

AI in European Healthcare:

Policy Recommendations for Optimising Added Value
and Safe, Ethical and Inclusive Adoption



EASAC & FEAM

AI in European Healthcare

September 2026

ISBN 978-3-7001-5315-3

This report can be found at
www.easac.eu & www.feam.eu

EASAC – the European Academies Science Advisory Council – is formed by the national science academies of EU Member States, Norway, Switzerland, the United Kingdom, and the Academia Europaea (a pan-European academy). EASAC's 29 member institutions collaborate with each other in giving advice to European policy-makers. In its entirety, EASAC provides a strong means for the collective voice of European science to be heard.

EASAC's mission reflects the view that science is central to many aspects of modern life and that an appreciation of the scientific dimension is a prerequisite to wise policy-making. This view underpins the work of many science academies at the national level. Given the importance of the European Union as an arena for policy-making, academies have recognised that the scope of their advisory functions needs to extend beyond the national domain and cover the European level. Therefore, European academies formed EASAC in 2001 so that they can speak in a strong voice at EU level.

Through EASAC, the academies provide collective, independent, strictly evidence-based advice about scientific aspects of policy issues to those who make or influence policy and legislation in the EU institutions and

in EU Member States. EASAC aims to deliver advice that is comprehensible, relevant, and timely.

EASAC's main mission is to inform both national and European Union policy-makers on evidence to take policy decisions – on questions of relevance regarding environment, energy, biosciences, and public health. EASAC's work processes are open and transparent, and its results are independent of any commercial or political bias.

The EASAC Council has 29 individual members—one highly experienced scientist nominated by each member academy. The Council agrees the initiation of projects, appoints members of working groups, provides peer review for drafts, and endorses reports for publication.

EASAC is mostly funded by the member academies and has no commercial or business sponsors. EASAC's working group experts devote their time free of charge. EASAC is supported by a secretariat hosted by the Austrian Academy of Sciences in Vienna. To find out more about EASAC, visit the website – www.easac.eu – or contact the EASAC Secretariat at secretariat@easac.eu.

FEAM – the Federation of European Academies of Medicine – brings together 25 national Academies of Medicine, Pharmacy and Veterinary Science, or the medical divisions of national Academies of Sciences, from across the WHO European region. Together, FEAM's member Academies represent over 5,000 leading scientists from across the whole biomedical spectrum. FEAM also collaborates with partner Academies from beyond the WHO European region, extending the reach of its network and enriching its scientific advice with a broader international perspective. In its entirety, FEAM provides a strong means for the collective voice of European biomedical science to be heard.

FEAM's mission is to promote cooperation between its member Academies; to provide them with a platform to formulate and express their common position on European matters concerning human and animal medicine, biomedical research, education, and health; and to extend to the European authorities the advisory role that these Academies exercise in their own countries on those matters.

Guided by this mission, FEAM works towards a broader vision: a Europe in which biomedical policy is underpinned by the best scientific advice the continent has to offer, and in which the health, safety and wealth of European citizens are advanced through research, nurtured by a creative and sustainable environment for medical research and training.

Drawing on its network of member and partner Academies, FEAM accesses the best of Europe's

biomedical scientific expertise, complemented by international collaboration, in carrying out its work. FEAM covers the full range of biomedical and health-related disciplines, with recent activity spanning artificial intelligence in healthcare, One Health, vaccination policy, xenotransplantation, medical waste, and the medicines ecosystem, among others. Its activities include producing evidence-based reports, scoping papers and policy statements; convening expert working groups and selection committees; organising the FEAM Forum, FEAM Annual Lecture and other events to brief policy-makers and stakeholders; and issuing statements on topical biomedical and health issues.

FEAM is a partner in EU-funded projects including SAPEA, which supports the European Commission's Scientific Advice Mechanism, and has previously contributed to projects such as PERISCOPE. FEAM's work processes are open and transparent, and its output is independent of any commercial or political bias.

The Federation is governed by a Board and a Council, with representatives nominated by each member Academy, and is guided in specific areas of work by standing Committees, including on One Health, Science and Education, and Advisory Ethics. FEAM is funded principally through the contributions of its member Academies and is supported by a Secretariat based at the Palais des Académies in Brussels. To find out more about FEAM, visit the website – www.feam.eu – or contact the Secretariat at info@feam.eu.

Suggested citation:

EASAC–FEAM (2026) Artificial Intelligence in European Healthcare. Cole-Hunter T, Abboud L, Knottnerus A, Martí-Bonmatí L, Barua I, Boman M, Fialová E, Mihalaş GI, Komadina R, Korja M, Lipschuetz M, Lovis C, Moingeon P, Noble A, Nunes R, Pencheva T, Raidou R, Ríos Insua D, Rückert D, Salmon I, Samaržija M, Steyerberg E, Tadić P, Vollmer S. EASAC–FEAM report, ISBN 978-3-7001-5315-3, www.easac.eu and www.feam.eu

The text of this work is licensed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>) which permits use, sharing, adaptation, distribution, and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and source, provide a link to the Creative Commons licence, and indicate any modifications.
Images are not covered by this licence.

This report and other project outputs can be viewed at: <https://easac.eu/publications> and <https://www.feam.eu/feam-publications/>
Some rights reserved.

© EASAC Secretariat c/o Austrian Academy of Sciences 2026

EASAC Secretariat
c/o Austrian Academy of Sciences
Dr. Ignaz Seipel-Platz 2
1010 Vienna
Austria

Telephone: +43 1-51581-1208
E-mail: secretariat@easac.eu
Web: www.easac.eu

FEAM Secretariat
Palais des Académies
Rue Ducale 1
1000 Brussels
Belgium

Telephone: +32 493-88-98-59
Email: info@feam.eu
Web: www.feam.eu



This report focuses on the following United Nations Sustainable Development Goals: goal 3 (good health and well-being), goal 4 (quality education), goal 9 (industry, innovation, and infrastructure), goal 10 (reduced inequalities), goal 12 (responsible consumption and production), goal 13 (climate action), goal 16 (peace, justice and strong institutions), and goal 17 (partnerships for the goals).

Cover image and graphics: Communication Works – Froning, Reise GmbH / Felix Sorau
Text copy-edited and typeset in Frutiger by The Clyvedon Press Ltd, Cardiff, United Kingdom.

Contents

Cross-references in blue type to sections, tables, figures, boxes, or references contain hyperlinks.

	<i>page</i>
Foreword	6
Executive summary	7
Urgent key messages	8
EU policy recommendations	8
Infographic	10
1 Introduction and background	11
2 Use cases, opportunities, and developments	15
3 Ensure reliable, safe, equitable, and trustworthy artificial intelligence in healthcare, through rigorous analytical and clinical validation	18
4 Data quality is a cornerstone of safe, effective, and trustworthy healthcare	26
5 Responsible artificial intelligence use and innovation require robust, multi-layered security strategies	30
6 Embed ethics into artificial intelligence design and development from an early stage	34
7 Artificial intelligence solutions in healthcare require engagement of key stakeholders	39
8 EU regulation and legislative framework for artificial intelligence in healthcare	49
9 Educational requirements to integrate artificial intelligence in European healthcare	56
10 General conclusions and recommendations: proactive policies for responsible artificial intelligence integration into European healthcare	59
Contributors	62
Glossary	63
Abbreviations	67
Appendix	68
Table A1 A structured comparative analysis of the major international frameworks about AI in healthcare	68
Table A2 A comparative snapshot by region/global frameworks of how ethical issues in AI applications in healthcare align with UNESCO's November 2021 Recommendation on the Ethics of Artificial Intelligence	70
Table A3 Continent-focused overview of policy priorities tailored to UNESCO's November 2021 Recommendation on the ethics of AI	71
Table A4 A concise comparison of how the US and China differ in ethical issues around AI in healthcare	72
Table A5 Example checklist aligning reporting, regulatory and ethical considerations for AI in healthcare	73
References	75

Foreword

Artificial intelligence (AI) is rapidly transforming healthcare, offering unprecedented opportunities to improve diagnosis, treatment, and the efficiency of health systems across Europe. At the same time, its integration into clinical practice raises complex questions around safety, ethics, governance, and equity that demand careful and coordinated policy responses.

This joint report by the European Academies Science Advisory Council (EASAC) and the Federation of European Academies of Medicine (FEAM) comes at a decisive moment for European health policy. As the European Union moves from developing regulatory frameworks – most notably the AI Act and the European Health Data Space – towards implementation, the focus must now shift to ensuring that AI delivers real, measurable benefits for patients and society, while maintaining public trust and safeguarding fundamental rights across Member States. A key challenge now lies in translating these frameworks into practice across Member States, ensuring consistent application and meaningful impact.

The evidence presented in this report makes clear that AI in healthcare cannot be approached as a purely technological innovation. Its success depends on robust clinical validation, high-quality and representative data, strong governance and oversight, and the meaningful engagement of healthcare professionals, patients, and citizens. Without these conditions, there is a risk that AI will exacerbate existing inequalities, introduce new forms of risk, or fail to deliver on its promise.

At the same time, Europe has a unique opportunity to lead globally in the development and deployment of trustworthy, human-centred AI in healthcare, and to position itself as a standard-setter in its governance. By aligning scientific excellence with strong ethical principles and effective governance, the European Union can ensure that innovation serves the public good and strengthens resilient, equitable health systems.

This report provides a set of actionable recommendations to support policy-makers, regulators, healthcare leaders, and other stakeholders in navigating this transition. It emphasises the need for a coordinated European approach that combines rigorous evaluation, interoperable data ecosystems, sustained investment in skills and capacity, and clear accountability structures.

We hope that this report will contribute to informed decision-making and constructive dialogue across Europe, helping to ensure that AI in healthcare is developed and deployed in ways that add value and are safe, ethical, and inclusive.

We thank all the members of the Working Group nominated by FEAM and EASAC member academies who generously contributed their time and expertise to the success of this project and the preparation of this comprehensive report.

Ferry Breedveld, FEAM President
Lise Øvreås, EASAC President

Executive summary

Artificial intelligence (AI)-assisted tools and systems are increasingly being applied and offer substantial potential for innovation across many scientific and societal fields. At the same time, appropriate policy guidance and regulation are crucial to ensure that the benefits of innovation are realised while effectively addressing the risks associated with implementing AI applications.

Given this overall challenge of leveraging AI innovations for the benefit of society while simultaneously managing potential risks, AI developments in healthcare and public health bring specific opportunities, questions, and requirements. Therefore, with this report, the European Academies Science Advisory Council (EASAC) and the Federation of European Academies of Medicine (FEAM) aim to present a high-level, accessible overview of the opportunities and risks of AI systems in healthcare and focused recommendations. The intention is to assist European policy-makers – and other policy stakeholders – in promoting responsible innovation by formulating and implementing evidence-based policies, while ensuring appropriate methodological, ethical, security, and regulatory safeguards.

The report therefore frames AI in healthcare as both an opportunity and a governance challenge. Its focus is on professional health and public-health uses of AI in Europe, including clinical care, population health, health-system management, and selected policy-relevant parts of medicines development, rather than on every possible consumer or wellness application of AI.

Understanding AI systems

For the purposes of this report, and given that the intended audience includes European Union (EU) policy-makers, we use the definition of an ‘AI system’ set out in the EU AI Act, [Regulation \(EU\) 2024/1689](#), Article 3(1): ‘A machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.’

This definition is part of the AI Act, which entered into force on 1 August 2024, with its provisions applying in phases from 2025. It is intended to help determine whether a software system falls under the scope of the AI Act or not. The European Commission (EC) gives non-binding guidelines clarifying how the definition should be applied in practice. The guidelines intend to support developers, providers, and users, including healthcare workers, in understanding whether their

systems qualify, and are subject to its rules, as AI systems under the EU AI Act. This definition applies to healthcare; however, there are different regulatory implications from it. Certain healthcare AI systems, including regulated medical-device systems subject to third-party conformity assessment, are classified as high-risk under the AI Act and are consequently subject to stringent safety, accuracy, governance, and oversight requirements ([EC 2025](#)). We view this risk-driven definition as essential for balancing regulation (risk mitigation) and innovation (benefit), ensuring that the regulatory approach fosters innovation.

For healthcare, the AI Act should be read together with the wider EU legal and regulatory ecosystem, including the Medical Device Regulation, the In Vitro Diagnostic Medical Device Regulation, the General Data Protection Regulation, and the European Health Data Space. The applicable obligations depend on intended use, clinical risk, data use, and whether the system functions as, or forms part of, a regulated medical device.

Drivers of change in AI system use

AI in healthcare is predicted to increase in application owing to converging drivers such as rising healthcare demand, workforce shortages, digitalisation of health records, and advances in computational power and data analytics. These pressures are prompting a shift towards more data-driven, personalised, and efficient healthcare models. At the same time, policy initiatives across the EU, alongside increased public and private investment, are accelerating AI integration. Together, these trends are reshaping expectations around healthcare delivery, prompting urgent attention to readiness, governance, and equity in implementation.

Patients and risk with AI systems

AI systems have both positive and negative effects on patients, offering potential improvements in diagnosis, treatment precision, and access to care, while also introducing new risks related to data privacy, algorithmic bias, and reduced human oversight. Unequal data representation can lead to disparities in care quality, and opaque decision-making processes may undermine patient trust. To safeguard patient welfare, it is essential that AI systems are transparent, and subject to rigorous clinical validation, with strong mechanisms for patient consent, safety, and redress. Health is a field which would benefit from small-data AI-driven systems (AI models designed to learn and operate effectively using relatively small, well-curated datasets rather than very large datasets), emphasising the need to support research and development in the field of small-data approaches.

A patient-centred approach should be operational rather than rhetorical. Patient organisations and patient representatives can contribute to priority setting, co-design, validation endpoints, procurement criteria, governance boards, communication materials, and post-deployment feedback. Evaluation should include patient-reported outcomes, patient experience, quality of life, accessibility, and the distribution of benefits and harms across patient groups.

Regional/sub-sectoral differences in the use of AI systems

The use of AI systems varies across Europe and healthcare sectors, reflecting disparities in digital infrastructure, investment capacity, regulatory readiness, and workforce expertise. While some regions and institutions lead in AI adoption, particularly in radiology, pathology, and administrative support, others lag relative to this because of resource constraints or limited technical capabilities. These differences risk deepening existing inequalities in healthcare access and quality. EU-level coordination and targeted support are needed to ensure balanced development and equitable AI integration across all Member States and health system segments.

EU-level action should therefore support structured mutual learning across Member States, including identification, comparison, and dissemination of best practices for responsible AI implementation. This is especially important where differences in infrastructure, workforce capacity, data readiness, and procurement expertise risk widening inequalities in access to beneficial AI-enabled care.

Current EU policies for AI systems

Current EU policies prioritise ethical, trustworthy, and human-centric AI systems, as reflected in initiatives such as the EU AI Act, the European Health Data Space, and Horizon Europe research funding. These frameworks aim to ensure safety, transparency, and equity while promoting innovation. However, gaps remain in harmonising implementation across Member States,

aligning health-specific regulations, and addressing sector-specific challenges. Continued policy refinement and cross-sector collaboration are essential to translate these principles into effective, context-sensitive actions within healthcare systems.

Policy shifts and integrated frameworks for AI systems

A proactive approach focusing on data protection and privacy, and energy and carbon footprint, is increasingly shaping AI-related policy developments (OECD 2025). To support sustainable and trustworthy AI in healthcare, integrated frameworks are needed that align digital health, environmental, ethical, and industrial strategies. Such coherence will help balance innovation with societal safeguards, ensuring that AI deployment contributes to broader EU goals, including climate neutrality, resilience, and social equity. Harmonised, forward-looking policies are key to managing complexity and fostering responsible AI adoption.

Interoperability and sustainability should be treated as implementation conditions, not technical afterthoughts. Procurement, funding, and evaluation frameworks should encourage use of recognised standards such as HL7 (Health Level Seven), FHIR (Fast Healthcare Interoperability Resources), and DICOM (Digital Imaging and Communications in Medicine) where relevant, assess cost-effectiveness and opportunity costs, and consider the energy and carbon footprint of compute-intensive AI systems.

Urgent key messages

This report aims to provide recommendations for embedding responsible AI practices into European healthcare, ensuring that AI deployment is managed proactively as part of a sustainable, anticipatory healthcare strategy, rather than merely responding to external factors such as emergencies or market inevitabilities. In the box below, we give the three urgent key messages of this report.

- 1. Responsible AI integration must be proactive, not reactive.** AI systems in healthcare should be guided by forward-looking policies that anticipate long-term impacts rather than reacting to crises or market pressures. Embedding responsibility requires early governance, inclusive design, and strategic investment to ensure AI strengthens – not destabilises – health systems.
- 2. Equity and trust are essential for sustainable AI adoption.** Without targeted action, AI systems risk widening health disparities and undermining public trust. EU policies must prioritise equitable access, transparency, and community engagement to ensure AI systems support universal healthcare goals and democratic values.
- 3. Coordinated EU action is crucial for safe, scalable innovation.** Fragmented approaches to AI-systems governance hinder safe implementation and slow progress. The EU must promote harmonised standards, interoperable cross-border data sharing, and collaborative oversight to enable robust, ethical, and scalable AI applications across Member States.

EU policy recommendations

This report provides policy options that are a mix of regulatory, institutional, financial, and capacity-building measures to help embed responsibility, trust, and sustainability into healthcare AI across the EU. The

report advocates for fit-for-purpose regulation such as regulatory sandboxes (controlled, time-limited environments in which innovators can test new tools under regulatory supervision, with safeguards for users and patients, while regulators learn how existing rules

apply and where guidance may need clarification) to support safe innovation and avoid over-regulation that stifles progress.

Below are 10 strategic policy recommendations that support responsible AI systems integration into

European healthcare, aligned with this report's goals and linked to the operational recommendations of each chapter. The strategic policy recommendations are elaborated on in [Chapter 10 \(General conclusions and recommendations; see page 59\)](#), with suggestions to make options actionable (also [Figure 1](#)).

Recommendation 1. Establish robust evaluation of AI's added value in healthcare. Mandate transparent evaluation of AI's net benefit – clinical outcomes, patient safety, security, equity, quality of life and cost-effectiveness – before and after deployment.

Recommendation 2. Develop comprehensive guidelines for responsible AI deployment. Develop EU-wide guidelines covering methodological quality, bias, external validity, cost-effectiveness, and environmental impacts.

Recommendation 3. Broaden ethical principles to encompass societal and long-term dimensions. Update ethical frameworks to address societal effects, institutional responsibilities, global impacts, and long-term (including intergenerational) consequences.

Recommendation 4. Address implementation-related challenges in AI deployment. Invest in interoperable infrastructure, harmonise rules across Member States, train staff, and support workflow integration with a focus on usability and patient safety.

Recommendation 5. Advance AI towards a learning health system. Support secure data platforms, feedback loops, and adaptive improvement with governance that ensures transparency, accountability, and trust.

Recommendation 6. Foster interoperability and data sharing ecosystems. Incentivise cross-border data ecosystems using common standards, strong privacy safeguards, and clear data stewardship.

Recommendation 7. Invest in training, education, and information for AI in healthcare. Fund interdisciplinary training and continuous professional development for clinicians, developers, and decision-makers to build competence and responsible use.

Recommendation 8. Engage stakeholders and protect public trust. Require participatory design, clear communication, and notification when AI systems are used, and accessible recourse so systems reflect societal values and maintain patient and public trust.

Recommendation 9. Define strategic research priorities for AI in healthcare. Prioritise research on transparency, bias mitigation, clinical validation, integration in practice, equity, and long-term system effects.

Recommendation 10. Strengthen governance and oversight structures. Establish clear accountability, independent monitoring, auditing, and corrective mechanisms to ensure compliance and safe performance over time.

Suggestions to make policy options actionable for AI in European healthcare

Illustrative actions that could help translate the ten policy recommendations into coordinated EU and Member State action.
Many actions will depend on context and require ongoing lifecycle governance.

	Recommendation	Action / implementation priority	Relevant to	Possible policy tools / mechanisms
 GOVERNANCE & ETHICS	R1	Support evaluation of added value before and after deployment; include clinical benefit, safety, security, equity, cost-effectiveness and patient impact.	European Commission; EMA/HTA bodies; national health authorities; payers; professional bodies; patient representatives.	EU evaluation framework; HTA templates; common metrics; real-world evidence registries; equity and safety indicators.
	R2	Promote EU-wide guidance for responsible deployment, covering bias, external validity, usability, environmental impact and accountability.	European Commission; Member States; CEN/CENELEC/ISO; regulators; professional societies; healthcare providers.	Implementation guidance; harmonised procurement clauses; deployment checklists; cost and environmental assessment criteria.
	R3	Consider broadening ethics frameworks to include societal, institutional, global and long-term impacts, not only individual rights.	European Commission; ethics bodies; research funders; academics; patient and civil-society organisations.	Societal impact assessment; embedded ethics plans; grant/procurement conditions; structured public deliberation.
	R4	Address implementation barriers: invest in infrastructure, workflow integration, local validation, procurement capacity and staff readiness.	Member States; regional health authorities; healthcare providers; procurers; clinicians; vendors and SMEs.	National implementation plans; infrastructure funding; sandbox pilots; procurement frameworks; change-management toolkits.
	R5	Strengthen learning health-system loops linking secure data, monitoring, feedback and continuous improvement.	European Commission/EHDS bodies; Member States; healthcare providers; health data access bodies; research infrastructures.	Secure processing environments; feedback loops; post-market data pipelines; governance charters; learning networks.
	R6	Enable interoperable cross-border data ecosystems using common standards, stewardship and privacy safeguards.	European Commission/EHDS bodies; national health data access bodies; EHR vendors; standards bodies; data custodians.	EHDS implementation; HL7 FHIR, DICOM and OMOP alignment; metadata/provenance rules; common data models.
	R7	Support role-specific AI literacy for clinical, technical, managerial, regulatory and policy workforces.	Professional regulators; universities; medical and public-health schools; CPD bodies; employers; ministries.	Competency frameworks; curriculum standards; CPD/licensing requirements; training toolkits; regulator/procurer training.
	R8	Promote co-design, public communication, transparency and accessible routes for challenge, explanation and recourse.	Healthcare providers; patient organisations; civil society; data protection authorities; health ministries; developers.	Co-design protocols; public AI registers; patient information; transparency standards; complaint and recourse routes.
	R9	Identify strategic research priorities for validation, bias mitigation, explainability, cybersecurity, workflow integration and long-term effects.	European Commission/Horizon Europe; national funders; research councils; academia; industry/SMEs; public-private partnerships.	Targeted calls; reference datasets; evaluation networks; translational pilots; challenge programmes; comparative implementation studies.
	R10	Strengthen clarity on governance, accountability, independent oversight, incident reporting, audit and corrective-action pathways.	European Commission; Member States; national competent authorities; notified bodies; provider boards; oversight bodies.	Oversight mandates; risk-management systems; incident reporting; audit trails; post-market surveillance; corrective-action protocols.

Note: AI = artificial intelligence; CEN = European Committee for Standardization; CENELEC = European Committee for Electrotechnical Standardization; CPD = continuing professional development; DICOM = Digital Imaging and Communications in Medicine; EHDS = European Health Data Space; EHR = electronic health record; EMA = European Medicines Agency; EU = European Union; HL7 = Health Level Seven International; FHIR = Fast Healthcare Interoperability Resources; HTA = health technology assessment; ISO = International Organization for Standardization; OMOP = Observational Medical Outcomes Partnership; SMEs = small and medium-sized enterprises.

Figure 1 AI in European healthcare: suggestions by EASAC-FEAM Working Group to make options actionable.

1 Introduction and background

The European Academies Science Advisory Council (EASAC) and the Federation of European Academies of Medicine (FEAM) produced this report on artificial intelligence (AI) in healthcare in close collaboration. EASAC and FEAM share common interests that align closely on this topic, reflecting their respective missions to provide science-based policy advice and promote health and well-being across Europe:

- 1. Scientific integrity and evidence-based policy:** both EASAC and FEAM are committed to providing the best possible advice so that policies can be grounded in robust, peer-reviewed scientific evidence. AI in healthcare presents significant opportunities and risks, requiring clear, impartial assessments. Jointly, they can ensure that policy recommendations reflect the latest scientific understanding and maintain public trust.
- 2. Equity and access to healthcare:** AI has the potential to improve healthcare delivery and access, particularly in underserved or resource-constrained settings. Both organisations share an interest in ensuring that such technologies do not exacerbate existing health inequalities but instead contribute to equitable healthcare across Europe.
- 3. Patient safety and ethical standards:** FEAM's medical expertise complements EASAC's broader scientific and policy perspective, allowing an approach to AI that prioritises patient safety, data privacy, informed consent, and ethical governance. These principles are critical in guiding the responsible development and deployment of AI tools in clinical practice.
- 4. Strengthening European health systems:** both EASAC and FEAM aim to provide science-based recommendations for resilient and sustainable European health systems. AI offers tools for efficiency and improved outcomes, but integration into health systems must be carefully managed. EASAC- and FEAM-nominated independent working-group members can jointly assess health system readiness and the infrastructure needed for AI to be beneficial and trustworthy.
- 5. Fostering innovation while managing risk:** there is a shared interest in supporting scientific and technological innovation while also ensuring appropriate regulatory oversight. This includes ensuring algorithmic transparency and accountability and ensuring that AI complements rather than replaces professional judgement in medicine.

Given that the primary intended audience of this EASAC–FEAM report is policy-makers, while also being useful to others that engage with the European Union (EU) AI Act (such as patient and consumer organisations, and educational institutions), the report tries to make high-level points with use case studies, without being technically over-detailed, on improving the following: (1) techniques (such as magnetic resonance image reconstruction or robotic assistance); (2) diagnosis (such as lesion detection, segmentation ([Turbé et al. 2023](#)) and characterisation); (3) decision support (such as predicting disease risk); and (4) clinical pathways (operational efficiency).

The report is intended primarily for EU and national policy-makers, regulators, health-system leaders, public-health authorities, and scientific academies. Its scope covers AI systems used in professional healthcare, public health, and health-related research, including hospital data, primary-care data, registry-linked data, cohort studies, and population-health data where these are relevant to clinical or public-health decision-making.

All in all, it is clear that AI is fundamentally redefining healthcare and the roles of health professionals, changing workflows, generating evidence, and modifying the functioning and potentially also the structure of healthcare systems:

- Diagnostics are shifting from human-first to AI-augmented deep learning pattern recognition, exceeding clinician performance by learning subtle, high-dimensional patterns.
- Clinical decision-making is moving towards continuous, data-driven personalisation, integrating electronic health records (EHRs), laboratories, genomics, images, and real-time vitals to generate risk predictions and updated treatment suggestions as new data arrives.
- Drug discovery and translational AI research are accelerating and becoming more computationally steered, shortening the path from target identification to candidate molecules by predicting protein structures, screening the huge chemical space, optimising lead compounds, and repurposing drugs.
- Healthcare delivery is being reorganised around automation and generative interfaces. Generative AI and natural language processing are transforming 'care logistics': summarising records, drafting notes, coding/billing support, patient messaging, and virtual assistants for monitoring or navigation.

- Regulation, ethics, and safety science are being forced to evolve in real time as AI systems are not ‘fixed devices’ and can drift with new data, amplify bias, or hallucinate in generative settings, fostering life-cycle monitoring, transparency, and human-oversight mandates.

Thus, the application and integration of AI can extend widely across public health and primary, secondary, and tertiary care settings, clinical support (e.g. pathology), shared care/decision-making, patient empowerment, and e-health, with the potential to better integrate across these healthcare boundaries, underpinned by the necessary transformation and connection of large-scale digital data platforms. With the ability to analyse large volumes of health data at speed to detect complex patterns, the promise of AI systems is for higher accuracy, improved health outcomes, and greater time and monetary efficiency. Alongside opportunities, there are associated challenges and risks that should be appropriately addressed, such as data quality, AI models’ external validity, algorithmic and governance issues, as well as (embedded and multi-scale) ethical issues, including data protection, security, equity, and non-discrimination.

FEAM and EASAC have previously worked at the intersection of AI, digital technologies, and health data. In a joint report with ALLEA in 2021, EASAC and FEAM formulated recommendations to improve health data sharing for research outside the EU/EEA ([EASAC–FEAM–ALLEA 2021](#)). FEAM has also formulated recommendations about health data use and the implementation of the EU General Data Protection Regulation (GDPR). A white paper published in 2023 summarised the main outcomes of a discussion about sustainable AI to drive global health looking at opportunities within and beyond the EU ([FEAM 2023](#)).

1.1 Policy context

The aim of this report is to draw lessons and good practices relevant to European policy-making. Therefore, the focus is on existing EU-level policies. AI systems present opportunities for patients, healthcare professionals, healthcare systems, industry, and startups. To ensure that AI systems benefit each stakeholder group, and that Europe is competitive in this area, the benefits of AI must be enabled and promoted, alongside regulation, guidelines, and education to prevent and address potential risks and costs. For instance, since end-users (patients, healthcare professionals) are unlikely to be able to examine the validity of the AI within the product/service offered, an adequate regulation and guidance framework governing the development and deployment of AI is crucial to ensure that it is optimally harnessed for health gain, that

patients are not harmed, and privacy and public trust are maintained.

The convergence of two sectors (AI advancing at a rapid pace, and the health sector) poses many questions, including how the AI systems are designed and implemented, and who holds responsibility throughout the lifetime of the product/service. Among the issues arising from the emerging AI technology are the following: (1) training, testing, validation and standardisation of algorithms; (2) data governance—ensuring datasets are representative, reliable, and interoperable across different health systems to prevent ‘data silos’; (3) privacy, personal data, and security; (4) ethical and regulatory issues; and (5) impacts and evidence.

Some policy gaps in these areas are best addressed at the supranational and national level (e.g. data privacy, certification of algorithms, models, and devices), while others could be addressed at the local level, including delivery of practical solutions (e.g. swarm and federated learning) and automation of workflows and patient pathways.

Recent international examples of repurposing healthcare-related personal data for non-health purposes illustrate issues that are relevant to EU AI-in-healthcare policy, as follows:

- ‘Chilling effect’ on healthcare access. Immigration enforcement policies led to immigrant communities avoiding healthcare services out of fear, resulting in substantial declines in routine care use ([Young et al. 2023](#)).
- Avoidance of public benefits because of fear. The 2019–2020 ‘public charge’ rule expansion triggered withdrawals from Medicaid and CHIP (Children’s Health Insurance Program) even among eligible immigrants, driven by concerns that personal health data would be used for immigration enforcement ([Bustamante et al. 2022](#)).
- Erosion of data representativeness threatening AI fairness. Trust loss leads to selective non-participation and incomplete datasets—compromising the validity, fairness, and generalisability of AI models trained on such biased data ([Young et al. 2022](#)).
- Conflict with GDPR’s purpose limitation and data minimisation. Deploying health data for enforcement purposes directly contradicts GDPR’s Article 5 requirements on purpose limitation and minimisation, potentially undermining legal compliance and citizen trust ([Haley et al. 2020](#)).

- Necessity of privacy-enhancing technologies in data governance.
To mitigate risks and reinforce trust, the literature strongly advocates technical safeguards – such as federated learning, anonymisation, differential privacy, and privacy-by-design – in AI systems using health data (Touw *et al.* 2021).

In short, AI-driven healthcare systems, EU policy, and regulation must mandate strict purpose limitation, enforce data minimisation, and require integration of privacy-enhancing technologies to safeguard public trust, data integrity, fairness, and GDPR compliance.

In terms of data governance and integrity, there should be an emphasis on protecting scientific integrity and preventing post-hoc data manipulation for political purposes. Guidelines or a code of conduct around data collection, reuse, and storage could be put forward. Data privacy and post-marketing surveillance of AI solutions must be addressed explicitly.

1.2 Aims and scope of the report

In view of the opportunities, potential risks, and contextual policy analysis described in the previous sections, the aims of this report can be further elaborated as follows:

1. Explore how the added value of AI-incorporated health products and services can be evaluated for adoption in healthcare; salient areas include randomised controlled clinical trials, other relevant prospective designs, and health-systems research. The report's description of use cases may also serve to illustrate their added value.
2. Determine where new guidance or regulation is desirable, while preserving innovation, and make recommendations on implementation of the EU AI Act and European Health Data Space or on revisions to other relevant legislation.
3. Formulate recommendations dealing with data issues and regulation of data protection (GDPR): issues include de-identification and use of secure processing environments, transparency, engendering public trust, lay communication, equity of health access, and model explainability, as well as the importance of consent governance (prospective and/or retrospective); the data federation and interoperability barriers across healthcare systems.
4. Formulate recommendations related to education and capacity-building needs: upskill the broader workforce in digital awareness and simultaneously build this into early education, and identify knowledge gaps and research priorities.

