Artificial Intelligence In Medicine 173 (2026) 103346

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Learning Health Systems provide a glide path to safe landing for AI in health

Vasa Curcin [a,*], Brendan Delaney [b], Ahmad Alkhatib [b], Neil Cockburn [c], Olivia Dann [a], Olga Kostopoulou [b], Daniel Leightley [a], Matthew Maddocks [a], Sanjay Modgil [a], Krishnarajah Nirantharakumar [a], Philip Scott [d], Ingrid Wolfe [a], Kelly Zhang [b], Charles Friedman [e]

[a] King's College London, Strand, London, WC2R 2LS, United Kingdom
[b] Imperial College London, Exhibition Road, London, SW7 2AZ, United Kingdom
[c] University of Birmingham, Birmingham, B15 2TT, United Kingdom
[d] University of Wales Trinity Saint David, College St, Lampeter, SA48 7ED, United Kingdom
[e] University of Michigan, 500 S State St, Ann Arbor, MI, 48109, United States of America

ARTICLE INFO

Keywords:
Artificial Intelligence
Learning Health Systems
Implementation science
Health data science

ABSTRACT

Artificial Intelligence (AI) holds significant promise for healthcare but often struggles to transition from development to clinical integration. This paper argues that Learning Health Systems (LHS)—socio-technical ecosystems designed for continuous data-driven improvement—provide a potential "glide path" for safe, sustainable AI deployment. Just as modern aviation depends on instrument landing systems, the safe and effective integration of AI into healthcare requires the socio-technical infrastructure of LHSs, that enable iterative development and monitoring of AI tools, integrating clinical, technical, and ethical considerations through stakeholder collaboration. They address key challenges in AI implementation, including model generalizability, workflow integration, and transparency, by embedding co-creation, real-world evaluation, and continuous learning into care processes. Unlike static deployments, LHSs support the dynamic evolution of AI systems, incorporating feedback and recalibration to mitigate performance drift and bias. Moreover, they embed governance and regulatory functions—clarifying accountability, supporting data and model provenance, and upholding FAIR (Findable, Accessible, Interoperable, Reusable) principles. LHSs also promote "human-in-the-loop" safety through structured studies of human-AI interaction and shared decision-making. The paper outlines practical steps to align AI with LHS frameworks, including investment in data infrastructure, continuous model monitoring, and fostering a learning culture. Embedding AI in LHSs transforms implementation from a one-time event into a sustained, evidence-based learning process that aligns innovation with clinical realities, ultimately advancing patient care, health equity, and system resilience. The arguments build on insights from an international workshop hosted in 2025, offering a strategic vision for the future of AI in healthcare.

1. Introduction

A health system becomes a Learning Health System (LHS) when it acquires the ability to continuously and systematically learn from its activities, and then apply the knowledge gained to improve the health of the individuals and populations it serves. The concept, first articulated by the U.S. Institute of Medicine (now National Academy of Medicine), envisions *"a system in which science, informatics, incentives, and culture are aligned for continuous improvement and innovation, with best practices seamlessly embedded in the care process, patients and families as active participants, and new knowledge captured as an integral by-product of the care experience"* [1,2]. Within an LHS every clinical interaction is an opportunity to learn by capturing data, analyzing them for insights, implementing interventions, and employing the measured outcomes of

* Corresponding author.
E-mail addresses: vasa.curcin@kcl.ac.uk (V. Curcin), brendan.delaney@imperial.ac.uk (B. Delaney), a.alkhatib@imperial.ac.uk (A. Alkhatib), n.cockburn@bham.ac.uk (N. Cockburn), olivia.dann@kcl.ac.uk (O. Dann), o.kostopoulou@imperial.ac.uk (O. Kostopoulou), daniel.leightley@kcl.ac.uk (D. Leightley), matthew.maddocks@kcl.ac.uk (M. Maddocks), sanjay.modgil@kcl.ac.uk (S. Modgil), krishnarajah.nirantharakumar@kcl.ac.uk (K. Nirantharakumar), philip.scott@uwtsd.ac.uk (P. Scott), ingrid.wolfe@kcl.ac.uk (I. Wolfe), kelly.zhang@imperial.ac.uk (K. Zhang), cpfried@umich.edu (C. Friedman).

https://doi.org/10.1016/j.artmed.2025.103346
Received 26 July 2025; Received in revised form 27 December 2025; Accepted 28 December 2025
Available online 31 December 2025

these interventions as new data to form an ongoing cycle of improvement. LHSs rely on socio-technical infrastructure to provide people, policies, technologies, and processes as services to enable these improvement cycles to function continuously and with economies of scale [3].

Health and health care, as domains of human endeavour, have in recent years become a major focus for application of AI approaches in widely varying types and sizes [4]. There is broad consensus that the number of models developed and validated with retrospective data greatly exceeds the number that has found their way into routine deployment [5]. This is for good reason. Just as it is much easier to get an airplane off the ground than to land it safely in all weather conditions, it is much easier to develop and validate an AI model with retrospective data than it is to deploy the model in a real-world healthcare setting. The current state of AI can be seen, by analogy, as a sky overcrowded with models that were relatively easy to get off the ground, now seeking social and technical infrastructure that will enable them to safely "land" in the environment of health care, creating opportunities to improve human health while also providing evidence on effectiveness, economic viability and safety.

It is important to note that this pattern is not inherently problematic: in many scientific and methodological domains, only a subset of innovations is intended for, or suitable for, real-world deployment. Early-stage methodological research necessarily produces far more ideas than ultimately reach clinical use. The challenge we highlight is therefore not the volume of unused models per se, but the absence of robust socio-technical infrastructure to support those models that are intended for translation into practice.

This paper posits that LHSs – inherently data-driven, iterative, and collaborative – provide the translational infrastructure needed for safely and most effectively deploying AI interventions in health, care, and wellbeing. It follows that AI interventions should consider LHSs as a potential glide path to a successful landing, and moreover, that institutions seeking to realize the full promise of AI should take steps to adopt LHS principles and methods, and build LHS infrastructure.

To support our proposition, we demonstrate how LHS principles address the technical lifecycle of AI, from development to deployment and monitoring, while also addressing ethical, organisational, and governance needs. The sections below describe the LHS concept and infrastructure, survey current AI applications in health, and examine how LHSs can manage the AI lifecycle and governance. In this way we describe the glide path by which AI can deliver significant and lasting benefits for human health.

While we describe an idealised LHS in which these capabilities operate smoothly, real-world LHS maturity is highly variable. Many institutions implement only subsets of these features. Our aim is to illustrate the potential value of LHS-aligned processes for AI, even when implemented incrementally.

It is also important also to note, as will be discussed in more detail, that this paper presents a singular view of LHS that emphasises infrastructure and, as such, is most concordant with the integration of LHS and AI. There are, at this writing, several co-existing views of LHS [6], none of which have achieved the level of maturity and standardization of aviation's instrument landing system, and there are multiple barriers to LHS development that must be acknowledged [7]. It is possible, therefore, that recognized potential value of LHS-AI integration may drive development of LHS infrastructure as much as it drives successful AI deployment. Furthermore

2. Learning Health Systems

LHSs are healthcare ecosystems built to continuously learn and improve. At their core is a cyclical process that links clinical practice ("performance") with the discovery and implementation of new knowledge (Fig. 1). In each cycle, data from routine practice are systematically collected, analyzed to generate insights, and then translated into changes in care, which in turn produce new data – enabling ongoing improvement [8].

