdom National Health Service Artificial Intelligence Lab



Kathrin Cresswell¹✉, Robin Williams², Sheena Dungey³, Stuart Anderson⁴, Miguel O. Bernabeu¹, Hajar Mozaffar⁵, Xiao Yang^{2,4}, Varun Sai¹, Sara Bea¹ & Sally Eason³

Internationally, health systems are investing in Artificial Intelligence (AI) to improve safety, quality, and efficiency, yet many efforts remain localised and do not progress beyond early development stages. In 2019, National Health Service (NHS) England and the Department of Health and Social Care launched the AI Lab to accelerate safe AI adoption. We conducted a mixed-methods evaluation of the AI Lab, analysing 1021 documents and 85 stakeholder interviews. The AI Lab made important contributions to national AI policy, regulation, and capability building, and positioned the United Kingdom as a global leader in AI deployment for health. Despite progress, implementation and scaling were hindered by shifting objectives, limited capacity, and systemic misalignment with service needs. Some AI technologies demonstrated high return on investment and improved clinical processes. Lessons from the AI Lab highlight critical socio-organisational factors, gaps in scaling support, and the need for sustained coordination to realise the long-term benefits of AI in health and social care systems.

The National Health Service (NHS) Artificial Intelligence Laboratory (hereafter AI Lab) was established in 2019 as a major government initiative with an initial funding allocation of £250 million¹. 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 delivery². Building on the vision of data-driven transformation, and 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^{3,4}.

The AI Lab began at a time when many nations had launched national AI strategies that involved significant public resource allocation for AI research across sectors⁵. Many of these focus on setting up dedicated AI research institutes, national AI research resources, establishing centres of excellence, and promotion of multi-sector collaboration^{5–20}. Key areas of interest include fundamental research; human–AI collaboration; ethical, legal, and societal implications; AI safety and security; development of shared public datasets and environments; measurement and evaluation; workforce needs; and expanding public–private partnerships. Various government programmes have funded AI research and development, but the productivity impacts of such interventions are not clear²¹. This may partly be due to a difficulty measuring longer-term downstream impacts.

In addition, AI programmes do not pay sufficient attention to promoting scaling and to advancing evaluation methods surrounding these rapidly evolving technologies. There is also very limited understanding of how to bridge the gap between research and implementation. In contrast to other AI initiatives, the AI Lab focused on embedding research on the implementation, deployment and scaling of AI initiatives in real-world settings. It remains unique in its scale and systematic approach to these issues.

Progress in the digital transformation of health and care has been slow and uneven²². 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 time²³. It was within this context that the AI Lab was envisioned to support the wider development and deployment of AI technologies at scale across the health and care system²⁴.

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) health technology and AI sectors (Box 1)²⁵. 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

¹Usher Institute, University of Edinburgh, Edinburgh, UK. ²Institute for the Study of Science, Technology and Innovation, University of Edinburgh, Edinburgh, UK.

³NHS Arden GEM Commissioning Support Unit, Warwick, UK. ⁴School of Informatics, University of Edinburgh, Edinburgh, UK. ⁵Business School, University of Edinburgh, Edinburgh, UK. ✉e-mail: Kathrin.Cresswell@ed.ac.uk

Box 1 | Structure and institutional setup of the AI Lab

- In 2019, the National Health Service England (NHSE), 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 (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.
- 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; community building; research and evidence; and governance and ethics.
- The AI Lab was set up as a part of NHSX, a joint unit bringing together teams from DHSC (responsible for setting national health and care policy), NHSE (overseeing, planning and funding healthcare services), and NHS Improvement (supporting providers to improve quality, finance and performance) to drive digital transformation. In 2022, NHSX was merged into the NHS England Transformation Directorate.
- The AI Lab's structure included broader support for strategy and policy development and programme management as well as delivery programmes.
- The AI Lab work involved collaboration with a wide range of institutions and stakeholders, including Government bodies, health and care providers, academics, technology companies, regulatory bodies, and the public.
- During the period of this evaluation the AI Lab project closed successfully and members transitioned into the National AI Team that aims to continue supporting AI adoption and implementation across the NHS in England.

Boris Johnson and his Chief Advisor Dominic Cummings²⁶. The AI Lab was jointly delivered by NHS England (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)²⁷. 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)²⁸. Additionally, the AI Lab focused on governance, helping to shape AI policy, regulation, and ethics within the health and care sector. Box 2 illustrates the different components of the AI Lab.

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 the early stages of development, which is why these applications were not considered by the AI Lab. At the time of writing there are no approved medical devices that deploy generative AI.

Our evaluation of the 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 an external independent review. Our evaluation explored the AI Lab as a mechanism to support and accelerate systemic change through AI in health and care settings. As part of this systemic change, the AI Lab provided research funding for evaluations, which in turn led us to obtain valuable insights into methodological best practice.

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

Results

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

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

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

An ambitious experimental initiative in a complex ecosystem

The 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 UK Government's priorities included innovation and scaling of technologies, but this often conflicted with NHS provider structures.