The scope of the report is generally to do the following:

- Highlight AI's transformative potential in improving diagnostics, drug discovery, development, and repurposing, patient care, public health, population health, and engagement. Impacts on practice and healthcare include consequences for patients and the doctor–patient relationship, with a need to consider ethical implications and the human factor in AI integration.
- Recognise generative AI's role in clinical decision-making and doctor-patient interaction, fostering trust and compliance.
- Address how proportionate, high-quality regulation can enable responsible innovation, reduce uncertainty, and support European competitiveness.
- Address the need for transparency about costs and a good balance between costs and the public return on investments made in AI technologies.
- Recognise the interdisciplinary nature of AI in healthcare, which includes insights from social sciences, behavioural sciences, and other relevant fields.

1.3 AI modalities and use cases

Reflecting on all of the above, understanding the AI landscape requires appreciating key distinctions. AI can be categorised by data complexity: monomodal AI relies on a single data type (e.g. image-based diagnostics) while multimodal AI integrates complex inputs (e.g. laboratory results, images, EHRs). Medicine is seen as these two pillars, being the local science of medicine and the global science of medicine. There is also a clear distinction between deterministic and stochastic models, and between foundation models and application-specific models. In terms of data sparsity and local adaptation, AI systems trained on global datasets may not reflect local epidemiology or rare conditions. An emphasis is needed on Bayesian contextualisation and representative data inputs. As such, this report advocates for model transparency, local validation, and clear statements about training data limitations. In the next chapter, several use cases are presented to illustrate AI opportunities, applications, challenges, and potential risks in various relevant healthcare fields.

For orientation, the report distinguishes four broad functional areas: AI for clinical decision support and diagnosis; AI for patient engagement, monitoring, and virtual care; AI for health-system, administrative, and public-health functions; and AI for health research and selected parts of medicines development.

These categories overlap, but they help policy-makers identify the relevant risks, evidence standards, and institutional responsibilities.

Across these functional areas, interoperability standards are central to implementation. Standards such as HL7 (Health Level Seven) and FHIR (Fast Healthcare Interoperability Resources) for health-data exchange and DICOM (Digital Imaging and Communications in Medicine) for imaging data can support portability,

auditability, secondary use, and cross-border learning under the European Health Data Space.

Embodied AI and robotics are relevant to healthcare, especially in surgery, rehabilitation, logistics, and assistive care. This report does not provide a full robotics assessment, but its principles on validation, accountability, human oversight, cybersecurity, and patient-centred implementation apply where robotic systems incorporate AI components.

2 Use cases, opportunities, and developments

The use case categories in this chapter include diagnosis, treatment optimisation, workflow efficiency, public health applications, and administrative tools.

2.1 Medical imaging and diagnostic services

Artificial intelligence (AI) holds considerable promise in clinical imaging, enhancing both the acquisition and interpretation of medical images. It can optimise the reconstruction of three-dimensional tomographic data, such as magnetic resonance imaging and computed tomography, thereby reducing scan times, improving image quality, and minimising radiation exposure. By integrating imaging with patient data, AI systems assist clinicians in diagnostic decision-making and personalised treatment planning, including radiotherapy for specific cancer types. Applications also encompass the detection, segmentation, and characterisation of lesions and tumours, the assessment of malignancy and aggressiveness, and the prediction of treatment response through the integration of imaging, genomic, and clinical information. AI-powered surgical and interventional systems contribute real-time guidance and anticipate potential complications, enhancing procedural precision and patient safety. The integration of these technologies into clinical workflows can strengthen diagnostic accuracy, reduce errors, and support more efficient, evidence-based, and patient-centred care. AI applications include assessment of the risk of developing a disease, early diagnosis, diagnostic prediction models, and models to reduce diagnostic errors. Examples of validated applications include multimodal AI in cancer diagnosis. Challenges in validation studies include addressing biases,

representativeness, and data heterogeneity in case studies. There is a lack of standardised, open-source, comprehensive, high-quality datasets that could be used for external validation. Consequently, comparisons of performance metrics between commercial AI solutions are currently difficult to conduct using large datasets.

For patients, the key issue is not only whether an imaging model performs well in aggregate, but whether it reduces missed or delayed diagnosis, avoids unnecessary follow-up, and performs equitably across cancer types, age, sex, ethnicity, disability, and socio-economic groups. Reporting should therefore translate technical limitations into patient-level consequences, including missed cancers, delayed treatment, overdiagnosis, and unequal access to high-performing tools.

2.2 Virtual patient care: patient engagement and compliance

AI-based apps and generative AI have great potential, particularly for personalised patient care and diagnostics, patient engagement and adherence, prognostic prediction, and treatment decision support. If well implemented, they can also help bridge public trust gaps in AI ([Moore et al. 2025](#)). Owing to the inherent generative nature of generative AI models, however, the reliability of these models has to be critically assessed.

Patient-facing and clinician-facing AI tools should be co-created with health professionals and patients from the start. Co-creation is not simply a training exercise

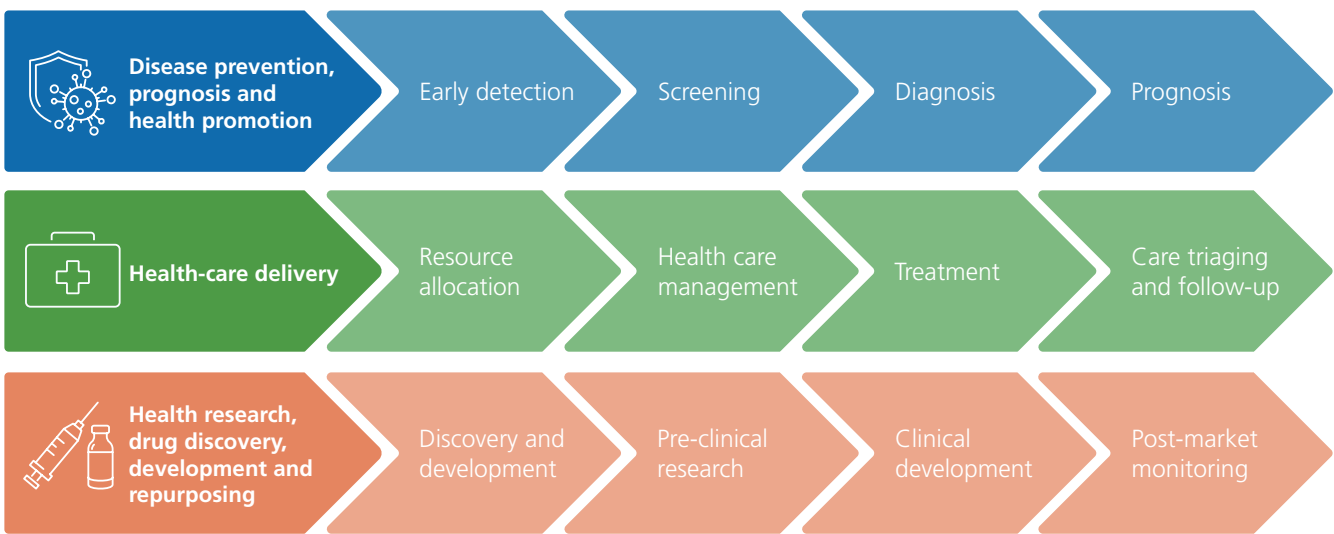


Figure 2 A general spectrum of artificial intelligence (AI) research and development in healthcare delivery and therapeutic development (adapted from Figure 1 of [WHO \(2023\)](#)).

after development: it helps define the clinical problem, workflow fit, communication needs, accessibility requirements, and the points at which human judgement and patient preference must remain decisive.

AI-powered predictive models based on diverse patient data sources – such as electronic health records (EHRs), genomic data, omics profiles, imaging, and clinical parameters – are not yet widespread; however, they are expected to revolutionise personalised medicine. By integrating these data streams, AI may enable clinicians to better identify patient-specific risks, tailor treatment strategies, and optimise resource allocation. Predictive tools have been developed to guide clinical decisions across various medical domains, including obstetrics, oncology, cardiology, and infectious diseases, improving outcomes and reducing unnecessary interventions (Huang *et al.* 2023). More generally, a scoping review identified actionable guidance for those closely involved in AI-based prediction model development, evaluation, and implementation (de Hond *et al.* 2022).

2.3 Medical/health research and drug discovery and development

In 2024, the European Medicines Agency (EMA) produced a reflection paper on the use of AI in the medicinal product life cycle, describing their current experience in a field where scientific knowledge is fast evolving (EMA 2024). The paper concluded that the quickly developing field of artificial intelligence/machine learning (AI/ML) shows great promise for enhancing all phases of the medicinal product life cycle, with direct applications of currently established regulatory principles, guidelines, and best practices to data management, governance, and statistical stringency. The report recommends that all organisations should reciprocally integrate data science competence with the respective fields within medicines development, manufacturing, and pharmacovigilance. However, the EMA warns that new risks are introduced when using AI models with exceptionally large numbers of trainable parameters arranged in non-transparent architectures, needing to be mitigated both during model development and deployment to ensure the safety of patients and integrity of clinical trial outcomes.

Within the scope of this report, drug discovery and medicines development are treated as policy-relevant health research use cases rather than as a complete assessment of pharmaceutical research and development. Relevant AI applications include target identification, toxicology prediction, trial design and recruitment, pharmacovigilance, repurposing, and prioritisation of areas of unmet need. These uses should be assessed not only for commercial productivity but also for affordability, public return on investment,

transparency of evidence, and contribution to European scientific competitiveness.

2.4 Rehabilitation

AI is rapidly transforming the field of medical rehabilitation by enabling more personalised, efficient, and predictive care. Recent studies highlight a range of use cases, including AI-driven predictive modelling for patient discharge planning in neurological and orthopaedic rehabilitation, achieving high accuracy in real-world clinical settings (Buscarini *et al.* 2025). AI technologies are also being integrated into rehabilitation through wearable devices, telerehabilitation platforms, and mobile health applications, enhancing remote monitoring, therapy personalisation, and patient engagement (Choukou and Barresi 2025). Furthermore, machine learning and brain–computer interfaces are being explored to optimise neurorehabilitation, offering new avenues for early diagnosis, adaptive treatment, and improved functional outcomes (Calderone *et al.* 2024). These developments emphasise the growing potential of AI to transform rehabilitation by improving accessibility, tailoring interventions, and ultimately enhancing the quality of life for individuals with acute or chronic conditions.

2.5 Population health

Guevara and colleagues investigated the optimal methods for using large language models (LLMs) to extract social determinants (Ashery *et al.* 2025) of health from narrative text in electronic health records, which provide critical yet sparsely documented information to support research and clinical care (Guevara *et al.* 2024). They found that fine-tuned models were less likely than ChatGPT to change their prediction when ethnicity and gender descriptors were added to the text, suggesting less algorithmic bias. Their findings demonstrate that, on the one hand, there is potential for LLMs to improve real-world evidence on social determinants of health and assist in identifying patients who could benefit from resource support. On the other hand, such generative AI tools may now be used by the public for obtaining health and safety advice, and symptom checking. However, expert assessment has shown that caution is needed to protect public safety, as information may be incorrect or potentially harmful (Oviedo-Trespalacios *et al.* 2023).

2.6 Public health practice

AI is increasingly being integrated into public health practice, offering transformative opportunities across disease surveillance, health promotion, and population health management. A 2023 narrative review outlines how AI has been applied to spatial modelling, epidemic forecasting, misinformation control, and public health surveillance—particularly during the COVID-19 pandemic, where it supported contact tracing, rapid

testing, and predictive modelling (Olawade *et al.* 2023). AI tools have also been used to enhance the precision and scalability of public health interventions, such as targeting vaccination campaigns or predicting outbreaks on the basis of environmental and social data. Additionally, a 2024 commentary in *Preventing Chronic Disease* emphasises the importance of embedding ethical and equity considerations into AI deployment, warning that without inclusive data practices and transparent algorithms, these technologies risk reinforcing existing health disparities (Dankwa-Mullan 2024). As AI continues to evolve, its role in public health is expected to expand, particularly in areas such as behavioural health analytics, environmental health monitoring, and real-time policy evaluation.

2.7 Administrative applications

AI is increasingly being leveraged to streamline and enhance administrative processes in healthcare, offering significant opportunities for efficiency and cost reduction. A 2025 study published in the *Journal of Medical Systems* highlights how generative AI is being used to automate medical documentation, manage revenue cycles, and support healthcare marketing and public relations, thereby reducing clinician burnout and improving operational workflows (Bhuyan *et al.*

2025). Similarly, a 2024 article in *Health Affairs Scholar* explores AI's role in Medicaid administration, where it is being applied to automate eligibility determinations, reduce improper payments, and alleviate administrative burdens that account for a substantial portion of healthcare spending (Cho and Miller 2024). These developments emphasise AI's potential to transform healthcare administration by automating repetitive tasks, enhancing decision-making, and enabling more proactive and data-driven management of healthcare systems.

In view of the use cases presented above—which span clinical, technical, public- and population-health, practical, research, and administrative applications—we conclude that the report's aims and scope adequately cover the relevant opportunities, challenges, and risks. These issues are elaborated in the following chapters.

A useful policy test for each use case is whether the AI application addresses a real clinical, public-health, or system problem; whether it has been co-created with intended users; whether evidence shows added value over current practice; and whether deployment improves safety, equity, sustainability, efficiency, or patient experience rather than merely adding technical complexity.

3 Ensure reliable, safe, equitable, and trustworthy artificial intelligence in healthcare, through rigorous analytical and clinical validation

This chapter examines the key principles, methodologies, and policy considerations required to ensure reliable, safe, equitable, and trustworthy artificial intelligence (AI) systems in healthcare. It distinguishes between analytical and clinical validation, discusses technical and real-world evaluation challenges, and situates validation within evolving international regulatory and governance frameworks.

The chapter further addresses emerging issues including multimodal and generative AI systems, explainability, human factors, life-cycle monitoring, and post-market surveillance. Finally, it provides policy-oriented recommendations to support robust validation practices and responsible implementation of healthcare AI across diverse clinical settings.

3.1 Introduction

Motivation

Validation is a critical step in developing and deploying AI systems in healthcare, as it provides the evidence needed to build trust, safety, and effectiveness (Tao *et al.* 2019). Without rigorous validation, even highly sophisticated algorithms can produce biased, unreliable, or clinically irrelevant results that may compromise patient outcomes and lead to preventable patient harm. Even more than traditional healthcare technologies with fixed and predictable behaviour, healthcare AI must meet rigorous standards to ensure that decisions affecting diagnosis, treatment, and resource allocation are accurate and equitable. AI models can be sensitive to data shifts, institutional context, and population differences, making robust validation essential. Importantly, validation also provides the evidence base needed for regulatory approval, reimbursement, and clinician adoption. In regulated environments, validation is not optional but a legal and ethical prerequisite for clinical deployment. In developing AI systems, datasets are typically partitioned into training, validation, and test sets. The *training set* is used to fit the model, the *validation set* to optimise parameters, and the internal *test set* to provide an initial performance estimate. Early planning of validation – including clear distinction between analytical and clinical validation, predefined metrics, and target populations – supports robust model design and prevents costly downstream errors. Neglecting this foundational planning can lead to mismatches in validation goals, poor generalisability, or regulatory rejection.

Analytical versus clinical validation

Validation occurs at two complementary levels: analytical and clinical. *Analytical validation* evaluates whether the algorithm performs as expected in controlled conditions, typically during development through internal testing with held-out datasets. It examines aspects such as accuracy, calibration, robustness, reproducibility, and risk controls aligned with medical-device standards (e.g. ISO 14971 risk management). Conversely, *clinical validation* tests the algorithm in real-world healthcare environments, focusing on external validation across diverse populations and settings to confirm generalisability and practical utility. This distinction is critical because a model can excel in controlled tests; yet, it can also fail in broader clinical practice because of variations in patient demographics, workflow, or data collection. Both validation levels are thus essential for establishing trustworthiness in healthcare AI. Planning for both validation levels in advance helps define appropriate data sources, validation cohorts, and performance endpoints, minimising the risk of model drift and ensuring that validation activities remain clinically meaningful. Both validation levels must be explicitly linked to the intended clinical use and level of autonomy of the AI system.

Positioning within the AI regulatory setting

Validation must also be understood within broader regulatory and quality frameworks. Agencies such as the US Food and Drug Administration (FDA) (FDA 2025a), the European Medicines Agency (EMA 2024), and the UK's Medicines and Healthcare products Regulatory Agency (MHRA) are creating standards that balance safety with innovation, emphasising transparency, reproducibility, and continuous quality assurance in healthcare AI. In the European context, the European Union (EU) AI Act (EU 2024) classifies most medical AI systems as high-risk and introduces legally binding requirements for validation, post-market monitoring, and human oversight (including Article 72 obligations for post-market surveillance). Other emerging global initiatives, such as Singapore's Health Sciences Authority and Pan-Canadian AI for Health guidance, highlight the increasing international effort to harmonise AI validation standards. Adhering to these standards – including version control, traceability, and life-cycle monitoring, and quality management systems (e.g. ISO 13485) – supports ongoing compliance and mitigates risks such as safety lapses, accountability gaps, or unrecognised model drift.

3.2 Analytical validation: concepts, methods, and challenges

Definition and scope

Analytical validation refers to the process of evaluating whether an AI system performs as intended under controlled conditions. It represents the first assurance layer, performed well before clinical deployment. Its goal is to ensure that models are technically sound, reliable, and appropriately benchmarked before they reach healthcare settings. Analytical validation also serves as an early warning system for algorithmic bias and technical limitations, allowing developers to iterate before patient exposure. Ensuring high data quality and representativeness at this stage – through diverse, well-documented datasets and robust pre-processing – is critical to prevent algorithmic bias, discriminatory outcomes, or unreliable model behaviour in practice.

Internal validation

A key part of analytical validation is internal validation, where model performance is assessed on data from the same source population as the training set, but never on the exact same datapoints. Typically, data are split into training, validation, and test sets, or handled by *k*-fold cross-validation, to make the most of limited datasets. This helps prevent overfitting and ensures performance metrics reflect genuine predictive ability. Avoiding data leakage is also essential. Data leakage occurs when information from the test or validation sets inadvertently leaks into the model training, leading to artificially inflated performance estimates. Examples include reusing the same patient encounter across multiple subsets or allowing pre-processing steps to incorporate knowledge from the test set. Implementing robust data versioning, software life-cycle controls (e.g. ISO 62304), and automated validation pipelines can further mitigate leakage risk. Clear documentation and strict data separation are critical to maintain trust in analytical validation. Failure to prevent leakage or document provenance undermines reproducibility and credibility.

Metrics

The metrics selected to evaluate performance depend on the clinical context and the type of prediction task (Hicks *et al.* 2022). Commonly used measures include accuracy, sensitivity (true positive rate), and specificity (true negative rate). Receiver operating characteristic curves and the area under the curve provide a global assessment of discriminative ability across thresholds, while calibration metrics measure the agreement between predicted probabilities and observed outcomes. From a regulatory and policy perspective, it is increasingly important to justify metric selection in relation to clinical risk and intended use, rather than relying solely on generic performance indicators. Fairness metrics are increasingly incorporated to ensure

equitable performance across subgroups (Hazirbas *et al.* 2022). Visualisation and visual analytics can augment this by providing intuitive, interactive representations of model behaviour, uncertainty, and performance, supporting better understanding of the inner workings of models (Alicioglu and Sun 2022). Emerging practices include uncertainty quantification and risk stratification to identify when models should defer to human judgement. Robust evaluation frameworks (Kocak *et al.* 2025) that account for the differential costs of false positives and false negatives, including both performance and fairness metrics, help prevent misinterpretation of results and reduce vulnerability to adversarial attacks or instability.

Challenges

Despite methodological rigour, challenges remain in analytical validation (Templin *et al.* 2024). Dataset quality plays an important role: incomplete or noisy records, mislabelling, and heterogeneity in clinical documentation can impact model reliability. Bias and underrepresentation risk poor generalisation, and repeated measures from the same patients or institutions can inflate performance estimates. Addressing these issues requires careful dataset curation and commitment to diversity and representativeness. AI systems in healthcare additionally face data privacy and security risks, as sensitive patient information can be exposed through breaches or improper handling (Murdoch 2021). Moreover, generative AI may produce hallucinations or inaccurate outputs and is vulnerable to adversarial mis-prompting or ‘jailbreaking’, which can potentially compromise clinical decisions (Omar *et al.* 2025b). Addressing these issues requires robust cybersecurity measures, strict access controls and model testing, continuous monitoring, and clear human oversight protocols to ensure outputs are safe, accurate, and ethically sound. Comprehensive stress-testing for robustness, adversarial resilience, and cybersecurity is thus an essential element of model evaluation and security assurance. These aspects will be discussed in Chapter 5.

Quality considerations

Transparency in data provenance, documentation of preprocessing pipelines, reproducibility of results, and adherence to quality management standards (e.g. ISO 13485) are increasingly viewed as essential for trustworthy analytical validation. Open protocols, independent replication, and standardised reporting further strengthen credibility. Additionally, model interpretability and explainability supported by effective visualisation and clear guidance (Musleh *et al.* 2025) can help build clinician trust, while helping stakeholders understand outputs, identify potential errors, and make informed decisions, bridging the gap between technical validation and clinical usability (Buhler *et al.* 2025).

Embedding explainability and human factors into analytical validation – through interpretable outputs such as SHapley Additive exPlanations (SHAP) values (Lundberg and Lee 2017), counterfactuals (Carvalho *et al.* 2019), workflow assessments (Topol 2019), and usability testing (ISO 62366) – supports clinician trust and mitigates the risk of misuse or cognitive overload.

3.3 Clinical validation: concepts, methods, and challenges

Definition and scope

Clinical validation determines whether an AI system demonstrates *real-world clinical utility* and *safety*. While analytical validation establishes that a model works as intended in controlled environments, clinical validation goes further by asking whether it meaningfully improves healthcare outcomes and can be safely introduced into routine practice. This validation ensures that algorithms are clinically relevant and technically sound. Clinical validation must explicitly address deployment context, workflow integration, and regulatory obligations for real-world use. Neglecting clinical validation risks deploying algorithms that perform well in controlled settings but fail under real-world variability, leading to inequitable or unsafe patient care (You *et al.* 2025).

External validation

External validation evaluates performance on patient populations and clinical settings distinct from those used for model development. This is critical for assessing generalisability, since healthcare systems vary widely in patient demographics, disease prevalence, and diagnostics/care delivery practices. Numerous studies have shown that models trained in academic medical centres often perform poorly when transferred to real-world settings. In the EU, external validation often relies on multicentre observational studies, registry-based studies, or pragmatic trials, reflecting practical and ethical constraints on large-scale prospective randomised controlled trials. Validating AI-enabled clinical systems across multiple heterogeneous sites ensures reliable and generalisable performance in real-world settings (You *et al.* 2025). Employing federated learning methods during multi-site validation can also protect patient privacy and enhance generalisability without data sharing (Dayan *et al.* 2021).

Clinical methodologies

To generate robust evidence, researchers are increasingly adopting clinical trial methodologies for AI evaluation (Gomis-Pastor *et al.* 2024). Randomised controlled trials remain the gold standard, allowing clear assessment of causal impact on outcomes. Pragmatic trials test algorithms in routine care environments and provide

insight into real-world usability and effectiveness. Finally, adaptive trial designs allow intra-study modifications, better reflecting the dynamic nature of AI.

Endpoints

The choice of endpoints is another crucial consideration (Ciani *et al.* 2021). While surrogate endpoints such as diagnostic accuracy are useful in early testing, ultimate validation should focus on patient-centred outcomes, such as morbidity, mortality, or patient quality of life. Including clinical and health-system impact endpoints (e.g. cost-effectiveness, efficiency, and quality of care) provides a more complete understanding of an AI's real-world benefits and risks. Regulators increasingly expect justification of endpoints in relation to clinical benefit, risk mitigation, and policy priorities such as cost-effectiveness and equity.

Where AI systems materially affect care pathways, validation should include patient-centred endpoints where feasible, such as patient-reported outcomes, patient experience, quality of life, time to diagnosis, treatment delay, access, safety events, and distribution of benefit across groups. These endpoints help avoid over-reliance on technical accuracy metrics that may not translate into better care.

Challenges

Variability across institutions in data quality and clinical workflows can obscure results, while interoperability issues limit portability of AI solutions. Ethical considerations (including transparency, informed consent, and equitable performance) must be addressed to ensure that validation efforts uphold patient rights and public trust. Neglecting ethical and cultural context risks inequitable care, patient harm, reputational damage, and loss of patient trust. It is also important to address integration challenges, such as ensuring that AI recommendations are interpretable, actionable, and do not disrupt clinical decision-making. One must also be aware of post-deployment drift, where AI performance may degrade over time because of changes in populations, protocols, or equipment. Here, we anticipate that scenario-based simulations to evaluate interactions between AI systems and clinicians, ongoing post-deployment monitoring to detect and address model drift, and transparent reporting of methodologies, metrics, and limitations to support reproducibility, regulatory compliance, and stakeholder trust will be valuable.

3.4 Factors impacting validation outcomes

Datasets

Validation quality depends heavily on dataset characteristics. Whether arising from historical inequities, imbalanced data collection, or selective

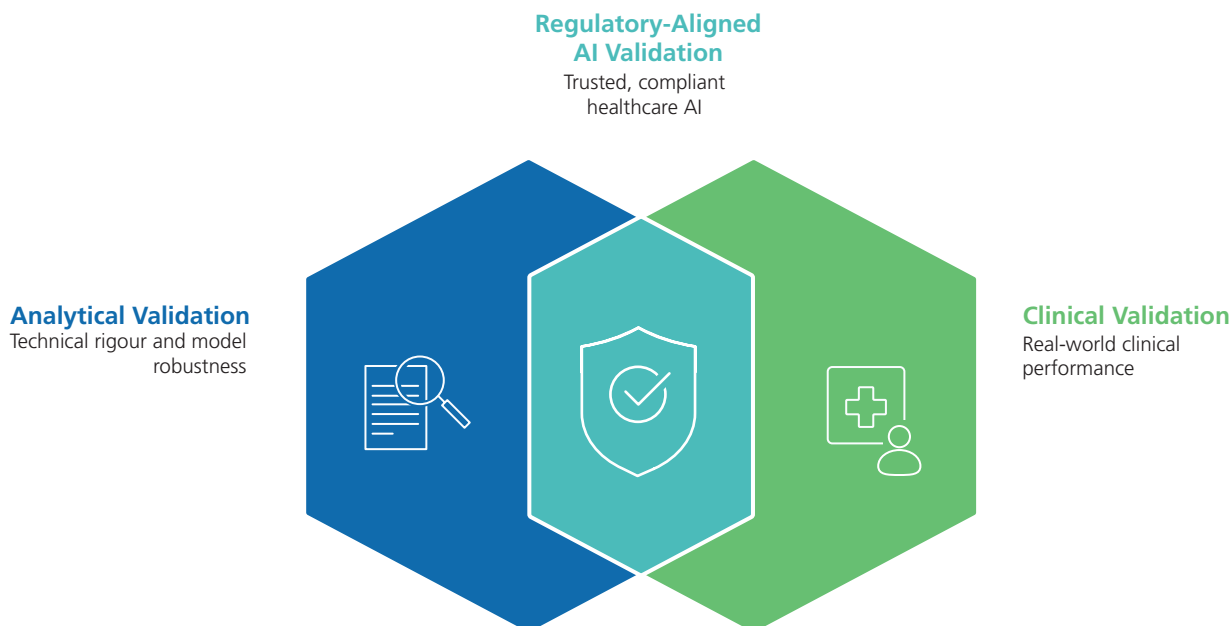


Figure 3 Aspects of integrated validation for reliable AI in healthcare: analytical validation, clinical validation, and in-between regulatory-aligned AI validation.

inclusion of certain populations, bias in datasets can undermine both analytical and clinical validation (Cross *et al.* 2024). For example, algorithms trained predominantly on data from middle-aged, White, urban populations may underperform in rural, minority, or paediatric groups. Beyond demographic diversity, clinical diversity – representing varying disease prevalence, comorbidities, and treatment patterns – is equally critical to ensure model generalisability. Bayesian approaches and uncertainty quantification can help identify when predictions are less reliable for underrepresented groups, guiding targeted evaluation, adaptation, or human oversight. Ensuring demographic diversity is therefore critical to fairness and equity, as well as to building models that are representative of the populations they are intended to serve. Equally important is data volume: small datasets risk overfitting and reduce statistical power, while large, heterogeneous datasets increase robustness and reproducibility. However, assembling such datasets can be expensive, time-consuming, and logistically challenging, particularly across multiple sites or rare patient populations. Incorporating synthetic or augmented data can also enhance dataset completeness, provided it is carefully validated against real-world distributions. Transparent reporting of dataset composition, curation processes, and limitations enhances accountability and allows users to judge potential risks of bias.

Regulatory environment

The regulatory environment also strongly influences validation practices (Aboy *et al.* 2024). In the EU, current regulatory frameworks substantially constrain fully autonomous, continuously learning AI systems in

clinical practice. Commercial systems must be locked at the time of approval, meaning they cannot dynamically update as new data becomes available. While this ensures stability and traceability, it also raises concerns about non-adaptivity, since models may degrade in performance over time as clinical practices evolve, patient populations shift, or when deployed across different geographical regions or healthcare settings. This trade-off emphasises the importance of planning for post-market surveillance and structured update pathways to maintain clinical efficacy and safety. Regulators and developers must therefore weigh the trade-offs between safety and adaptability.

Closely linked to this issue is software life cycle and version control. Given the rapid pace of algorithmic advancement, consistent and transparent methods for updating AI systems are essential. Concepts such as stage-gating approvals, version tracking, and kitemarking provide structured pathways for certifying updates, but implementation remains uneven. Integration of automated auditing tools and standardised documentation protocols can improve traceability and simplify regulatory compliance (EPRS 2022). Without clear standards, stakeholders face uncertainty about whether updates preserve or compromise safety. Embedding reproducibility and traceability into every stage of the life cycle is crucial to maintaining public trust.

Stakeholders

Regulators prioritise patient safety, transparency, and compliance with ethical and legal norms. Developers often emphasise innovation, speed, scalability, and

competitive advantage. Clinicians seek tools that integrate smoothly into workflows and enhance decision-making without creating additional liability. In addition to accuracy and calibration, patients value equity, explainability, and assurances that algorithms perform as intended. Policy-makers, meanwhile, focus on population-level outcomes, cost-effectiveness, equitable access, and ensuring broader health system goals. Reconciling these perspectives requires dialogue and shared frameworks that balance innovation with responsibility. Bridging these perspectives requires multi-stakeholder advisory panels, co-design initiatives, and structured feedback loops to align priorities and mitigate conflicts.

3.5 Regulatory and policy landscape

International collaborations

The validation of AI systems in healthcare does not occur in isolation but within a complex global regulatory environment. International collaboration has been critical in shaping standards and aligning practices. Agencies such as the European Medicines Agency (EMA), the US Food and Drug Administration (FDA), the UK's Medicines and Healthcare products Regulatory Agency (MHRA), and Japan's Ministry of Health, Labour and Welfare have organised workshops, surveys, and pilot projects to address the technical and policy challenges of AI validation. For instance, the FDA's 'Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)' outlines a 'predetermined change control plan' to manage modifications in AI/ML-based SaMDs (FDA 2019). Similarly, the EMA's reflection paper discusses the use of AI throughout the medicinal product life cycle, emphasising the need for adaptive regulatory approaches (EMA 2024). Collaborative initiatives such as the International Medical Device Regulators Forum also aim to harmonise standards and facilitate cross-border deployment of AI systems. The need to balance safety and regulatory oversight with the promotion of innovation that could improve healthcare delivery is widely recognised in all these efforts. The World Health Organization (WHO) has further contributed by issuing ethical and governance frameworks for AI in health, with a focus on equity and transparency (WHO 2025).

Beyond pre-market authorisation, international regulators are increasingly converging on the need for structured post-market surveillance and real-world performance monitoring of AI systems. Frameworks emerging from the FDA, EMA, and International Medical Device Regulators Forum emphasise life-cycle oversight, including continuous data collection, performance re-evaluation, and mechanisms for incident reporting once AI systems are deployed in clinical practice. These efforts signal a shift from static approval

models towards ongoing regulatory stewardship of AI-enabled medical technologies.