Crucially, LHS are not just about analytics or technology – they are socio-technical systems. This means they encompass people (patients, families, clinicians, researchers, administrators), processes (policies, governance, workflows), cultures, and technology (electronic health records, data warehouses, AI tools) working in concert. One key feature of LHS is the formation of a multi-stakeholder learning community that supports the process and gets readjusted at the start of each cycle. This community unites all relevant parties – for example, doctors, nurses, data scientists, implementation scientists, patients and their families, management and senior leadership teams, and health IT staff – who work collaboratively to identify problems and co-create solutions. Their shared goal is to improve a specific aspect of health (a "health problem of interest"), and they remain engaged through the entire cycle. This continuity of collaborative execution distinguishes LHS from traditional quality improvement efforts, where different teams of people might separately handle data analysis and implementation.

Another defining aspect of LHS is embracing uncertainty and learning from failure, e.g. when an intervention underperforms. Rather than assuming what better interventions are upfront, an LHS acknowledges knowledge gaps and undertakes rigorous discovery (e.g. analyzing data, for example via AI, to find what might work) before implementing any changes. This scientific mindset is built into clinical operations. Over time, multiple learning cycles can operate concurrently, supported by a socio-technical infrastructure that scales learning throughout an organization. Examples of such infrastructure include integrated data systems, interoperability standards, and policies for data governance

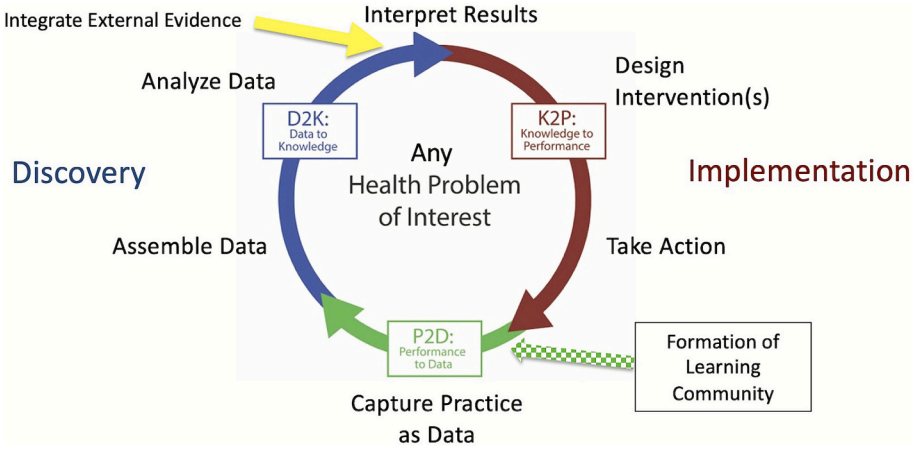

Fig. 1.

and ethics. Indeed, the vision is that a mature LHS will have many learning loops running (for different conditions or processes), all enabled by shared data and technology services.

3. AI applications in health

Artificial Intelligence (AI) offers unprecedented and multiple opportunities to improve health and care [9], from enhancing diagnostics to personalizing treatments [10]. Yet, bridging the gap between promising AI models and real-world impact has proven challenging [11,12]. Many AI systems that excel in retrospective studies or controlled settings fail to translate into routine practice, due to issues like data drift, workflow misalignment, additional data entry requirements, lack of clinician and patient trust, and the structural challenges of implementing change in health systems. For example, a fundamental problem is the "last mile" of implementation – integrating AI into complex, socio-technical health and care systems that have inherent variability, regulation and strict safety requirements [13]. This "last mile" is analogous to the final approach of an aircraft to a runway.

AI has rapidly become a focal point in digital health, with applications spanning nearly every domain of biomedicine. Machine learning algorithms, especially deep learning models, now approach or exceed human-level performance in certain tasks like medical image interpretation [14]. For instance, AI models can detect diabetic retinopathy in retinal photos or classify skin lesions from images with high accuracy [15]. In radiology and pathology, AI aids in finding subtle anomalies on X-rays, CT scans, or biopsies [16]. Beyond imaging, predictive models are used for prognostics – estimating risk of outcomes such as sepsis, hospital readmission, or disease complications – by mining patterns in electronic health record (EHR) data [17]. There are demonstrated use cases across drug discovery (e.g. AI systems identifying new drug candidates), virtual health assistants for patient triage or differential diagnosis, robotics in surgery, and personalized medicine approaches combining genomics with AI [18]. Natural language processing (NLP) algorithms can sift through clinical notes to flag patients who meet criteria for clinical trials or who need follow-up care [19]. Meanwhile, optimization and scheduling algorithms improve operational efficiencies like patient flow and staffing [20]. In mental health, conversational AI chatbots are starting to provide cognitive behavioural therapy exercises or triage advice [21]. Indeed, AI is seen as a key enabler for healthcare's "quintuple aim" (enhancing patient experience, improving population health, reducing costs, improving provider work life, and achieving health equity) [22]. In the short term, AI can automate high-volume repetitive tasks (like image screening). In the longer term, AI is anticipated to facilitate precision medicine – supporting training, education and tailoring care based on a patient's unique data profile.

Despite this promise, real-world adoption of AI in healthcare remains limited. While hundreds of AI systems have been published—including methodological models, simulation/training tools, and patient-facing clinical decision support systems—only a smaller subset are designed for, or appropriate for, routine clinical deployment [23]. These categories serve different purposes, and low deployment rates should not be interpreted as failure for models whose aims are purely methodological or educational – thus lack of translation can reflect disciplinary or organisational priorities rather than technical shortcomings. However, the reasons for low uptake of models intended for frontline deployment are manifold. Generalizability is a major concern: an algorithm trained in one setting often performs worse when deployed elsewhere due to differences in patient populations or data coding [24]. Integration challenges frequently arise – AI tools must seamlessly fit into clinical workflows and EHR systems, which is non-trivial (poor integration was a key factor in the limited success of earlier decision support systems) [25]. Transparency and trust issues also hinder uptake: clinicians may be reluctant to rely on "black box" algorithms whose reasoning they cannot interrogate, especially if an AI might make errors that harm patients. Some AI interventions may produce marginal benefits at a vastly increased resource cost [26]. Threats to autonomy, intrinsic motivation, professional pride and skill are also important concerns when AI takes over complex judgement tasks [27].

Ethical issues such as bias, AI reflecting or even amplifying racial or gender disparities present in training data [28], and privacy concerns with patient data further complicate deployment. Moreover, healthcare regulators were designed for medicine safety, which in the UK have an average usage lifespan of 37 years, but the average medical device lifecycle is 18 months [29]. Health system functions including regulation, financial incentive and reimbursement structures, as well as the clinicians and managers that need to use new technologies, need to adapt to AI-based services and the rapid lifecycles of AI.

In practice, many health AI projects stall after proof-of-concept due to this socio-technical gap. Researchers have dubbed this the "last mile problem" of AI in health – moving from model development to sustained clinical use. It is here that LHSs offer a way forward. By design, an LHS provides an ecosystem that addresses data quality, workflow integration, continuous evaluation, and stakeholder engagement – precisely the factors that determine whether an AI succeeds or fails in practice (Table 1).