The question of how to accelerate adoption and scaling within the NHS is a recurring topic among Number 10 [10 Downing Street, the official residence and executive office of the Prime Minister of the United Kingdom] 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

Government's priorities were also not always viewed as reflecting system-level priorities, as the following quote illustrates.

...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

Box 2 | 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
- Fostering a community of practice of AI practitioners in health and care by providing opportunities for co-production

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 the 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 the 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, but no Phase 5 projects were funded
- Projects were implementation-oriented innovation projects designed to support the scaling of existing AI technologies. The objective was to facilitate the integration of AI tools into practice.
- Much of the funding provided by the AI Lab was aimed at exploring and attempting to resolve the well-documented uneven performance and

occasional decline in performance of AI models in “real-world” operational healthcare contexts. The larger-scale projects funded by the AI Lab typically explored the technology transition from Technology Readiness Level (TRL) 7 or 8 to TRL 9. This stage of transition is typically excluded from state-funded research and innovation activity. For example, European Union Research and Innovation Actions generally support transitions in the TRL range of 2–6, while Innovation Actions support transitions in the range of TRL 4–8.

- Biggest and most expensive component of the AI Lab

Skunkworks projects

- Focused on rapidly developing and testing short-term, exploratory, proof-of-concept projects (lasting between 6 and 8 weeks) using agile development
- The AI Lab team worked with NHS organisations and data scientists to build product prototypes, mostly open-source and publicly available
- Injected AI capability into organisations for a short time and helped create their own capability
- Facilitated 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

AI ethics initiative

- Researched issues arising from bias and discrimination in algorithms and how to safeguard against this
- Governing the use of data
- Workforce training to build capability and confidence

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. The following quote illustrates that the AI Lab was initially envisaged as a series of experiments.

...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.

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 a serious disruption. The pandemic shifted strategic priorities and diverted resources, but it also created new opportunities for

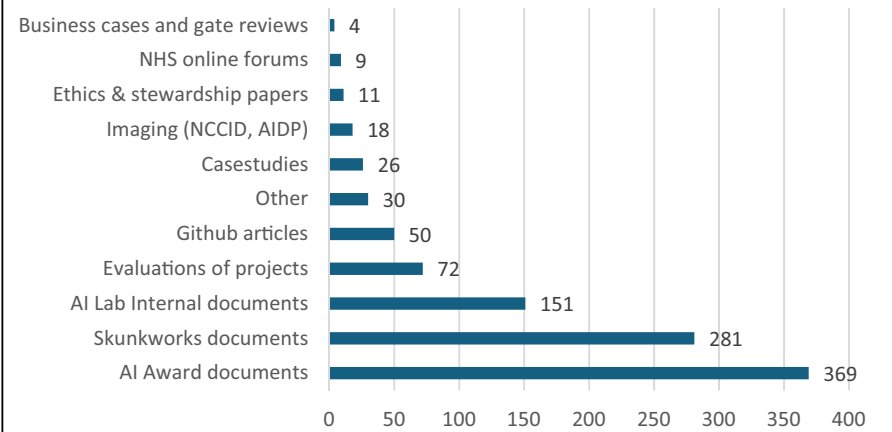
AI, such as the development of the NCCID. More widely, COVID-19 also focused the NHS on data, automated interpretation of data and the benefits of sharing infrastructure in enabling the response of the health system. This shifted attitudes towards digitalisation of health and care delivery, coupled with rapid product development by vendors, and opened the minds of many to the potential benefits of the work of the AI Lab. As evidenced in the following quote, many resources were diverted during the pandemic, including those of provider organisations and those within the AI Lab itself.

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... 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'd 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

Table 1 | Overview of the dataset

Data Collected																							
85 interviews	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%)	Number of interviewees in the interview	Number of interviews	1	76	2	8	3	1														
Number of interviewees in the interview	Number of interviews																						
1	76																						
2	8																						
3	1																						
12 observations	Two project progress meetings, four AI Lab board meetings, six workshops																						
1021 documents	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:																						
	Subject area of documents reviewed <table><tr><td>Business cases and gate reviews</td><td>4</td></tr><tr><td>NHS online forums</td><td>9</td></tr><tr><td>Ethics & stewardship papers</td><td>11</td></tr><tr><td>Imaging (NCCID, AIDP)</td><td>18</td></tr><tr><td>Casestudies</td><td>26</td></tr><tr><td>Other</td><td>30</td></tr><tr><td>Github articles</td><td>50</td></tr><tr><td>Evaluations of projects</td><td>72</td></tr><tr><td>AI Lab Internal documents</td><td>151</td></tr><tr><td>Skunkworks documents</td><td>281</td></tr><tr><td>AI Award documents</td><td>369</td></tr></table>	Business cases and gate reviews	4	NHS online forums	9	Ethics & stewardship papers	11	Imaging (NCCID, AIDP)	18	Casestudies	26	Other	30	Github articles	50	Evaluations of projects	72	AI Lab Internal documents	151	Skunkworks documents	281	AI Award documents	369
Business cases and gate reviews	4																						
NHS online forums	9																						
Ethics & stewardship papers	11																						
Imaging (NCCID, AIDP)	18																						
Casestudies	26																						
Other	30																						
Github articles	50																						
Evaluations of projects	72																						
AI Lab Internal documents	151																						
Skunkworks documents	281																						
AI Award documents	369																						

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, begun to understand AI markets, and facilitated learning on how to develop, implement, and evaluate AI technologies. The following excerpt shows how

the AI Lab may be viewed as a series of natural experiments that allow learning and contribute to the empirical evidence base surrounding AI implementation and adoption processes.