Guidelines and frameworks

These efforts have also produced guidelines and frameworks that address both technical validation methods and broader policy implications. For example, FDA has released guidance on Good Machine Learning Practice, principles for reproducibility and transparency, and frameworks for evaluating adaptive algorithms and for ensuring the development of safe and effective AI/ML medical devices (FDA 2025b). Similarly, the MHRA's strategic approach to AI regulation aims to ensure patient safety while fostering industry innovation into 2030 (MHRA 2024). In the European context, the EU AI Act further reinforces these requirements by classifying most medical AI systems as high-risk and mandating risk management, data governance, human oversight, and post-market monitoring. Article 72 strengthens obligations for post-market surveillance and incident reporting, effectively extending validation responsibilities beyond market entry and into routine clinical use. In parallel, internationally recognised standards, such as ISO 13485 (quality management systems), ISO 14971 (risk management), ISO 62304 (software life-cycle processes), and ISO 62366 (usability and human factors engineering), provide concrete, auditable requirements that directly inform validation activities for AI-based medical devices. Risk analysis methods such as Failure Mode and Effects Analysis are increasingly used to operationalise these standards, enabling systematic identification, prioritisation, and mitigation of algorithmic and workflow-related risks before and after deployment.

Gaps

In some use cases, such as AI tools for medical-imaging triage, the link between validation, diagnostic accuracy, and patient safety is direct, and validated products have received regulatory clearance. By contrast, rapidly advancing areas such as multimodal large language models are developing faster than current guidance can accommodate. Therefore, a tension remains between ensuring patient safety and efficacy and enabling timely innovation and adoption. Policy-makers continue to debate how best to evaluate dynamic learning systems, address data privacy concerns, and ensure equity in access to validated AI tools. Regulatory sandboxes and controlled real-world testing environments are increasingly considered to bridge this gap, allowing limited clinical deployment under enhanced oversight while generating prospective validation evidence. In parallel, the development of visualisation, visual analytics, and structured guidance is increasingly recognised not only as a usability aid but also as a regulatory enabler, supporting model explainability, clinician understanding, defensible

decision-making, and safe, informed adoption of complex AI systems.

3.6 Emerging issues in validation

Multimodal AI systems

The rise of multimodal AI systems (Soenksen *et al.* 2022), which integrate diverse data types such as medical imaging, electronic health records, genomics, and clinical notes, introduces new challenges for validation. These systems offer the potential for more comprehensive and accurate predictions, but their complexity complicates both analytical and clinical validation. Performance must be assessed not only for each individual data modality but also for their combined interactions, which can introduce subtle biases or errors that are difficult to detect. Federated learning (Zhang *et al.* 2021) could be a way to address ethical concerns in healthcare by enabling AI to train on decentralised patient data, keeping sensitive information private. Only model updates are shared, which can support compliance with the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR).

Emerging foundation and general-purpose AI models, including large language and multimodal models trained across broad domains, introduce additional validation challenges related to transferability across clinical contexts, emergent behaviour, unintended capabilities, and unclear validation boundaries. Their broad applicability and rapid adaptation potential complicate conventional validation approaches that are typically designed for narrowly defined clinical tasks.

Human factors

Human factors play a central role in the safe deployment of AI in healthcare. Clinicians remain the primary decision-makers, and there is a risk of over-reliance on algorithmic outputs, particularly when AI systems are perceived as highly accurate. Validation should therefore include assessment of AI integration into clinical workflows and its impact on clinician decision-making. Usability studies, trust calibration, and workflow simulations help identify cognitive overload or over-reliance on AI outputs. Structured evaluations ensure AI is an effective decision support tool, providing actionable insights while preserving human oversight. To this end, *explainability* (i.e. how clearly an AI's decisions can be understood by humans) and *interpretability* (i.e. how understandable the AI model's internal logic is) are critical. Models must provide actionable insights rather than opaque predictions, and guidance should be built into analytics to support safe and secure decision-making (see Chapter 5 for further details). Standards for interpretability should include both global model behaviour and case-specific reasoning to support clinician trust. Success also relies on collaboration

among healthcare providers, data scientists, and ethicists to create ethical, effective, and practically applicable AI solutions in clinical practice (McLennan *et al.* 2022), keeping in mind potential biases that arise (Hanna *et al.* 2025).

Assistive and autonomous AI systems

The distinction between assistive and autonomous AI systems further influences validation thresholds. Assistive tools, which augment clinician decisions, may tolerate lower levels of validity if oversight is continuous. On the other hand, fully autonomous systems, responsible for decisions without human intervention, require more stringent validation, including robust evidence from multi-site clinical studies and clearly defined safety monitoring mechanisms.

Validation thresholds should be determined by intended use, risk, degree of autonomy and the strength of human oversight. Assistive tools may justify proportionate evidence requirements only where risks are low and oversight is meaningful; autonomous or highly consequential systems require stronger prospective and multi-site evidence.

Acculturation

Acculturation in healthcare is also essential. Healthcare professionals and patients must be educated about AI's capabilities, limitations, and appropriate use. Structured training programmes, scenario-based simulations, and patient education initiatives can improve adoption, trust and confidence, and responsible use. Engagement strategies such as co-design workshops with patients and clinicians can further enhance adoption and trust in AI tools (Zicari *et al.* 2021).

Legal aspects

Finally, legal aspects intersect with all these considerations. Liability frameworks, informed consent requirements, and data privacy regulations shape both validation strategies and real-world deployment. However, liability allocation and accountability frameworks for AI-enabled clinical decision-making remain an evolving area across regulatory systems. In the European context, emerging frameworks such as the EU AI Act (EU 2024) further emphasise risk-based classification, transparency, human oversight, and post-market monitoring requirements for high-risk medical AI systems, directly influencing validation and governance practices. Ensuring compliance with legal and ethical standards is necessary to protect patients and support sustainable implementation. Continuous monitoring of regulatory changes, including evolving AI-specific legislation, is also critical to maintain alignment with best practices.

3.7 Recommendations

Recommendation 1. Establish rigorous validation planning

Action: clearly distinguish analytical validation from clinical validation, and define validation strategies early in the development process. Validation plans should specify appropriate datasets, evaluation metrics, and target populations, and include post-deployment monitoring and feedback loops from the outset.

Rationale: early and well-defined validation frameworks ensure that AI systems demonstrate clinically meaningful performance and can be safely integrated into healthcare practice.

Alignment with overall recommendations: R1 (evaluation of AI's added value), R2 (guidelines for responsible deployment), R9 (research priorities).

Recommendation 2. Ensure high-quality and representative datasets

Action: curate diverse, high-quality datasets that minimise data leakage and systematic bias. Document data provenance and preprocessing methods, and ensure datasets adequately represent marginalised and underrepresented populations, supported by systematic bias audits.

Rationale: poorly curated or unrepresentative datasets can lead to biased models, discriminatory outcomes, and unreliable clinical performance.

Alignment with overall recommendations: R1, R2, R6 (data ecosystems), R9.

Recommendation 3. Apply robust model evaluation and security testing

Action: evaluate models using appropriate validation protocols (e.g. train/validation/test splits, cross-validation) and clinically meaningful metrics, including calibration analysis. Systems should be stress-tested for robustness, adversarial vulnerabilities, and cybersecurity risks. For generative or adaptive systems, evaluation should explicitly assess hallucinations, prompt sensitivity, and unintended behaviour using scenario-based simulations.

Rationale: robust evaluation and security testing are essential to prevent unsafe deployment and ensure reliable system performance in real-world healthcare settings.

Alignment with overall recommendations: R1, R2, R4 (implementation challenges), R10 (governance and oversight).

Recommendation 4. Conduct external and multi-site validation

Action: validate AI systems across multiple institutions, healthcare settings, and patient populations to ensure generalisability. Where appropriate, leverage federated learning or privacy-preserving methods to support cross-site validation. Complement this with post-market surveillance to detect performance drift. When randomised controlled trials are infeasible, use pragmatic or registry-based prospective studies. Validation approaches should be proportionate to the intended clinical use, deployment context, and risk classification of the system.

To avoid unrealistic burdens, especially for small and medium-sized enterprises and public-sector innovators, policy-makers should support shared validation infrastructures, reference datasets, regulatory sandboxes, and transparent but feasible documentation requirements.

Rationale: multi-site validation helps ensure AI systems remain reliable across diverse clinical environments and reduces the risk of hidden biases or performance degradation.

Alignment with overall recommendations: R1, R5 (learning health systems), R6, R9.

Recommendation 5. Implement structured risk analysis and failure management

Action: apply structured risk analysis frameworks (e.g. Failure Mode and Effects Analysis) to identify technical, workflow, and cross-modality failure modes. Establish mitigation, fallback, and escalation mechanisms to manage potential risks and support regulatory safety assessments.

Rationale: systematic risk management is necessary to identify edge cases, prevent failure propagation, and demonstrate compliance with safety standards.

Alignment with overall recommendations: R1, R2, R4, R10.

Recommendation 6. Establish deployment governance and post-market monitoring

Action: define clear governance frameworks for clinical deployment, including accountability structures, human oversight mechanisms, and escalation pathways. Implement continuous post-market monitoring, incident reporting, and periodic model revalidation, aligned with regulatory requirements such as those under the EU AI Act. Post-market monitoring should include durable audit trails, version histories, model-change documentation, performance drift monitoring, incident reporting and mechanisms for decommissioning or

retraining systems that no longer perform safely or equitably.

Rationale: ongoing monitoring and governance are essential to detect performance drift and maintain patient safety in real-world use.

Alignment with overall recommendations: R5, R10, R4.

Recommendation 7. Promote explainability and human-centred design

Action: provide interpretable and transparent outputs (e.g. explainability tools such as SHapley Additive exPlanations (SHAP) or counterfactual explanations) and evaluate their effects on clinical workflows, cognitive load, and user trust. Implement human-centred design approaches, including clinician and patient co-design and usability testing.

Rationale: explainable systems improve clinician trust, support safe decision-making, and facilitate effective integration into healthcare practice.

Alignment with overall recommendations: R4, R8 (stakeholder engagement), R3 (ethical frameworks).

Recommendation 8. Ensure regulatory compliance, transparency, and life-cycle management

Action: adhere to applicable regulatory frameworks (e.g. EU AI Act, EMA, FDA standards) and maintain version control, traceability, reproducibility, and life-cycle monitoring. Developers should engage regulators early when designing adaptive or continuously learning systems, and clearly document system assumptions, limitations, and intended use.

Rationale: strong life-cycle management and regulatory alignment ensure accountability and reduce risks associated with continuously evolving AI systems.

Alignment with overall recommendations: R2, R4, R10.

Recommendation 9. Assess broader clinical, ethical, and societal impacts

Action: evaluate AI systems not only for technical performance but also for clinical outcomes, cost-effectiveness, equity impacts, privacy, consent, and cultural context. Assess how AI deployment affects healthcare systems and public trust over time.

Rationale: responsible AI deployment requires consideration of broader ethical, societal, and health-system impacts beyond model accuracy alone.

Alignment with overall recommendations: R1, R3, R8, R9.

Recommendation 10. Foster multi-stakeholder collaboration and education

Action: promote ongoing collaboration among clinicians, developers, ethicists, regulators, and patients throughout the AI life cycle. Establish structured AI literacy and training programmes, particularly for clinicians, while also improving patient understanding of AI-enabled healthcare.

Rationale: inclusive collaboration improves system design, strengthens trust, and supports responsible adoption of AI in healthcare.

Alignment with overall recommendations: R7 (training and education), R8 (stakeholder engagement), R4 (implementation).

4 Data quality is a cornerstone of safe, effective, and trustworthy healthcare

This chapter proposes a practical, standards-based framework for measuring, characterising, labelling and operationalising data quality across critical domains of medical data, particularly clinical tabular variables—including diagnoses, treatments, laboratory results and follow-up information—and imaging data, including radiology, nuclear medicine and pathology, encompassing DICOM series and associated annotations. It considers alignment with European institutions and infrastructures supporting federation and interoperability, with the aim of improving consistency and comparability across healthcare institutions and federated datasets. It also anticipates the implementation of Article 78 of the European Health Data Space Regulation, which provides for a Union data quality and utility label for datasets made available through health data access bodies, thereby offering a forward-compatible approach to emerging regulatory requirements (EU 2025).

High-quality data are the foundation of trustworthy medicine. In daily clinical practice, accurate, complete, and timely data underpin every stage of the care continuum—from diagnosis and treatment decisions to follow-up and outcome evaluation. Poor-quality data can lead to misdiagnosis, inappropriate interventions, and inefficient resource allocation, directly affecting patient safety and care continuity. Ensuring data quality at the point of capture (primary use) safeguards the reliability of the clinical record and supports effective communication among healthcare professionals.

Data quality is necessary but not sufficient for trustworthy medicine. Clinical judgement, patient context, professional accountability, and the therapeutic relationship remain essential, especially where artificial intelligence (AI) outputs are uncertain, incomplete, or difficult to explain. Better data alone cannot make care trustworthy.

In secondary uses – such as clinical research, artificial intelligence (AI) development, and health system management – data quality determines the validity, reproducibility, and generalisability of analytical results. Incomplete or biased datasets compromise model performance, introduce hidden confounders, and weaken the evidence base that informs clinical and policy decisions. As healthcare becomes increasingly data-driven, the integrity of both primary and derived data flows becomes critical to maintaining trust in AI-enabled diagnostics, predictive models, and decision-support tools.

Robust data quality governance – integrating technical, procedural, and ethical dimensions – is thus essential to

ensure that medical data remain fit for purpose across their entire life cycle: from initial collection to secondary exploitation. The convergence of clinical, regulatory, and AI-readiness requirements demands harmonised standards and transparent assessment methods capable of bridging operational healthcare realities with research-grade expectations.

To ensure FAIR (Findable, Accessible, Interoperable, and Reusable) data principles, data repositories can integrate the RDA FAIR Data Maturity Model for comprehensive evaluation, complemented by FAIR-EVA for automated FAIRness assessments (Aguilar Gomez and Bernal 2023). Together, these components enable a systematic, transparent and scalable approach to assuring data quality in AI-driven medical research and clinical applications. The FAIR principles provide a framework for making data findable, accessible, interoperable and reusable across research and clinical contexts (Wilkinson *et al.* 2016). For secondary use within the European Health Data Space, TEHDAS recommended a common data-quality framework, including the publication of dataset descriptions and information on their quality and utility (TEHDAS 2023c). Article 78 of the EHDS Regulation establishes a Union data quality and utility label: datasets made available through health data access bodies may carry the label, while datasets collected and processed with Union or national public funding must carry it (EU 2025). Such a label should transparently document the key attributes that determine a dataset's usability, including the following:

1. Metadata describing content, structure, and context.
2. Data dictionaries and applied standards (e.g. FHIR (Vorisek *et al.* 2022)).
3. Provenance and lineage information showing how data were generated, processed, and transformed.
4. Coverage details such as population, timeframe, and modalities.
5. Access conditions and provision times, defining how and how quickly data can be obtained.
6. Modification history and links to related datasets.

To assess these aspects in a structured and comparable way, the RDA FAIR Data Maturity Model can be used. It provides measurable indicators to evaluate each FAIR dimension at the catalogue level – for example, within a data repository or federation – and can be extended

Table 1 Technical data quality dimensions

Dimension	Definition (from DAMA's dimensions of data quality 2020)	We recommend for the EHDS
Completeness 	The degree to which all required variables are present for all observed cases.	✓
Uniqueness 	No entity exists more than once within the dataset.	✓
Validity 	The degree to which dataset values comply with predefined internal rules and external reference standards.	✓
Timeliness 	The degree to which the period between the time of creation of the real value and the time that the dataset is available is appropriate. (Not assessable on anonymised datasets.)	✓
Consistency 	The degree to which related dataset values adhere to predefined allowed or required relationships across attributes within a record, across records within a data file, between data files, and across time for the same entity.	✓
Accuracy 	The degree of correspondence between dataset values and real values.	✓
Integrity 	The degree to which dataset values remain intact and unaltered throughout their life cycle, ensuring protection against corruption, unintended modification, or loss during data capture, storage, processing, and exchange.	✗

to deeper layers of data access and reuse. Automated evaluation can be performed with FAIR-EVA, a tool implementing RDA FAIR metrics to test compliance programmatically (Wilkinson *et al.* 2019; Aguilar Gomez and Bernal 2023). Automated tools such as FAIR-EVA can assess many machine-readable FAIR indicators. However, their coverage and interpretation vary, and some domain- and context-specific aspects—particularly semantic interoperability, the appropriateness of controlled vocabularies and ontologies, and adherence to specialised standards—may require complementary expert assessment using a structured checklist (RDA FAIR Data Maturity Model Working Group 2020; Aguilar Gomez and Bernal 2023).

The EHDS guidance and European Commission FAQs summarise the scope and expectations for FAIR-aligned data quality labels, emphasising transparency, comparability, and readiness for cross-border data exchange (TEHDAS 2023b). Integrating FAIR assessments into data quality frameworks ensures that datasets are not only technically sound but also discoverable, trustworthy, and reusable for both clinical and research purposes.

4.1 Data quality dimensions

Seven technical quality dimensions are important when dealing with healthcare datasets: completeness, uniqueness, validity, timeliness, consistency, accuracy, and integrity. Documentation and utility aspects (metadata descriptors, use of standards, provenance, coverage, data access service level agreements, and data maturity level) complement these for the final label.

4.2 Core dimensions and thresholds of data quality

Different metrics and thresholds can be applied when measuring data quality. Data quality in healthcare relies on the combined assurance of technical validation, semantic coherence control, and operational reliability. We have drafted the following seven dimensions to define a practical, standards-based framework applicable to imaging data (DICOM and video), clinical audio records, and structured clinical (tabular) datasets. Each dimension includes measurable thresholds that indicate when data can be considered ‘fit for purpose’, ensuring reliability for both primary clinical use and secondary research or AI development.

These quality dimensions and thresholds (Tables 1 and 2) provide a unified framework for operationalising data quality across radiological and clinical domains. By defining explicit, measurable targets for completeness, uniqueness, validity, consistency, accuracy, and integrity, healthcare institutions can ensure that both primary

data collection and secondary reuse – including AI training and validation – are grounded in trust, interoperability, and regulatory compliance.

The possible quality issues detected and proposed solutions are illustrated in Table 2. While the technical

Table 2 Operationalising data quality for AI in healthcare: mapping issues to standards-based solutions

Dimension	Typical problems: imaging (DICOM)	Typical problems: clinical (tabular)	Recommended solutions	Cautions/strategic notes
Completeness	Missing mandatory image series; missing 'Type 1/1C' tags (patient/modality info); missing radiology reports.	Missing diagnosis dates or treatments; missing lab units; gaps in follow-up records.	1. Re-extract from source. 2. Enforce series templates. 3. Tabular: use validated imputation (e.g. MICE or k-NN) for non-critical values.	Define a 'minimum viable dataset'. Flag and audit all imputed values. Threshold: $\geq 95\%$ for mandatory fields.
Uniqueness	Duplicate study identifiers (UIDs); repeated frames in a series; file cloning.	Duplicate patient IDs; multiple records for a single event without version control.	1. De-duplicate using deterministic rules (keep the most recent/complete). 2. Imaging: Use hash-based file verification.	Zero tolerance for duplicate IDs in released datasets. Log every merge/drop for the audit trail.
Validity	Illegal pixel ranges; de-identification leaks (burned-in text); invalid segmentation attributes.	Impossible dates (birth after death); invalid clinical codes (non-standard ICD/SNOMED); free-text in coded fields.	1. Conformance testing (DICOM/FHIR). 2. Privacy scanners. 3. Cross-validate against terminology servers (SNOMED CT, LOINC).	Set defect targets $< 0.5\%$. Personal health information leaks must trigger an immediate security stop.
Consistency	Discrepancy between acquisition date and patient age; modality/sex conflicts.	Temporal logic errors (e.g. treatment starts before diagnosis); incompatible units across tables.	1. Implement longitudinal rule engines. 2. Synchronise timestamps using system provenance. 3. Enforce referential integrity rules.	Document all business rules. Aim for $\geq 99\%$ pass rate; justify exceptions (e.g. emergency overrides) in metadata.
Accuracy	Quantitative values (SUV, ADC) outside expected range; documented versus actual slice spacing.	Mismapped concepts; transcription errors from paper notes; 'label noise' in diagnostic outcomes.	1. Follow QIBA profiles. 2. Geometry validation. 3. Spot-audit against source EHR. 4. Calibrate natural language processing/extraction pipelines.	Warning: do not impute primary outcomes or critical measurements. Record the 'conditions of measurement' (e.g. scanner type).
Integrity	Corrupted pixel data; missing frames; broken references between series and annotations.	Checksum mismatches; 'orphan' records (test results without a linked patient).	1. File integrity checksums. 2. Enforce Foreign Key integrity at the ETL (extract, transform and load) stage. 3. Re-export from PACS if corrupted.	Release only datasets with zero corruption. Maintain a manifest of all processing runs to ensure reproducibility.
FAIR principles	Missing acquisition metadata; no provenance for AI annotations.	No data dictionary; opaque local codes; absent usage licences or cohort definitions.	1. Publish metadata records (DCAT-AP). 2. Map local codes to OMOP/FHIR. 3. Use automated tools (e.g. FAIR-EVA) for assessment.	Quality is not just about the data, but the documentation. Labels must include coverage, standards, and access terms.

Abbreviations: ADC, apparent diffusion coefficient; AI, artificial intelligence; DCAT-AP, Data Catalogue Vocabulary Application Profile for data portals in Europe; DICOM, Digital Imaging and Communications in Medicine; EHR, electronic health record; ETL, extract, transform and load; FAIR, findable, accessible, interoperable and reusable; FAIR-EVA, FAIR Evaluator, Validator and Advisor, an automated tool for assessing adherence to FAIR principles; FHIR, Fast Healthcare Interoperability Resources; ICD, International Classification of Diseases; ID, identifier; k-NN, k-nearest neighbours; LOINC, Logical Observation Identifiers Names and Codes; MICE, multiple imputation by chained equations; OMOP, Observational Medical Outcomes Partnership, commonly referring to the OMOP Common Data Model; PACS, picture archiving and communication system; QIBA, Quantitative Imaging Biomarkers Alliance; SNOMED, Systematized Nomenclature of Medicine; SNOMED CT, Systematized Nomenclature of Medicine Clinical Terms; SUV, standardized uptake value; Type 1/1C, DICOM data-element requirement categories in which Type 1 elements are mandatory and must contain a valid value, while Type 1C elements are mandatory under specified conditions; UID, unique identifier.

parameters below (e.g. DICOM tags) are essential for biomedical engineers, policy-makers should focus on the 'Recommended solutions' column to understand the infrastructure required for safe AI.

4.3 Recommendations

Recommendation 1. Clarify and enforce rules for secondary use of health data

Action: establish clear and consistent EU-wide rules for the secondary use of health data for research and AI development, with strong safeguards for privacy and security. Developers and deployers of health AI should be required to document dataset provenance, legal basis for data use, permissions, and key dataset characteristics used in training and validation. In parallel, the EU should accelerate data harmonisation, interoperability standards, and secure data-sharing mechanisms to enable responsible innovation without weakening data protection.

Rationale: transparent data governance and clear rules for secondary use are essential to build trust, ensure legal compliance, and enable AI innovation to move safely from research into routine care.

Alignment with overall recommendations: R6 (interoperability and data ecosystems), R10 (governance and oversight), R1 (evaluation of AI's value).

Recommendation 2. Align EU guidance with real-world healthcare system capabilities

Action: ensure that EU guidance on AI in healthcare explicitly reflects the operational realities of hospitals and healthcare providers, including the cost, time, and uneven availability of high-quality clinical data. Where higher standards for data collection, validation, or evaluation are required, policies should include realistic implementation timelines and targeted support mechanisms to enable healthcare systems to meet these expectations.

Rationale: AI policy that assumes capabilities not yet available in healthcare systems risks slowing adoption, creating compliance burdens, and widening disparities between healthcare providers.

Alignment with overall recommendations: R4 (implementation challenges), R1 (evaluation), R7 (training and capacity building).

Recommendation 3. Develop federated data infrastructure as a strategic capability

Action: support federated data infrastructures that allow datasets across hospitals, regions, and Member States to be analysed securely while data remain locally governed. The EU should designate clear leadership and accountability for federation initiatives at both EU and national levels, and provide sustained funding to scale successful pilot projects into operational infrastructures.

Rationale: federated data approaches allow large-scale analysis and AI development while preserving data privacy, institutional autonomy, and public trust.

Alignment with overall recommendations: R6 (data sharing ecosystems), R5 (learning health system), R10 (governance structures).

Recommendation 4. Standardise clinical data definitions and semantics

Action: mandate common definitions and semantic standards for key clinical measures – including what is measured, how it is measured, and when measurements occur – and require the use of agreed terminologies, ontologies, and data formats across Member States. Lessons from successful standardisation initiatives such as DICOM in medical imaging should inform similar efforts across broader clinical data domains.

Rationale: standardised clinical data definitions are essential for interoperability, valid cross-site analysis, and the safe evaluation and deployment of AI systems across healthcare systems.

Alignment with overall recommendations: R6 (interoperability and data ecosystems), R1 (evaluation of AI's added value), R9 (research priorities).

5 Responsible artificial intelligence use and innovation require robust, multi-layered security strategies

5.1 A new technology entailing new threats

The previous chapters mentioned many successful applications of artificial intelligence (AI) within the healthcare domain. Several of them were truly game-changing, modifying health practice as we know it in several respects. Yet along with all its potential enormous benefits, AI brings along risks to healthcare that we must learn to navigate if we want to fully realise such benefits for the benefit of all.

In the past 2 years, the World Economic Forum has ranked adverse outcomes of AI technologies, including misinformation and disinformation, among the worst global threats in the short and medium term in its Global Risks Report ([Elsner et al. 2025](#)): their potential to disrupt geopolitical stability, national security, and, more in line with the current report, public health is large. Indeed, numerous significant AI risks to healthcare systems have already been reported, for example in the AI incident database ([Responsible AI Collaborative 2024](#)) or the International AI Safety Report ([Bengio et al. 2025](#)). Complementing this, a 2024 article in *Cluster Computing* presents a systematic review of 99 studies on data breach mitigation in healthcare, highlighting the growing use of AI alongside blockchain and encryption technologies to protect electronic health records and personal health data ([Nemec Zlatolas et al. 2024](#)). The study emphasises that healthcare remains one of the most targeted sectors because of the high commercial value of its data, and stresses the need for scalable, adaptive cybersecurity mechanisms. These findings collectively point to the critical need for proactive, interdisciplinary approaches to secure AI systems and maintain public trust in digital healthcare.

Moreover, AI is increasingly being used with other health technologies, for example synthetic biology, potentially giving rise to further biosecurity threats based on using AI to generate new biochemical weapons ([Urbina et al. 2022](#)). Similarly, a 2024 chapter in *Cybersecurity and Artificial Intelligence* discusses how AI can both enhance and challenge medical record security: while AI-driven anomaly detection and biometric authentication can strengthen defences, the complexity of these systems also increases the risk of unauthorised access if not properly governed ([Herzog 2024](#)). These findings emphasise the dual-edged nature of AI integration in healthcare and the urgent need for robust, multi-layered cybersecurity strategies that evolve alongside technological innovation.

5.2 Understanding the new threats

Such issues affect all four major tasks in healthcare, namely prognosis, diagnosis, treatment, and clinical workflow. Some examples include the following:

- During the COVID-19 pandemic, opposition to vaccination and lockdowns was strongly promoted through AI-enabled misinformation campaigns and was uncritically adopted in some communities.
- Diagnostic errors are another concern. Flawed AI recommendations can lead to unsafe or ineffective treatments. For example, AI tools used to analyse facial microexpressions – increasingly proposed for mental health triage or pain detection – have been shown to misclassify emotions across cultural and neurological contexts, leading to diagnostic errors or discriminatory outcomes ([Barrett et al. 2019](#)). Similarly, AI-driven symptom checkers may overlook atypical cases of stomach ulcers, especially when trained on non-representative datasets, delaying treatment or producing false reassurance ([Semigran et al. 2015](#)). When these tools operate without real-time clinical judgement, the risk that healthcare systems will embed and amplify biases or blind spots increases—contributing to cumulative digital fragility.
- Importantly, fraudsters and terrorists may take advantage of this by fooling the systems. Indeed, AI-powered hospital systems have a larger attack surface ([Herzog 2024](#)), making them targets for ransomware. Attackers may also exploit the interface between image-acquisition and diagnostic systems by subtly altering images, as explained below.
- Similarly, through repeated use or misuse, AI systems can unintentionally reveal sensitive patient data, requiring AI use to be monitored in the same way as other sensitive IT systems.
- Indeed, cyberattacks on health centres emphasise the importance of cybersecurity, especially with cloud-based AI systems. There are also specific susceptibilities (e.g. risks affecting home-based patient monitoring and management systems, as well as hospital systems). The increasing reliance on AI in healthcare has amplified the risks of data breaches and cyberattacks, including raising urgent concerns about the security of personal health information, which is further analysed in [section 6.2](#).

5.3 Dissecting the core security issues in AI for healthcare

For virtually every opportunity AI can offer, there may also be a risk. Importantly, these risks are not just another form of cyber threat. They are specific and it is crucial that we understand the underlying reasons:

- Numerous AI systems are highly complex and capable of learning and adapting over time. This makes them more unpredictable and difficult to control than traditional cyber systems which are often more static and predictable.
- AI threats often target not just systems or networks but also the data that trains AI models and the decisions made: if you train algorithms with biased data, this will lead to biased recommendations, perpetuating unfairness, as further detailed in [Chapter 6](#).
- AI can be used to automate and scale attacks at an unprecedented speed, allowing adversaries to execute large-scale manipulations with minimal effort. Traditional cyber threats often require more manual, labour-intensive tactics.
- AI's role in controlling critical systems, as many in healthcare are, entails that a successful attack can have far-reaching consequences, ranging from physical harm and financial instability to reputational damage ([Camacho et al. 2025](#)).
- Moreover, this growing critical dependence on AI applications in healthcare introduces a complex array of threats, particularly as these systems become integral to clinical decision-making, diagnostics, and patient management. A 2023 literature review in the *Asian Journal of Medicine and Health* categorises AI-related risks into clinical data vulnerabilities, technical system failures, and socio-ethical concerns, emphasising that overreliance on AI without robust safeguards can lead to cascading failures in care delivery during cyberattacks or system malfunctions ([Muley et al. 2023](#)). This is especially concerning in high-stakes environments such as emergency medicine or intensive care, where AI-driven tools are increasingly used for triage and monitoring. A 2025 scoping review in *BMC Medical Ethics* further warns that as AI systems become embedded in healthcare infrastructure, the ethical and operational consequences of system outages or data breaches become more severe, potentially compromising patient safety and institutional trust ([Ratti et al. 2025](#)).

- Besides, AI tools and expertise are now widely accessible: causing harm using AI is at the fingertips of virtually everybody, even more with the emergence of AI-based Crime-As-A-Service platforms ([Kaloudi and Li 2020](#)).
- IT systems tend to be proprietary and operate in silos: their review, repair and updating tend to be costly and time-consuming. As an example, in 2013, most hospitals were using the ninth version of the International Classification of Diseases (ICD) system—even though a revised version (i.e. ICD-10) was released as early as 1990. The difficulties in updating hospital software infrastructure can create many vulnerabilities with the use of modern tools such as machine learning/deep learning systems.

Importantly, AI systems are subject to novel types of attack using so-called adversarial examples, in which attackers subtly alter data in a systematic manner to fool a machine-learning system. One notorious example in medicine ([Finlayson et al. 2019](#)) refers to image-classification systems that can recognise images of tumours and even improve expert radiologists' performance, but can be completely fooled by such adversarial examples. Therefore, with no system outage, no ransomware, no alert, the AI would be simply making wrong clinical recommendations leading to incorrect treatment plans, patient harm, and loss of credibility. Similarly, a 2024 study titled 'Systematically assessing the security risks of AI/ML-enabled connected healthcare systems' highlights how integrating AI with peripheral devices and communication technologies (e.g. Bluetooth) introduces new vulnerabilities that adversaries can exploit to manipulate clinical data or disrupt care delivery ([Elnawawy et al. 2024](#)). For instance, the study demonstrates how adversarial data can be injected into an ML-enabled glucose monitoring system through insecure communication channels.

These types of attack have affected all types of data and numerous AI algorithms ([Ríos Insua et al. 2023](#)), including large language models (LLMs) through prompt attacks ([Lian 2026](#)). Attacks such as this essentially alter the data input to an AI algorithm. They can be easily automated and have potentially devastating consequences in the health domain. Even more concerning, these attacks can originate from anywhere in the world, making transnational regulation and control incredibly challenging, given that they are hard to detect. Besides, adversaries are smart and constantly evolve and improve their AI capabilities. And adversarial examples are just a subset of the broader problem of machine learning models exhibiting unpredictable and overly confident behaviour outside the training distribution ([Xia et al. 2025](#)). Moreover, the healthcare system aggressively adopts machine learning models to improve disease

detection and patient treatment without explicitly addressing these reported adversarial attacks, which could have motivations (Finlayson *et al.* 2019) ranging from insurance fraud, creative billing, or forcing device approval.