A deeper issue that affects the translation of methodological work into practice, particularly in multidisciplinary teams, is the absence of shared incentives, language, and cross-disciplinary alignment between model development and clinical implementation. Thus the collaboration enabled by the LHS needs to be backed by the wider institutional recognition of success and reward that goes beyond traditional domain

Table 1
Main challenges associated with the "last mile" deployment of AI-based solutions and opportunities provided by LHSs.

Challenge		Opportunity
Generalizability and data drift	Models trained in one setting often underperform elsewhere due to variations in population, coding practices, and clinical workflows.	LHSs embed feedback loops that monitor performance over time, enabling regular model updates and adaptation to local contexts.
Transparency and trust ("black-box" concern)	Lack of transparency and accountability reduces clinician and patient confidence in AI outputs.	Provenance tracking, open documentation, and participatory validation build trust and accountability.
Regulatory and lifecycle misalignment	Traditional approval pathways were designed for static medical devices, not continuously learning systems with short update cycles.	LHSs support iterative approval and oversight processes, aligning with evolving concepts of "continuously learning" AI regulation.
Responsibility and accountability ambiguity	When AI influences decisions, liability among clinicians, developers, and healthcare organizations is often unclear.	System-level provenance and reproducibility solutions allow for detail of human and software agents that participated in individual tasks.
Data governance and provenance limitations	Inconsistent recording of data lineage and model versioning complicates reproducibility and safe model updates.	Shared socio-technical infrastructures in the LHS enables consistent provenance capture and reporting
Cultural and organisational inertia	Risk-averse environments and lack of a learning culture inhibit experimentation, feedback, and iterative improvement.	LHS can be used as a framework to manage risk in introducing new data-driven initiatives
Model monitoring and maintenance gaps	Few institutions have structures for continuous evaluation, post-deployment surveillance, and performance recalibration.	LHSs embed ongoing evaluation and improvement cycles ("data → knowledge → performance → data"), enabling AI to evolve safely over time.

metrics.

4. Connecting AI and Learning Health Systems to promote model deployment and continuous improvement

Deploying AI in healthcare is not a one-off event but a lifecycle that spans development, implementation, evaluation, maintenance, and evolution. LHSs are uniquely equipped to manage this lifecycle in a sustainable way. In the sections that follow, we will illuminate the specific connections between them.

4.1. Co-creation and user-centered design

LHS cycles start with a multistakeholder learning community focused on a health problem. This naturally promotes and enables co-creation of AI solutions with people within health and care systems, patients, and AI end-users, as well as establishing whether such solutions are feasible in the given context. A recent review of LHS models shows the scope for more detailed stakeholder modelling to ensure equity [30]. Instead of tech companies or data scientists building algorithms in isolation, the LHS framework invites clinicians, patients, and other users to be part of the design and validation process. Such co-design was identified as a key enabler in nearly half of studies on clinical AI implementation [31]. By capturing user requirements and domain knowledge early, co-creation ensures the AI addresses real clinical needs and fits the local context. For example, if an LHS community decides to develop an AI tool to predict patient deterioration, healthcare professionals would contribute to defining what "deterioration" means in practice, what warning signs are actionable, and how alerts should be presented, and patients' and carers' perspectives would also be taken into account. This example points to the more general challenge (one that is especially salient for LHS) of how to account for diverse users who may have different conceptualisations of the domain, and more broadly, differing interests and perspectives that may not align. LHS could thus also benefit from AI support for multistakeholder deliberation and sense-making [32].

This participatory approach to development of AI tools improves the relevance, usability, and trustworthiness of said tools. The collective sense-making that occurs in participatory approaches also creates a shared purpose that helps drive adoption, bringing the whole learning community along the journey and leading to better implementation as well as better design [33]. It also streamlines workflow integration – since the future beneficiaries helped design the AI, they are more likely to adopt it and less likely to be surprised by its behavior. In essence, the LHS turns AI development into a *cooperative learning process* between humans and machines, rather than a vendor delivering a static product.

It is important to note that the level of required co-design varies by the type of AI tool. Systems used for internal operations—such as bed-capacity forecasting or scheduling optimisation—primarily require engagement with technical and operational experts rather than patients or frontline clinicians. In such cases, issues arise less from insufficient patient co-design and more from inadequate alignment with the expertise of operations researchers or systems engineers.

4.2. Planning for model evaluation and monitoring

Before an AI model is deployed into practice, there must be a plan for how its real-world performance will be evaluated and monitored. In the context of LHS, this planning process is collaborative and forward-looking. Through the co-creation process, clinicians, data scientists, administrators, and patients jointly define what "success" means and how it should be measured once the model is in use. Considerations could include evaluation on different subgroups in a heterogenous population, plans for managing data quality issues, and clinician behavior. Critically, evaluation of AI tools in deployment is often much more challenging than evaluating on a static dataset.

One key challenge is the delayed availability of outcome labels for an AI model. For example, if a diagnostic AI tool suggests an incorrect diagnosis to a patient with cancer based on a scan, the error may not come to light for months – or ever – if the patient does not return for a follow-up. Another subtle but significant challenge lies in the potential feedback loops created when AI predictions influence treatment decisions. Consider a model that uses an ECG to assess whether a patient is at risk for a condition that can only be confirmed by an echocardiogram [34]. If only patients predicted as high-risk receive an echocardiogram, then over time, the resulting dataset used to evaluate the model will become increasingly biased. Low-risk predictions may rarely be confirmed or refuted, making it difficult to identify false negatives. In other words, when models shape what data is collected later on, it can mask errors and distort performance estimates.

The LHS paradigm is uniquely positioned to plan for and manage these complexities, even though it does not automatically eliminate them. Evaluation plans may include a combination of quantitative metrics (e.g., accuracy, calibration, clinical utility), qualitative feedback (e.g. user satisfaction, clinician trust), and operational indicator (e.g. AI tool usage patterns, override patterns) – crucially, planned across cycles, acknowledging potential delays or drop-outs in outcome availability for certain actions. An LHS ensures that the evaluation is not a one-time audit, but an ongoing, structured process that ensures a responsible use of AI in healthcare. Specifically, bias arising from downstream actions of model deployment affecting data used in re-training may be accounted for by data provenance and the ability of an LHS to monitor the entire cohort of patients over longer time periods. Furthermore, this multi-cycle capability also allows incorporation of methods to address selective labels and verification bias.

4.3. Managing model evolution

Once an AI approach is deployed, the work is not done – models must be updated over time. LHS are built for this continuous updating. Rather than freezing an algorithm after deployment, an LHS treats each use of the AI as an opportunity to improve it. For instance, the outcomes and errors of an AI's prediction(s) can be fed back as new training data (the "performance to data" part of the cycle) to recalibrate the model. This addresses problems such as model drift, where an AI's accuracy degrades as clinical practice or patient populations change. Apart from re-training, updates may include recalibration, threshold adjustment, feature review, changes in clinical workflow, or full redevelopment. In the LHS paradigm, model evolution can be governed by the learning community: they set criteria for when the model should be retrained, modified, or replaced, based on ongoing performance metrics or organisational criteria.