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

Table 2 | Overview of findings^a

An ambitious experimental initiative in a complex ecosystem (majority of interviews with key stakeholders reflecting retrospectively on experiences, e.g., lessons learned and evaluations, plus limited evidence from documentary sources).
• 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 (all interviews, some evidence in documents, but not always explicit).
• 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 the whole ecosystem.
Generating real-world evidence to support market authorisation and large-scale adoption (AI Awards documents, including progress reports and evaluations).
• 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 is uncertain. £44M in benefits recorded.
• Significant first real-world testing in care settings was generated, including evidence for routes to implementation. Eight out of ten projects contributed (or are in the process of contributing) to the scientific literature.
Scaling of technologies and pathways to procurement, reimbursement, deployment and operation (interviews with suppliers and AI Lab staff working on commercialisation).
• 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.

^aInformation in brackets refers to sources of evidence for these findings.

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. The quote below highlights the important contribution of the AI Lab in relation to regulation.

...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. The excerpt below captures the difficulties encountered with the AI Deployment Platform, which was initially planned to be rolled out

nationally, but this did not materialise. Nevertheless, the work provided important learning opportunities.

...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 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... 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 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.

Interview 9, DHSC

Turbulent macro-environmental influences and political drivers

The AI Lab's strategy and delivery were 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. As evidenced in the following quote, the drivers underlying the establishment of the AI Lab at the Government level were to position the UK at the forefront of using AI in health and care.

The work was done and ultimately with Boris Johnson [the then Prime Minister] and Matt Hancock's [the then Health Minister] 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 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. NHSX, the initial delivery unit for the AI Lab, was dissolved in 2022. This led to staff changes, uncertainty, and restructuring within the programme. Departures and associated disruptions meant that painfully acquired expertise and knowledge dissipated. The effects on the ground of changes at the centre are illustrated by the following quote.

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

Changes in government were at the core of changes in the AI Lab's strategic direction.

...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 election's 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 drove a shift away from experimentation toward a focus on evidencing the delivery of tangible benefits. Ring-fenced funding was reallocated, and planned programmes were cut.

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 £180 m down to £75 m 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 was escalated by requests from multiple disparate stakeholders with different reporting requirements and timeframes across departmental boundaries, driven by the need to demonstrate the value of investment. As a major government programme, the AI Lab was subject to high accountability requirements supported by reported evidence.

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. Government funding models with fixed timelines and costs potentially inhibit innovations that require more flexible models.

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 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. Organisations collaborating in the delivery of the AI Lab were often competing:

... 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.

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. Supplementary Table 1 and Table 3 summarise data extraction for the Phase 4 Award projects.

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. There were 10 Phase

Table 3 | Benefits realised for completed Phase 4 projects

Award	Benefit category	Benefit (as reported by Technology Specific Evaluation Team (TSET) evaluations for a modelled patient cohort size)	Number of patient participants in project (as reported by TSET evaluations)	Average benefit per patient (estimated from TSET reported benefit)	Estimated benefit per project (based on participant numbers)
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 quality-adjusted life years, 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 are associated with more accurate imaging reading, reducing the need for further tests	1471	£36	£53,489
Project 7, Phase 4	c. Efficiency –care longer term	Subsequent lifetime management savings to the NHS associated with more accurate imaging reading reducing activity in pathways of care	1471	£32	£46,516
Project 7, Phase 4	d. Patient health and wellbeing	Not possible to derive due to insufficient evidence generated as part of the TSET evaluation	1471		
Project 6, Phase 4	a. Efficiency –point of delivery	Clinician reporting time saved associated with more automated triaging in diagnostic pathway	597	£7	£4293
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	–£2178
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 the diagnostic pathway	597	£4	£2249
Project 6, Phase 4	d. Patient health and wellbeing	Not possible to derive due to insufficient evidence generated as part of the TSET evaluation			

4 projects 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 10 projects included in the analysis amounted to £22.5 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 (3/10). There were also instances (2/10) 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 a varying staff skill mix. One project (1/10) delivered a retrospective study only, limited by timeline and budget constraints, which impacted 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 (2/10) were able to demonstrate significant improvements in 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 Return on Investment (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 in longer-term care and improved patient outcomes modelled as a five-year (discounted) saving to social care and an increase in quality-adjusted life years (QALYs)²⁹. 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 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 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 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 the level of burden and unmet need with significant monetary implications and patient outcomes.