5.4 Pointing towards solutions

The above findings highlight the urgent need for resilient, transparent, and security-grounded AI governance frameworks that account for the risks of critical dependence in healthcare settings. The question is then how we ensure safer and more secure AI systems as we face a true arms-race in the AI cybersecurity domain, which is becoming increasingly autonomous, raising the risk that human decision-makers – essential for ethical and legal oversight – are systematically excluded ('out of the loop'). The EU's critical health infrastructure is especially vulnerable, given its complex digital interdependencies and high-value personal data. Historical cases such as autonomous drone operations (e.g. Libya in 2020) and close-call military escalations triggered by automated systems (e.g. the Petrov incident in 1983) illustrate how removing human judgement can nearly trigger catastrophic outcomes—parallels applicable to future AI-enabled disruptions in healthcare (Dubber and Lazar 2025).

As individuals, we should be aware of AI-related risks, help others become aware of them, and adopt good AI-safety practices when interacting securely with AI-based systems. This should include a healthy scepticism to verify data and information, including health information, to prevent manipulation, and even contribute to citizen oversight by encouraging the reporting of abuse or vulnerabilities that we detect in such systems.

Institutions and corporations should require AI system developers to design systems that are robust against adversarial attacks by integrating defence techniques from the field of adversarial machine learning, which are being gradually developed. This requires being permanently aware of the latest AI and cyber risk trends as well as promoting the training of response teams on core concepts of adversarial risk management. Importantly, beyond relatively recent cyber insurance products, insurance companies are slowly introducing AI-related products and instruments. Some of these will soon be tailored to risks associated with adversarial attacks; this may support a more responsible adoption of AI, particularly in healthcare (Stern *et al.* 2022).

Governments also have an important role as well. As a minimal demand, we should expect compliance with human rights and other international acts, and this should lead to establishing strong legislative frameworks. Among others, one good example is the EU AI Act, further discussed in Chapter 8. For many,

it has been perceived as an excessive burden and a deterrence to AI innovation. But the Act should instead be viewed as a stimulus for innovation in a different and more responsible direction.

In particular, this entails supporting and investing in research on secure machine learning, including the development and promotion of countermeasures to protect AI systems from attacks. Moreover, EU policies must therefore embed mandatory 'human-in-the-loop' controls, rigorous auditing of AI in cyber-defence applications, and sector-specific safeguards aligned with the AI Act and the Cyber Resilience Act to prevent both sudden and cumulative failures across, in particular, healthcare systems. All this is especially relevant within the health sector.

It also calls for international collaboration, particularly in relation to sharing data and information to better understand global threats. Of course, this is far from simple in this very strategic and geopolitically relevant domain, but earlier examples of international cooperation in nuclear energy and biochemical weapons might pave the way for this endeavour.

5.5 Privacy and security of personal data

The ability to demonstrate that anonymised data cannot be re-identified, and that all personal data are held securely with strict access restrictions, will be important for retaining public trust. A role for enforceable regulation to minimise risk should be considered. The integration of AI into healthcare systems introduces significant security risks and privacy concerns, particularly about the protection of personal health data. A 2025 study published on *arXiv* provides a comprehensive overview of the vulnerabilities AI systems face, including adversarial attacks, data leakage, and insufficiently tested threat models (Chang *et al.* 2025). The authors emphasise that current research is fragmented and lacks coordination with biomedical domains, which hampers the development of robust, context-aware defences. Similarly, a 2021 article in *BMC Medical Ethics* highlights the growing risk of re-identification of anonymised health data owing to advanced AI algorithms, raising concerns about patient consent and the ethical use of data in public-private partnerships (Murdoch 2021). Collectively, these publications emphasise the urgent need for proactive, interdisciplinary strategies to secure AI systems in healthcare and uphold patient confidentiality.

5.6 Recommendations

Recommendation 1. Recognise and address the distinct cybersecurity risks of AI in healthcare

Action: policy-makers, healthcare providers, and developers should explicitly recognise that AI systems

introduce novel security risks beyond those of traditional IT systems, including model manipulation, adversarial attacks, training data poisoning, and prompt-based vulnerabilities in generative models. EU cybersecurity and health authorities should develop guidance and risk assessment frameworks tailored to AI-specific threats in healthcare environments.

Rationale: AI-driven systems can be compromised in ways that differ significantly from traditional software systems, potentially affecting diagnostic outputs, clinical decision support, and patient safety. Understanding these risks is essential for designing effective safeguards.

Alignment with overall recommendations: R1 (evaluation of AI's added value), R4 (implementation challenges), R10 (governance and oversight).

Recommendation 2. Promote responsible innovation that integrates security and safety

Action: encourage responsible innovation frameworks that integrate security, safety, and regulatory compliance into the AI development life cycle without unnecessarily slowing technological progress. This includes adopting structured approaches to secure development, risk management, and testing in healthcare AI projects.

Rationale: secure development practices can coexist with innovation if governance frameworks promote early risk identification, iterative testing, and responsible deployment rather than reactive security fixes after systems are deployed.

Alignment with overall recommendations: R2 (responsible deployment guidelines), R4 (implementation challenges), R9 (research priorities).

Recommendation 3. Implement security-by-design for healthcare AI systems

Action: require security-by-design principles in the development and deployment of machine learning

systems used in healthcare. This includes secure data pipelines, robust model architectures, adversarial resilience testing, and alignment with regulatory frameworks such as the EU AI Act and relevant cybersecurity standards.

Rationale: embedding security protections from the earliest design stages is essential for preventing vulnerabilities and ensuring safe operation in critical sectors such as healthcare.

Alignment with overall recommendations: R2 (responsible deployment guidelines), R4 (implementation challenges), R10 (governance and oversight).

Recommendation 4. Strengthen data sovereignty and governance over healthcare AI systems

Action: establish strong data sovereignty and governance controls for AI systems handling patient data. This should include the ability to audit AI models, enforce strict access controls, and ensure that AI inference and storage occur on-premises or within hospital-controlled cloud environments where feasible. Governance frameworks should clearly define accountability for risk ownership and oversight, ensuring that decisions about data control remain institutional and regulatory rather than purely technological.

Rationale: without strong governance mechanisms, AI deployment risks shifting control of sensitive health data and critical decision-support systems away from healthcare institutions and public oversight bodies.

Alignment with overall recommendations: R6 (data ecosystems and sharing), R8 (public trust), R10 (governance and oversight).

6 Embed ethics into artificial intelligence design and development from an early stage

Ethics provides goals and interesting technical objectives with a long-term perspective; ethics provides values and these create expectations, allowing for coordination and collaboration; ethics in artificial intelligence (AI) has optimised the positive impact of AI on our lives, and is not just a 'naysayer'. AI is made by humans; it interacts with humans; it changes the world by acting; it is a socio-technical system and a transformative power that impacts society and the planet. Ultimately, AI is not value neutral and will inevitably interact with instrumentalised ethical constraints.

Ethics guidelines should ideally be grounded in universally accepted frameworks of values and principles, as well as in mechanisms that protect fundamental human rights and freedom, while also promoting equitable access to AI applications in health. A major challenge, however, is that any international normative framework for the ethical and societal implications of AI must be built on shared views about acceptable uses of private health data, respect for human rights and freedom, and, more broadly, the ethical values that should govern the use of AI in health. At present, significant differences exist at the national level between Europe, the USA, and China (see [Appendix](#) for comparative tables, [Tables A1–A4](#)).

With this in mind, the 2021 UNESCO Recommendation on the Ethics of Artificial Intelligence could serve as an inspiration for European guidelines to support the design, development, and implementation of AI tools. Such guidelines would help assess ethical impact, ensure responsible governance and stewardship, and establish appropriate data policy rules, while also encouraging international collaboration. Similarly, in 2024 the WHO highlighted six core principles to guide the ethics and governance of AI in health: protecting autonomy; promoting human well-being, safety, and the public interest; ensuring transparency and explainability; fostering responsibility and accountability; ensuring inclusiveness and equity; and promoting environmentally sustainable and responsive AI.

Real-world cases have demonstrated that AI can perpetuate or even amplify bias and inequity in healthcare delivery, particularly when models are trained on non-representative datasets or deployed without adequate oversight. A widely cited example is the commercial algorithm used in US hospitals to allocate care management resources, which significantly underestimated the health needs of non-White patients compared with White patients. This occurred because

the algorithm used healthcare spending as a proxy for health status, inadvertently encoding systemic disparities in access to care ([Hasanzadeh et al. 2025](#)). A global review published in *PLOS Digital Health* further revealed that most AI models in clinical medicine are trained on datasets from high-income countries, particularly the USA and China, and are authored predominantly by male data scientists rather than clinicians ([Celi et al. 2022](#)). This lack of diversity in both data and development teams limits the generalisability of AI tools and risks reinforcing global health inequities.

To address these challenges, several evidence-based strategies have been proposed. The 2025 review in *npj Digital Medicine* emphasises the importance of embedding bias recognition and mitigation throughout the AI life cycle: from data collection and model design to deployment and post-market surveillance ([Hasanzadeh et al. 2025](#)). Key actions include the following:

- Diversifying datasets to reflect the demographic and clinical diversity of target populations.
- Conducting external validation across different settings and subgroups to assess generalisability.
- Implementing bias audits and fairness metrics as part of regulatory and institutional review processes.
- Engaging interdisciplinary teams, including ethicists, clinicians, and community stakeholders, to ensure that AI tools are contextually appropriate and ethically sound.

These strategies aim to ensure that AI in healthcare promotes equity rather than exacerbating existing disparities. Relevant ethical and social principles include non-discrimination and fairness; transparency and interpretability (notably limited in LLMs); validation across diverse demographic and geographical settings; equity in data representation; and algorithmic accountability. Generally, ethical considerations should also address societal and systemic impacts, not only individual-level rights.

6.1 Do not harm, and respect human dignity

Ethical considerations surrounding AI in healthcare are increasingly critical, particularly about the principles of 'do no harm' and respect for human dignity. A 2025 scoping review published in *BMC Medical Ethics*

emphasises that as AI technologies – ranging from machine learning to large language models – are integrated into real-world healthcare settings, traditional ethical frameworks often fall short in addressing the nuanced and context-specific challenges they pose (Ratti *et al.* 2025). These include risks of algorithmic bias, erosion of patient autonomy, and the depersonalisation of care. Similarly, a 2024 commentary in *Preventing Chronic Disease* emphasises the importance of embedding health equity and ethical responsibility into AI deployment, warning that without deliberate safeguards, AI could exacerbate existing disparities and undermine trust in healthcare systems (Dankwa-Mullan 2024). These publications advocate for a shift from abstract ethical principles to more grounded, context-aware approaches that prioritise transparency, accountability, and the preservation of human dignity in all AI-driven healthcare applications.

6.2 Respect health-data privacy and patient confidentiality

The ethical protection of health data, privacy, and patient confidentiality in AI-driven healthcare is a growing concern as digital technologies become more embedded in clinical practice. A 2024 review in *Applied Sciences* explores the tension between leveraging AI for improved patient outcomes and safeguarding sensitive health information, emphasising the importance of techniques such as Differential Privacy and encryption to maintain confidentiality while enabling data utility (Williamson and Prybutok 2024). The study also highlights the challenges of aligning AI systems with legal frameworks such as the GDPR and cybersecurity legislation, and highlights the need for robust oversight to prevent algorithmic bias and ensure informed consent. Complementing this, a 2025 article in *AI & Society* presents qualitative insights from patients, caregivers, and healthcare professionals involved in AI-based monitoring for Parkinson disease, revealing deep concerns about data ownership, commercialisation, and privacy within home environments (Bavli *et al.* 2025). These findings emphasise the necessity of interdisciplinary governance and patient-centred design to ensure that AI in healthcare respects individual rights and maintains public trust.

6.3 Enforce fairness standards and non-discrimination

The biases in AI models caused by non-representative datasets must be tackled to ensure fair healthcare applications. Points of consideration include the following:

1. Use of different meanings of discrimination: a diagnostic system should discriminate as much as possible between those with and without disease;

and those with higher and lower risk; etc. This is ‘discrimination for the good’. This may also imply discrimination related to income, work, sexual behaviour, genome, and even ethnicity-related risks. Of course, personal data protection should be ensured in all cases.

2. Misuse of data for criminal purposes or other forms of discrimination against individuals or groups (Broeders *et al.* 2017).
3. An independent review commissioned by the UK Government ‘reveals how existing biases and discrimination in society can unwittingly be incorporated at every stage of the life cycle of the devices, and then magnified in algorithm development and machine learning. The evidence for adverse clinical impacts of these biases is currently patchy, though indicative. Seven of (their) recommendations are therefore focused on actions to enable the development of bias-free AI devices, with the voices of the public and patients incorporated throughout.’ (UK Government 2024, p. 3).

Studies have collectively highlighted the importance of designing AI systems with fairness, transparency, and inclusivity to prevent the reinforcement of existing inequities in healthcare delivery. Published examples of bias in healthcare AI include the following:

4. Obermeyer *et al.* (2019) in *Science* analysed a widely used commercial algorithm in US healthcare systems designed to identify patients for high-risk care management. The algorithm used healthcare costs as a proxy for health needs, which led to significant racial bias: non-White patients were considerably sicker than White patients at the same risk score. This resulted in fewer non-White patients being referred for additional care. The bias stemmed from using healthcare costs as a proxy for health needs, because costs reflect systemic disparities in healthcare access and spending. Adjusting the algorithm to use actual health needs instead of cost reduced this bias by 84%.
5. Omar *et al.* (2025a) in *Nature Medicine* evaluated LLMs in clinical decision-making and found that these models exhibited sociodemographic biases. Specifically, LLMs were more likely to provide less accurate or less appropriate medical recommendations for patients from marginalised groups, including racial minorities and individuals with lower socio-economic status. The findings suggest that these biases are embedded in the training data and can influence diagnostic and therapeutic decisions, potentially exacerbating health disparities.

6. [Omiye et al. \(2023\)](#) in *NPJ Digital Medicine* demonstrated that LLMs can propagate race-based medicine—clinical practices that incorrectly use race as a biological determinant. The study found that LLMs reproduced outdated and scientifically unsupported race-based clinical guidelines, such as adjusting kidney function estimates based on race. This perpetuates harmful disparities and emphasises the need for critical oversight in the medical content these models are trained on and generate.

6.4 Foster transparency and interpretability

Transparency and interpretability are fundamental to trustworthy AI in healthcare, enabling clinicians and patients to understand the reasoning behind algorithmic decisions and to detect biases or errors. However, reviews indicate persistent gaps: many healthcare AI systems remain ‘black boxes’, lacking descriptive post-hoc interpretability essential for patient-centric deployment ([Shafik et al. 2026](#)). The EU’s Artificial Intelligence Act (effective August 2024) makes specific provisions for transparency: high-risk medical decision-support systems must implement either ad hoc explainability or post-hoc clarification methods to satisfy ethical scrutiny ([Freyer et al. 2024](#)). Complementing legal mandates, multidisciplinary frameworks advocate a layered transparency approach – incorporating model-agnostic explanations, audit trails, and clear taxonomies – to align AI practice with European values of accountability, equity, and patient autonomy ([Kiseleva et al. 2022](#)).

Transparency should be practical and role-specific. Clinicians, patients, managers, and regulators do not all need the same information, but each needs enough explanation to understand intended use, limitations, uncertainty, accountability and routes for challenge or escalation.

6.5 Validation across diverse demographic and geographical settings

Validation of AI systems across diverse demographic and geographical settings is critical to ensure equity and reliability in EU healthcare. The European Parliament highlights that AI models developed in limited contexts can embed systemic biases – related to sex, ethnicity, age, or region – which risk perpetuating healthcare inequities unless validated in different populations ([Lekadir et al. 2022](#)). Clinical research confirms that external validation – testing prediction models on geographically and demographically distinct cohorts – is essential to assess reproducibility and generalisability ([Ramspek et al. 2021](#)). A recent review in *BMC Medical Informatics and Decision Making* notes that many AI tools, including well-known diagnostic systems, performed poorly when deployed in real-world clinical environments with different equipment, staff expertise,

or patient profiles—emphasising the need for externally validated, calibrated models before rollout ([Pennestri et al. 2025](#)).

A landmark study in COVID-19 mortality prediction employed transfer learning and demonstrated that models validated across racial and ethnic groups yielded significantly improved and fairer performance ([Gu et al. 2025](#)). Similarly, research on foundational language models has shown that pre-trained systems often favour Western or English-speaking populations, performing poorly on content originating from South or Southeast Asia ([Dunn et al. 2024](#)). To address such disparities systematically, the ‘model cards’ framework has been proposed, which mandates reporting of model accuracy broken down by demographic groups – such as race, gender, and region – to promote transparency and accountability ([Mitchell et al. 2019](#)).

Furthermore, EU regulatory frameworks such as the AI Act and the EU Medical Device Regulation increasingly require multi-site clinical evaluation, with high-risk AI (such as in medical diagnostics) subject to thorough external validation as a condition for certification ([Li et al. 2025](#)). To support this, the EMA’s DARWIN EU initiative is building federated real-world data infrastructure across Member States, which can underpin robust validation in diverse European settings ([EU 2025](#)). Collectively, this body of evidence supports a policy approach that requires multi-country, multi-demographic validation to promote safety, fairness, and public trust in AI-based healthcare tools across the EU.

6.6 Architect AI systems with built-in ethical protocols to proactively ensure equitable patient care

Clarification of ethical issues is necessary early in the design of AI tools, as this helps address potential concerns that may arise from their use. Embedded ethics requires collaboration between developers and ethicists to ensure that ethical considerations are integrated throughout the development process. These guidelines should define the ethical frameworks to be applied and provide shared standards for addressing anticipated risks and issues, such as bias, lack of transparency, or data privacy. They should also include provisions for training users to understand potential ethical issues associated with applications of AI tools in health.

Embedded ethics should also be reflected in operational recommendations: ethics expertise should be integrated into project teams, procurement criteria, validation planning, deployment governance, and post-market review, rather than being consulted only at the end of development.

Co-designing AI systems in healthcare with input from the social sciences is increasingly recognised as essential for advancing health equity. A 2024 article in *Frontiers in Digital Health* argues that AI has the potential to address complex health disparities by integrating insights from social determinants of health – such as race, socio-economic status, and environment – into its design and deployment (Green *et al.* 2024). However, realising the potential of ethics by design requires interdisciplinary collaboration, particularly with social scientists who can help ensure that AI tools are contextually aware and ethically grounded. A 2023 scoping review protocol published in *BMJ Open* further emphasises the importance of embedding equity, diversity, and inclusion (EDI) principles throughout the AI life cycle, noting that failure to do so risks reinforcing existing inequities and marginalising underrepresented populations (Nyariro *et al.* 2023). These findings emphasise that co-design with social scientists is not merely a methodological preference but a critical strategy to ensure that AI in healthcare is inclusive, just, and responsive to the needs of diverse communities.

The concept of multiscale embedded ethics is gaining traction as a vital framework for ensuring health equity in the development and deployment of AI in healthcare. A 2024 study published in *Science and Engineering Ethics* presents a comprehensive toolbox for integrating ethical and social science analysis into healthcare AI research, emphasising that ethical reflection must occur at every stage of technology development—from design to implementation (Willem *et al.* 2025). This approach, known as embedded ethics, involves interdisciplinary collaboration with social scientists, ethicists, and affected communities to proactively identify and address issues such as algorithmic bias, misrepresentation, and inequitable outcomes. The study outlines practical methods including stakeholder analysis, ethnographic research, and focus groups to ensure that ethical concerns are not treated as afterthoughts but are structurally embedded into AI innovation. Complementing this, the 2024 commentary in *Preventing Chronic Disease* emphasises the importance of applying an equity lens to AI deployment, warning that without deliberate ethical integration, these technologies risk reinforcing existing disparities (Dankwa-Mullan 2024). Together, these publications advocate for a multiscale, participatory approach to ethics that aligns technological advancement with social justice and inclusive healthcare.

6.7 Recommendations

Recommendation 1. Establish a common risk-tier framework for health AI

Action: develop a harmonised risk-tier framework for healthcare AI, including multimodal and general-purpose models, that applies proportionate

regulatory requirements according to system risk and level of autonomy. Ethical considerations should be embedded from the earliest design stages to ensure that safety, fairness, and accountability are integral to system development.

Rationale: risk-based regulation enables innovation while ensuring that higher-risk applications receive stronger oversight and safeguards, particularly in clinical decision-making contexts.

Alignment with overall recommendations: R2 (guidelines for responsible deployment), R3 (ethical frameworks), R10 (governance and oversight).

Recommendation 2. Mandate transparency in AI use and data provenance

Action: require healthcare organisations and AI developers to provide clear and accessible disclosure of AI use, including the intended purpose of systems, underlying data sources, and decision boundaries or limitations. Transparency requirements should extend to both developers and healthcare providers deploying AI systems.

Rationale: transparency improves accountability, supports clinical decision-making, and strengthens public trust in the use of AI in healthcare.

Alignment with overall recommendations: R1 (evaluation of AI value), R8 (stakeholder engagement and trust), R10 (governance and oversight).

Recommendation 3. Ensure human oversight, accountability, and workforce preparedness

Action: require human oversight mechanisms and clearly defined accountability structures for AI-assisted clinical decisions. Healthcare systems should also ensure that medical education and clinician training programmes prepare healthcare professionals to understand AI tools and integrate them safely into clinical workflows.

Rationale: maintaining meaningful human oversight helps prevent over-reliance on automated systems and ensures that clinicians remain responsible for patient care decisions.

Alignment with overall recommendations: R4 (implementation challenges), R7 (training and education), R10 (governance and oversight).

Recommendation 4. Enforce robust data governance standards

Action: implement strict data governance frameworks that ensure informed consent, privacy protection, data minimisation, access controls, and auditable data use for AI development and deployment in healthcare. These

frameworks should align with existing European data protection regulations and emerging AI governance standards.

Rationale: strong data governance is essential to protect patient rights, ensure compliance with legal frameworks, and maintain trust in digital health systems.

Alignment with overall recommendations: R6 (data sharing ecosystems), R8 (public trust), R10 (governance and oversight).

Recommendation 5. Establish continuous post-market monitoring and bias oversight

Action: implement ongoing post-market surveillance systems for healthcare AI, including structured incident reporting, bias monitoring, and mechanisms for rapid remediation when performance or safety concerns emerge.

Rationale: AI systems may evolve or degrade after deployment owing to changing clinical contexts, data distributions, or system updates; continuous monitoring is therefore essential to maintain safety and effectiveness.

Alignment with overall recommendations: R1 (evaluation of AI value), R5 (learning health systems), R10 (governance and oversight).

Recommendation 6. Invest in capacity-building and international collaboration

Action: support capacity-building initiatives and international cooperation to ensure that AI governance frameworks, technical expertise, and regulatory capabilities are developed across healthcare systems globally, promoting equitable access to safe and trustworthy AI technologies.

Rationale: AI governance challenges are global in nature; collaboration and shared learning help reduce inequalities in access, strengthen regulatory capacity, and improve collective oversight of rapidly evolving technologies.

Alignment with overall recommendations: R7 (training and education), R8 (stakeholder engagement), R9 (research priorities).

7 Artificial intelligence solutions in healthcare require engagement of key stakeholders

Policy-makers would benefit from understanding that the development, internal and external validation, and cost-effectiveness assessments of artificial intelligence (AI) solutions in healthcare require active engagement from diverse stakeholder groups to ensure these systems are ethical, effective, equitable, reliable, cybersecure and aligned with real-world needs. Key stakeholders include patients, citizens, healthcare workers, healthcare leaders, health policy-makers, regulators, manufacturers, community organisations, payers and insurers, and academic and research institutions. Below, we outline the roles, key questions for engagement, and methods to incorporate their perspectives into the successful deployment of AI solutions in a clinical setting.

7.1 Definition of stakeholder groups

Patients: individuals who use healthcare services and are directly affected by AI solutions.

Citizens: the wider public, whose indirect but significant voice influences the establishment of broader trust in healthcare AI solutions.

Healthcare workers: clinicians (doctors, nurses, etc.), allied health professionals, and administrative staff who interact with AI solutions in clinical settings.

Healthcare leaders: individuals or groups in senior decision-making roles within healthcare organisations, focusing on the strategic direction, adoption, and implementation of AI technologies.

Health policy-makers: government officials and politicians shaping healthcare policies and AI integration.

Regulators: agencies, including third-party organisations such as notified bodies, which are responsible for approving and monitoring AI solutions to ensure safety, efficacy, and compliance.

Manufacturers: companies and developers designing and producing AI healthcare technologies.

Community organisations: organisations representing underserved or marginalised populations and working to ensure equitable AI deployment.

Payers and insurers: entities funding healthcare services, influencing AI adoption through reimbursement policies.

Academic and research institutions: entities both developing AI solutions and conducting studies on AI's impact and effectiveness in healthcare.

7.2 Key questions for engaging and involving stakeholder groups

A robust European AI-in-healthcare ecosystem requires proactive engagement of various stakeholder groups. This translates into, for example, mobilising patients, healthcare professionals, and technology companies in both the development and deployment phases of AI solutions. Ethical-technology frameworks highlight that early engagement across diverse stakeholder groups supports shared responsibility, cultivates trust, and aligns innovations with real-world clinical needs ([Subramanian et al. 2024](#)). It cannot be overstated how crucial it is that AI developers embed explainability, accountability, and fairness during the design stage, while healthcare workers and patients should validate AI tools in clinical practice, guiding their integration and monitoring unintended effects. This kind of development requires a complex ecosystem of intertwined responsibilities across technical, clinical, and regulatory domains—echoing calls for co-responsibility frameworks and structured accountability regimes, as recommended in EU research reviews and technological ethics agendas ([Trocin et al. 2023](#)). Embedding stakeholder engagement throughout the AI life cycle ensures that Europe's AI-powered health solutions remain ethically sound, clinically effective, and socially legitimate.

To enable policy-makers at different levels to participate in decision-making about AI regulation and its implementation, they should be aware of the critical questions and their responses from the perspective of various stakeholder groups. From these groups' perspectives, these same questions are also essential for understanding and defining each stakeholder's role, for instance in co-development. Addressing these questions and discussing them among different groups fosters stakeholder engagement in achieving a common goal.

Patients

What are the unmet needs of patients? What are the priority health conditions the AI solution should address? How can the AI solution be made accessible and understandable to patients with diverse literacy levels? What are the primary threats and fears experienced by patients about the concept of using AI solutions in treatment pathways?

Citizens

How can public trust in the fairness and reliability of the AI solution be enhanced? Does it matter to citizens whether the developer and manufacturer of the AI solution is a public or private organisation? Are citizens

confident that the AI solution protects their data? What additional measures should the adopter of the AI solution take, in the view of citizens, to increase trust in such applications? Do citizens feel that public trust could be strengthened through education and transparency?

Healthcare workers

What is the clinical problem the AI solution aims to solve? How should the AI solution be integrated into every clinical workflow without increasing burden? Do healthcare workers find the proposed AI solution reliable and clinically relevant? How does the AI solution affect healthcare workers' workload, job satisfaction, efficiency, and patient interactions? What training or support do healthcare workers need to adopt the AI solution effectively?

Healthcare leaders

Have the suppliers ensured that the AI solution's performance is validated properly against local real-world data, including checks for biases and generalisability? What mechanisms can be used to monitor staff well-being and address any workflow challenges during clinical use? What continuous monitoring metrics, beyond technical accuracy, have been implemented for evaluating the AI solution's impact on outcomes and cost-efficiency? What feedback loops would help to monitor the AI solution's performance and staff engagement post-implementation? How have leaders ensured that the organisation possesses sufficient in-house expertise and understanding of AI procurement and its clinical application? If the AI solution frees up human resources for other purposes, how will these resources be utilised?

Health policy-makers

What policies are needed to guide ethical AI development in healthcare? Is the AI solution aligned with regional, national, and international legislation? How do AI systems impact healthcare costs, access, and quality at a system level, according to the policy-makers? How can policy-makers reliably balance innovation with patient safety and equity? How should ethical, regulatory and cybersecurity concerns be continually assessed, and how should policies be revised in response?

Regulators

What standards should the AI solution meet for cybersecurity, safety, efficacy, and fairness? Are validation processes robust enough to ensure compliance with regulatory requirements and high clinical accuracy? How do regulators monitor the AI solution post-deployment for ongoing safety and performance? How can regulators collaborate with developers to streamline approvals without compromising rigour?

Manufacturers

How can manufacturers incorporate end-users' feedback into the AI solution design? Are manufacturers providing transparent data on the AI solution's performance and limitations? How do manufacturers address feedback from real-world AI use to continuously improve solutions? Are manufacturers aligning with clinical and regulatory requirements from the very beginning of the solution development? Are manufacturers aware of the timelines, stakeholders, and costs associated with regulatory processes?

Community organisations

How can the AI solution address the needs of marginalised or underserved communities? Is the AI solution tested for fairness across diverse populations? Does the AI solution reduce or exacerbate health disparities? How can these groups be empowered to participate in the AI solution development?

Payers and insurers

How can the AI solution demonstrate cost-effectiveness and value for reimbursement? Does the AI solution meet payer requirements for evidence-based outcomes? How does the AI solution impact healthcare costs and coverage decisions? How can payers align reimbursement policies with AI innovation?

Academic and research institutions

What research is needed to inform evidence-based AI development? How can academic studies validate AI systems' clinical and societal impact? What long-term studies are needed to assess AI's effectiveness and safety? How can academic expertise inform practical AI deployment?

Academic researchers and research institutions, including hospitals, must recognise a fundamental difference in AI research compared with most conventional clinical and preclinical studies. When the objective of the research is to develop a medical AI device intended for clinical use, the product will require conformity assessment and CE marking under the Medical Device Regulation (MDR), as well as an established manufacturer and a placing-on-the-market organisation. Not all European research institutions possess the capability and the ISO-compliant quality management systems necessary to act as a manufacturer of AI-based medical devices. Even when such quality systems are in place within research institutions, the issues of manufacturer liability and the ability to function as a commercial placing-on-the-market entity on the open market very rapidly become major obstacles. Similarly, few publicly funded healthcare organisations are realistically able to operate in this commercial environment. If these

regulatory and commercial realities are not taken into account already in the very early phases of AI medical device development, even an excellent AI solution is extremely unlikely to ever reach routine clinical use. For this reason, it is essential for academic researchers and research institutions to establish collaboration with commercial partners capable of acting as the legal manufacturer of AI-based medical devices right from the earliest stages of development.

In short, strong cross-sector collaboration between stakeholder groups throughout and right from the earliest stages is key to the quality of innovation and, ultimately, to the development of clinically meaningful applications. According to the European Patients' Forum (a European patient umbrella organisation), 'patient engagement throughout the life cycle of AI solutions could ensure a realistic assessment of how they can enhance patient wellbeing, improve accessibility, and serve patients' needs. Moreover, the involvement of patients could result in an increased understanding of digital health solutions, which in turn could positively affect patients' acceptance of and trust in such tools.' (Data Saves Lives 2022).

Patient organisations and professional users should have defined roles in needs assessment, co-design, validation priorities, procurement decisions, user-facing communication, governance, and post-deployment monitoring. This moves stakeholder engagement from consultation to shared responsibility for safe and useful implementation.

- Healthcare professionals. The Standing Committee of European Doctors (CPME) formulated recommendations in 2024 (CPME 2024).
- Consumers. The European Consumer Organisation (BEUC) developed some early policy recommendations in 2020 (BEUC 2020).

7.3 Methods for incorporating views of stakeholder groups

Policy-makers at different levels would probably benefit from employing various methods to deepen their understanding of the effective use of AI solutions in healthcare. These methods can enable policy-makers to obtain valuable additional insights from different stakeholder groups. To ensure that stakeholder groups' perspectives are integrated into the backbone of AI policies, regulation and implementation, the following methods could be employed from the early stages of policy-making and development processes.

1. Consultations. Conduct one-on-one or small-group interviews with representatives from each stakeholder group to gather in-depth insights. For example, interview patients from diverse

backgrounds to understand their concerns about AI in healthcare.