Crucially, LHS provides the regulatory, governance and infrastructure to do this safely. The iterative loop (data → knowledge → practice) means that after each model update (knowledge), an intervention or evaluation occurs (practice), and only if the updated AI shows improved or at least non-inferior performance does it become fully adopted. This controlled evolution is aligned with emerging regulatory concepts of adaptive AI systems. In fact, regulators like the United States FDA are exploring lifecycle-based oversight via mechanisms like Predetermined Change Control Plans,[1] which specify in advance the conditions under which a model may be updated. These approaches support structured, transparent, evidence-driven evolution, mirroring LHS formal learning cycles of planning, change, and evaluation that ensure the changes are evidence-based with documented provenance. If performance drops or unintended consequences emerge, the LHS can quickly revert or adjust the model in the next cycle. This alertness to emergent change stands in

[1] https://www.fda.gov/medical-devices/software-medical-device-samd/predetermined-change-control-plans-machine-learning-enabled-medical-devices-guiding-principles.

contrast to static deployments, where an AI might quietly become unsafe due to model drift before anyone notices, e.g. as a consequence of seasonal variability in data. In a mature LHS with multiple cycles, AI models become *living components* of care pathways, continuously learning from new data under the watch of the learning community and data scientists.

An LHS, by definition, embeds continuous evaluation into its routine operation – *"new knowledge is captured as an integral by-product of the care experience"*. Every time an AI-driven intervention is applied, the LHS measures outcomes and compares them against expectations. This creates a constant feedback loop to refine the AI and its integration into practice, either continuously or through a series of discrete steps. Traditional clinical trials or validations provide only a snapshot (critically before deployment if this is a regulatory approval study) while a LHS enables ongoing real-world evaluation (akin to post-market surveillance in regulatory terms). For example, suppose an AI algorithm for sepsis early warning is rolled out in a hospital. In an LHS approach, the system would track metrics such as true/false alert rates, sepsis mortality, clinician response times, and any adverse events. These data are analyzed (perhaps by another AI or by the quality improvement team) to assess the AI's clinical utility and safety continuously [35]. If, say, false alarms are too frequent and causing alert fatigue, the learning community might decide to tweak the sensitivity threshold or incorporate an additional data input – effectively refining the model or its usage protocol. This would then be tested in the next cycle and so on. Over time, the AI tool either improves or is retired if it cannot meet the desired outcomes, but importantly, this decision is driven by evidence gathered during routine practice. The LHS thus prevents the scenario of an AI being deployed and "forgotten." It institutionalizes an evaluate-and-improve mentality, similar to the DevOps/MLOps approach in software where systems are constantly monitored and iterated, indeed regulation of AI does build upon these software quality approaches. In healthcare, such agility is rarely present outside of an LHS context. Through continuous improving, AI remains *fit for purpose*, and the health system avoids stagnation with outdated algorithms.

As AI capabilities evolve toward more agentic behaviours—including autonomous task execution, workflow orchestration, or proactive recommendations—model evolution becomes more than a technical update problem. Agentic systems require mechanisms to monitor goal-directed behavior, prevent unintended action sequences, and ensure alignment with clinical governance. LHSs offer a natural setting for such oversight, since their structured cycles of evidence generation, evaluation, and stakeholder governance are well-suited to managing dynamic, semi-autonomous AI tools.

5. Connecting Learning Health Systems and AI for regulation and governance

In addition to managing the more technical challenges discussed above, LHSs provide a framework for addressing regulatory and ethical challenges of AI in healthcare, such as ensuring clarity of responsibility, transparency of algorithms, tensions between commercial and open approaches, implementation accountability, provenance of data and models, and the role of humans-in-the-loop. How LHSs approach these challenges is deeply rooted in consensus LHS core values propounded in 2012 [36] and, more recently in the LHS Core Commitments put forward by the National Academy of Medicine [37]. We discuss how these LHS principles help navigate these issues:

5.1. Transparency of responsibility

When AI systems assist in clinical decisions, it can become unclear who is responsible for the outcomes – the clinician user, the organization deploying the AI, or the developer of the algorithm. A LHS, by virtue of its collaborative structure, can make responsibility more transparent. The learning community overseeing an AI project within an LHS brings together all stakeholders to define roles and boundaries explicitly (e.g. who validates the model, who approves its use, and who responds to its recommendations). This shared governance means that responsibility is acknowledged at each stage: data collection (usually the health system's responsibility), model development (data scientists and developers), and clinical decision-making (clinicians guided by hospital policies). For instance, a hospital LHS might establish a committee (including clinicians, AI specialists, and ethicists) that must sign off on any AI-derived protocol change, thereby clearly assigning accountability. Such structures help avoid the "responsibility vacuum" that can occur with AI. Indeed, researchers implementing diagnostic LHS have flagged *"medico-legal responsibility for generated evidence"* as a significant challenge to be proactively addressed. By tackling this in the LHS governance (e.g. having legal/risk managers in the learning community discussions), the duties and liabilities of each party are delineated before deployment. Transparency is further enhanced by LHS documentation practices – every learning cycle produces artifacts (analysis reports, decision logs, implementation plans) that can be audited. This makes it clear *why* and *by whom* a certain AI-informed decision was made, a critical feature for accountability and regulatory compliance. Thus, embedding an LHS approach to the development and maintenance of hazard logs as required for medical devices (part of the NHS Data Security and Protection Toolkit in the UK) can assist safe deployment of AI.

Since LHS emphasises learning and sharing knowledge, there is a philosophical alignment with open-source principles. Transparency is valued because it enables collective learning – an opaque algorithm is antithetical to the spirit of an LHS. Moreover, studies on AI enablers have noted that *open-source software can improve transparency and accountability by allowing experts to identify vulnerabilities*. Hybrid approaches that blend open-source and commercial models, e.g. licensed extended versions with additional features, may help software companies to balance transparency with income generation to develop mature AI products.

5.2. Responsibility of implementation

Introducing an AI tool into clinical practice is an active intervention that requires oversight. Outside a LHS setting, this responsibility should fall to the designated Clinical Safety Officer, often leading to duplication and variability between organizations. In contrast, a LHS explicitly manages implementation as part of the learning cycle (the K2P phase, "Knowledge to Performance"). This means the learning community takes collective responsibility for *how* an AI is deployed – including training staff, integrating into workflows, setting guidelines for use, managing hazard logs and monitoring initial results [38]. Through shared implementation responsibility, the LHS helps cultivate trust and clinicians, patients and carers see that a reliable support system stands behind the AI, not just a vendor. It also ensures there is a defined responsible party to take action if the AI underperforms, causes harm or misfires. For example, an on-call data scientist to fix a bug or a clinician lead to issue a notice to stop using the tool if a safety issue arises. The LHS helps to clarify the responsibilities set out in a system's DCB 0129 and DCB 0160 in the NHS, and equivalent on other health systems

5.3. Provenance, trust, and FAIRness

Provenance – the record of how data and models have been processed – is an essential component of accountability and trust. By capturing provenance throughout the research and implementation workflow, we embed mechanisms to verify trust in the system, for example through standards such as W3C PROV [39]. This is particularly relevant for AI, where complex data pipelines and model training processes can otherwise be opaque. In an LHS, every step of model development and deployment can be logged: which data were used for training, how they were pre-processed, which version of the algorithm was applied, who reviewed the outputs, and how the model was integrated into the clinical system. Such provenance metadata embedded

into the LHS creates a traceable audit trail. If an error or bias is discovered later (for example, the AI is less accurate for a certain subgroup of patients), one can trace back to see if the training data lacked diversity or if a certain parameter tweak led to the issue.