Ultimately, our analysis indicates that a project’s success in generating evidence for large-scale commissioning or deployment depended both on technology maturity/penetration and appropriateness of the evaluation programme. Supplementary Table 1 summarises the dimensions of appropriateness of the 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 Awards, we will now describe findings from Phase 3 Awards. Supplementary Table 2 summarises data extraction for the Phase 3 Award projects that had been 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 the feasibility of technology integration with existing clinical pathways (Project 5) to randomised controlled trials aimed at establishing the

effectiveness of interventions compared to standard care (Project 3). The latter matched the scope of a Phase 4 Award, which underscores the difficulty in clearly distinguishing between the two phases.

A considerable proportion of Phase 3 Award projects (7/10) reported sufficient evidence of having achieved the main objective of this phase, which was to support the 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 from 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 the 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 10 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 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. 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 the 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 coordination difficulties stemming from national procurement guidance and the inherent complexity of NHS procurement, which is managed independently by hospitals with diverse needs. The following statement exemplifies the convoluted procurement processes in provider organisations.

...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 [hospitals] 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

Suppliers highlighted the lack of a route to procurement and the impact on suppliers.

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 of America (USA) 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 the UK Medicines and Healthcare products Regulatory Agency (MHRA) on adapting regulatory processes to take account of AI and with the National Institute for Health and Care Excellence (NICE) on adapting the Evidence Standards Framework to establish a quality and cost-effectiveness framework including AI, in the UK, regulatory complexity was reported to be greater than in other countries (such as the USA).

Significant public funding for mature product development, implementation, and rollout mobilised expectations about RoI for the taxpayer. This presented novel challenges for NHSE/DHSC, which 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. This participant outlines the perceived lack of a robust commercialisation process.

... 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 a strong background in 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. This participant involved in contracting described the delicate balance between government intervention to promote a thriving supplier ecosystem and adversely influencing the market.

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 being sufficiently 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. Establishing a need before developing technological solutions is essential.

... 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. The following excerpt illustrates the challenges surrounding evidence and contextual variations.

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

The 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 the 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 development³⁰.

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 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.

As the programme was extensive and complex, several challenges emerged in the evaluation (Box 3).

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 were 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. 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.

Box 3 | Overview of key challenges of the evaluation

The AI Lab 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 assessing 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.

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 a review of evidence generated through the 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 further relied on gatekeepers to facilitate access to documents and participants. While this approach opened many doors and ensured we had all the 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 were limited, and tools were not routinely integrated into service delivery. There had been little opportunity to enhance the 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 been 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.

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^{31–38}.

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 headcount) 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. Specific targeted applications have shown promise in particular settings and we now know where benefits can be anticipated^{39–42}. Strategic decision makers 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, and balancing enthusiasm for innovation with the slow emergence of robust evidence⁴³.

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.

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 leadership⁴⁴; building necessary capacity, capability, and culture; maintaining a systems thinking perspective; clearly articulating evidence surrounding benefits; financial resources⁴⁵; harnessing the power of adopter communities; and managing supplier relationships^{46,47}. The AI Lab began to address some of these factors, such as promoting evidence generation, connecting stakeholders, and experimentation with building capacity and capability.

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 the coordination of strategic initiatives
 Uncertainty surrounding the sustainability of developed and implemented systems
 Need for a shift in focus from short-term gains to sustainable long-term digitalisation strategies

Contracts

Contractual re-negotiations over time leading to changing objectives and relationships
 Limited contact between developers and provider organisations
 Commercial interests may result in limited sharing of 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 the biggest benefits of deployment can be expected
 Quantifiable benefits are 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)
 The level of support required by provider organisations is often underestimated

Evaluation

Evaluations of complex transformation initiatives are challenging
 Evaluations must be systematic, independent, and ongoing, 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 the local level
 Issues with clinical engagement
 Lack of commitment from organisational leaders and end-users

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 contexts⁴¹.

Post-market surveillance of AI is a key element in controlling risk, but the costs and requirements for 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 administration⁴⁸. 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 settings⁴¹. To address this, service redesign focusing on prevention, population health and integrated care is crucial in the development of AI tools.

The market for health AI presents additional challenges. Small, quickly evolving startups have been drawn to the clinical applications market, often lacking understanding of the particularities of the potentially challenging health regulation and procurement landscapes. National coordination of public sector AI has been argued to be important in this context, but there is limited understanding of how this may be operationalised in health and care⁴⁹. 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 product⁵⁰.

Strategic shifts, such as the move toward platform deployment models trialled by the AI Lab, reflect efforts to scale AI and streamline implementation and procurement processes⁵¹. However, although promising, the AI Lab experience has shown the difficulties associated with concerted approaches to the procurement of platforms. 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 organisations⁵².