2. Workshops. Organise multi-stakeholder workshops to facilitate dialogue. For example, host a workshop with ground-level healthcare workers, patients, and manufacturers.
3. Bilateral meetings. Arrange targeted meetings between specific groups (e.g. regulators and manufacturers) to address technical or policy-specific issues. For example, meet with regulators to discuss validation standards for AI algorithms.
4. Surveys and questionnaires. Distribute surveys to large stakeholder groups such as citizens or healthcare workers to collect quantitative and qualitative feedback. For example, survey citizens on their trust in AI-driven diagnostics.
5. Public comment periods. Consider opening a public comment website to gather broad input from citizens and other stakeholder groups on proposed AI systems or policies. For example, solicit public feedback to draft guidelines for AI transparency.
6. Advisory panels. Establish a multi-stakeholder advisory panel with representatives from each stakeholder group to provide ongoing input throughout the AI (policy) implementation life cycle. For example, create an advisory panel with patients, ethicists, and healthcare professionals to guide AI policies.
7. Participatory pilot sessions. Create environments and opportunities for end-users to test AI solutions before wider deployment. For example, engage healthcare workers in a few hospitals in agile testing of AI solutions.
8. Public engagement campaigns. Launch campaigns to educate and involve citizens in discussions about AI in healthcare. For example, use active social media platforms to gather public opinions on AI.

7.4 Foster Europe's unique model of AI and public health convergence

AI is a technology advancing at a rapid pace, while public health is diverse and itself comprises several subsectors. The convergence of AI and public health sectors in Europe involves integrating AI technologies into healthcare systems to enhance efficiency, diagnostics, treatment, and preventive measures. This integration is driven by EU-wide initiatives aiming to position Europe as a leader in trustworthy AI while addressing unique challenges in a regulated environment.

Europe's approach emphasises perhaps the strictest regulations to ensure AI systems in public health are safe and effective. For instance, AI systems for assessing health risks or allocating resources in public health fall under high-risk categories, mandating risk management and human oversight. Moreover, AI in public health must avoid perpetuating inequalities, such as biased algorithms in epidemiological modelling or resource allocation. In Europe, accountability is emphasised, ensuring that AI assists rather than replaces human decision-making in clinical settings. These policies, together with stringent regulatory rules, may create compliance hurdles particularly for startups, potentially lagging global competitors.

In Europe, all citizens should be able to access electronic health records by 2030 under the Digital Decade Policy Programme 2030 (<https://digital-strategy.ec.europa.eu/en/policies/artificial-intelligence-health>). This policy programme positions Europe as a potential leader in global AI-public health convergence. This convergence can accelerate preventive care, personalise treatments, and optimise resource allocation, potentially lowering healthcare costs through AI efficiencies while enhancing global market competitiveness for European tech firms and startups.

Universal access to electronic health records (EHRs) provides a foundation for AI by aggregating anonymised, high-volume data across 27 Member States, supporting federated learning models that preserve privacy while enabling cross-border AI applications. This creates a 'data union' effect, where AI can mine EHRs for insights without centralising sensitive information, outpacing regions with siloed data. For instance, universal EHR systems enable real-time public health surveillance, as seen in AI tools processing EHRs for epidemiological forecasting during outbreaks. Moreover, AI can analyse EHR trends to generate automated, personalised follow-up protocols, reducing no-show rates and improving adherence. Regarding chronic diseases, AI could be exploited to predict chronic disease exacerbations (e.g. diabetes). Similarly, continuous monitoring initiatives benefit from EHR integration with wearables and IoT devices, where AI can process real-time data for anomaly detection. This convergence allows population-level surveillance, such as AI mining aggregated EHRs to track public health trends (e.g. flu outbreaks), providing Europe with stronger crisis-response capabilities than those of less digitised regions.

In contrast to, for example, China and the USA, Europe's multilingual landscape can be addressed through AI-powered translation embedded in EHR platforms, ensuring equitable access. This offers a competitive edge by enabling seamless data sharing in diverse populations, reducing errors in multicultural healthcare

settings, and fostering AI innovations in natural language processing tailored to European languages, unlike monolingual-dominant systems elsewhere.

Most importantly, patient empowerment tools can emerge from EHR access, with AI enabling self-management tools that provide actionable insights. This creates opportunities for new tools, such as virtual health coaches relying on EHR trends and wearables, giving Europeans an advantage in consumer health tech markets through privacy-focused, AI-enhanced platforms that comply with stringent regulations.

The convergence of AI and public health sectors in Europe, while promising, faces significant hurdles that could impede straightforward integration. Constrained public sector budgets and a shortage of skilled personnel are major barriers to developing groundbreaking AI solutions within healthcare. Moreover, a shortage of AI specialists and healthcare professionals trained in digital tools hampers development and deployment. Similarly, the healthcare sector's limited expertise in developing interoperable, user-centric digital platforms hinders the creation of Europe-wide AI solutions for citizens. Inconsistent data quality and lack of interoperability within and across Member States' EHR systems hinder AI model accuracy and scalability. Regulatory complexity, data interoperability issues, and public trust concerns further complicate progress. These facts weaken Europe's competitive edge in citizen-focused health tech.

Owing to the limited resources, experience, and expertise at the disposal of mid-sized and particularly small businesses (i.e. startup companies), the European Union could procure needed AI services for public health convergence from three or more European big tech companies through structured public procurement processes. Such procurement processes should be designed to promote competition and mitigate vendor lock-in risks. This approach aligns with EU strategies to address challenges such as limited funding, human resources, and know-how by outsourcing development of Europe-wide platforms, while leveraging the Data Act (effective September 2025) to enforce easier switching between providers, eliminating data transfer fees and reducing dependency on a single vendor. By fostering a multi-vendor ecosystem, it probably enhances innovation and resilience in public health AI. Moreover, this multi-vendor strategy is encouraged in the EU's cloud outsourcing guidelines, promoting diversification to avoid high switching costs and enhance AI innovation in healthcare. However, risks include potential data privacy breaches and over-reliance on non-EU big tech. This aspect has become exceptionally pronounced in 2025 amid escalating world political turmoil, particularly involving the USA. With US firms such as Amazon, Microsoft, and Google dominating more than two-thirds

of the EU cloud market, any US policy shifts – such as retaliatory measures or abrupt service changes – could halt European initiatives. Moreover, public trust in US big tech companies is notably low in Europe in 2025, driven by geopolitical tensions, data privacy scandals, and concerns over digital sovereignty, which could lead to strong public opposition to AI-public health convergence initiatives reliant on these firms. Overall, while feasible and strategically beneficial, success depends on rigorous tender processes and enforcement of anti-lock-in measures to fully address public health convergence challenges.

In summary, Europe's unique model of AI and public health convergence leverages universal EHR access by 2030 under the Digital Decade, creating a standardised, privacy-secure data ecosystem through the EHDS and GDPR. This enables scalable AI applications such as real-time surveillance, personalised follow-ups, and multilingual EHR tools, fostering innovation in a regulated environment. The EU AI Act helps ensure that high-risk AI systems are developed and used ethically, prioritising trust and equity, giving Europe a competitive edge in global health tech through robust data infrastructure and cross-border interoperability.

7.5 Develop clinically meaningful AI solutions

When developing clinically impactful AI innovations for European healthcare, all stakeholder groups should be considered from the outset of the development process. In doing so, the same questions outlined in [section 7.2](#) 'Key questions for engaging and involving stakeholder groups' should be posed to stakeholder groups.

The main end-users of clinical AI solutions and devices in Europe are healthcare professionals and patients. Clinical AI solutions in Europe are designed to enhance healthcare delivery and patient outcomes, primarily as assistive tools under strict regulatory frameworks such as the EU AI Act and Medical Device Regulation (MDR). For healthcare professionals, clinical AI's main use cases include diagnostics and decision support, predictive analytics, and workflow optimisation (for example triage tools), enhancing accuracy and efficiency while maintaining clinician oversight. For patients, clinical AI supports self-management and even self-diagnosis, empowering proactive health management.

From the viewpoint of healthcare professionals, clinically meaningful AI solutions are those that augment clinical practice by improving patient outcomes, enhancing diagnostic accuracy, and streamlining workflows so that clinicians stay in the loop. In Europe, clinical AI solutions must be evidence-based, demonstrating real-world efficacy through rigorous clinical validation, such as real-world data studies, to ensure they provide safer and more standardised care. For instance, AI solutions that detect critical diseases in imaging studies are often seen

as meaningful, particularly if they reduce diagnostic delays and errors. Such clinical AI solutions improve patient outcomes. In terms of streamlining workflows, particularly AI triage solutions, designed to prioritise urgent cases by flagging abnormalities in imaging studies, have become widely used in radiology. However, AI triage solutions in radiology raise concerns about low or no sensitivity to a range of emergency findings, which can often lead to disruptions of non-AI-assisted standard workflows. In more detail, AI triage solutions often focus only on a few specific conditions, missing multiple other critical and urgent findings. When AI triage solutions miss non-targeted emergencies, these cases are not flagged as urgent, and they remain in the standard workflow queue. In other words, they are not automatically prioritised to the top of the radiologist's worklist. In busy European hospitals, where radiologists review hundreds of scans daily, these unflagged emergency cases may wait hours or even days, potentially delaying critical diagnoses and treatment. Therefore, AI triage solutions for medical emergencies, in particular, require robust training and auditing to ensure clinical safety in Europe's diverse healthcare settings.

Developing a 100% sensitive AI solution for clinical diagnostics – meaning one that never misses a true positive case – is fundamentally unattainable and impractical owing to a combination of technical, regulatory, and liability factors. From the technical point of view, for example, human biology and medical data are inherently 'messy' and variable, making 100% sensitivity impossible. From the regulatory viewpoint, achieving 100% sensitivity would require exhaustive testing across infinite scenarios, which is logistically unfeasible. Finally, a 100% sensitive AI would probably need to operate as a standalone system (independent of human input), shifting full liability to the manufacturer for any errors, thus posing enormous legal and financial risks. All these aspects emphasise the role of AI solutions as assistive tools, rather than replacements, to uphold clinician accountability and patient trust.

Recently, concerns have emerged highlighting that the use of AI in clinical practice may, over time, lead to a decline in clinicians' expertise ([Berzin and Topol 2025](#)). Berzin and Topol wrote that 'concern is mounting that off-loading clinical tasks and reasoning will lead to loss of skills (deskilling), adopting errors or bias from AI (mis-skilling), or failure to achieve competence (never-skilling).' When AI is used to help doctors to diagnose and detect various diseases, doctors may become too reliant on AI. For example, a study of gastroenterologists found a 20% decline in their ability to independently detect polyps and abnormalities during colonoscopies after becoming perhaps too accustomed to AI assistance ([Budzyn et al. 2025](#)). AI should serve as a supportive tool to enhance clinical decision-making,

not replace human expertise at any phase. To ensure clinicians maintain their expertise, it may be prudent to prioritise AI solutions that validate or challenge a clinician's initial diagnosis. For instance, in diagnostic imaging, it may be advantageous in certain situations to favour AI solutions that can be utilised only after a radiologist's initial image interpretation. By operating in this manner, the likelihood of clinical deskilling for radiologists is reduced, while diagnostic accuracy may improve. In any case, even when AI automates certain steps of the clinical work, continuous training and hands-on practice are essential to prevent the erosion of critical clinical skills (deskilling).

AI solutions designed to predict diseases, conditions, or mortality far into the future hold significant potential but also face substantial challenges in becoming clinically significant. These solutions often aim to leverage vast datasets – genetic, clinical, lifestyle, and environmental factors – to forecast health outcomes. While they can offer some insights for preventive medicine and personalised care, their accuracy and practical utility are constrained by several factors, particularly when considering competing risks of diseases and death. Competing risks introduce uncertainty, as they alter the probability of the predicted event occurring. Moreover, such AI solutions rely on historical data, which may not fully capture future variables such as new treatments, environmental changes, or lifestyle shifts. Long-term predictions are inherently less reliable owing to the dynamic nature of health determinants. Predicting serious conditions or death far in the future raises ethical questions about psychological impact, informed consent, and potential misuse (e.g. by insurers or employers). Overly deterministic predictions could also cause anxiety or fatalism. Most importantly, validating long-term predictions is challenging, as it requires decades of follow-up data. Therefore, most solutions are validated on shorter-term outcomes, making their long-term accuracy highly speculative. AI predictions may be more reliable and actionable for shorter timeframes (months instead of years), where interventions can be planned and competing risks are less dominant.

AI solutions for patients are increasingly available through mobile apps, wearables, and web platforms, focusing on empowering self-management (e.g. tracking and adherence), self-diagnosis, and proactive health management (e.g. predictive alerts and personalised insights). These tools are primarily designed for direct consumer use. The AI solutions for patients (e.g. symptom checkers, chronic disease trackers) vary in their regulatory status under the MDR. The manufacturer's stated intended purpose determines whether the MDR applies. If an app claims to diagnose, monitor, or treat medical conditions, it is probably a medical device. Wellness-focused apps avoid this by emphasising lifestyle over medical outcomes.

MDR-compliant tools integrate better with healthcare systems, providing more reliable data for shared decision-making. Patients should verify CE marking for clinical reliability and consult healthcare professionals to ensure safe use, especially for diagnostic or treatment decisions.

In summary, developing clinically meaningful AI solutions for healthcare professionals and patients in Europe requires aligning with clinical needs, regulatory requirements (e.g. EU AI Act, Medical Device Regulation (MDR)), and ethical principles to enhance healthcare delivery and patient outcomes. These solutions are primarily assistive tools, designed to support decision-making, improve efficiency, and empower patients while maintaining clinician accountability and patient trust. Engagement with end-users (doctors, patients) during design ensures usability and relevance. AI should augment, not replace, human expertise. For doctors, this means tools that validate or refine diagnoses (e.g. AI image analysis performed after the radiologist's initial interpretation). Finally, AI models need regular recalibration to reflect new medical knowledge, population changes, and user feedback.

7.6 Understand learning healthcare systems: from private innovation to public altruism

A learning health system (LHS) refers to a healthcare framework where knowledge generation is seamlessly integrated into everyday clinical practice. In an optimal scenario, this creates continuous cycles of data collection, analysis, evidence-based improvements, and innovation to enhance patient outcomes, efficiency, and equity. LHS emphasises embedding research and learning directly into care delivery, rather than treating them as separate activities. In essence, it transforms healthcare from a static model into a dynamic, adaptive one that 'learns' from real-world data and experiences. The EU envisions learning health systems as a means of addressing systemic challenges such as ageing populations, chronic disease burdens, workforce shortages, and rising costs.

AI is a cornerstone of operationalising LHS, acting as the 'engine' for turning vast health data into actionable insights. AI enables continuous data collection from routine care, analysis for patterns, generation of new knowledge (e.g. predictive models), and translation of that knowledge back into practice. In more detail, AI can automate daily or even hourly data mining from EHRs and other population data, accelerating research translation. However, continuously learning AI systems also require ongoing monitoring for accuracy, reliability, and biases.

EU legislation, particularly the EU AI Act (Regulation (EU) 2024/1689) and the Medical Device Regulation

(MDR, Regulation (EU) 2017/745) along with the In Vitro Diagnostic Medical Device Regulation (IVDR, Regulation (EU) 2017/746), explicitly contemplates and permits continuous learning (or adaptive) AI systems also for clinical medicine, provided they undergo rigorous pre-market conformity assessments and ongoing post-market monitoring. These systems are not required to be ‘frozen’ (i.e. static and non-updating) upon deployment. Instead, the framework emphasises controlled adaptability to maintain best performance and compliance, recognising that true continuous learning can enhance real-world utility. However, such systems introduce risks such as performance drift, bias amplification, or unintended outputs. When continuous learning occurs within predefined boundaries that do not alter the device’s intended purpose or risk profile without reassessment, the EU does approve such systems.

When using LHS, providers must implement a proportionate system to collect, analyse, and act on real-world data (e.g. EHR data from hospitals) throughout the system’s life cycle. This includes logging operations for traceability and evaluating ongoing compliance with accuracy, robustness, and bias-mitigation standards. Moreover, continuous risk assessment, including bias testing and human-in-the-loop interventions to override AI decisions, must be performed. Systems must be transparent (e.g. explainable outputs) and cyber-resilient. In summary, EU legislation approves continuously learning clinical AI through a life-cycle approach, mandating monitoring to safeguard accuracy, reliability, and fairness.

How can proprietary advancements scale to serve broader public health goals? For example, a private health tech company can invest heavily in research and development, using proprietary datasets from partnered hospitals and complying with MDR for CE marking as a Class IIb medical device. The AI operates within a continuously learning health system, adapting to new data under a Predetermined Change Control Plan (PCCP), as per AI Act requirements. LHS is integrated into national disease screening programmes across EU Member States. It becomes accessible in public hospitals, reducing disparities in disease detection for underserved populations, such as low-income communities. Governments collaborate with the company to subsidise access. The company releases non-proprietary components (e.g. bias detection frameworks) to the open-source community, fostering trust and enabling smaller developers to build complementary tools.

In another example, a startup can create LHS for managing chronic diseases such as diabetes. The system uses machine learning to analyse wearable

device data (e.g. glucose monitors) and EHRs, providing personalised treatment recommendations (e.g. insulin adjustments, lifestyle changes). It is marketed to private clinics or insurers. Built with proprietary algorithms, the LHS complies with GDPR for data privacy and AI Act for transparency (e.g. explainable recommendations). Continuous learning refines predictions based on patient outcomes, with post-market monitoring per MDR Annex IX. The LHS is adopted by public health systems to manage diabetes at scale, reducing hospital admissions and empowering patients in underserved areas with free or subsidised access.

LHS within the EU should embed patients, citizens, and healthcare workers at its core—leveraging frameworks such as the 2025 European Health Data Space (EHDS) and the EU AI Act to align private innovation with public altruism. Under EHDS Regulation (EU) 2025/327, citizens gain rights to access, contribute to, and control their digital health records, while national Health Data Access Bodies enable secondary uses of anonymised data in the public interest (EFPIA 2024). To foster trust, the system must adopt dynamic consent models – allowing individuals to adjust their preferences over time – and shared decision-making tools that empower users in clinician-led AI-supported care (Kaye *et al.* 2015). Co-design initiatives, such as those promoted by the European Patients’ Forum, have been shown to improve the usability, transparency, and ethical alignment of AI tools (Banerjee *et al.* 2022). Simultaneously, workforce initiatives such as the EU-funded Susa project train clinicians in AI ethics, regulation, and digital skills from the outset—ensuring that technology remains human-centred (Cookson 2025). Finally, international consensus frameworks such as FUTURE-AI reinforce the importance of fairness, explainability, and continuous feedback throughout the AI life cycle (Lekadir *et al.* 2025). By situating LHS governance within these overlapping policies and standards, Europe can transform individual data contributions and professional engagement into a collective asset—advancing both innovation and equitable care.

Secondary use of healthcare data legislation in the EU facilitates data handover to private companies for beneficial purposes, with robust safeguards to prevent misuse. This supports transitions from private innovation to public altruism in LHS, as discussed previously. A European LHS must ensure that patient data are treated as a shared societal asset, prioritising collective benefit over individual gain. Under the EU’s Data Governance Act (DGA) and EHDS, data altruism also enables voluntary sharing of health data for public-interest research – without monetary reward – through authorised, not-for-profit intermediaries (Lalova-Spinks *et al.* 2023). Ethical debates increasingly emphasise data ownership and benefit-sharing, with frameworks such as the Collaborative Healthcare Data Ownership

(CHDO) model advocating shared stewardship reflecting patients', clinicians', and researchers' collective interests (Liu and Guo 2024). Although individual incentives (e.g. personalised feedback) may be limited or infeasible, citizens' contributions must yield societal returns – such as improved public health outcomes, enhanced AI diagnostics, or more equitable care delivery – which aligns with widespread public expectations for fairness. To realise this, support for data altruism must be matched by transparent governance and benefit-sharing frameworks – including democratic data cooperatives, secure processing environments, and clear consent mechanisms – to foster trust, accountability, and legitimacy (TEHDAS 2023a). These mechanisms ensure that individual data contributions are transformed into communal value, reinforcing the EU's commitment to data solidarity, equity, and innovation for the public good.

7.7 Healthcare organisations should partner with companies for medical device manufacturing

The imbalance between hospitals holding vast amounts of healthcare data and private companies possessing the technical expertise and resources to develop commercial AI solutions is a critical issue. Clinical AI solutions (e.g. cancer detection, sepsis prediction) often start as private innovations but need hospital data to scale altruistically across public systems. The imbalance complicates this dialogue and collaboration. This dynamic shapes the development, deployment, and equitable scaling of AI in healthcare, impacting innovation, access, and ethical considerations. Hospitals and public health systems in the EU hold extensive, high-quality health data (e.g. electronic health records, imaging, genomic data), critical for training robust AI models for diagnostics, predictive analytics, or personalised medicine. However, data are often siloed because of privacy concerns, interoperability issues, or institutional reluctance to share with private entities. Private companies (e.g. health tech startups and tech giants) have the computational infrastructure, AI expertise, commercialisation expertise and funding to develop sophisticated models but lack direct access to diverse, real-world datasets. Without access to hospital data, companies rely on limited or synthetic datasets, reducing model generalisability and increasing bias risks. Companies must invest heavily to negotiate data access agreements or build synthetic data pipelines, raising costs that may limit startup participation and favour large corporations.

The imbalance between hospitals' data and companies' know-how/resources creates bottlenecks but also drives innovation through collaboration. Additionally, companies often lack the clinical context needed to address relevant healthcare challenges, resulting in less effective AI models unless healthcare professionals are actively involved as key stakeholders from the start of development. Most importantly, the regulatory,

practical, and resource constraints in the EU make it highly challenging for public hospitals to take on the role of a medical device manufacturer compared with private companies. In other words, while public hospitals can technically act as manufacturers of clinical AI tools under EU law, it is rare and impractical owing to regulatory, resource, and operational barriers. Private companies are far better positioned to fulfil this role, making them the default manufacturers in practice.

University hospitals in particular, which are often at the forefront of research and innovation, frequently encounter a frustrating 'innovation trap'. This challenge arises when university hospitals conduct excellent and cutting-edge research on AI tools (e.g. diagnostic algorithms for rare diseases) but face insurmountable barriers to distributing these solutions to other institutions or scaling them for broader use. This issue is particularly acute in the EU, where university hospitals generate high-quality AI solutions and large, validated datasets from consecutive clinical cases but lack the infrastructure and resources to commercialise and disseminate their innovations. This creates a 'valley of death' between research innovations and deployable clinical tools. At best, university hospitals can develop AI tools for in-house use under MDR Article 5(5) exemptions, but owing, for example, to a lack of funding, legal teams and marketing resources, and to management-system, IP-licensing and cybersecurity-compliance requirements, in-house AI tools cannot be deployed and distributed outside the university hospital. This further discourages investment in academic AI research and development, exacerbating the hospital-company imbalance, as university hospitals innovate but struggle to scale solutions for deployment, hindering EU-wide AI progress. To overcome this, university hospitals could partner with companies for manufacturing. Without public-private hybrids, university hospitals typically turn into isolated innovators, hindering the altruistic scaling of AI from research to routine care.

7.8 Recommendations

Recommendation 1. Prioritise AI solutions addressing high societal disease burden (patients)

Action: Policy-makers should prioritise the development, procurement, and reimbursement of AI solutions targeting conditions associated with high disease burden and substantial long-term healthcare costs, ensuring that innovation focuses on areas with the greatest potential public health impact.

Rationale: Aligning AI development priorities with population health needs helps ensure that innovation delivers meaningful patient and societal benefits, rather than primarily commercial advantages.

Alignment with overall recommendations: R1 (evaluation of AI value), R9 (strategic research priorities).

Recommendation 2. Improve transparency and quality assurance for citizen-facing AI tools (citizens)

Action: National authorities and relevant organisations should curate and publish independently evaluated lists of high-quality AI tools for health-related use by citizens, particularly for applications not classified as medical devices and therefore subject to lighter regulatory oversight.

Rationale: Independent evaluation and transparent guidance can help citizens identify trustworthy tools while reducing the risks associated with unregulated or poorly validated AI applications.

Alignment with overall recommendations: R8 (stakeholder engagement and trust), R10 (governance and oversight).

Recommendation 3. Strengthen AI education and interdisciplinary collaboration for healthcare professionals (healthcare workers)

Action: Expand education and training on AI development, validation, and clinical deployment as a core component of modern medical education and continuing professional development. Policy-makers should actively promote interdisciplinary collaboration between clinicians, data scientists, and AI specialists.

Rationale: Clinicians must understand AI systems to evaluate their reliability, integrate them into clinical workflows, and maintain safe patient care.

Alignment with overall recommendations: R7 (training and education), R4 (implementation challenges).

Recommendation 4. Encourage critical and participatory adoption of AI solutions (healthcare leaders)

Action: Healthcare leaders should apply critical scrutiny to AI products and marketing claims and avoid top-down implementation strategies that lack engagement from frontline healthcare professionals. AI adoption strategies should prioritise clinician involvement, workflow integration, and organisational readiness.

Rationale: AI deployment without clinician engagement or adequate evaluation often leads to poor adoption, ineffective systems, and wasted resources.

Alignment with overall recommendations: R4 (implementation challenges), R8 (stakeholder engagement).

Recommendation 5. Maintain balanced and strategic AI regulation at the EU level (health policy-makers)

Action: EU policy-makers should maintain robust but proportionate regulatory frameworks that prevent unsafe or ineffective AI systems from entering clinical use while supporting responsible innovation within the European ecosystem. Policy-makers should also encourage the development and adoption of European AI technologies through targeted financial and legislative incentives.

Rationale: Strategic regulation ensures patient safety and strengthens Europe's capacity to compete globally in the development of trustworthy AI technologies.

Alignment with overall recommendations: R2 (responsible deployment guidelines), R10 (governance and oversight), R9 (research priorities).

Recommendation 6. Expand regulatory capacity and expertise across Europe (regulators)

Action: Policy-makers should strengthen the capacity of regulatory authorities and notified bodies, including increasing the number of designated organisations and investing in AI and cybersecurity expertise. They should also support harmonisation of regulatory assessment practices across Member States through greater collaboration and transparency.

Rationale: Limited regulatory capacity currently creates bottlenecks that slow the safe introduction of AI-based medical technologies.

Alignment with overall recommendations: R10 (governance and oversight), R4 (implementation challenges).

Recommendation 7. Facilitate early regulatory engagement and regulatory sandboxes (manufacturers)

Action: Encourage manufacturers to engage early with notified bodies and regulatory sandboxes, enabling clearer guidance on regulatory expectations and smoother conformity assessment processes. Policy-makers should also provide capacity-building and financial support for notified bodies to deliver timely and consistent regulatory advice.

Rationale: Early regulatory engagement helps reduce uncertainty and supports efficient development of safe and compliant AI systems.

Alignment with overall recommendations: R2 (responsible deployment guidelines), R10 (governance and oversight), R9 (research priorities).

Recommendation 8. Strengthen community participation to reduce health inequalities (community organisations)

Action: Policy-makers should ensure that communities most affected by AI-enabled healthcare systems are involved as active partners from the earliest stages of design and deployment, through co-design workshops, participatory data governance, and transparent communication on data use and risk–benefit trade-offs.

Rationale: Community engagement strengthens trust, ensures equitable deployment of AI technologies, and helps prevent the widening of health inequalities.

Alignment with overall recommendations: R3 (ethical frameworks), R8 (stakeholder engagement), R1 (evaluation of AI value).

Recommendation 9. Link reimbursement to evidence of real-world clinical value (payers and insurers)

Action: Reimbursement of AI solutions should be contingent on high-quality evidence of clinical benefit, safety, and cost-effectiveness, supported by peer-reviewed research and health-economic analyses. Policy-makers should promote value-based reimbursement models, including outcome-based

contracts and staged reimbursement linked to demonstrated performance.

Rationale: Evidence-based reimbursement ensures that healthcare budgets support AI solutions delivering measurable benefits to patients and health systems.

Alignment with overall recommendations: R1 (evaluation of AI value), R5 (learning health systems).

Recommendation 10. Promote sustainable collaboration between academia and industry (academic and research institutions)

Action: Policy-makers should encourage long-term collaborative ecosystems between academic institutions and industry partners, enabling AI innovations developed in research environments to progress through validation, regulatory approval, and clinical deployment.

Rationale: Translating AI innovations from research into clinical practice requires sustained partnerships that combine scientific expertise with industrial development and regulatory experience.

Alignment with overall recommendations: R9 (research priorities), R5 (learning health systems), R6 (data ecosystems).

8 EU regulation and legislative framework for artificial intelligence in healthcare

Artificial intelligence (AI) offers substantial opportunities to improve healthcare, but its deployment also raises important implications for patient safety, trust and fundamental rights (EC 2025). In addition to AI systems used directly in professional healthcare services, two further use cases should be distinguished. The first one concerns AI deployed within the health sector for facility administration and human-resources management (for example, workforce planning, rostering, and performance oversight). The second covers systems that support certain forms of healthcare but are offered by commercial entities rather than healthcare providers, such as consumer wearables and related platforms. This distinction is decisive for the applicable regulatory regime and the obligations that follow. This regulatory section is limited to the use of AI in professional healthcare services. It intentionally excludes AI used in health-sector administration and AI-enabled services offered directly to consumers by commercial entities, because these uses may be governed, at least in part, by different regulatory requirements, standards and liability rules, including those arising under employment and consumer law.

8.1 Regulation of AI

AI is subject to relatively new European regulation: Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on AI and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (AI Act) (hereinafter: 'AI Act' or 'AIA') introduces a risk-based categorisation of systems.

For healthcare AI, legal analysis should distinguish the AI Act from the Medical Device Regulation, the In Vitro Diagnostic Medical Device Regulation, General Data Protection Regulation (GDPR), European Health Data Space (EHDS) rules on secondary use and data access, and relevant liability frameworks. These regimes interact, but they do not perform the same function.

8.1.1 The AI Act groups AI systems into four risk tiers

AI systems are grouped into four risk tiers (Barichella 2023). In the first tier, unacceptable-risk systems are prohibited under Article 5 AIA (e.g. manipulative or exploitative techniques that can cause harm, social scoring by public authorities, and certain uses of remote biometric identification). In the second tier, high-risk systems are defined in Articles 6–7 AIA. High-risk systems are divided into two categories: (1) AI that is

a safety component of, or itself is, a product covered by the EU product-safety regimes listed in Annex I AIA since they pose a risk to health and life (such as medical devices and in vitro diagnostic devices), and (2) AI used in specific use cases listed in Annex III AIA (e.g. biometrics, critical infrastructure, education, employment, access to essential services including credit, law enforcement, migration/border control, justice and democratic processes). According to Annex III, an AI system is not treated as high-risk if it does not pose a significant risk to health, safety, or fundamental rights (including not materially influencing decisions). This exception applies when the system (1) performs only a narrow procedural task, (2) merely improves the outcome of a previously completed human activity, (3) detects decision-making patterns or deviations without replacing or influencing an already completed human assessment and is subject to proper human review, or (4) carries out a preparatory task for an assessment tied to an Annex III use case. However, any Annex III system that profiles natural persons is always considered high-risk. High-risk AI system providers must meet stringent obligations (risk management, data governance, technical documentation, logging, transparency to users, human oversight, accuracy, robustness, and cybersecurity). The third tier presents limited-risk systems, such as generative AI systems or emotion-recognition systems trigger transparency duties under Article 50 AIA (for example, telling people they are interacting with AI and labelling AI-generated or manipulated content). Finally, minimal-risk systems face almost no mandatory obligations, except Article 4 AIA (AI literacy). The intended purpose and mode of deployment are decisive for classification (Boine and Rolnick 2023). While these are generally defined during development, the provider should state clearly in the technical documentation what the system has been trained for and the contexts in which it is intended to be used.

The AIA recognises persons and entities involved in the AI value chain. In healthcare uses of AI, the main actors are system providers and system deployers. The AIA also mentions persons affected by AI systems. Providers include (1) model providers, who train and release general-purpose or domain models (e.g. medical language or imaging models), and (2) downstream/system providers, who adapt, fine-tune, or integrate those models into specific AI systems with an intended medical purpose and place them on the market (Hopkins et al. 2025). Deployers are the professional users – healthcare providers such as individual doctors, hospitals, clinics, labs, or telemedicine platforms – that configure, supervise, and operate the AI system within

clinical workflows. They are required to keep records of operation, ensure appropriate staff training, conduct regular monitoring of the system, report malfunctions and adverse incidents to the provider, and guarantee human oversight. Some of these obligations – such as ensuring human oversight – may be contractually transferred back to the AI provider. Affected persons are individuals impacted by the system's outputs – in healthcare, patients – especially where the outputs cause or risk harm to health or to fundamental rights (e.g. privacy, non-discrimination, fair treatment).