Understanding the provenance of our models is also key to overcoming the issue of **model collapse**. This concept, sometimes also referred to as the **autophagia** of AI, denotes the progressive degradation in performance, diversity, and fidelity of AI models when successive generations are trained—directly or indirectly—on outputs generated by their predecessors rather than on clean, human-authored data. The term draws from the analogy of low-background steel in nuclear science: just as post-1945 steel is contaminated by fallout, Internet content post-2022 is increasingly "polluted" by AI-generated material [40].

LHS emphasises robust data and knowledge management practices. The FAIR principles (Findable, Accessible, Interoperable, Reusable) have been advocated to maximize the utility of health data [41]. In an LHS, routine clinical data (from EHRs, devices, etc.) are continuously captured as a by-product of care and then made available for analysis in a privacy-conscious manner. Achieving this requires data interoperability across different sources and institutions – a challenge that LHS initiatives tackle by adopting common data standards and shared repositories. By ensuring that data are FAIR, LHS make it easier to train and update AI models on comprehensive, real-world datasets. For example, a hospital network functioning as an LHS might implement standardized coding and open APIs that allow AI developers to reliably pull anonymized patient data for model development (with appropriate governance). Additionally, LHS data practices emphasize data quality and provenance, meaning each data point's origin and context are tracked. This is crucial for **AI-ready data**, a concept closely aligned with the FAIR principles, as models are highly sensitive to garbage-in/garbage-out; an LHS will thus include processes to clean and validate data continuously. By underpinning AI with a strong data foundation, LHS reduces the risk of model bias and drift. Indeed, large-scale learning networks (such as those in some national LHS efforts) treat data as a shared asset for learning, which accelerates AI development while maintaining rigor in how data are used [42].

In LHSs, knowledge management is as important as data management. AI models can be viewed as exemplars of knowledge that can be represented as FAIR Digital objects [43], bringing many of the same benefits to model management that accrue to data by achieving the FAIR principles. Standards such as W3C DCAT provide foundational vocabularies for these metadata descriptions with specialised extensions, including Health-DCAT-AP, developed by the European Health Data Space initiative, to allow datasets, registries, biobanks, AI models, and associated digital services to be discovered, linked, and reused **safely and lawfully** across national and institutional boundaries. The movement to Mobilize Computable Biomedical Knowledge–with chapters in North America and the U.K and new chapters forming in continental Europe and Australasia—champions ecosystems of models and algorithms conforming to the FAIR principles [44].

Provenance also supports reproducible research, meaning that other sites or researchers can understand exactly how an AI result was obtained and attempt to reproduce or validate it. In regulatory terms, this aligns with requirements like the FDA's 21 CFR Part 11, which mandate the auditability of software used in clinical decisions [45]. Another relevant example is the ISO/DTS 23494–1, a biotechnology information standard, providing consistent documentation of the life-cycle of related research objects from the acquisition of a specimen to analytical procedures and downstream data processing and analysis [46].

By embedding provenance capture into the LHS's data/AI pipeline, we can capture data that can then be used to develop methods and tooling (e.g. dashboards, audit trail viewers) to explain and justify AI-driven decisions in the health system when needed [47]. Such a mechanism also helps avoid "algorithm creep," where, over time no one remembers how or why the model does what it does; in an LHS, that institutional memory is preserved in the provenance logs. This level of transparency is a strong antidote to the black-box criticism of AI and is invaluable for governance, as it allows independent audits and continuous quality assurance of the AI process. While this may sometimes be seen as healthcare inertia slowing down rapid technology advancement, it is essential to ensuring these innovations are implemented in a sustainable manner.

5.4. Human-in-the-loop

A commonly touted principle for safe AI in healthcare is to keep a *"human in the loop,"* On the one hand, this principle can be interpreted as addressing the so called 'value alignment problem' - how to ensure that the AI support for decision making and planning is aligned with the evolving values, interests and preferences of human stakeholders. This issue is of particular relevance to LHS given: 1) the ever-increasing use of large language models for advice giving [48]; 2) the inherent diversity of stakeholders and hence the need to account for competing interests, and the fact that preferences may evolve and be shaped in response to the outcomes of interventions. Alignment can thus be supported by designing AI-stakeholder interactions so as to accommodate human inputs that relate to preferences, while simultaneously leveraging AI capabilities for information retrieval, analysis and arguments that guide shaping and elicitation of stakeholder preferences [49,50].

On the other hand, a more narrow interpretation of the *human-in-the-loop* principle mandates that clinicians retain final decision authority rather than allowing fully automated decisions [51]. Patients and the public also prefer a hybrid system rather than a doctor-only or AI-only approach [52]. While this is important, there is a risk that the human-in-the-loop paradigm becomes a fig leaf that obscures accountability. If an AI recommendation contributes to harm, and in the absence of shared governance such as the one promoted by LHS, the developer might blame the clinician for not overriding it, while the clinician might argue they trusted the system's regulatory-approved advice, thus sharing the blame [53]. LHS can not only establish the responsibility boundaries, but also treat human-AI interaction as part of the learning process. Instead of assuming the presence of a human automatically ensures safety, an LHS will rigorously study how humans and AI actually work together (the *"human–AI team"* dynamics). For example, the LHS might track when clinicians follow or contradict AI advice and the outcomes of each scenario [54]. This can reveal if the "human oversight" is effective or if, in practice, users either over-rely on the AI, leading to a form of automation bias, or ignore a useful tool. The learning community can then adjust training or system design accordingly – perhaps tightening the conditions under which the AI can act without human confirmation or conversely, simplifying the user interface so clinicians pay attention at the right moments. The LHS thus does not take human-in-the-loop for granted; it treats it as a factor to be studied and optimized. Moreover, by having a collective forum (the learning community) discuss incidents and near-misses, the LHS ensures that accountability is shared and lessons are learned, rather than individual clinicians being unfairly blamed for systemic issues.

LHSs offer robust support for AI governance through promoting transparency (through open data/model practices and provenance tracking), clarifying accountability (through defined roles and continuous oversight), and improving trust (through co-creation, open review, and demonstrated safety in practice). By aligning the deployment of AI with an organization's learning and quality processes, LHS ensures that ethical principles and regulatory requirements are not an afterthought but an integral part of the AI lifecycle. This synergy addresses the oft-cited concerns about AI – from unclear liability to opaque algorithms – within the operational workflow of healthcare. The result is a more responsible innovation, where AI can be introduced and scaled in a manner that is transparent to users, acceptable to regulators, and ultimately safer for patients.

These issues intersect with emerging discussions around Sovereign AI, the principle that nations or health systems should retain strategic

control over critical data assets, model development pathways, and the computational infrastructure that underpins them. As AI capabilities become increasingly central to clinical operations and population health planning, LHSs provide a natural governance environment to mitigate risks associated with dependence on opaque, externally controlled AI ecosystems.