There is a broader literature on government intervention in innovation, including in sectors outside of AI and healthcare. For example, policy and innovation studies have explored how public investments can shape ecosystem development and institutional readiness for adopting new technologies. Ecosystem development can be promoted through financing research and development and boosting competitiveness; directly funding research institutions and basic science; promoting collaboration and structural change; creating regional hubs and talent pipelines; and by supporting adoption and fostering an innovation ecosystem^{53–56}.

Public investments and policy can support institutional readiness and adoption by stimulating research and development investment in firms; serving as early customers through procurement; investing in human capital and training; establishing regulatory frameworks and guidelines; and by helping to address implementation challenges^{54,57}.

As other countries are now actively pursuing system-level approaches to promote the development, adoption, and implementation of AI in health and social care, there are significant opportunities to learn from the experiences of the AI Lab. International strategies, which like the UK before

the AI Lab were geared towards increasing economic competitiveness and placed a strong emphasis on early stage AI research and development have recently started to recognise the importance of translating AI research into real world contexts^{5–20}.

It is unlikely that there will be another AI Lab in the UK, as needs and structures are evolving. The AI Lab presented the beginning of a much larger transition financed by wider operational funds of the NHS. There are, however, important lessons for early systemic programmes seeking to promote the safe implementation and adoption of AI in health and social care. These include an initial focus on developing both qualitative and quantitative baselines and establishing the most pressing system needs before any development and implementation activity. Programmes need to link together various disparate stakeholder groups, including supply, regulation, academia, health and care settings, and policy. A networking and community-building approach is therefore essential. As many stakeholders (e.g. technology supply, clinical researchers) are already focused on early-stage technology development, there is an important need for such programmes to focus on scaling and sustainability of systems. This will also need to involve the generation and curation of empirical evidence promoted as part of these activities. An ethos of ethical and safe development and deployment is essential throughout.

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 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 the 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 technologies⁵⁸. 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).

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 market⁵⁹. There may be scope for government to be responsible for overall planning and commercial companies to focus on delivery⁶⁰.

In addition to insights on how to orchestrate systemic change, our detailed review of evaluation methods has also allowed us to obtain valuable insights into methodological best practice. 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 UK Government guidance on impact evaluation of AI interventions (which aims to promote high-quality evaluation to improve public services, and ensure accountability for public spending)⁶¹. 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.

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

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.
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 health system 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.

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 momentum⁶². 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 implications⁶³.

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.

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 the 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 the 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.

Methods

We received ethical approval from the Research Ethics Committee at the University of Edinburgh School of Social and Political Science (reference number 3228). We obtained written informed consent from each of the participants.

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

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 a 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 to ensure representation of different stakeholders involved in AI Lab-related activities. We sampled a wide range of stakeholder viewpoints, including tool developers, managers, implementers, strategic decision makers, regulators, evaluators, academic experts, and professional bodies. Gatekeepers, who were the leaders of the AI Lab and also commissioned our research, provided us with an initial list of key stakeholders covering all AI Lab 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 examine 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 the ongoing strategy. Sample topic guides are provided in Box 6.

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.

Table 4 | Lessons for various audiences (* signifies the lessons identified directly by interviewees, whereas the others were identified by us as a direct result of qualitative data analysis)

Lessons for major digitalisation programmes in other areas – government	• Funding of work through discrete programmes—lack of strategic integration between these—failure to build cumulatively on programmes • From projects to programmes and from local to system innovation—longer timeframes are 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 for monitoring that captures progress towards the desired outcome • Seek to mitigate effects of volatile macro-environments (e.g., changing ministers and 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 needs • 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, but they need to be captured* • 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 is now mandated by the 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's 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 it is hard for suppliers to get into the market, issues cannot be generalised 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 is essential) • Tension between managing uncertain processes and risk around increasing flow of innovations and DHSC/NHSE's 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 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 technology over multiple cycles—with ever-shortening development cycles • Active involvement of the 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, balance between 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 the desire for a 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 (new approaches are needed)* • Key is to manage expectations of various stakeholders (that may 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* • 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)