If a deployer changes the intended purpose of an AI system so that it falls into the high-risk category, under Article 25(1)(c) AIA, that deployer is reclassified as the provider of that system and must fulfil all provider obligations (conformity assessment, CE marking, technical documentation, quality management, post-market monitoring, etc.). Article 25(1)(b) AIA covers substantial modifications to an already high-risk system with the same consequence (the modifier becomes the provider). Any healthcare provider that substantially modifies the AI system may, in these circumstances, be treated as an AI system provider.

8.1.2 AI systems in healthcare are mostly high or limited risk

In healthcare, AI may be deployed in multiple ways, and the applicable regulation follows the intended use. Where the AI system is used as, or as a safety component of, a medical device or an in vitro diagnostic device (under Regulation (EU) 2017/745 and Regulation (EU) 2017/746), it is classed as a high-risk system under the first category in Article 6 of the AIA. Classification in this category presents no difficulty, since providers of these systems are subject to the stringent requirements of the abovementioned regulations, including a certification requirement.

Other AI systems used in healthcare that are not medical devices must be assessed against their concrete application to determine the appropriate category. Depending on the use case, a system may still be qualified as high-risk under Annex III AIA—for example, when used by or on behalf of public authorities to assess individuals' eligibility for essential public support and services (including healthcare), to set priorities for dispatching first-response services (police, fire, or medical), or in emergency healthcare patient triage systems. The question is what precisely counts as emergency healthcare. Does routine emergency-department triage in a hospital constitute emergency-healthcare triage within the meaning of Annex III, or only when the system affects clinical urgency ranking? Clinical urgency ranking could involve mass-casualty events, but it remains uncertain whether a standard hospital emergency ward would qualify as such. Overall, the full automation of any triage process using AI

remains a highly challenging and inherently high-risk endeavour. AI-based triage tools cannot be trained on – nor can they realistically encompass – the full spectrum of possible clinical presentations, especially rare, atypical, or complex disease manifestations. As a result, fully automated triage carries a substantial risk of under-triage, where patients requiring urgent or emergent intervention may be inappropriately assigned lower priority levels. On such occasions, under-triage can lead to dangerous delays in time-sensitive care, potentially resulting in increased morbidity, mortality, or serious adverse events. Although many contemporary AI systems are designed primarily as decision-support tools (augmenting rather than replacing clinical judgement) and often exhibit a safety-preserving bias towards over-triage, complete replacement of human triage by autonomous AI remains premature and unsupported by the current evidence base. For these reasons, widespread implementation of fully autonomous AI triage should be approached with extreme caution, prioritising hybrid models with mandatory human oversight to safeguard patient safety. The classification of the triage system greatly influences the obligations of both providers and deployers. On the other hand, AI systems used for patient communications, such as chatbots, are typically limited risk under Article 50 AIA, which imposes transparency duties, including informing patients that they are interacting with an AI system.

By contrast, automated processing of medical documentation does not usually fall into the high-risk category or limited risk category. However, if an additional service is layered on top – for example, a chatbot providing advice to patients – this would be considered a limited-risk system, carrying its own set of obligations for both providers and healthcare institutions.

The AI Act's categorisation of AI systems may prove problematic in practice. A single system may be capable of multiple uses, its deployment context may evolve over time, and repurposing by users or third parties may be difficult to predict or control. Yet, compliance obligations must be fulfilled at the stages of design, training, and testing, namely before the system's concrete use case is fully determined. This may create legal uncertainty and may lead either to over-compliance or to under-compliance.

8.2 Patients' 'AI consent' should be required in an experimental context but not in routine care

It is not settled whether the use of AI systems in healthcare requires a separate patient consent ([Hurley et al. 2025](#)). Informed consent should be required at least when the system is used experimentally. Informed consent is expressly required for participation in regulatory sandboxes under Article 61 AIA.

A principled approach is to distinguish experimental contexts – including research prototypes, pilot deployments, model training on patient data, and pre-deployment testing – from routine clinical practice. Where AI is used experimentally, or where patient data are used to train or test models beyond direct care, patient involvement should be conditioned on informed consent that clearly explains the purpose, methods, foreseeable risks and benefits, data flows (including de-identification and sharing), and the right to refuse or withdraw consent without prejudice to care. This specific consent is comparable to participation in clinical trials (Nikhilesh 2023).

By contrast, when AI systems are used as part of routine, non-experimental care within approved clinical workflows and under professional oversight, a specific ‘AI consent’ should generally not be required. In these settings, the legal and ethical baseline can be satisfied through informed consent to treatment, coupled with transparency about the use of decision-support tools and the clinician’s responsibility for decisions, so that patients are clearly informed whether and how an AI tool is being used, and which tool it is. An exception remains for procedures or devices for which specific and device-bound informed consent is otherwise mandated by applicable law or medical standards.

Patient notification should be proportionate to the role of the AI system. Routine use of AI may not require a separate consent process in every case, but patients should be clearly informed when AI materially supports diagnosis, triage, treatment selection, or care management, and they should know how to ask questions or seek human review.

Informed consent to the provision of healthcare services must not be confused with consent to the processing of personal data. In healthcare, the processing of patient data typically rests on Article 6(1)(c) GDPR – the controller’s legal obligation – together with Article 9(2)(h) GDPR. No separate GDPR consent for data processing is therefore required. A patient’s informed consent authorises the medical intervention itself; it is not, and should not be treated as, a data-processing consent.

8.3 AI and data protection

8.3.1 GDPR Articles 5, 6, 9 and 89 are especially relevant for AI systems

A separate and highly important area is the protection of personal data and data subject to medical confidentiality. Healthcare data fall within the special category of data (health data) and are therefore subject to Article 9 GDPR as well as the general requirements of Article 6 GDPR. Their processing is permitted for the provision of healthcare; however,

this permission cannot automatically be extended to AI providers that wish to reuse such data to further train or improve their systems. In such cases, healthcare institutions should address the contractual conditions with providers in detail—ideally prohibiting any onward transfer or permitting sharing only where data have been anonymised so that the GDPR no longer applies. Decisions must also consider the nature of the system and the purpose for which it is used. From a functional perspective, it remains true that the more high-quality data a system can access, the more accurate and reliable its results tend to be.

Under Article 6 GDPR, the controller must identify a lawful basis for each processing purpose. In healthcare, this will typically be compliance with a legal obligation incumbent on the controller under national law to provide healthcare services (Article 6(1)(c) GDPR); in some contexts, processing may alternatively rest on the performance of a task carried out in the public interest (Article 6(1)(e) GDPR). Because health data are a special category, a corresponding condition in Article 9 is also required—most commonly Article 9(2)(h) GDPR (health or social care, including the management of healthcare systems) or Article 9(2)(i) GDPR (public interest in public health), also subject to national safeguards. Where an AI provider seeks to use patient data for secondary purposes such as model retraining, a separate lawful basis and Article 9 GDPR condition are generally needed; the ‘provision of care’ ground used by the healthcare institutions does not automatically cover the AI system provider’s own improvement purposes. In practice, controllers often choose between anonymisation (removing the data from the GDPR’s scope) or obtaining a suitable lawful basis for secondary use and documenting a compatible-use assessment. Roles and responsibilities should be fixed in an Article 28 GDPR data-processing agreement (processor role) when the improvement is provided only in favour of a particular controller and not all deployers of the AI system.

Controllers must also observe the data protection principles under Article 5 GDPR, notably purpose limitation (process only for specified, explicit and legitimate purposes) and data minimisation (collect and process only what is adequate, relevant, and necessary). With AI systems, the minimisation principle can seem paradoxical: statistically, broader and more heterogeneous datasets improve model accuracy and reliability, yet the GDPR requires strict necessity in relation to the stated purpose. This tension is resolved not by collecting ‘as much as possible’, but by demonstrating – through a documented necessity and proportionality assessment – that each category of data is genuinely required to achieve the clinical or operational objective. Practically, controllers should

(1) narrow the purpose statement to the concrete clinical function; (2) curate datasets to remove attributes that are not materially contributory; (3) prefer pseudonymisation, aggregation and privacy-preserving techniques where feasible; (4) set retention limits aligned to validation and audit needs; and (5) implement strong access controls and logging. Given the likely high risk to rights and freedoms, a Data Protection Impact Assessment under Article 35 will often be required before deployment, referencing the AI Act's risk-management and human-oversight controls.

Finally, Article 89 GDPR provides a framework for processing for scientific research, subject to appropriate safeguards such as pseudonymisation, strict access control, and purpose limitation compatible with research aims. When relying on Article 89, controllers still need a lawful basis under Article 6 and Article 9 conditions for special-category data—commonly Article 9(2)(j) (research in the public interest). Article 89 may also permit limited derogations from certain data-subject rights where exercising those rights would seriously impair the research and where such derogations are laid down in law with suitable protections. In all cases, transparency obligations remain: patients should be informed, in clear and accessible terms, of the purposes of processing, the roles of the healthcare provider and any AI provider, and whether any secondary use (such as model improvement) will occur, together with the choices and safeguards available to them.

8.3.2 Article 10 of the AI Act complements requirements of the GDPR

The AI Act data-processing obligations are set in Article 10 (data governance). For high-risk AI systems that make use of model training with data, providers must develop the system based on training, validation, and testing datasets that satisfy Article 10's quality and governance criteria. Datasets must be relevant, sufficiently representative, and – as far as possible – free of errors and complete in view of the intended purpose, and must exhibit appropriate statistical properties, including, where applicable, with respect to the persons or groups for whom the system is intended. Providers must also ensure that datasets consider, to the extent required, geographical, contextual, behavioural, or functional characteristics of the setting of use.

Article 10 recognises an exceptional possibility for providers to process special categories of personal data (e.g. health data) strictly to the extent necessary to detect and correct bias in relation to a high-risk system, subject to appropriate safeguards and in addition to GDPR requirements. These safeguards operate alongside, and do not replace, the controller's obligations under GDPR. Data governance duties complement the GDPR's purpose-limitation and data-minimisation principles. In practice, the apparent

tension between minimisation and AI performance should be addressed. It is also important to mention that the AI Act's allowance does not replace GDPR.

8.3.3 The European Health Data Space

The European Health Data Space (Regulation (EU) 2025/327) draws a clear line between primary use and secondary use of electronic health data. Primary use covers data processing for the direct provision of healthcare—ensuring patients can securely access and share their electronic health records across borders through MyHealth@EU, and requiring interoperable electronic health record (EHR) systems, strong security, and auditability within care pathways. Secondary use enables re-use of electronic health data for defined public-interest purposes but only under a permit issued by national health data access bodies and within secure processing environments where users may extract only non-personal (anonymised) outputs. Under the European Health Data Space, training, testing, and evaluation of algorithms – including AI systems – is an expressly permitted purpose of secondary use of electronic health data. Health data are not available under an open-data regime. AI providers must apply to a health data access body and obtain a data permit, showing necessity for a listed purpose and a GDPR legal basis. Results must be made public within 18 months and contain only anonymous data. Individuals have a right to opt out of secondary-use processing of their personal electronic health data, which Member States must enable through an accessible mechanism.

The European Health Data Space is focused on high-quality health data for training, validating, and testing AI systems in healthcare, while the AI Act deals mainly with a legal framework for high-risk AI systems. According to Recital 68 AIA, the European Health Data Space should facilitate trusted, non-discriminatory sources of high-quality data and the training of AI algorithms on those data sets, in a privacy-preserving, secure, timely, transparent, and trustworthy manner, and with an appropriate institutional governance. However, it is not clarified how high-quality data are defined and ensured within European Health Data Space governance and how European Health Data Space can deliver the necessary volume and diversity of data to reduce bias and, above all, ensure the quality of AI system outputs (Rissetto 2021).

8.4 Automated decision-making

Automated decision-making means decisions about a specific person that are made solely or partly by algorithm and that have legal or comparably significant effects on them. In healthcare, automated decision-making can arise in algorithmic triage and risk scoring (prioritising emergency patients), and in clinical decision-support systems that propose diagnoses or

treatments (Khosravi *et al.* 2024). In healthcare, as AI-driven automated decision-making expands, the right of patients to an explanation of adverse clinical decisions remains only partly settled. EU law addresses such decisions under GDPR Article 22 and Article 86 AIA: GDPR targets solely automated decisions producing legal or similarly significant effects, while the AI Act covers decisions made with specified high-risk systems in Annex III (not necessarily fully automated) and applies only insofar as EU law does not already regulate the explanation right. An affected person in healthcare can invoke Article 86 AIA when a decision about them is taken by a deployer (e.g. a hospital or insurer) on the basis of a high-risk AI system's output and the decision produces legal effects or similarly significant effects that they consider adverse to their health, safety, or fundamental rights.

The affected persons may request from the deployer a clear, meaningful explanation of the AI system's role in the decision process and the main elements of the decision taken. Explanations should set out the procedures and principles used to reach the outcome, but overly technical disclosures can undermine the effective exercise of rights (Wasserman-Rozen *et al.* 2024), and even deployers may lack insight into model behaviour—an asymmetry noted by the Court of Justice of the European Union (C-634/21, SCHUFA), which places certain duties on the AI system provider acting as a personal data processor. Under the AI Act, the explanation right attaches only to certain high-risk systems; GDPR, by contrast, provides a right to human intervention, whereas the human-oversight duties under the AI Act are largely preventive rather than remedial (Sarraf 2025). Human oversight is no cure-all given automation bias and capability limits (Biermann *et al.* 2022). Finally, these rights apply only where decisions have legal or similarly significant effects; many health uses – such as triage support or automated evidence gathering – may fall outside, particularly where varying degrees of AI involvement complicate classification (Demková 2021), except where preparatory outputs effectively determine outcomes (Palmiotto 2024).

In healthcare, automated systems typically support rather than replace clinicians, so decisions are only partly autonomous as required under Article 86 AIA. However, Article 86 applies only to the Annex III standalone high-risk categories and does not extend to other high-risk AI systems that are regulated as products—such as medical devices. It is also uncertain whether decisions made within clinical care fit the notion of 'legal or similarly significant effects'. Many algorithm-supported outcomes (e.g. suggested diagnoses and treatment) do not have such effects.

The remedy of Article 86 AIA is regulatory and procedural, not a civil-liability regime: Article 86 AIA

does not grant damages, and the AI Act does not establish rules on civil liability. Any claim for injury from clinical care remains under separate frameworks—notably national medical-liability law (contract/tort) where standards of fault, causation, and compensation are assessed independently of Article 86 AIA. Under the GDPR, a patient could in principle claim compensation for material and non-material damage (Article 82 GDPR), but this provision may be triggered only for decisions that are made solely by automated means. In healthcare, AI generally plays a supportive role in clinical judgement (Khosravi *et al.* 2024).

8.5 Liability lies with the healthcare provider

Liability for damage is of fundamental importance in the use of AI in healthcare. AI systems can significantly influence diagnostics, treatment decisions, and the organisation of healthcare. If it were not clearly established who is liable for potential errors or harm caused by an AI system, patients would be left in uncertainty and might be unable to exercise their right to compensation. Some AI systems can learn or adapt and make autonomous decisions. Even the provider is often unable to predict a system outcome or explain why the system generated a particular output, namely why it decided as it did (Bathae 2018).

In healthcare, the healthcare provider should be liable to the patient for harm resulting from the use of tools incorporated into its services – including AI systems – regardless of fault. The burden sits with the healthcare provider to ensure that systems it deploys are appropriate, validated, and monitored, and that risks are addressed through procurement, quality management, and insurance. The liability that is not based on fault follows from the complexity of AI systems. The hard question is what counts as a 'defect' in an AI system. An algorithm may be technically sound yet still generate harmful outcomes because it ingests poor-quality data, is applied to an out-of-scope population or is misinterpreted by downstream processes (Griffin 2021). These grey zones make strict no-fault liability particularly important: they protect the patient while leaving room to sort out recourse between AI system providers and healthcare providers. In practice – and for the foreseeable future – clinicians must still confirm diagnoses and treatment plans *lege artis*, and the healthcare provider must ensure that AI outputs are validated, context-appropriate, and subject to effective human oversight. If liability were not assessed in this way, a patient could easily end up in a situation where they could not obtain compensation because the system is not proved defective.

In addition to individual liability, the conformity of medical devices is strictly regulated. Medical devices under the MDR are auditable through conformity assessment, technical documentation review,

post-market surveillance, vigilance obligations, and inspections by notified bodies or competent authorities. If a medical device includes AI and falls under the AI Act as a high-risk AI system, it must also be auditable through AI-specific documentation, logging, risk management, data governance, human oversight, robustness, cybersecurity, and post-market monitoring. For AI medical devices, the AI Act assessment is generally integrated into the MDR conformity assessment, so the manufacturer must be able to demonstrate compliance with both regulatory regimes.

The recast of the Product Liability Directive – Directive (EU) 2024/2853 – modernises strict (no-fault) liability for ‘defective products’, expressly extending the notion of a product to software and AI. The Directive allocates strict liability to actors who place a product on the market or put it into service; it does not regulate the professional obligations of healthcare providers towards patients. In typical care pathways, the patient’s legal relationship is with the healthcare provider, which uses AI-enabled devices or software within a service. The manufacturer stands outside that treatment contract. Accordingly, disputes about harm in clinical care are ordinarily resolved under healthcare/treatment liability frameworks (e.g. institutional, organisational, or clinical duty of care), not under the Product Liability Directive—unless a separate product-defect claim is brought directly against an economic operator in the product chain. In other words, the Product Liability Directive may enable recourse by patients (or by providers through indemnity) against manufacturers for defective AI products, but it does not set or displace the rules that govern provider responsibility to patients in routine clinical practice.

Where AI systems are partly explainable or operate as black boxes, accountability should not depend solely on technical interpretability. Procurement, deployment, and monitoring arrangements should specify intended use, human oversight, audit logs, incident reporting, responsibility for updates, and routes for patient complaint or redress.

8.6 Recommendations

Recommendation 1. Clarify consent requirements for clinical and experimental uses of AI

Action: For routine, non-experimental clinical use of AI within approved clinical workflows and under professional oversight, policy-makers should rely on standard informed consent to treatment, supplemented by transparent disclosure that an AI tool is being used and that clinical responsibility remains with the healthcare professional. For experimental or developmental uses – including pre-deployment testing, pilots, regulatory sandboxes, or training and validation activities beyond direct care – a specific informed consent process should be required. This consent should

clearly describe the purpose, methods, foreseeable risks and benefits, data flows, and the right of individuals to refuse or withdraw without prejudice.

Rationale: Distinguishing routine clinical use from experimental deployment helps protect patient rights while avoiding unnecessary administrative burdens in everyday clinical practice.

Alignment with overall recommendations: R3 (ethical frameworks), R8 (public trust), R10 (governance and oversight).

Recommendation 2. Maintain clear liability structures for AI-enabled clinical care

Action: Ensure that primary liability towards patients for harm arising from AI-enabled care remains with the healthcare provider, while requiring appropriate validation, monitoring, and human oversight for deployed AI systems. Policy-makers should encourage the use of contractual and indemnity arrangements that enable healthcare providers to seek recourse from manufacturers where system defects contribute to harm.

Rationale: Clear allocation of responsibility is essential to maintain patient protection and ensure that accountability is preserved within clinical care pathways.

Alignment with overall recommendations: R1 (evaluation of AI value), R4 (implementation challenges), R10 (governance and oversight).

Recommendation 3. Clarify the classification of high-risk AI systems in healthcare

Action: Provide clearer guidance on which healthcare AI deployments fall under Annex III of the EU AI Act and should therefore be treated as high-risk systems. Regulatory clarification should ensure consistent classification across Member States and reduce uncertainty for developers, healthcare providers, and regulators.

Rationale: Ambiguity around high-risk classification can slow innovation, create regulatory uncertainty, and hinder responsible deployment of healthcare AI systems.

Alignment with overall recommendations: R2 (guidelines for responsible deployment), R4 (implementation challenges), R10 (governance and oversight).

Recommendation 4. Guarantee patient rights to meaningful explanations of AI-assisted decisions

Action: Patients should have a right to obtain a meaningful explanation of AI-assisted outcomes when they believe that the outcome has negatively affected their health or fundamental rights. This right should

apply irrespective of the system's classification or intended autonomy level, ensuring transparency and recourse for affected individuals.

Rationale: Access to explanations strengthens accountability, supports patient autonomy, and helps maintain trust in AI-enabled healthcare systems.

Alignment with overall recommendations: R3 (ethical frameworks), R8 (stakeholder engagement and trust), R10 (governance and oversight).

Recommendation 5. Promote privacy-preserving data governance for multi-institutional AI training

Action: When training AI systems on multi-institutional patient data, policy-makers should require the use

of privacy-preserving governance mechanisms, such as secure data environments, federated learning approaches, and robust de-identification protocols. Clear best practices for data sharing should also be developed to encourage responsible data sharing across institutions, improving model performance while maintaining strong privacy protections.

Rationale: Clear and trusted data-sharing frameworks are essential for developing robust AI systems while safeguarding patient privacy and maintaining public trust.

Alignment with overall recommendations: R6 (data sharing ecosystems), R5 (learning health systems), R10 (governance and oversight).

9 Educational requirements to integrate artificial intelligence in European healthcare

This chapter addresses the critical educational transformations required to support successful artificial intelligence (AI) integration in European healthcare. With an estimated 12 million healthcare professionals and 450 million citizens across the EU, comprehensive education and engagement strategies are essential for realising AI's benefits while mitigating risks. To understand what mandatory AI competency requirements for healthcare professionals would entail is a priority, including curriculum reform and continuous professional development. Educational initiatives addressing digital health literacy, transparent AI disclosure, and inclusive design for vulnerable populations are another priority. Finally, the current and potential future role of experience with AI in interdisciplinary collaboration needs investigation, bringing together clinicians, developers, ethicists, patients, and the public as a whole. Action needs to be taken now. Without immediate intervention, Europe risks widening a two-tier healthcare system divided by AI literacy, undermining both quality of care and public trust. A digital divide already exists, and AI may widen it. These recommendations provide parts of a roadmap for positioning Europe as a global leader in ethical, inclusive AI healthcare education.

9.1 Mastering the machine

A strategic European approach must integrate AI education into medical and public health curricula from the beginning, equipping future clinicians and allied healthcare professionals and policy-makers with the foundational knowledge and critical thinking skills necessary to navigate AI-augmented practices (Gazquez-Garcia *et al.* 2025). As AI technologies increasingly complement the work of health professionals, with diagnostic support, clinical decision tools, and administrative automation becoming embedded in practice, medical education curricula must be updated to reflect evolving roles. Professional education should not only highlight how to use AI responsibly but also emphasise the unique human skills that are at risk of diminishing over time with the use of AI, including clinical decision-making and judgement, empathy, and contextual reasoning (Budzyn *et al.* 2025). Greater understanding is needed of how to integrate and evaluate AI-based tools in real-world settings, learn from their use, and use that knowledge to design better systems; education of those who use and work with the tools is therefore essential. This is where engineering and computer science, clinical medicine, and the social sciences are needed to work together. Continued professional development is

essential to help existing staff adapt to new workflows, maintain professional autonomy, and safeguard the quality of patient care. Moreover, Europe must foster interdisciplinary training and capacity-building, bringing together clinicians, data scientists, ethicists, and health administrators in collaborative learning environments. Educational models that emphasise transdisciplinary frameworks must incorporate problem-based or patient-based learning, such as in-depth analyses of clinical problems, with AI tools. A forward-looking European strategy mandates the integration of data science and AI from the very start of higher education (Hicke *et al.* 2026), particularly within medical programmes, to ensure early and meaningful impact even for those with no previous academic exposure (Ma *et al.* 2024). There is no substitute for hands-on experience.

Europe must act decisively to prevent a potential crisis of competence. Medical licensing bodies across Member States should be consulted about how to best require demonstrable AI literacy as a prerequisite for practice, with standardised competency assessments developed at the EU level. They should be encouraged to increasingly include AI in their clinical guidelines. National medical councils could for example encourage investigations into mandating minimum hours of AI ethics and application training in all healthcare professional curricula, proportionate to specialty-specific needs. Failure to update professional education at this critical juncture risks creating a two-tier system where AI-literate practitioners dominate clinical decision-making and research opportunities, potentially marginalising colleagues without access to quality training and widening the gap between well-resourced and under-resourced healthcare institutions. The European Commission could encourage the establishment of clear benchmarks for AI integration in medical education, with funding mechanisms that incentivise rapid adoption.

Educational strategies should include structured, ongoing professional development to keep pace with evolving AI technologies. This may involve certification programmes, micro-credentials, workshops, and refresher courses. Clinicians should have opportunities to update their skills as new AI tools are introduced, ensuring competence and safe adoption across their careers. Furthermore, institutions must assess whether AI education improves knowledge, skills, and decision-making. Competency frameworks and assessment metrics should measure technical proficiency, ethical reasoning, and practical application

in clinical workflows. In Sweden, the seven universities offering Doctor of Medicine training have divided a total of 16 course modules among them, with each university responsible for between one and four modules. The modules are shared through a Moodle server, where course material remains available for at least a year. This collaborative structure fosters shared understanding and allows the shared course material to be updated continuously.

9.2 Re-defining the healer

Because AI can raise concerns among medical students about job security, professional identity, and future opportunities, medical education must address these directly. Training could incorporate AI literacy. Curricula could encourage mentorship programmes and structured dialogues between students, educators, and AI developers to sustain motivation and commitment to the field (Rincon *et al.* 2025). Educators should also be supported in the design of the curriculum itself. The extent to which that is possible in practice is dependent on culture, financial stability, infrastructure, and many other factors. Specific initiatives could include AI-Health Innovation fellowships, student-led AI ethics committees (which could be made mandatory in all medical schools), and industry-academia exchange programmes with rotations. Success measures of such initiatives could then encompass medical school application rates, student satisfaction scores, graduate employment in AI-health roles, and traditional innovation outputs (papers, patents, startups, etc.).

9.3 A guardian of public trust

As large language models (LLMs) are increasingly deployed in patient-facing roles, from triage chatbots to treatment explanations, their current deployment lacks standardised transparency measures, creating risks for patient autonomy and informed consent. Studies show that while LLMs can accurately summarise clinical guidelines and convert complex jargon into more easily understandable language, they remain susceptible to generating incorrect information (Peng *et al.* 2023), potentially causing direct patient harm. The absence of mandatory disclosure requirements allows AI systems to shape healthcare decisions invisibly, undermining informed consent and patient autonomy. Users may not realise when AI is involved in generating clinical advice, and some (although not all) feel uneasy if AI is used without explicit clinician oversight (Busch *et al.* 2025). This trust deficit, if left unaddressed, threatens the entire digital health transformation agenda.

Beyond awareness, patients should be involved in co-design and governance of AI systems. Mechanisms such as citizen panels, focus groups, or participatory workshops can ensure AI tools meet patient needs, enhance patient trust, and address societal expectations.

Educational initiatives should account for varying levels of digital literacy across populations (Kimiafar *et al.* 2023). Materials must be accessible to vulnerable groups, including older adults, non-native speakers, or those with limited access to technology, to avoid widening health disparities. European healthcare providers should be legally required to disclose AI involvement in clinical pathways, with standardised signage, digital indicators, and verbal notification protocols. Without immediate regulatory intervention, AI deployment will amplify existing health inequities, creating 'AI deserts' where disadvantaged populations with limited digital literacy, non-native speakers, elderly patients, and economically marginalised communities receive demonstrably inferior care compared with tech-savvy populations. The digital divide in healthcare is not merely about access to devices; it is about who has the knowledge and agency to question AI recommendations, seek human oversight, and advocate for alternative approaches.

For positive engagement, AI health-literacy programmes could be arranged as community workshops in the 24 official EU languages, with cultural adaptation for diverse communities. Trust-building initiatives, public AI demonstration centres, patient success story campaigns to balance alarmist reports of AI shortcomings and hallucinations, and transparent error reporting systems, are essential. Inclusive design principles are also a prerequisite for trust.

9.4 Integrating ethics and training developers

Ethical considerations must be embedded into the design and deployment of AI from the outset. This requires integrating ethicists into development teams alongside clinicians, engineers, and data scientists to anticipate issues such as data and AI bias, accountability, and transparency. Providing developers with structured ethics training covering fairness and inclusivity may help ensure that technical decisions align with healthcare priorities and societal expectations. Emerging evidence from co-creation workshops, ethics frameworks, and pilot training initiatives suggests that structured ethics engagement can improve the early identification of fairness, transparency, and validation issues in AI system design and testing; however, stronger empirical evidence is still needed to show that ethics training directly reduces real-world AI bias incidents (Lopez-Ramos *et al.* 2025).

Training should include awareness of data governance, privacy and consent regulation, and AI governance frameworks. Understanding legal obligations ensures AI deployment is safe, compliant, and trustworthy. Education programmes should address the diversity of societal norms, language, and cultural expectations across Europe. Developers and clinicians should have access to simulation environments,

virtual patient scenarios, and sandboxed AI systems (Kaur *et al.* 2022) to practice deployment, identify potential failures, and understand human–AI interactions in realistic settings.

9.5 Recommendations

Recommendation 1. Mandate AI literacy for clinical licensing and professional practice

Action: National medical boards should introduce mandatory AI literacy requirements for professional licensing and continuing medical education, ensuring that clinicians understand the principles of AI development, validation, limitations, and safe clinical deployment.

Rationale: Clinicians must be able to critically evaluate AI-enabled tools to ensure safe patient care and responsible use in clinical workflows.

Alignment with overall recommendations: R7 (training and education), R4 (implementation challenges), R1 (evaluation of AI's value).

Recommendation 2. Establish transparency and public disclosure standards for healthcare AI

Action: The European Commission should develop standardised transparency requirements obliging healthcare providers and AI developers to disclose where and how AI systems are used in healthcare, including their intended purpose, limitations, and governance arrangements. Operational recommendations should also require developer-facing ethics training, including patient safety, privacy, informed consent, bias, accountability, transparency, and embedded ethics.

Rationale: Clear transparency standards strengthen accountability, enable informed decision-making by clinicians and patients, and support public trust in healthcare AI.

Alignment with overall recommendations: R8 (stakeholder engagement and trust), R10 (governance and oversight), R1 (evaluation of AI's value).

Recommendation 3. Harmonise AI competencies in health professional education

Action: The European Commission and Member States, working through the Education, Youth, Culture and Sport Council, should coordinate the standardisation of AI-related competencies across medical and health professional education programmes, ensuring that

curricula across Member States include foundational training in AI, digital health technologies, and data governance.

Rationale: Harmonised training standards will help reduce disparities in digital health capability across Member States and ensure that future healthcare professionals are equipped to work with AI technologies.

Alignment with overall recommendations: R7 (training and education), R4 (implementation challenges).

Recommendation 4. Establish regulatory and innovation sandboxes for healthcare AI

Action: Research institutions, in collaboration with regulators and healthcare providers, should create regulatory and innovation sandboxes that enable AI systems to be tested in controlled real-world environments. These platforms should support responsible experimentation, evaluation, and regulatory learning.

Sandboxes and AI labs should also be used as co-creation environments where clinicians, patients, developers, ethicists, regulators, and procurement teams can test whether a system is useful, usable, safe, and scalable in real workflows.

Rationale: Sandbox environments facilitate innovation while allowing regulators and developers to understand risks, refine standards, and improve oversight mechanisms before large-scale deployment.

Alignment with overall recommendations: R9 (research priorities), R2 (responsible deployment guidelines), R4 (implementation challenges).

Recommendation 5. Strengthen public engagement and AI awareness initiatives

Action: Member States should implement public engagement and communication initiatives to improve citizens' understanding of AI in healthcare, enhance transparency, and strengthen trust in AI-enabled health technologies.

Rationale: Public understanding and engagement are essential for maintaining trust and ensuring that AI systems reflect societal values and expectations.

Alignment with overall recommendations: R8 (stakeholder engagement and trust), R3 (ethical frameworks).

10 General conclusions and recommendations: proactive policies for responsible artificial intelligence integration into European healthcare

Artificial intelligence (AI)-assisted tools and systems are increasingly being applied, offering high potential for important innovations in many scientific and societal fields including healthcare. At the same time, appropriate policy guidance and regulation are crucial to ensure that the full benefits of innovation can be realised while effectively addressing relevant risks that arise when implementing AI applications. Given this overall challenge of leveraging AI innovations for the benefit of society while simultaneously managing potential risks, AI developments in healthcare and public health bring specific opportunities, questions, and requirements.

Therefore, while the European Union (EU) has translated discussion into action in several ways, such as the EU AI Act and the European Health Data Space, this report has provided an overview of AI-related developments in healthcare and public health, discussed relevant next steps and issues to be addressed, and has translated scientific evidence and insights into advice for further policy development. In this context, starting from use cases – illustrating important opportunities, developments, and potential risks and requirements in medical, public health, and administrative applications – the following topics were considered: analytical and clinical validation, data quality and data sharing, security risks and cyber threats, ethical considerations, involvement of key stakeholders, the EU regulation and legislative framework for AI in healthcare, and education and public engagement.