6. Way forward

While many existing frameworks focus on evaluating or deploying individual AI models, the distinctive contribution of this paper is to articulate a systems-level, reusable socio-technical approach. Rather than addressing adoption on a model-by-model basis, we argue for institutionalising AI deployment within an LHS, enabling repeatable, scalable, and cumulative learning across multiple AI tools. Now we outline five key steps to make this a reality:

6.1. Integrate AI initiatives into LHS frameworks, recognizing that LHS, too, is a work in progress

Healthcare organizations should embed AI within a formal LHS framework rather than handling them as isolated IT implementations. This means establishing multidisciplinary learning communities for each major AI intervention, responsible for guiding the project from inception through continuous monitoring. Clinicians, data scientists, IT, patients, and leadership must all have a seat at the table. By treating each AI deployment as a learning cycle, organizations will naturally address design, validation, implementation, and evaluation in one cohesive process rather than silos. This framework should also take account of the relevant parts of risk classification and control actions for software as a medical device regulation (ISO14971 and IEC62304).

In this analysis, the authors have described a "frozen", specific, and implicitly mature version of an LHS. This was by intention to offer the clearest portrayal of the potential benefits of AI and LHS integration. While the current state of LHS concepts and methods reflect 18 years of continuous development as reflected in part by a growing literature [55], their deployment within institutions is incomplete and slowed by a wide range of challenges [56,57]. It is important therefore, for those pursuing the integration proposed here, to assess the elements of LHS infrastructure that exist in an environment to be sure they exist in a sufficiently mature form to meet the demands AI will place upon them.

6.2. Invest in standards-based data infrastructure and FAIR data and knowledge practices

A critical enabler for both LHS and AI is a strong data backbone. Health systems (and their partners in government and industry) should invest in interoperable EHR systems, data warehouses, and registries that adhere to FAIR principles. The aim is to have common data models and exchange standards so that data from different sources can be pooled for machine learning and outcomes analysis. This poses numerous challenges. Although standards such as SNOMED CT [https://www.snomed.org/] and HL7 FHIR [https://www.hl7.org/fhir/] are now quite mature, their implementation remains inconsistent and has not yet achieved the necessary level of semantic interoperability. Common internal data models such as openEHR [https://openehr.org/] have very little adoption by major EHR vendors and data aggregation formats like OMOP [https://www.ohdsi.org/data-standardization/] risk loss of context from rich clinical data and need aligning with data catalog standards such as W3C DCAT and Health-DCAT-AP. Another important aspect of this is utilising bound identifiers for the type and version of AI models used, supporting monitoring and transparency, relevant terms exist in the SNOMED-CT (UK version) 'clinical observation' hierarchy. It also involves data governance that balances openness with privacy – for example, using federated learning or de-identified datasets within a secure data environment to allow AI training across institutions without exposing sensitive information. Provenance standards such as W3C PROV and ISO/TS 23494–1:2023 allow dataset histories and audit traces to operate in a distributed environment. National and regional networks can amplify this by linking LHS across sites, creating learning networks where AI models and insights are shared for mutual benefit.

6.3. Establish continuous model monitoring and maintenance

Just as hospitals have pharmacovigilance programs to monitor drug safety, they should create ***AI-vigilance*** programs for deployed algorithms. Within the LHS, dedicate a team (or extend the duties of the learning community) to routinely review AI performance metrics, bias indicators, and user feedback. This team would manage model updates in a controlled way – analogous to software updates, but with clinical validation at each step. Regulators and payers should support this by allowing mechanisms for rapid update approval and reimbursement models that recognize the ongoing effort of maintaining AI systems (rather than a one-time purchase). In essence, make continuous learning a contractual and regulatory expectation for any AI used in patient care. This will enforce that AI systems remain safe and effective as conditions change.

6.4. Cultivate an ethical, responsible and learning culture

Technology alone cannot create an LHS– the culture of the organization must value learning, transparency, and patient-centric innovation. Leadership should promote policies that encourage reporting of AI failures or near-misses without fear of blame (a just culture), echoing how morbidity and mortality conferences function for learning from clinical errors. Ethical principles like equity, accountability, and patient engagement should be baked into AI projects from the start. Concretely, this could involve establishing an ethics review board for algorithmic tools, including patient representatives to voice concerns and preferences. It also means training clinicians about the basics of AI, not just how to use a particular tool but how to critically appraise and question it [58]. Over time, a learning culture will normalize the idea that AI in healthcare is always under evaluation and subject to improvement – much like any drug or clinical practice might be.

6.5. Encourage open collaboration and knowledge sharing

The ethos of an LHS is inherently collaborative and cumulative. Stakeholders should therefore publish and share methodologies and outcomes of AI implementations (successes and failures alike) in peer-reviewed literature or public forums, contributing to the global learning community. Initiatives like open-source algorithms, public challenge datasets, or shared benchmarking of AI on common tasks can accelerate collective progress [3]. Funding agencies and journals could incentivize this by requiring that AI tools coming out of publicly funded research be made available for evaluation in other LHS settings. In order to deliver a "human-in-the-loop" approach, we have to ensure the humans do not lose their skills, knowledge, and intuition, and our training programs should be adapted to reflect that ambition in the presence, or with assistance, of AI technologies.

7. Summary

Realizing the vision of LHSs is our best strategy for a future where AI plays a very strong role in transforming health and care rather than losing their traction after pilot studies. LHS offers the mechanisms and infrastructure to continuously align AI tools with clinical reality, correct their course when needed, and prove their value (or lack thereof) with rigorous outcome data. By following LHS principles – co-creation, continuous learning, and shared governance – healthcare can create a glide path for AI that avoids the booms and busts of hype cycles and

instead achieves steady, responsible innovation, moving from model-by-model adoption to system-based, reusable socio-technical adoption.

The journey forward will require commitment from all quarters: healthcare providers must embrace data-driven experimentation; AI developers must engage with healthcare's complexities; regulators must adapt to iterative improvement paradigms; and patients must be partners in the process. The reward for this collective effort is profound: a healthcare system that learns as it delivers care, constantly getting better, smarter, and more just – with AI as an ally rather than a threat. None of this implies that creating and operating LHSs is easy [7], but there is increasing evidence of effective real-world LHS implementation with an AI focus [59].

Returning to our metaphor in closing, it is unimaginable how modern aviation could function safely and efficiently without instrument landing systems (ILS). Both ILSs and LHSs are socio-technical systems built on infrastructure consisting of training people, consensus policies, established processes, and proven digital technologies. To us, it is equally unimaginable how AI can support healthcare safely and efficiently without being embedded in LHSs.

CRediT authorship contribution statement

Vasa Curcin: Writing – review & editing, Writing – original draft, Conceptualization. **Brendan Delaney:** Writing – review & editing, Writing – original draft, Conceptualization. **Ahmad Alkhatib:** Writing – review & editing, Writing – original draft. **Neil Cockburn:** Writing – review & editing, Writing – original draft. **Olivia Dann:** Writing – review & editing, Writing – original draft. **Olga Kostopoulou:** Writing – review & editing, Writing – original draft. **Daniel Leightley:** Writing – review & editing, Writing – original draft. **Matthew Maddocks:** Writing – review & editing, Writing – original draft. **Sanjay Modgil:** Writing – review & editing, Writing – original draft. **Krishnarajah Nirantharakumar:** Writing – review & editing, Writing – original draft. **Philip Scott:** Writing – review & editing, Writing – original draft. **Ingrid Wolfe:** Writing – review & editing, Writing – original draft. **Kelly Zhang:** Writing – review & editing, Writing – original draft. **Charles Friedman:** Writing – review & editing, Writing – original draft, Conceptualization.