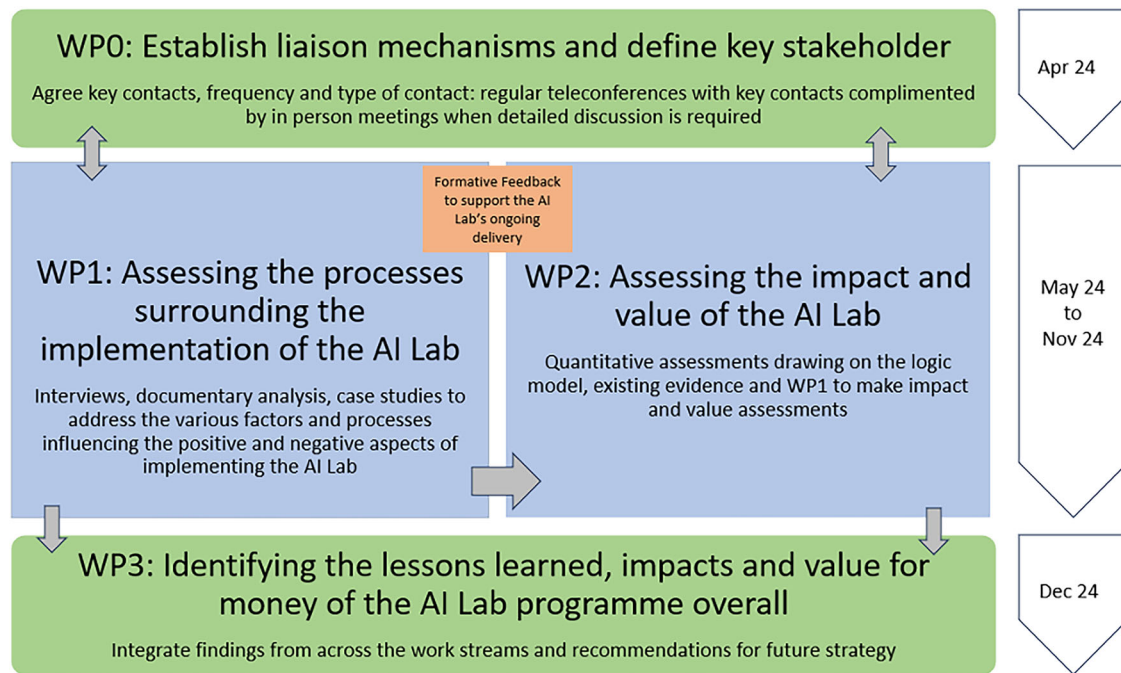


Fig. 1 | Overview of the methods.

Box 6 | 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?

1. Initial vision and to what extent this has been realised
2. Nature of programme activities and innovations
3. Expected and realised benefits
 - Impact on safety, quality, and efficiency of care
 - Impact on development processes of AI
 - Impact on the implementation of technology into practice
 - Impact on regulations and policies
 - Impact on market development
 - Value for money
 - Any other types of impact

4. Any unintended/unanticipated consequences (positive and negative)
5. Challenges and potential ways to address these
6. Differences between AI Lab components

Lessons learned

1. 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?)
2. How could impacts and value be maximised in the future?
3. Lessons learned and implications for future initiatives

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 Supplementary Tables 3 and 4.

Analysis

Transcripts and documents were uploaded into NVivo 12 software to facilitate coding. We used the Technology, People, Organizations, and Macroenvironmental (TPOM) factors framework as a coding framework to ensure that various technological, social, organisational, and macro-environmental dimensions were represented in the analysis⁶⁴. 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 in 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. Benefits were quantified by TSET evaluators based on a modelled cohort size. We derived an average benefit per patient from these estimates and extrapolated this figure to the actual number of patients who participated in each study, thereby estimating the total benefit realised during the study period. This approach assumes that the average benefit per patient, as determined from the modelled cohort, is representative of the actual study participants. It also presumes a linear relationship between the number of patients and the total benefit. While this method provides a practical means of estimating realised benefits, it is important to acknowledge these assumptions and the potential for associated uncertainties in the extrapolated results.

Whilst the approach to, and focus of, measuring economic benefit for Phase 4 was variable across all projects, those that 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 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 discussed 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.

Data availability

Anonymised metadata is available from the corresponding author upon request. Raw data are not publicly available to preserve individuals’ privacy under the European General Data Protection Regulation.

Received: 15 March 2025; Accepted: 16 June 2025;