This has made clear that the pervasive integration of AI within the healthcare ecosystem is fundamentally redefining healthcare and the roles of health professionals and citizens. At the heart of this transition is the cultivation of a digitally fluent workforce, a goal that must be pursued through modernisation of curricula and continuous professional education to ensure that AI literacy remains a career-long competency. This educational framework provides the necessary foundation for robust, science-based decision-making processes, where the development and deployment of AI-assisted tools are meticulously evaluated and monitored for both positive and negative health and societal impacts. In this way – based on thorough, comprehensive, and purpose-appropriate validation – we can promote the adoption of promising technologies while identifying those not meeting the standards and/or requiring further refinement to meet clinical standards.

Specifically, validation should include rigorous testing and clearly defined performance metrics, with transparency in data provenance and preprocessing to support reproducibility and trustworthiness. Validation strategies should also incorporate human factors, including clinician workflow integration, cognitive load assessment, and explainability of AI outputs. In addition, validation frameworks should be able to address the continuously emerging AI challenges while remaining compliant with evolving EU regulatory standards. To support this evolution, research and development must be significantly leveraged to facilitate secure, high-integrity data sharing, ensuring that fundamental European values, such as privacy, equity, transparency, and sovereignty, are embedded by design into the architectural core of every innovation.

Finally, these advancements must be anchored by a proactive regulatory framework that not only incentivises adoption but also enforces the clinical principles of reliability and damage prevention and reparation. By translating these principles into the digital domain, we safeguard all stakeholders and uphold the trust essential to the future of an AI-augmented healthcare system.

To conclude this report, EASAC and FEAM have formulated three overarching key messages and 10 policy recommendations. These together constitute a call to embed the responsible use of AI systems in European healthcare. To this end, AI applications must be managed proactively as part of a sustainable, anticipatory healthcare strategy, rather than simply reacting to external factors such as emergencies or market developments.

The recommendations could be implemented through a staged roadmap rather than as a static list. These strategic policy recommendations are suggestions to make options actionable.

Near-term actions include clarifying scope and responsibilities, updating procurement criteria, setting minimum validation and auditability expectations, and strengthening patient and professional involvement.

Medium-term actions include shared validation infrastructures, regulatory sandboxes, secure processing environments, cross-Member-State best-practice exchange, and support for small and medium-sized enterprises and public-sector innovators to meet proportionate evidence and transparency requirements.

Longer-term actions include embedding AI literacy and ethics training in health-professional and developer education, monitoring patient outcomes

and inequalities, evaluating cost-effectiveness and environmental sustainability, and maintaining adaptive governance as technologies and evidence evolve.

Three key messages

1. **Responsible AI integration must be proactive, not reactive.** AI systems in healthcare should be guided by forward-looking policies that anticipate long-term impacts rather than reacting to crises or market pressures. Embedding responsibility requires early governance, inclusive design, and strategic investment to ensure AI strengthens, not destabilises, health systems.
2. **Equity and trust are essential for sustainable AI adoption.** Without targeted action, AI systems may widen health disparities and undermine public trust. EU policies must prioritise equitable access, transparency, community engagement, and sovereignty to ensure AI systems support universal healthcare goals and democratic values.
3. **Coordinated EU action is crucial for safe, scalable innovation.** Fragmented approaches to AI-systems governance hinder safe implementation and slow progress. The EU must promote harmonised standards, interoperable cross-border data sharing, and collaborative oversight to enable robust, ethical, and scalable AI applications across Member States.

The 10 recommendations for EU policy-makers, as summarised in the Executive summary, are based on the interrelated, more specific conclusions and recommendations in the chapters, which covered scientific and methodological requirements, security

risks, ethical considerations, stakeholder involvement, the EU regulation and legislative framework, and education and public engagement. These policy recommendations are self-explanatory and are given below.

Ten recommendations for EU policy-makers on AI in Healthcare

Recommendation 1. Establish robust evaluation of AI's added value in healthcare. EU policy-makers should promote comprehensive evaluation frameworks to assess AI's net impact in healthcare, weighing clinical benefits against potential risks and adverse effects. This requires transparent, evidence-based assessments of effectiveness, safety, security, equity, and cost-efficiency, ensuring that AI applications deliver meaningful improvements without compromising ethical or quality standards.

Recommendation 2. Develop comprehensive guidelines for responsible AI deployment. EU policy-makers should support the creation of detailed guidelines to address key challenges in deploying AI in healthcare, including methodological rigour, data bias and representativeness, external validity, safety and security, cost-effectiveness, and environmental sustainability. Digital tools can aid in standardising evaluations and promoting accountability.

Recommendation 3. Broaden ethical principles to encompass societal and long-term dimensions. EU policy-makers should ensure that ethical frameworks for AI in healthcare go beyond individual rights to include societal, institutional, and global considerations, as well as long-term impacts. Ethics guidance must address collective well-being, intergenerational equity, and the broader social implications of AI deployment.

Recommendation 4. Address implementation-related challenges in AI deployment. EU policy-makers should ensure coordinated strategies to overcome implementation barriers in AI healthcare applications. This includes investing in interoperable digital infrastructure, harmonising regulatory frameworks across Member States, fostering workforce training, and supporting integration into clinical workflows. A focus on real-world usability, transparency, and patient safety is essential to maximise benefits and minimise disruption.

Recommendation 5. Advance AI towards a learning health system. EU policy-makers should support the development of AI-driven learning health systems that continuously improve through real-time data integration, feedback loops, and adaptive decision-making. This requires investment in secure data infrastructures, secure AI algorithms, cross-sector collaboration, and governance models that prioritise transparency, accountability, and public trust.

Recommendation 6. Foster interoperability and data sharing ecosystems. To unlock AI's full potential, the EU should incentivise interoperable data infrastructures that facilitate secure, cross-border data sharing. Policy-makers ought to promote common standards and protocols, while reinforcing privacy safeguards and data stewardship, to support large-scale, responsible AI innovation.

Recommendation 7. Invest in training, education, and information for AI in healthcare. EU policy-makers should prioritise comprehensive training and education programmes to equip healthcare professionals, technical experts, and decision-makers with the skills needed to develop, evaluate, and responsibly implement AI. Interdisciplinary curricula and continuous professional development are essential to foster trust, competence, and ethical awareness across the health ecosystem.

Recommendation 8. Engage stakeholders and protect public trust. Active engagement with patients, clinicians, communities, and civil society throughout AI development and deployment is essential. EU policy-makers should require transparent communication, participatory design, and robust recourse options to ensure AI systems reflect societal values, uphold public trust, and respect patient autonomy. EU policy frameworks ought to prioritise public outreach initiatives aimed at highlighting the significance of data sharing and countering unfounded concerns about AI through fact-based communication.

Ten recommendations for EU policy-makers on AI in Healthcare (continued)

Recommendation 9. Define strategic research priorities for AI in healthcare. EU policy-makers should support targeted research to address critical gaps in AI for healthcare, including algorithmic transparency, the development of safe and secure algorithms, bias mitigation, clinical validation, and system integration. Priorities should also encompass interdisciplinary approaches, long-term impacts, and equity considerations to ensure that AI innovations align with public health goals and societal values.

Recommendation 10. Strengthen governance and oversight structures. EU policy-makers should establish clear governance frameworks, including independent oversight bodies, to monitor AI applications in healthcare. These structures should define roles and responsibilities, ensure regulatory compliance, and enable mechanisms for accountability, auditing, and corrective action when needed.

Contributors

EASAC Director of the Biosciences and Public Health Programme

Thomas Cole-Hunter

FEAM Executive Director

Louise Abboud

Authors: Working Group members (with nominating academy and network)

1. André Knottnerus, Royal Netherlands Academy of Arts and Sciences, EASAC (Co-chair)
2. Luis Martí-Bonmatí, Spanish Royal National Academy of Medicine, FEAM (Co-chair)
3. Ishita Barua, The Norwegian Academy of Science and Letters, EASAC
4. Magnus Boman, The Royal Swedish Academy of Sciences, EASAC
5. Eva Fialová, Czech Academy of Sciences, EASAC
6. Gheorghe Ioan Mihalas, Romanian Academy of Medical Sciences, FEAM
7. Radko Komadina, The Slovenian Academy of Medicine, FEAM
8. Miikka Korja, Council of Finnish Academies, EASAC
9. Michal Lipschuetz, The Israeli National Academy of Science in Medicine, FEAM
10. Christian Lovis, Swiss Academy of Medical Sciences, FEAM
11. Philippe Moingeon, The French Academy of Pharmacy, FEAM
12. Alison Noble, The Royal Society, EASAC
13. Rui Nunes, Portuguese Academy of Medicine, FEAM
14. Tania Pencheva, The Bulgarian Academy of Sciences, EASAC
15. Renata Raidou, The Austrian Academy of Sciences, EASAC
16. David Ríos Insua, The Spanish Royal Academy of Sciences, EASAC
17. Daniel Rückert, The UK Academy of Medical Sciences, FEAM
18. Isabelle Salmon, The Royal Academies for Science and the Arts of Belgium, EASAC
19. Miroslav Samaržija, The Croatian Academy of Sciences and Arts, EASAC
20. Ewout Steyerberg, Royal Netherlands Academy of Arts and Sciences, FEAM
21. Predrag Tadić, The Academy of Medical Sciences of the Serbian Medical Society, FEAM
22. Sebastian Vollmer, The German National Academy of Sciences Leopoldina, EASAC

Peer reviewers

1. Antonella Cardone, Cancer Patients Europe (invited by FEAM)
2. Ana Gavrovska, The Academy of Medical Sciences of the Serbian Medical Society (invited by FEAM)
3. Paul Hoff, Swiss Academy of Medical Sciences (invited by FEAM)
4. Mário Silva, Academy of Sciences of Lisbon (invited by EASAC)
5. Eline Slagboom, Royal Netherlands Academy of Arts and Sciences (invited by FEAM)
6. Ján Szöllös, Slovak Academy of Sciences (invited by EASAC)
7. Alberto Tozzi, Accademia Nazionale di Medicina (invited by FEAM)

Acknowledgement of AI use

Generative AI tools (OpenAI's GPT-5.6 Pro or earlier) were used for technical support in the production of this report, including assistance with harmonising text and formatting references. These tools were not used for drafting the text or for analysis or interpretation of evidence. All AI-generated suggestions were critically reviewed by the authors and the main contributors, who bear full responsibility for the accuracy and integrity of the report.

Working group composition and timetable

This report is the result of multiple discussions between FEAM and the EASAC Biosciences and Public Health Panel from 2023 to 2026.

An EASAC-FEAM joint scoping paper was finalised in January 2023 and presented at a conference ('Sustainable AI to drive Global Health') in Gothenburg in May 2023, co-organised by FEAM under the auspices of the Swedish Presidency of the Council of the European Union and with interventions from the European Parliament rapporteur for the European Health Data Space, the DG SANTE legal lead on AI, and members of the Swedish Parliament and Health Authority.

EASAC and FEAM's member academies nominated persons to form a Working Group composed of experts acting in an individual capacity (see the list of Working Group members above). The Working Group, chaired by André Knottnerus (EASAC) and Luis Martí-Bonmatí (FEAM), met several times online from 2024 to 2026 and in-person in December 2024 at FEAM in Brussels.

A final version of the report was reviewed by peers also nominated by member academies (see the list of Peer reviewers above) in March 2026.

Glossary

Term	Definition
Artificial intelligence (AI)	In this report, artificial intelligence is used as a shorthand for AI systems and related computational methods, understood in line with the EU AI Act definition of an AI system. In healthcare, AI systems may generate predictions, content, recommendations, or decision-support outputs for purposes such as diagnosis, triage, risk prediction, treatment planning, clinical documentation, research, public health, and health-system management. Their use requires appropriate validation, data governance, human oversight, transparency, equity safeguards, and accountability.
Accountability	The clear assignment of responsibility for the design, development, validation, deployment, monitoring, and consequences of an AI system. In healthcare, accountability may be shared among developers, manufacturers, healthcare institutions, clinicians, regulators, and other actors, but should not be allowed to become diffuse or unclear.
Adversarial attack	An attempt to manipulate an AI system's inputs, training data, or operating environment to produce incorrect, unsafe or unintended outputs. In healthcare, adversarial attacks may create risks for diagnosis, treatment decisions, data integrity, and patient safety.
AI Act	The EU regulation establishing harmonised rules for the development, placing on the market, deployment and use of artificial intelligence systems. It follows a risk-based approach, with specific obligations for high-risk AI systems, including many healthcare-related applications.
AI-specific consent	A non-standard term referring to specific consent for the use of AI in care or research. It should be used carefully, because separate AI-specific consent is not always legally required in routine care, although patient information, transparency and the opportunity to ask questions may still be needed.
AI system	A machine-based system that can operate with some degree of autonomy and generate outputs such as predictions, recommendations, classifications, or decisions that influence real or virtual environments.
Algorithmic bias	Systematic error or distortion in an AI system that can produce unfair, inaccurate or unequal outcomes for particular groups, settings or populations. Bias may arise from data, model design, deployment context, or unequal access to the benefits of AI.
Analytical validation	The process of demonstrating that a system accurately and reliably measures, detects, classifies, or predicts what it is technically intended to measure, detect, classify or predict. Analytical validation is distinct from clinical validation, which asks whether the system is useful and safe for its intended clinical purpose.
Anonymisation	Processing data so that individuals can no longer be identified, either directly or indirectly, using means reasonably likely to be available. Properly anonymised data are no longer personal data, but anonymisation can be difficult to guarantee in rich health datasets.
Automated decision-making	A decision-making process in which a system makes or materially shapes a decision with limited or no human involvement. In healthcare, automated decision-making raises particular issues for safety, accountability, transparency, patient rights, and the role of professional judgement.
Clinical decision support	Tools or systems that provide clinicians, patients, or other users with information, alerts, recommendations, risk estimates, or other outputs to support healthcare decisions. Clinical decision support may be rule-based or AI-based.
Clinical validation	The process of demonstrating that an AI system is clinically meaningful and performs adequately for its intended medical purpose, target population, and care setting. Clinical validation should consider clinical benefit, safety, usability, equity, and the consequences of errors.
Co-design / co-creation	The involvement of intended users and affected communities, such as clinicians, patients, carers, data scientists, and health system managers, as partners in the design, testing, implementation and evaluation of AI systems. Co-design can improve usability, trustworthiness and uptake.
Cybersecurity	The protection of digital systems, networks, software and data from unauthorised access, misuse, disruption, or attack. In healthcare AI, cybersecurity is closely linked to patient safety, data protection, system reliability, and public trust.
Data altruism	The voluntary sharing of data by individuals or organisations for objectives of general interest, such as health research, public health, innovation, or policy development. Data altruism requires appropriate safeguards, transparency, and governance.
Data federation	An approach that enables data from multiple sources or organisations to be queried, analysed or used together without necessarily moving all data into a single central database. In healthcare, data federation can support collaborative research and AI development across hospitals, registries, or research institutions while reducing some privacy and governance risks associated with centralising sensitive health data.

Term	Definition
Data governance	The rules, roles, processes, and safeguards that determine how data are collected, accessed, shared, used, protected, audited, and deleted. In healthcare AI, data governance is central to privacy, security, data quality, accountability, and public trust.
Data quality	The extent to which data are fit for their intended purpose. Relevant dimensions include accuracy, completeness, consistency, timeliness, representativeness, provenance, interoperability, and clinical relevance. Poor data quality can undermine safety, fairness, and reliability.
Dataset provenance	Information about the origin, collection, processing, transformation, custodianship, and permitted uses of a dataset. Provenance is important for assessing reliability, bias, representativeness, accountability, and suitability for AI development or validation.
Deep learning	A type of machine learning based on artificial neural networks with many layers. Deep learning is widely used for image recognition, speech recognition, natural language processing and complex prediction tasks, but can be difficult to interpret and requires careful validation.
Dynamic consent	A model of consent that allows individuals to make, review, and change preferences about data use over time, often through digital tools. Dynamic consent may support more flexible engagement in research or learning health systems, but it also creates practical and governance challenges.
Electronic health record	A digital record of a person's health information, usually generated and maintained in healthcare settings. It may include diagnoses, medications, laboratory results, imaging, procedures, clinical notes, and care plans.
Embedded ethics	The integration of ethical analysis, stakeholder engagement and reflection into the design, development, deployment, and monitoring of AI systems, rather than treating ethics as a separate end-stage review.
Equity	Fair distribution of healthcare opportunities, access, risks, and benefits across individuals, groups, and regions, with attention to existing disadvantage and health inequalities. In AI, equity requires attention to unequal data quality, representation, access, and impact.
European Health Data Space	The EU framework intended to improve access to and use of electronic health data for healthcare delivery, research, innovation, policy-making, and regulation, while maintaining safeguards for privacy, security, and fundamental rights.
Explainability	The ability to provide understandable reasons or evidence for an AI system's output, especially where that output affects clinical decisions, patient care or accountability. Explainability may involve technical methods, user-facing explanations, documentation, or contextual information about limitations.
External validation	Testing an AI model on independent data that were not used for model development, preferably from different sites, populations, or time periods, to assess whether performance generalises beyond the original development setting.
FAIR principles	Principles for making data findable, accessible, interoperable, and reusable. In health data governance, FAIR does not mean unrestricted openness, but rather well-described and appropriately governed access and use.
Fairness	The principle that AI systems should avoid unjustified discrimination or unequal performance across individuals or groups. Fairness may require technical, ethical, legal, and social judgement, and different fairness measures may sometimes conflict.
Federated learning	A machine learning approach in which models are trained across multiple sites or datasets without moving raw data to a central location. Model updates, rather than the underlying data, are shared. Federated learning can support privacy-preserving collaboration but still requires governance, security, and validation.
Foundation model	A large AI model trained on broad datasets that can be adapted to many different tasks, including through fine-tuning or prompting. Foundation models can support many downstream applications but may also carry inherited biases, opacity, and safety risks.
Generalisability	The ability of an AI system to maintain acceptable performance when used in populations, institutions, countries, or clinical settings beyond those in which it was developed or validated.
Generative AI	AI systems that generate new content, such as text, images, code, audio, video, or synthetic data, based on patterns learned from training data. In healthcare, generative AI may support documentation, communication, education, research, and workflow tasks, but requires safeguards against error, bias, and hallucination.
Hallucination	A plausible-sounding but incorrect, unsupported or fabricated output generated by an AI system, especially a large language model. In healthcare, hallucinations can create safety risks if users rely on them without verification.

Term	Definition
Health data	Data relating to a person's physical or mental health, including information generated through healthcare services, diagnostics, medical devices, public health systems, research, registries, or patient-generated sources.
High-risk AI system	An AI system that falls within categories defined by the EU AI Act as posing significant risks to health, safety, or fundamental rights. High-risk systems are subject to stricter requirements for data governance, documentation, transparency, human oversight, accuracy, robustness, and cybersecurity.
Human oversight	Measures that allow appropriately qualified people to understand, monitor, challenge, override, or stop the use of an AI system when needed. In healthcare, human oversight should support, not undermine, professional judgement and patient safety.
Informed consent	A person's voluntary agreement to a healthcare intervention or research participation after receiving adequate information about its purpose, benefits, risks, and alternatives. In the AI context, informed consent should be distinguished from transparency, data protection consent and general patient information duties.
Interoperability	The ability of systems, organisations, or datasets to exchange information and use it meaningfully across technical, semantic, organisational, and legal boundaries. Interoperability is essential for high-quality data use, cross-border healthcare, and AI development.
Interpretability	The extent to which humans can understand how a model reaches its outputs. Interpretability often refers to the model's internal logic, while explainability may also include external explanations of model behaviour or output.
Large language model (LLM)	A type of AI model trained on very large collections of text and other data to understand, summarise, generate, and transform language. In healthcare, LLMs may support clinical documentation, literature synthesis, patient communication, coding, triage support, and administrative tasks. Their outputs require careful verification because they may be inaccurate, biased, incomplete, or fabricated.
Learning health system	A health system that systematically uses data generated through care, research, and public health activity to improve practice, policy, and outcomes over time. AI may support learning health systems, but only where data governance, evaluation, professional responsibility, and public trust are strong.
Machine learning	Machine learning is a type of artificial intelligence in which computer systems learn patterns from data to improve performance on a task without being explicitly programmed for every rule. In healthcare, machine learning can help identify patterns in medical images, laboratory data, genomic data, or patient records to support diagnosis, risk prediction, service planning, or treatment selection. Machine learning systems depend on the quality, diversity, and relevance of the data they are trained and tested on, which affects reliability, fairness, and applicability across patient populations.
Medical device	A product, including software, intended by the manufacturer for a medical purpose such as diagnosis, prevention, monitoring, prediction, prognosis, treatment, or alleviation of disease. Some AI systems used in healthcare may qualify as medical devices under EU law.
Model drift	A decline or change in AI model performance over time because the data, clinical practice, patient population, disease patterns, or healthcare context differ from those seen during development or validation. Model drift is one reason why post-deployment monitoring is necessary.
Multimodal AI	AI that can process or combine more than one type of input, such as text, images, laboratory data, genomic data, sensor data, or clinical records. In healthcare, multimodal AI may support richer decision support but can increase complexity and validation challenges.
Natural language processing	AI methods for analysing, understanding, generating, or transforming human language. In healthcare, this can include summarising clinical notes, extracting information from records, supporting patient communication or analysing scientific literature.
Personal data	Information relating to an identified or identifiable natural person. Health data are usually a special category of personal data and require additional safeguards under data protection law.
Post-market surveillance	Continuing monitoring of a product or system after it has been placed on the market or deployed, to detect safety, performance, usability, cybersecurity, or other problems over time and to support corrective action where needed.
Pseudonymisation	Processing personal data so that direct identifiers are replaced or separated, while re-identification remains possible with additional information kept separately and securely. Pseudonymised data remain personal data.
Representativeness	The extent to which data used to develop, test or monitor an AI system reflect the populations, clinical settings, and real-world conditions in which the system will be used. Poor representativeness can lead to unsafe or inequitable performance.

Term	Definition
Secondary use of health data	The use of health data for purposes other than the direct care of the individual from whom the data were generated, such as research, innovation, public health, regulatory decision-making, health technology assessment, or health system planning.
Secure processing environment	A controlled digital environment in which authorised users can access and analyse sensitive data under strict safeguards, often with limits on data export, copying, and re-identification. Secure processing environments are important for governed secondary use of health data.
Software as a Medical Device (SaMD)	Software intended for one or more medical purposes that performs these purposes without being part of a hardware medical device. AI-based diagnostic, monitoring, or decision-support tools may fall into this category.
Transparency	The provision of clear and accessible information about an AI system's purpose, role, data sources, limitations, performance, and governance, so that users and affected people can understand how it is used and where responsibility lies.
Virtual care	The use of digital technology, such as video consultations, apps, digital tools, or remote monitoring, that allows a patient to receive care, advice, or monitoring without physically attending a healthcare setting. Virtual care is not necessarily AI-based, but it may incorporate AI-enabled triage, monitoring, communication, or decision-support tools.

Abbreviations

AI	Artificial intelligence
AICC	AI Code of Conduct
AI/ML	Artificial intelligence/machine learning
AIA	Artificial Intelligence Act; Regulation (EU) 2024/1689
ALLEA	European Federation of Academies of Sciences and Humanities ('All European Academies')
BEUC	European Consumer Organisation
DICOM	Digital Imaging and Communications in Medicine
EASAC	European Academies Science Advisory Council
EC	European Commission
EHDS	European Health Data Space
EHR	Electronic health record
EMA	European Medicines Agency
EU	European Union
FAIR	Findable, Accessible, Interoperable and Reusable
FDA	Food and Drug Administration, USA
FEAM	Federation of European Academies of Medicine
FHIR	Fast Healthcare Interoperability Resources
GDPR	General Data Protection Regulation
IFU	Instructions for use
LHS	Learning health system
LLM	Large language model
MDR	Medical Device Regulation; Regulation (EU) 2017/745
MHRA	Medicines and Healthcare products Regulatory Agency, UK
NAM	National Academy of Medicine
NMPA	National Medical Products Administration
OECD	Organisation for Economic Co-operation and Development
PIPL	Personal Information Protection Law
RDA	Research Data Alliance
SaMD	Software as a Medical Device
TEHDAS	Towards the European Health Data Space
TRIPOD	Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis
UNESCO	United Nations Educational, Scientific and Cultural Organization
WHO	World Health Organization

Appendix

Table A1 A structured comparative analysis of the major international frameworks on AI in healthcare

	UNESCO	EU AI Act	OECD AI Principles	GDPR/privacy standards	FDA/US tech-regulatory approaches
Purpose, scope, and health focus	Broad, rights-based ethics applicable across sectors, with explicit emphasis on health to ensure humane, equitable, and accountable AI deployment	Health AI is treated under risk-based regulation; high-risk health uses (e.g. clinical decision support) require conformity assessments, robust governance, and post-market monitoring.	Promote human-centred values, fairness, transparency, robustness; health AI should align with these while enabling public welfare	Focus on lawful processing of health data, data minimisation, purpose limitation, data subject rights; data governance deeply shapes how health AI is trained and deployed.	Emphasise safety, clinical validity, post-market surveillance, and transparent documentation; a risk-based medical-device lens for AI/ML in health.
Core ethical principles	Human rights, non-discrimination, autonomy, beneficence, justice, transparency, accountability, and participation	Strong emphasis on transparency, human oversight, explainability where feasible, and traceability; conform to regulatory requirements for high-risk AI.	Justice, fairness, non-discrimination, empowerment, and accountability; international cooperation to uphold these.	Data minimisation, purpose limitation, consent where appropriate, data subject rights; privacy-by-design is central.	Safety, effectiveness, analytical validity, clinical relevance; explainability and clinician oversight where it impacts patient care.
Data governance and privacy	Data quality, provenance, consent, privacy protections, and stewardship; cross-border considerations stressed; emphasis on avoiding misuse and bias in data.	Strong transparency and governance around data used in high-risk health AI; data governance, documentation, and auditability are required.	Fair data use, non-discrimination, accessibility, and governance; cross-border data flows with safeguards.	Rigorous data protection standards; strong rights for data subjects; impact on training data, model updates, and data sharing	Emphasis on data integrity, traceability, and post-market data collection to ensure ongoing safety/effectiveness
Transparency, explainability, and auditability	Advocacy for meaningful transparency; documentation of purposes, models, data sources, limitations; participatory governance and public accountability	Requires documentation (model cards, technical documentation), post-market monitoring; higher demand for explainability in high-risk use	Encourages clear disclosure and human-centred explanations; supports audits and governance dashboards.	Data processing transparency; 'right to explanation' debates exist, but focus is broader privacy rights and purpose specification.	Emphasis on risk-based disclosure, labeling, clinician-facing information, and traceability of model updates.
Safety, risk management, and governance in deployment	Systematic risk assessments, safeguards, human oversight, redress mechanisms, ongoing monitoring	Formal risk assessment, validation in clinical contexts, governance structures, post-market surveillance for high-risk health AI.	Risk-based thinking; governance that anticipates societal impacts; invites monitoring and accountability mechanisms.	Privacy risk considerations; Data Protection Impact Assessments where processing is high risk; data safety measures.	Pre-market clearance/approval for high-risk devices, continuous real-world performance monitoring, change-management for updates.

Table A1 (continued)

	UNESCO	EU AI Act	OECD AI Principles	GDPR/privacy standards	FDA/US tech-regulatory approaches
Human oversight and clinical integration	Emphasises human-centric care, clinician and patient autonomy, and appropriate human oversight in care decisions assisted by AI.	Requires human oversight for many high-risk health AI systems; design to integrate with clinical workflows and allow clinician intervention.	Supports human-centred design and the safeguarding role of professionals and users.	While not hardware-focused, supports safeguards that respect human oversight and decision-making rights.	Clinician-facing labelling and the need for clinician oversight in AI-enabled therapies or devices.
Equity, inclusion, and addressing bias	Explicit focus on equity, avoiding amplification of health disparities; inclusive design and participation of diverse populations.	Bias and fairness considerations embedded in risk management and data governance; non-discrimination requirements in care.	Emphasis on inclusive design and fairness; reduce barriers to benefit across populations.	Bias concerns addressed by data protection and fairness in processing; prompts for impact assessments in high-risk contexts.	Bias and fairness considerations by clinical validation in diverse populations; post-market surveillance for unequal performance.
Accountability and liability	Clear accountability across stakeholders; redress pathways for rights violations; governance mechanisms to assign responsibilities.	Legal accountability for providers, implementers, and developers; traceability and documentation support accountability.	Shared responsibility among designers, users, and regulators; emphasis on governance to hold actors to account.	Accountability by data controllers and processors; regulatory penalties for non-compliance; rights-based remedies for data subjects.	Accountability through regulatory clearance, post-market reporting, and corrective actions for AI-enabled devices.
International cooperation and capacity building	Strong emphasis on global governance, standards, and support for capacity building in diverse contexts (including low- and middle-income countries).	Regionally harmonised rules; mutual recognition and cooperation within the EU; global influence through export of standards.	Cross-country collaboration on standards, data governance, and shared evaluations.	Cross-border data protection with adequacy decisions; global impact on data practices.	Global engagement in harmonisation efforts (e.g. ICH guidance); supports international manufacturers and multi-country deployments.
Implementation challenges and practical steps	Balancing universal rights-based norms with diverse national contexts; resource constraints for capacity-building; ensuring meaningful participation.	Regulatory complexity; need for compliance infrastructure; ongoing post-market obligations.	Navigating diverse regulatory regimes while maintaining core ethical commitments.	Data protection compliance complexity; safeguarding data while enabling health AI innovation.	Regulatory rigour for evolving machine learning-based devices; managing updates without stifling innovation.

Table A2 A comparative snapshot by region/global frameworks of how ethical issues in AI applications in healthcare align with UNESCO's November 2021 Recommendation on the Ethics of Artificial Intelligence

	Alignment with UNESCO	Typical focus areas	Key takeaway
Europe (data protection, risk-based regulation, health AI governance)	Strong emphasis on human rights, privacy, non-discrimination, transparency, and accountability; risk-based oversight of AI used in health.	1. Data protection and privacy for health data (GDPR-compliant processing, data minimisation, consent, security). 2. Regulation of high-risk AI in health (requiring risk assessments, human oversight, documentation, and accountability for outcomes). 3. Governance mechanisms that encourage transparency and patient safety, with responsibility distributed among developers, manufacturers, and healthcare providers.	Europe operationalises UNESCO principles through privacy-by-design, risk-based regulation, and clear duty of care in clinical contexts.
United States (FDA-regulated health AI, sector-specific governance)	Emphasis on safety, efficacy, accountability, and transparency in health AI; respect for patient rights within regulatory boundaries; non-discrimination and bias considerations in clinical tools.	1. Regulatory approval and ongoing performance monitoring for AI/ML-based medical devices and software as a medical device (SaMD). 2. Emphasis on clinical validation, real-world performance, and post-market surveillance. 3. A more sector-specific (healthcare) approach with less overarching national data-rights framework than the EU; privacy protections occur primarily under a mosaic of federal/state laws.	UNESCO's call for safety, accountability, and human-centred design is reflected in US regulatory practice, with a strong emphasis on clinical-grade evidence and ongoing oversight.
OECD and global policy initiatives (harmonisation, shared principles)	Strong convergence on human-centred AI, transparency, accountability, inclusiveness, and safeguarding rights; emphasis on governance, data stewardship, and cross-border cooperation.	1. High-level policy principles for trustworthy AI, including health applications, risk management, and stakeholder engagement. 2. Guidelines for data governance, algorithmic transparency where feasible, and mechanisms for accountability. 3. Active promotion of international cooperation, capacity-building, and sharing best practices.	UNESCO's normative framework dovetails with OECD's strategic orientation towards principled, multi-stakeholder governance and cross-border collaboration in health AI.
Health-focused global guidance (WHO and allied health ethics initiatives)	Strong emphasis on ethics, equity, rights-based considerations, safety, and patient welfare; insistence on inclusive access and protecting vulnerable groups.	1. Ethics guidelines for digital health interventions, data stewardship, and privacy in health contexts. 2. Guidance on equitable access to AI-enabled health benefits, avoiding exacerbation of health disparities. 3. Recommendations for governance structures that involve patients, clinicians, researchers, and policy-makers.	UNESCO's themes on rights, equity, and safety are central in WHO-style health ethics discussions, reinforcing global health equity and patient protection.
Other national/regional exemplars (privacy, clinical governance, capacity-building)	General adherence to rights-based, transparency, and accountability principles; emphasis on governance, bias mitigation, and stakeholder inclusion.	1. National privacy laws and sector-specific guidelines for AI in health (data governance, consent, data sharing, and security). 2. Professional and clinical governance frameworks for AI tools used in diagnostics, imaging, and decision support. 3. Capacity-building, workforce training, and public engagement to understand AI in health.	National approaches often reflect UNESCO principles through privacy protections, clinical governance, and investment in AI literacy and ethics training.