Funding

Curcin is supported by EPSRC-funded King's Health Partners Digital Health Hub (EP/X030628/1). Delaney is supported by the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (NIHR203323), and the NIHR North-West London Patient Safety Research Collaboration (NIHR NWL PSRC, Ref. NIHR204292). Dann is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) [Grant reference number EP/Y035216/1] Centre for Doctoral Training in Data-Driven Health (DRIVE-Health) at King's College London, with additional support from the Evelina Children's Trust. Wolfe is supported by NIHR Applied Research Collaboration South London (ARC South London) at King's College Hospital NHS Foundation Trust (NIHR200152). The funders had no role in this publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The paper has evolved from the discussions at the international "Learning Health Systems and AI" workshop convened by Prof Charles Friedman, Prof Vasa Curcin, and Prof Brendan Delaney, held at King's College London in March 2025. The authors wish to acknowledge all the participants in the workshop who did not elect to be co-authors of this paper.

References

[1] Institute of Medicine. Best care at lower cost: the path to continuously learning health care in America. Washington (DC): National Academies Press; 2012.
[2] Etheredge LM. A rapid learning health system. Health Aff (Millwood) 2007;26(2):w107–18.
[3] Friedman CP, Lomotan EA, Richardson JE, Ridgeway JL. Socio-technical infrastructure for a learning health system. Learn Health Syst 2024;8(1):e10405. https://doi.org/10.1002/lrh2.10405.
[4] Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ 2023;23:689. https://doi.org/10.1186/s12909-023-04698-z.
[5] He M, Li Z, Liu C, Shi D, Tan Z. Deployment of artificial intelligence in real-world practice: opportunity and challenge. Asia Pac J Ophthalmol (Phila) 2020;9(4):299–307. https://doi.org/10.1097/APO.0000000000000301.
[6] Friedman CP, Greene SM. Four distinct models of learning health systems: strength through diversity. Learn Health Syst 2025;9(2):e70009. https://doi.org/10.1002/lrh2.70009.
[7] Reid RJ, Wodchis WP, Kuluski K, Lee-Foon NK, Lavis JN, Rosella LC, et al. Actioning the Learning Health System: an applied framework for integrating research into health systems. SSM - Health Syst 2024;2:100010. https://doi.org/10.1016/j.ssmhs.2024.100010.
[8] Friedman CP. What is unique about learning health systems? Learn Health Syst 2022;6(3):e10328.
[9] Higginson I, Shand J, Robert G, Boaz A, French C, N'Djai A, et al. Reimagining health and care to tackle the rising tide of inequity, multimorbidity, and complex conditions. NEJM Catal 2024;5(5):5. https://doi.org/10.1056/CAT.24.0266.
[10] Genovese A, Borna S, Gomez-Cabello CA, Haider SA, Prabha S, Trabilsy M, et al. From promise to practice: harnessing AI'S power to transform medicine. J Clin Med 2025;14(4):1225. https://doi.org/10.3390/jcm14041225.
[11] Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med 2019;25(9):1337–40. https://doi.org/10.1038/s41591-019-0548-6.
[12] Sendak MP, D'Arcy J, Kashyap S, Gao M, Nichols M, Corey K, et al. A path for translation of machine learning products into healthcare delivery. Clin Transl Med 2020;10(2):e35.
[13] Habli I, Sujan M, Lawton T. Moving beyond the AI sales pitch – empowering clinicians to ask the right questions about clinical AI. Future Healthc J 2024;11(3):100179. https://doi.org/10.1016/j.fhj.2024.100179.
[14] Yoo H, Lee SH, Arru CD, Doda Khera R, Singh R, Siebert S, et al. AI-based improvement in lung cancer detection on chest radiographs: results of a multi-reader study in NLST dataset. Eur Radiol 2021;31(12):9664–74. https://doi.org/10.1007/s00330-021-08074-7.
[15] Pinto-Coelho L. How artificial intelligence is shaping medical imaging technology: a survey of innovations and applications. Bioengineering 2023;10(12):1435. https://doi.org/10.3390/bioengineering10121435.
[16] Van Leeuwen KG, Schalekamp S, Rutten MJCM, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. Eur Radiol 2021;31(6):3797–804. https://doi.org/10.1007/s00330-021-07892-z.
[17] Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019 Apr 4;380(14):1347–58. https://doi.org/10.1056/NEJMra1814259 [PMID: 30943338].
[18] Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J 2021;8(2):e188–94. https://doi.org/10.7861/fhj.2021-0095.
[19] Jackson R, Kartoglu I, Stringer C, Gorrell G, Roberts A, Song X, et al. CogStack – experiences of deploying integrated information retrieval and extraction services in a large National Health Service Foundation Trust hospital. BMC Med Inform Decis Mak 2018;18(1):47. https://doi.org/10.1186/s12911-018-0623-9.
[20] Li X, Tian D, Li W, Hu Y, Dong B, Wang H, et al. Using artificial intelligence to reduce queuing time and improve satisfaction in pediatric outpatient service: a randomized clinical trial. Front Pediatr 2022 Aug 10;10:929834. https://doi.org/10.3389/fped.2022.929834. PMID: 36034568; PMCID: PMC9399636.
[21] Ong QC, Ang CS, Chee DZY, Lawate A, Sundram F, Dalakoti M, et al. Advancing health coaching: a comparative study of large language model and health coaches. Artif Intell Med 2024;157:103004. https://doi.org/10.1016/j.artmed.2024.103004.
[22] Nundy S, Cooper LA, Mate KS. The quintuple aim for health care improvement: a new imperative to advance health equity. JAMA 2022;327(6):521–2. https://doi.org/10.1001/jama.2021.25181.
[23] Kraljevic Z, Bean D, Shek A, Bendayan R, Hemingway H, Yeung JA, et al. Foresight – a generative pretrained transformer for modelling of patient timelines using electronic health records: a retrospective modelling study. Lancet Digit Health 2024;6(4):e281–90. https://doi.org/10.1016/S2589-7500(24)00025-6.
[24] Shen JH, Raji ID, Chen IY. The data addition dilemma [preprint]. In: arXiv; 2024. https://doi.org/10.48550/arXiv.2408.04154.
[25] Corrigan D, Munnelly G, Kazienko P, Kajdanowicz T, Soler JK, Mahmoud S, et al. Requirements and validation of a prototype learning health system for clinical diagnosis. Learn Health Syst 2017;1(4):e10026. https://doi.org/10.1002/lrh2.10026.