Published online: 17 July 2025

References

1. UK Government. *Health Secretary announces £250 million investment in artificial intelligence*. <https://www.gov.uk/government/news/health-secretary-announces-250-million-investment-in-artificial-intelligence> (2024).
2. Panch, T., Szolovits, P. & Atun, R. Artificial intelligence, machine learning and health systems. *J. Glob. Health* **8**, 020303 (2018).
3. Sheikh, A. et al. Health information technology and digital innovation for national learning health and care systems. *Lancet Digit. Health* **3**, e383–e396 (2021).
4. Sharma, M. et al. Artificial intelligence applications in health care practice: scoping review. *J. Med. Internet Res.* **24**, e40238 (2022).
5. Fatima, S., Desouza, K. C. & Dawson, G. S. National strategic artificial intelligence plans: a multi-dimensional analysis. *Econ. Anal. Policy* **67**, 178–194 (2020).
6. US National Science Foundation. National Artificial Intelligence Research Resource Pilot. <https://www.nsf.gov/focus-areas/artificial-intelligence/nairr> (2025).
7. Smart Nation Singapore. National Artificial Intelligence Strategy. <https://www.smartnation.gov.sg/files/publications/national-ai-strategy.pdf> (2025).
8. Agency for Digital Government. The Danish National Strategy for Artificial Intelligence. <https://en.digst.dk/strategy/the-danish-national-strategy-for-artificial-intelligence/> (2025).
9. Ira Alanko. The AuroraAI Programme. https://www.europarl.europa.eu/cmsdata/240174/AIDA%20workshop_28052021_Alanko.pdf (2025).
10. Ministry of Health and Welfare. Healthcare 4.0 Era for Healthy Korea. https://www.mohw.go.kr/board.es?mid=a20401000000&bid=0032&tag=&act=view&list_no=375996 (2025).
11. Palaniappan, K., Lin, E. Y. & Vogel, S. Global regulatory frameworks for the use of artificial intelligence (AI) in the healthcare services sector. *Healthcare* **12**, 562 (2024).
12. Council for Social Principles of Human-centric AI. Social Principles of Human-Centric AI. <https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/humancentricai.pdf> (2025).
13. Australian Alliance for Artificial Intelligence in Healthcare. A National Policy Roadmap for Artificial Intelligence in Healthcare. https://www.mq.edu.au/_data/assets/pdf_file/0005/1281758/AAAIH_NationalAgendaRoadmap_20231122.pdf (2025).
14. Secretary of State for Digitalization and Artificial Intelligence. National Strategy for Artificial Intelligence (Spain). <https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/ficheros/National-Strategy-on-AI.pdf> (2025).
15. Catalan Data Protection Authority. Automated Decision-making in Catalonia. https://apdcat.gencat.cat/web/.content/03-documentacio/intelligencia_artificial/documents/Informe-IA-Angles-Report-Final.pdf (2025).
16. Norwegian Ministry of Local Government and Modernisation. National Strategy for Artificial Intelligence (Norway). https://www.regjeringen.no/contentassets/1febbb2c4fd4b7d92c67ddd353b6ae8/en-gb/pdfs/ki-strategi_en.pdf (2025).
17. National Program for Artificial Intelligence. UAE National Strategy for Artificial Intelligence. <https://ai.gov.ae/wp-content/uploads/2021/07/UAE-National-Strategy-for-Artificial-Intelligence-2031.pdf> (2025).
18. Government of Canada. The Artificial Intelligence and Data Act (AIDA). <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document> (2025).
19. Ministerie van Economische Zaken en Klimaat. Strategic Action Plan for AI. <https://oecd.ai/en/work/documents/netherlands-strategic-action-plan-for-ai-2019> (2025).
20. Die Bundesregierung. Strategie Künstliche Intelligenz der Bundesregierung. <https://www.bmwi.de/Redaktion/DE/>

- Publikationen/Technologie/strategie-kuenstliche-intelligenz-der-bundesregierung.html (2025).
21. Jacob, B. A. & Lefgren, L. The impact of research grant funding on scientific productivity. *J. Public Econ.* **95**, 1168–1177 (2011).
 22. Cresswell, K., Sheikh, A. & Williams, R. Accelerating health information technology capabilities across England's National Health Service. *Lancet Digit. Health* **3**, e758–e759 (2021).
 23. Fennelly, O. et al. Successfully implementing a national electronic health record: a rapid umbrella review. *Int. J. Med. Inform.* **144**, 104281 (2020).
 24. NHS England. About the NHS AI Lab. <https://transform.england.nhs.uk/ai-lab/about-the-nhs-ai-lab/> (2024).
 25. Arora, A. et al. Innovation pathways in the NHS: an introductory review. *Ther. Innov. Regul. Sci.* **55**, 1045–1058 (2021).
 26. The Guardian. Boris Johnson pledges £250m for NHS artificial intelligence. <https://www.theguardian.com/society/2019/aug/08/boris-johnson-pledges-250m-for-nhs-artificial-intelligence> (2024).
 27. NHS England. NHS AI Lab. <https://transform.england.nhs.uk/ai-lab/> (2024).
 28. NHS England. AI in Health and Care Award winners. <https://transform.england.nhs.uk/ai-lab/ai-lab-programmes/ai-health-and-care-award/ai-health-and-care-award-winners/> (2025).
 29. HM Treasury. The Green Book. <https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-government/the-green-book-2020#a6-discounting> (2025).
 30. Eilers, K., Peters, C. & Leimeister, J. M. Why the agile mindset matters. *Technol. Forecast. Soc. Change* **179**, 121650 (2022).
 31. National Audit Office. Digital transformation in government. <https://www.nao.org.uk/wp-content/uploads/2024/02/digital-transformation-in-government.pdf> (2024).
 32. University of Edinburgh. Final report: Global Digital Exemplar (GDE) Programme. https://www.ed.ac.uk/sites/default/files/atoms/files/final_report_gde_evaluation_programme.pdf (2023).
 33. Cresswell, K. et al. The Long and Winding Road: NHS Care Records Service Evaluation. <http://www.cphs.mvm.ed.ac.uk/grantdocs/526%20%20Final%20report%20v31st%20Mar%20FINAL.pdf> (2011).
 34. Checkland, K. et al. National evaluation of the Vanguard new care models programme: Interim report: understanding the national support programme. https://www.research.manchester.ac.uk/portal/files/103375904/Interim_report_of_the_NCM_external_evaluation_final_v1.pdf (2019).
 35. Hayes, G. Independent Review of NHS and Social Care IT. <https://ntouk.files.wordpress.com/2020/11/nhs-and-social-care-it-review-2009.pdf> (2009).
 36. Wachter, R. Making IT work: Report on Health Information Technology in England. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/550866/Wachter_Review_Accessible.pdf (2016).
 37. National Audit Office. Digital transformation in the NHS. <https://www.nao.org.uk/wp-content/uploads/2019/05/Digital-transformation-in-the-NHS.pdf> (2020).
 38. National Audit Office. Evaluating government spending. <https://www.nao.org.uk/wp-content/uploads/2021/12/Evaluating-government-spending.pdf> (2021).
 39. Alami, H. et al. Understanding AI integration via NASSS framework. *BMC Health Serv. Res.* **24**, 701 (2024).
 40. He, J. et al. Practical implementation of AI in medicine. *Nat. Med.* **25**, 30–36 (2019).
 41. Ghassemi, M. et al. Machine learning in health: challenges and opportunities. *AMIA Summits Transl. Sci. Proc.* **2020**, 191 (2020).
 42. Yeo, K., Li, Z. & Lin, W. Barriers and facilitators to AI adoption. *BMC Med. Inform. Decis. Mak.* **21**, 54 (2021).
 43. Mittelstadt, B. D. et al. Ethics of algorithms. *Big Data Soc.* **3**, 2053951716679679 (2016).
 44. Tagscherer, F. & Carbon, C. C. Leadership for successful digitalization. *Sustain. Technol. Entrep.* **2**, 100039 (2023).
 45. Bevan, H. et al. Spreading and scaling innovation and improvement: understanding why the differences matter. *J. Innov.* **10**, 69–73 (2024).
 46. Hinder, S. et al. Promoting inter-organisational knowledge sharing. *PLoS ONE* **16**, e0255220 (2021).
 47. Johnson, M. et al. The managed prosumer in information infrastructures. *Inf. Commun. Soc.* **17**, 795–813 (2014).
 48. Cresswell, K. et al. Evaluation of healthcare digitalisation. *J. Gen. Intern. Med.* **38**, 3610–3615 (2023).
 49. Wirtz, B. W. et al. AI in the public sector. *Int. J. Public Adm.* **42**, 596–615 (2019).
 50. Mikhaylov, S. J. et al. AI in the public sector: collaboration opportunities. *Philos. Trans. A Math. Phys. Eng. Sci.* **376**, 20170357 (2018).
 51. Bounfour, A. *Platforms and Artificial Intelligence* (Springer, 2022).
 52. Sheikh, A. et al. 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* **343**, d6054 (2011).
 53. Lau, T. Size matters when adopting and scaling AI. In *Banking on (Artificial) Intelligence* (Palgrave Macmillan, Cham, 2025).
 54. Howell, S. T. Government intervention in innovation. *Annu. Rev. Financ. Econ.* **16**, (2024).
 55. Larédo, P., Köhler, C. & Rammer, C. The impact of fiscal incentives for R&D. In *Handbook of Innovation Policy Impact* 18–53 (Edward Elgar Publishing, 2016).
 56. Campbell, D. M., Stickney, B., Milat, A. & Thackway, S. Improving health systems through innovation in population health and health services research. *Public Health Res. Pract.* **28**, 2831814 (2018).
 57. Zabala, F. J. *Grow Your Business with AI: A First Principles Approach for Scaling Artificial Intelligence in the Enterprise*. (A press LP, Berkeley, CA, 2023).
 58. Health Foundation. *Priorities for an AI Health Care Strategy* (Health Foundation, 2025).
 59. Cresswell, K. et al. Concerted adoption for digital transformation. *J. Am. Med. Inform. Assoc.* **31**, 1211–1215 (2024).
 60. Mazzucato, M. *Mission Economy* (Penguin UK, 2021).
 61. HM Treasury. *The Magenta Book* (HM Treasury, 2025).
 62. Mozaffar, H. & Candi, M. Extending the process frontier of digital transformation: A flow-oriented perspective. *Inf. Syst. J.* **35**, 720–760 (2024).
 63. McKinsey & Company. *The State of AI in 2022* (McKinsey & Company, 2024).
 64. Cresswell, K., Williams, R. & Sheikh, A. Developing and applying a formative evaluation framework for health information technology implementations: qualitative investigation. *J. Med. Internet Res.* **22**, e15068 (2020).

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. 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.

Author contributions

K.C. led the drafting of the manuscript and led data collection and analysis for the AI Awards. R.W. led the infrastructure work, S.A. led the ethics and regulations work, H.M. led the knowledge sharing and communities of practice work, and S.D. and M.B. led the value and impact work. X.Y. extracted and analysed AI Award data, S.B. supported data collection and

analysis of Skunkworks and V.S. led on a case study. S.E. provided overall strategic guidance. All authors analysed the data and commented on drafts of the manuscript.

Competing interests

M.B., R.W. and K.C. were commissioned as the TSET for one of the Phase 4 AI Award projects. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-025-01805-w>.

Correspondence and requests for materials should be addressed to Kathrin Cresswell.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025