[26] Halvorsen N, Hassan C, Correale L, Pilonis N, Helsingen LM, Spadaccini M, et al. Benefits, burden, and harms of computer-aided polyp detection with artificial intelligence in colorectal cancer screening: microsimulation modelling study. BMJ Med 2025;4:e001446. https://doi.org/10.1136/bmjmed-2024-001446.
[27] Burton JW, Stein MK, Jensen TB. A systematic review of algorithm aversion in augmented decision making. J Behav Decis Mak 2020;33(2):220–39. https://doi.org/10.1002/bdm.2155.
[28] Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. J Glob Health 2019;9(2):010318.
[29] Chapman AM, Taylor CA, Girling AJ. Are the UK systems of innovation and evaluation of medical devices compatible? The role of NICE's Medical Technologies Evaluation Programme (MTEP). Appl Health Econ Health Policy 2014;12(4):347–57. https://doi.org/10.1007/s40258-014-0104-y.
[30] Cooper J, Haroon S, Crowe F, Nirantharakumar K, Jackson T, Fitzsimmons L, et al. Perspectives of health care professionals on the use of AI to support clinical decision-making in the management of multiple long-term conditions: interview study. J Med Internet Res 2025;27:e71980. https://doi.org/10.2196/71980.
[31] Kamel Rahimi A, Pienaar O, Ghadimi M, Canfell OJ, Pole JD, Shrapnel S, et al. Implementing AI in hospitals to achieve a learning health system: systematic review of current enablers and barriers. J Med Internet Res 2024;26:e49655. https://doi.org/10.2196/49655.
[32] Kirschner PA, Buckingham Shum SJ, Carr CS, editors. Visualizing argumentation: software tools for collaborative and educational sense-making. New York: Springer; 2012.
[33] Knowles SE, Ercia A, Caskey F, Rees M, Farrington K, Van der Veer SN. Participatory co-design and normalisation process theory with staff and patients to implement digital ways of working into routine care: the example of electronic patient-reported outcomes in UK renal services. BMC Health Serv Res 2021;21:706. https://doi.org/10.1186/s12913-021-06702-y.
[34] Elias P, Poterucha TJ, Rajaram V, Moller LM, Rodriguez V, Bhave S, et al. Deep learning electrocardiographic analysis for detection of left-sided valvular heart disease. J Am Coll Cardiol 2022;80(6):613–26. https://doi.org/10.1016/j.jacc.2022.05.029.
[35] Feng J, Phillips RV, Malenica I, Bishara A, Hubbard AE, Celi LA, et al. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare. NPJ Digit Med 2022;5(1):66. https://doi.org/10.1038/s41746-022-00611-y.
[36] Brooks D, Douglas M, Aggarwal N, Prabhakaran S, Holden K, Mack D. Developing a framework for integrating health equity into the learning health system. Learn Health Syst 2017;1(4):e10029. https://doi.org/10.1002/lrh2.10029.
[37] McGinnis JM, Fineberg HV, Dzau VJ. Shared commitments for health and health care: a trust framework from the learning health system. NAM Perspect 2024. https://doi.org/10.31478/202412c.
[38] Lim HC, Austin JA, van der Vegt AH, Rahimi AK, Canfell OJ, Mifsud J, et al. Toward a learning health care system: a systematic review and evidence-based conceptual framework for implementation of clinical analytics in a digital hospital. Appl Clin Inform 2022;13(2):339–54. https://doi.org/10.1055/s-0042-1743243.
[39] Curcin V. Embedding data provenance into the learning health system to facilitate reproducible research. Learn Health Syst 2017;1(2):e10019. https://doi.org/10.1002/lrh2.10019.
[40] Shumailov I, Shumaylov Z, Zhao Y, Papernot N, Anderson R, Gal Y. AI models collapse when trained on recursively generated data. Nature 2024;631(8022):755–9. https://doi.org/10.1038/s41586-024-07566-y.
[41] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 2016;3:160018. https://doi.org/10.1038/sdata.2016.18.
[42] Smith SE. Data sharing in health needs to be FAIR-ER than in science [submission]. Office of the National Data Commissioner (Australia); 2022.
[43] Flynn AJ, Conte M, Boisvert P, Richesson R, Landis-Lewis Z, Friedman CP. Linked metadata for FAIR digital objects carrying computable knowledge. Res Ideas Outcomes 2022;8:e94438. https://doi.org/10.3897/rio.8.e94438.
[44] Khan N, Rubin J, Williams M. Summary of fifth annual public MCBK meeting: mobilizing computable biomedical knowledge (CBK) around the world. Learn Health Syst 2023;7:e10357. https://doi.org/10.1002/lrh2.10357.
[45] U.S. Food and Drug Administration. Part 11, electronic records; electronic signatures — scope and application. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/part-11-electronic-records-electronic-signatures-scope-and-application;. [Accessed 24 July 2025].
[46] Wittner R, Holub P, Mascia C, Frexia F, Müller H, Plass M, et al. Toward a common standard for data and specimen provenance in life sciences. Learn Health Syst 2023 Apr 18;8(1):e10365. https://doi.org/10.1002/lrh2.10365. PMID: 38249839; PMCID: PMC10797572.
[47] Kale A, Nguyen T, Harris FC, Li C, Zhang J, Ma X. Provenance documentation to enable explainable and trustworthy AI: a literature review. Data Intell 2023;5(1). https://doi.org/10.1162/dint_a_00119.
[48] Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman M, et al. A comprehensive overview of large language models [preprint]. In: arXiv; 2023. https://doi.org/10.48550/arXiv.2307.06435.
[49] Russell S. Human compatible: AI and the problem of control. London: Allen Lane; 2019.
[50] Bezou-Vrakatseli E, Cocarascu O, Modgil S. Towards dialogues for joint human-AI reasoning and value alignment [preprint]. In: arXiv; 2024. https://doi.org/10.48550/arXiv.2405.18073.
[51] Mosqueira-Rey E, Hernández-Pereira E, Alonso-Ríos D, Vázquez-García A. Human-in-the-loop machine learning: a state of the art. Artif Intell Rev 2023;56:3005–54. https://doi.org/10.1007/s10462-022-10246-w.
[52] Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health 2021;3(9):e599–611.
[53] Pezzo MV, Pezzo SP. Physician evaluation after medical errors: does having a computer decision aid help or hurt in hindsight? Med Decis Making 2006;26(1):48–56.
[54] Nurek M, Kostopoulou O. How the UK public views the use of diagnostic decision aids by physicians: a vignette-based experiment. J Am Med Inform Assoc 2023;30(5):888–98. https://doi.org/10.1093/jamia/ocad019.
[55] Friedman CP, Greene SM, Rubin JC. Ten reasons why learning health systems will have a transformational effect on health and health care. Learn Health Syst 2025;9(4).
[56] Shaw L, Perrier M, Tahmasebian K, et al. Developing a codebook to characterize barriers, enablers, and strategies for implementing learning health systems from a multilevel perspective. Learn Health Syst 2025;9(2):e10452. https://doi.org/10.1002/lrh2.10452.
[57] McLachlan S, Dube K, Johnson O, et al. A framework for analysing learning health systems: are we removing the most impactful barriers? Learn Health Syst 2019;3:e10189. https://doi.org/10.1002/lrh2.10189.
[58] Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: a mapping review. Soc Sci Med 2020;260:113172. https://doi.org/10.1016/j.socscimed.2020.113172.
[59] Steel PAD, Wardi G, Harrington RA, et al. Learning health system strategies in the AI era. npj Health Syst 2025;2:21. https://doi.org/10.1038/s44401-025-00029-0.