Table A3 Continent-focused overview of policy priorities tailored to UNESCO's November 2021 Recommendation on the ethics of AI

	Alignment with UNESCO	Policy priorities	Key gaps	Actions and metrics
Africa	strong emphasis on rights, privacy, inclusion, and capacity-building; health data governance is emerging.	1. Adopt a continental health AI ethics framework aligned with UNESCO; harmonise high-level data protection and health data governance (privacy-by-design, consent, data minimisation). 2. Establish regional bodies for multi-stakeholder oversight (patients, clinicians, researchers, regulators). 3. Invest in digital health infrastructure and AI literacy; support locally led health AI pilots focused on underserved populations.	Fragmented national laws; limited, uneven data availability; scarce funding for governance and oversight; limited cross-border data exchange.	1. Quick wins: national AI ethics guidelines for health; regional data-sharing agreements with privacy safeguards; capacity-building programmes for regulators and clinicians. 2. Metrics: number of countries with health AI ethics guidelines; data protection impact assessments in health; proportion of health AI pilots with equity-focused design.
Americas	emphasis on safety, efficacy, transparency, equity; mixed regulatory maturity (Canada/USA strong, Latin America variable)	1. Promote regional standards for health AI evaluation, validation, and post-market surveillance; ensure non-discrimination in clinical AI tools. 2. Harmonise data governance and consent frameworks for health data, enabling safe cross-border use where appropriate. 3. Strengthen patient and clinician engagement in governance structures.	Fragmented regulations across countries; disparities in infrastructure and data availability; varying enforcement capacities.	1. Regional ethics charters and shared evaluation benchmarks; joint registries of health AI devices; targeted funding for ethics training in LMICs. 2. Metrics: number of cross-border data-use pilots; time-to-regulatory clearance for health AI; equity impact indicators in deployed tools.
Asia	strong diversity—principles are widely endorsed, but regulatory maturity varies; risk-based oversight increasing.	1. Accelerate national AI ethics and health data protection regimes; establish risk classifications and governance for health AI with clear accountability. 2. Create regional knowledge-sharing platforms, regulatory sandboxes, and joint validation standards. 3. Prioritise equity: ensure rural and low-resource populations benefit from health AI; address language and cultural appropriateness.	Wide heterogeneity in data protection laws; uneven clinical validation; potential state-driven governance dynamics; digital divide.	1. Shared region-wide guidelines for health AI evaluation; capacity-building for regulators; pilots in essential health services (telemedicine, triage, diagnostics) with independent audits. 2. Metrics: adoption rate of health-AI impact assessments; number of regulatory sandboxes; access indicators for AI-enabled health services in underserved areas.
Europe	strongest alignment; mature data protection (GDPR), explicit risk-based AI governance (EU AI Act), robust health governance.	1. Implement and enforce EU AI Act provisions for health AI; ensure cross-border data sharing with privacy safeguards. 2. Strengthen independent conformity assessment, transparency of clinical algorithms, and human oversight where appropriate. 3. Targeted measures to mitigate bias in diagnostic and treatment tools; protect vulnerable groups and ensure accessibility.	Need for consistent national implementation across Member States; improving real-world monitoring and explainability in clinical settings.	1. Concrete implementation roadmaps for Member States; ongoing post-market surveillance dashboards; accessibility and bias-reduction milestones. 2. Metrics: number of health AI products regulated; cross-border data flows compliant with GDPR-like standards; patient-reported outcome equity indicators.

Table A3 (continued)

	Alignment with UNESCO	Policy priorities	Key gaps	Actions and metrics
Oceania	strong privacy protections, emphasis on indigenous data rights, and equitable access in health.	1. Align national AI ethics with UNESCO; advance health AI that addresses remote and indigenous communities; protect data sovereignty and culturally appropriate data governance. 2. Invest in telehealth and remote diagnostics with culturally sensitive design and consent processes.	Smaller markets and limited local data; disparities between urban and remote/island communities; need for regional collaboration.	1. Pacific-wide guidelines for health data stewardship; targeted funding for rural/remote health AI pilots; governance models that include indigenous and community voices. 2. Metrics: uptake of health AI in remote communities; rate of culturally appropriate AI tool development; indigenous data sovereignty compliance indicators.

Table A4 A concise comparison of how the US and China differ in ethical issues around AI in healthcare

	USA	China
Privacy and data governance	Strong emphasis on patient privacy (Health Insurance Portability and Accountability Act, USA) and protections for protected health information; data sharing often balanced with consent and de-identification. State laws (e.g. California's CCPA/CPRA) add more privacy rights.	Privacy protections governed by the Personal Information Protection Law (PIPL) and related cybersecurity laws; tighter controls on data collection, storage, and cross-border transfer; emphasis on data security and sovereignty, with data localisation in many health contexts.
Informed consent and autonomy	Emphasis on patient autonomy and informed consent for data use and AI-assisted decisions; transparency about when AI is involved in care.	Consent frameworks exist but health AI deployment often operates within a coordinated, state-linked health system; emphasis on public health goals and governance may influence how consent is managed in practice.
Fairness, bias, and discrimination	Growing focus on bias in AI tools, disparate impact across racial/ethnic groups, and bias mitigation; regulatory and professional bodies advocate for fairness in clinical AI.	Bias concerns exist, but policy framing often centres on societal cohesion, equal access, and improving population health; bias mitigation is discussed within national AI ethics guidelines and health tech governance.
Transparency and explainability	Calls for transparency in AI decision support; explainability is valued, but many high-stakes medical AI systems are still 'black boxes' but may be accepted if they are rigorously validated and appropriately supervised.	Transparency and accountability are addressed within national ethics guidance and regulatory expectations; the degree of public-facing explainability varies by device and supervision level.
Accountability and liability	Liability frameworks (physician, hospital, or device maker) apply; regulatory expectations focus on safety, performance, and post-market surveillance; FDA regulates AI medical devices.	Accountability is framed within regulatory oversight of AI medical devices and health authorities; greater emphasis on governmental oversight and conformity with national standards and protocols.
Safety, clinical oversight, and human-in-the-loop	Emphasis on safety, risk management, and maintaining clinician oversight; FDA requires rigorous validation, clinical evidence, and ongoing monitoring for AI SaMDs.	Safety and reliability are central to regulatory approval and post-market surveillance; strong emphasis on supervised deployment in hospitals and alignment with national health goals.
Equity and access	Disparities in access to AI-enabled care and potential for unequal benefits; policy discussions focus on equitable access, payer coverage, and bias mitigation.	Large-scale adoption aims to reduce regional and rural disparities; policy supports expanding AI health tools to public health and primary care networks, with governance to ensure broad access.

Table A4 (continued)

	USA	China
Regulatory and governance approaches (how they address the ethics)	1. Regulatory framework: FDA oversees AI-based medical devices (AI/ML–SaMD) with a life-cycle approach; emphasis on safety, effectiveness, and post-market surveillance; concepts such as Good Machine Learning Practice and a predetermined change control plan for learning systems. 2. Privacy and data protection: Health Insurance Portability and Accountability Act (USA) governs protected health information; state privacy laws (e.g. CCPA/CPRA) add protections; data use often requires patient consent or de-identification; IRBs/ethical review for research use. 3. Standards and accountability: Professional bodies (AMA, ACM/IEEE) publish ethics guidelines; NIST is developing AI risk management standards; civil rights enforcement by OCR for discrimination concerns. Practical impact: Innovation is market-driven with strong regulatory gates for safety; greater emphasis on individual rights, consent, and clinician accountability; data sharing balanced against privacy laws.	1. Regulatory framework: Health AI devices are regulated by NMPA (the national medical-device regulator) with a risk-based approach; post-market surveillance and conformity with national standards are emphasised; national health policy supports rapid deployment of AI in hospitals and public health. 2. Privacy and data protection: PIPL and related laws regulate personal data processing, with strict controls on health data, data localisation, and cross-border transfers; compliance requires explicit protections and risk assessments. 3. Ethics and guidance: National-level ethics guidelines for AI (and sector-specific health AI guidelines) emphasise safety, societal benefit, and alignment with public interests; governance often features centralised oversight and standardised deployment in public health systems. Practical impact: Faster, more centralised deployment of AI in healthcare; strong data governance and cross-border transfer controls; ethics guidance is integrated into national strategy and regulatory approval processes.

Table A5 Example checklist aligning reporting, regulatory and ethical considerations for AI in healthcare

Below is a concise compliance checklist for AI in healthcare, harmonising three key frameworks: the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) + AI reporting guideline ([Collins et al. 2024](#)), the EU AI Act, and the National Academy of Medicine (NAM) AI Code of Conduct (AICC). By integrating transparent reporting (TRIPOD), legal-regulatory safeguards (EU AI Act), and ethical conduct (NAM AICC), this checklist supports trustworthy, compliant, and socially responsible AI deployment in EU healthcare. The checklist should be used in the following way: (1) complete each item during development and pre-deployment; (2) document compliance with TRIPOD + AI Act + NAM AICC; (3) review regularly (e.g. annually or after major updates), and (4) embed outputs in technical documentation, regulatory submissions, and public-facing materials (instructions for use (IFU), patient literature, clinical governance reports).

Principle	Checklist Item	References
1. Scope and risk classification	Define intended use, target population, and classify risk level under the EU AI Act (high-risk if diagnostic/support function).	Collins et al. (2024) , TRIPOD+AI items 1 and 3a–b; Regulation (EU) 2024/1689 , Article 6(1)–(1c) and Annex I; Regulation (EU) 2017/745 , Article 52 and Annex VIII, Rule 11; Regulation (EU) 2017/746 , Article 48 and Annex VIII.
2. Transparent reporting	Use TRIPOD+AI or TRIPOD-LLM checklists (Title, Abstract, Methods, Ethics) to ensure clarity in model design and evaluation.	Collins et al. (2024) , TRIPOD+AI items 1–27; Gallifant et al. (2025) , TRIPOD-LLM items 1–19.
3. Ethical commitments	Align with NAM AICC’s 6 commitments and 10 principles—e.g. accountability, equity, autonomy, privacy, reliability, environmental sustainability.	National Academy of Medicine (2025) , Tables 3-2 and 3-3.

Table A5 (continued)

Principle	Checklist Item	References
4. Data representativeness and bias	Describe dataset demographics, missingness, subgroups; evaluate bias metrics and mitigate unfairness.	Collins et al. (2024) , TRIPOD+AI items 3c, 5a, 7, 11, 14, 20b–c, 23a and 25–26; Regulation (EU) 2024/1689 , Articles 4a and 10; National Academy of Medicine (2025) , “Equitable” and “Ensure Equity”.
5. Explainability and clinician oversight	Provide model explainability; ensure human-in-the-loop or ability to override for high-risk decisions.	Regulation (EU) 2024/1689 , Articles 13–14; Lekadir et al. (2025) ; National Academy of Medicine (2025) , “Transparent” and “Accountable”; for LLM studies, Gallifant et al. (2025) .
6. Privacy and cybersecurity	Ensure GDPR/EHDS compliance, data minimisation, anonymisation, secure data practices; aligned with NAM privacy and safety principles.	Regulation (EU) 2016/679 , Articles 5, 25, 32 and 35; Regulation (EU) 2024/1689 , Article 15(5); Regulation (EU) 2025/327 , Chapter IV and Article 105; National Academy of Medicine (2025) , “Secure”.
7. Validation and monitoring	Conduct internal/external validation with performance metrics (CI); implement continuous monitoring and recalibration post-deployment.	Collins et al. (2024) , TRIPOD+AI items 12a, 12c–g, 16, 20c, 23a–b and 24; Regulation (EU) 2024/1689 , Articles 9, 15 and 72; National Academy of Medicine (2025) , “Safe”, “Effective”, “Adaptive” and “Monitor Performance”.
8. Stakeholder engagement and consent	Involve clinicians, patients, ethicists; document consent, conflicts of interest, explain use, funding sources.	Collins et al. (2024) , TRIPOD+AI items 17, 18a–b and 19; Regulation (EU) 2016/679 , Articles 6 and 9; National Academy of Medicine (2025) , “Engaged”, “Accessible” and “Engage Impacted Individuals”.
9. Reporting and audit trail	Archive code and models (as allowed); maintain audit logs; prepare IFUs tailored to user groups as per EU AI Act.	Collins et al. (2024) , TRIPOD+AI items 18c–f and 22; Regulation (EU) 2024/1689 , Articles 11–13 and 18–19 and Annex IV; Regulation (EU) 2017/745 , Annex I, sections 23.1 and 23.4; Regulation (EU) 2017/746 , Annex I, sections 20.1 and 20.4.1.
10. Environmental and resource stewardship	Report model training energy/resource use; commit to sustainable, efficient designs per NAM environmental principle.	National Academy of Medicine (2025) , “Efficient”; Regulation (EU) 2024/1689 , Articles 40(2) and 95(2)(b); for providers of general-purpose AI models, Article 53(1)(a) and Annex XI.

References

- Aboy, M., Minssen, T. and Vayena, E. (2024) Navigating the EU AI Act: implications for regulated digital medical products. *npj Digital Medicine* **7**, 237. <https://doi.org/10.1038/s41746-024-01232-3>
- Aguilar Gomez, F. and Bernal, I. (2023) FAIR EVA: bringing institutional multidisciplinary repositories into the FAIR picture. *Scientific Data* **10**, 764. <https://doi.org/10.1038/s41597-023-02652-8>
- Alicioglu, G. and Sun, B. (2022) A survey of visual analytics for explainable artificial intelligence methods. *Computers & Graphics* **102**, 502–520. <https://doi.org/10.1016/j.cag.2021.09.002>
- ALLEA, EASAC and FEAM (2021) International sharing of personal health data for research. Berlin/Brussels: ALLEA, EASAC and FEAM. https://allea.org/wp-content/uploads/2021/03/International-Health-Data-Transfer_2021_web.pdf
- Antenore, M., Camacho Rodriguez, J.M. and Panizzi, E. (2023) A comparative study of bot detection techniques with an application in Twitter COVID-19 discourse. *Social Science Computer Review* **41**(5), 1520–1545. <https://doi.org/10.1177/08944393211073733>
- Ashery, A.F., Aiello, L.M. and Baronchelli, A. (2025) Emergent social conventions and collective bias in LLM populations. *Science Advances* **11**(20), eadu9368. <https://doi.org/10.1126/sciadv.adu9368>
- Banerjee, S., Alsop, P., Jones, L. and Cardinal, R.N. (2022) Patient and public involvement to build trust in artificial intelligence: a framework, tools, and case studies. *Patterns* **3**(6), 100506. <https://doi.org/10.1016/j.patter.2022.100506>
- Barichella, A. (2023) Regulating artificial intelligence at the EU level: obstacles and prospects. Policy Paper No. 294. Paris: Jacques Delors Institute. https://institutdelors.eu/content/uploads/2025/04/PP294_Regulation_IA_Barichella_EN.pdf
- Barrett, L.F., Adolphs, R., Marsella, S. et al. (2019) Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest* **20**(1), 1–68. <https://doi.org/10.1177/1529100619832930>
- Bathae, Y. (2018) The artificial intelligence black box and the failure of intent and causation. *Harvard Journal of Law & Technology*. <https://jolt.law.harvard.edu/assets/articlePDFs/v31/The-Artificial-Intelligence-Black-Box-and-the-Failure-of-Intent-and-Causation-Yavar-Bathae.pdf>
- Bavli, I., Ho, A., Mahal, R. and McKeown, M.J. (2025) Ethical concerns around privacy and data security in AI health monitoring for Parkinson's disease: insights from patients, family members, and healthcare professionals. *AI & Society* **40**(1), 155–165. <https://doi.org/10.1007/s00146-023-01843-6>
- Bengio, Y., Mindermann, S., Privitera, D. et al. (2025) International AI Safety Report. *arXiv*. <https://doi.org/10.48550/arXiv.2501.17805>
- Berzin, T.M., Topol, E.J. et al. (2025) Preserving clinical skills in the age of AI assistance. *The Lancet* **406**(10513), 1719. [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(25\)02075-6/abstract](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(25)02075-6/abstract)
- BEUC (2020) Artificial intelligence in healthcare. Factsheet, BEUC-X-2020-052. Brussels: BEUC. https://www.beuc.eu/sites/default/files/publications/beuc-x-2020-052_artificial_intelligence_in_healthcare.pdf
- Bhuyan, S.S., Sateesh, V., Mukul, N. et al. (2025) Generative artificial intelligence use in healthcare: opportunities for clinical excellence and administrative efficiency. *Journal of Medical Systems* **49**(1), 10. <https://doi.org/10.1007/s10916-024-02136-1>
- Biermann, J., Horton, J.J. and Walter, J. (2022) Algorithmic advice as a credence good. *SSRN Electronic Journal*. <http://dx.doi.org/10.2139/ssrn.4326911>
- Boine, C. and Rolnick, D. (2023) General purpose AI systems in the AI Act: trying to fit a square peg into a round hole. Paper presented at WeRobot 2023, Boston, MA, USA. <https://www.bu.edu/law/files/2023/09/General-Purpose-AI-systems-in-the-AI-Act.pdf>
- Broeders, D., Schrijvers, E. and Hirsch Ballin, E. (2017) Big data and security policies: serving security, protecting freedom. WRR Policy Brief No. 6. The Hague: Netherlands Scientific Council for Government Policy. https://www.wrr.nl/site/binaries/site-content/collections/documents/2017/01/31/big-data-and-security-policies-serving-security-protecting-freedom/WRR_PB6_BigDataAndSecurityPolicies.pdf
- Budzyń, K., Romańczyk, M., Kitala, D. et al. (2025) Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. *The Lancet Gastroenterology & Hepatology* **10**(10), 896–903. [https://doi.org/10.1016/S2468-1253\(25\)00133-5](https://doi.org/10.1016/S2468-1253(25)00133-5)
- Bühler, K., Höllt, T., Schultz, T. and Vázquez, P.-P. (2025) AI-in-the-loop: the future of biomedical visual analytics applications in the era of AI. *IEEE Computer Graphics and Applications* **45**(2), 90–99. <https://doi.org/10.1109/MCG.2024.3517293>
- Buscarini, L., Romano, P., Cocco, E.S. et al. (2025) Enhancing patient rehabilitation outcomes: artificial intelligence-driven predictive modelling for home discharge in neurological and orthopaedic conditions. *Journal of NeuroEngineering and Rehabilitation* **22**(1), 117. <https://doi.org/10.1186/s12984-025-01654-4>
- Busch, F., Hoffmann, L., Xu, L. et al. (2025) Multinational attitudes toward AI in health care and diagnostics among hospital patients. *JAMA Network Open* **8**(6). <https://doi.org/10.1001/jamanetworkopen.2025.14452>
- Bustamante, A.V., Félix-Beltrán, L., Nwadiuko, J. and Ortega, A.N. (2022) Avoiding Medicaid enrollment after the reversal of the changes in the public charge rule among Latino and Asian immigrants. *Health Services Research* **57**(Suppl. 2), 195–203. <https://doi.org/10.1111/1475-6773.14020>
- Calderone, A., Latella, D., Bonanno, M. et al. (2024) Towards transforming neurorehabilitation: the impact of artificial intelligence on diagnosis and treatment of neurological disorders. *Biomedicine* **12**(10), 2415. <https://doi.org/10.3390/biomedicine12102415>
- Camacho, J.M., Couce-Vieira, A., Arroyo, D. and Ríos Insua, D. (2025) A cybersecurity risk analysis framework for systems with artificial intelligence components. *International Transactions in Operational Research* **33**(2), 798–825. <https://doi.org/10.1111/itor.70049>
- Carvalho, D.V., Pereira, E.M. and Cardoso, J.S. (2019) Machine learning interpretability: a survey on methods and metrics. *Electronics* **8**(8), 832. <https://doi.org/10.3390/electronics8080832>
- Celi, L.A., Cellini, J., Charnipon, M.-L. et al. (2022) Sources of bias in artificial intelligence that perpetuate healthcare disparities: a global review. *PLOS Digital Health* **1**(3), e0000022. <https://doi.org/10.1371/journal.pdig.0000022>
- Chang, Y., Liu, H., Lu, C. and Zhang, N. (2025) SoK: security and privacy risks of healthcare AI. *arXiv*, arXiv:2409.07415v2. <https://doi.org/10.48550/arXiv.2409.07415>
- Cho, T. and Miller, B.J. (2024) Using artificial intelligence to improve administrative process in Medicaid. *Health Affairs Scholar* **2**(2), qxae008. <https://doi.org/10.1093/haschl/qxae008>
- Choukou, M.-A. and Barresi, G. (2025) Artificial intelligence in rehabilitation: latest research and practice. *Frontiers Research Topic*. <https://doi.org/10.3389/fdgth.2025.1606088>
- Ciani, O., Grigore, B., Blommestein, H. et al. (2021) Validity of surrogate endpoints and their impact on coverage recommendations: a retrospective analysis across international health technology assessment agencies. *Medical Decision Making* **41**(4), 439–452. <https://doi.org/10.1177/0272989X21994553>
- Collins, G.S., Moons, K.G.M., Dhiman, P. et al. (2024) TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **385**, e078378. <https://doi.org/10.1136/bmj-2023-078378>
- Cookson, C. (2025) EU project launched to prepare health workers for a digital future. *Financial Times*, 4 March 2025. <https://www.ft.com/content/a56ef5a3-f5d8-446d-ae9b-f503cce20de7>
- CPME (2024) Deployment of AI in healthcare: sector-specific challenges and accelerators. CPME 2024/073 FINAL. Brussels: Standing Committee of European Doctors. <https://www.cpme.eu/>

- [api/documents/adopted/2024/11/cpme_ad_09112024_073.final.policy.on.deployment.of.ai.in.healthcare.pdf](#)
- Cross, J.L., Choma, M.A. and Onofrey, J.A. (2024) Bias in medical AI: implications for clinical decision-making. *PLOS Digital Health* **3**(11), e0000651. <https://doi.org/10.1371/journal.pdig.0000651>
- Dankwa-Mullan, I. (2024) Health equity and ethical considerations in using artificial intelligence in public health and medicine. *Preventing Chronic Disease* **21**, 240245. <https://doi.org/10.5888/pcd21.240245>
- Data Saves Lives (2022) How can we support the patient data journey throughout the healthcare system to build public trust? Data Saves Lives. <https://datasaveslives.eu/blog/how-can-we-support-the-health-data-journey-290722>. Accessed 11 June 2026.
- Dayan, I., Roth, H.R., Zhong, A. et al. (2021) Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature Medicine* **27**(10), 1735–1743. <https://doi.org/10.1038/s41591-021-01506-3>
- de Hond, A.A.H., Leeuwenberg, A.M., Hooft, L. et al. (2022) Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *npj Digital Medicine* **5**, 2. <https://doi.org/10.1038/s41746-021-00549-7>
- Demková, S. (2021) The decisional value of information in European semi-automated decision-making. *Review of European Administrative Law* **14**(2), 29–50. <https://doi.org/10.7590/187479821X16254887670874>
- Dunn, J., Adams, B. and Tayyar Madabushi, H. (2024) Pre-trained language models represent some geographic populations better than others. In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 12966–12976. Torino, Italy: ELRA and ICCL. <https://aclanthology.org/2024.lrec-main.1135/>
- Dubber, T. and Lazar, S. (2025) Military AI cyber agents (MAICAs) constitute a global threat to critical infrastructure. *arXiv*, arXiv:2506.12094. <https://doi.org/10.48550/arXiv.2506.12094>
- EC (2025) Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act). C(2025) 5053 final. Brussels: European Commission. https://ai-act-service-desk.ec.europa.eu/sites/default/files/2025-08/commission_guidelines_on_the_definition_of_an_artificial_intelligence_system_established_by_regulation_eu_20241689_ai_actenglish_nf2skcqftrfdggjavcodopcwz4_112455.PDF
- EFPIA (2024) Patient involvement in HTA. Brussels: EFPIA. <https://www.efpia.eu/news-events/the-efpia-view/blog-articles/patient-involvement-in-hta>. Accessed 11 June 2026.
- El nawawy, M., Hallajian, M., Mitra, G. et al. (2024) Systematically assessing the security risks of AI/ML-enabled connected healthcare systems. In: *Proceedings of the IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE 2024)*, pp. 97–108. <https://doi.org/10.48550/arXiv.2401.17136>
- Elsner, M. et al. (2025) Global Risks Report 2025. Geneva: World Economic Forum. <https://www.weforum.org/publications/global-risks-report-2025/>
- EMA (2024) Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle. Amsterdam: European Medicines Agency. https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf
- EU (2016) Regulation (EU) 2016/679 of 27 April 2016 (General Data Protection Regulation). *Official Journal of the European Union* **L119**, 1–88.
- EU (2017a) Regulation (EU) 2017/745 of 5 April 2017 on medical devices. *Official Journal of the European Union* **L117**, 1–175.
- EU (2017b) Regulation (EU) 2017/746 of 5 April 2017 on in vitro diagnostic medical devices. *Official Journal of the European Union* **L117**, 176–332.
- EU (2024) Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence. *Official Journal of the European Union*. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
- EU (2025) Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847. *Official Journal of the European Union*, L 2025/327, 5 March 2025. <https://eur-lex.europa.eu/eli/reg/2025/327/oj/eng>
- EU (2026) Regulation (EU) 2026/1744 of 8 July 2026 amending Regulation (EU) 2024/1689 as regards the simplification of the implementation of harmonised rules on artificial intelligence. *Official Journal of the European Union* 2026/1744, 24 July 2026. <https://eur-lex.europa.eu/eli/reg/2026/1744/oj/eng>
- FDA (2019) Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD). Silver Spring, MD: Food and Drug Administration. <https://apo.org.au/node/228371>
- FDA (2025a) Artificial intelligence in software as a medical device. Silver Spring, MD: Food and Drug Administration. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-software-medical-device>. Accessed 11 June 2026.
- FDA (2025b) Good Machine Learning Practice for Medical Device Development: Guiding Principles. Silver Spring, MD: Food and Drug Administration. <https://apjohnndim.com/wp-content/uploads/suggested-papers/AI%20and%20XAI%20Guidelines/Good%20Machine%20Learning%20Practice%20%28GMLP%29%20now%20aligned%20with%20an%20IMDRF%20final%20document%20%28Jan%202025%29.pdf>
- FEAM (2023) Sustainable AI to drive global health: a white paper summarizing opportunities of data and artificial intelligence in healthcare within the European Union and beyond. Brussels: FEAM. <https://www.feam.eu/wp-content/uploads/Sustainable-AI-to-Drive-Global-Health-white-paper-11-Sep.pdf>
- Finlayson, S.G. et al. (2019) Adversarial attacks on medical machine learning. *Science* **363**(6433), 1287–1289. <https://doi.org/10.1126/science.aaw4399>
- Freyer, N., Groß, D. and Lipprandt, M. (2024) The ethical requirement of explainability for AI-DSS in healthcare: a systematic review of reasons. *BMC Medical Ethics* **25**, 104. <https://doi.org/10.1186/s12910-024-01103-2>
- Gallifant, J., Afshar, M., Ameen, S. et al. (2025) The TRIPOD-LLM reporting guideline for studies using large language models. *Nature Medicine* **31**, 60–69. <https://doi.org/10.1038/s41591-024-03425-5>
- Gazquez-Garcia, J., Sánchez-Bocanegra, C.L. and Sevillano, J.L. (2025) AI in the health sector: systematic review of key skills for future health professionals. *JMIR Medical Education* **11**, e58161. <https://doi.org/10.2196/58161>
- Gomis-Pastor, M. et al. (2024) Clinical validation of digital healthcare solutions: state of the art, challenges and opportunities. *Healthcare* **12**(11), 1057. <https://doi.org/10.3390/healthcare12111057>
- Green, B.L., Murphy, A. and Robinson, E. (2024) Accelerating health disparities research with artificial intelligence. *Frontiers in Digital Health* **6**, 1330160. <https://doi.org/10.3389/fdgh.2024.1330160>
- Griffin, F. (2021) Artificial intelligence and liability in health care. *Health Matrix: The Journal of Law-Medicine* **31**, 65–106. <https://scholarlycommons.law.case.edu/healthmatrix/vol31/iss1/5/>
- Gu, T., Pan, W., Yu, J. et al. (2025) Mitigating bias in AI mortality predictions for minority populations: a transfer learning approach. *BMC Medical Informatics and Decision Making* **25**, 30. <https://doi.org/10.1186/s12911-025-02862-7>
- Guevara, M., Chen, S., Thomas, S. et al. (2024) Large language models to identify social determinants of health in electronic health records. *npj Digital Medicine* **7**, 6. <https://doi.org/10.1038/s41746-023-00970-0>
- Haley, J. et al. (2020) One in five adults in immigrant families with children reported chilling effects on public benefit receipt in 2019. https://www.urban.org/sites/default/files/publication/102406/one-in-five-adults-in-immigrant-families-with-children-reported-chilling-effects-on-public-benefit-receipt-in-2019_0.pdf
- Hanna, M.G., Pantanowitz, L., Jackson, B. et al. (2025) Ethical and bias considerations in artificial intelligence/machine learning.

- Modern Pathology* **38**(3), 100686. <https://doi.org/10.1016/j.modpat.2024.100686>
- Hasanzadeh, F., Josephson, C.B., Waters, G. et al. (2025) Bias recognition and mitigation strategies in artificial intelligence healthcare applications. *npj Digital Medicine* **8**, 154. <https://doi.org/10.1038/s41746-025-01503-7>
- Hazirbas, C., Bitton, J., Dolhansky, B. et al. (2022) Towards measuring fairness in AI: the Casual Conversations dataset. *IEEE Transactions on Biometrics, Behavior, and Identity Science* **4**(3), 324–332. <https://doi.org/10.1109/TBIOM.2021.3132237>
- Herzog, N.J., Celik, D. and Sulaiman, R.B. (2024) Artificial intelligence in healthcare and medical records security. In: Jahankhani, H., Bowen, G., Sharif, M.S. and Hussien, O. (eds), *Cybersecurity and Artificial Intelligence: Transformational Strategies and Disruptive Innovation*, pp. 35–57. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-52272-7_2
- Hicke, Y., Geathers, J., Rajashekar, N. et al. (2026) MedSimAI: simulation and formative feedback generation to enhance deliberate practice in medical education. *arXiv*, arXiv:2503.05793. <https://doi.org/10.48550/arXiv.2503.05793>
- Hicks, S.A. et al. (2022) On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports* **12**, 5979. <https://doi.org/10.1038/s41598-022-09954-8>
- Hopkins, A., Cen, S.H., Ilyas, A. et al. (2025) AI supply chains: an emerging ecosystem of AI actors, products, and services. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* **8**(2), 1266–1277. <https://doi.org/10.1609/aies.v8i2.36628>
- Huang, R., Li, H., Suomi, R. et al. (2023) Intelligent physical robots in health care: systematic literature review. *Journal of Medical Internet Research* **25**, e39786. <https://doi.org/10.2196/39786>
- Hurley, M.E., Lang, B.H., Kostick-Quenet, K.M. et al. (2025) Patient consent and the right to notice and explanation of AI systems used in health care. *The American Journal of Bioethics* **25**(3), 102–114. <https://doi.org/10.1080/15265161.2024.2399828>
- Kaloudi, N. and Li, J. (2020) The AI-based cyber threat landscape: a survey. *ACM Computing Surveys* **53**(1), article 20, 1–34. <https://doi.org/10.1145/3372823>
- Kaur, D.P., Singh, N.P. and Banerjee, B. (2022) A review of platforms for simulating embodied agents in 3D virtual environments. *Artificial Intelligence Review* **56**(4), 3711–3753. <https://doi.org/10.1007/s10462-022-10253-x>
- Kaye, J., Whitley, E.A., Lund, D. et al. (2015) Dynamic consent: a patient interface for twenty-first century research networks. *European Journal of Human Genetics* **23**(2), 141–146. <https://doi.org/10.1038/ejhg.2014.71>
- Khosravi, M., Zare, Z., Mojtabaiean, S.M. et al. (2024) Artificial intelligence and decision-making in healthcare: a thematic analysis of a systematic review of reviews. *Health Services Research and Managerial Epidemiology* **11**, 23333928241234863. <https://doi.org/10.1177/23333928241234863>
- Kimiafar, K., Sarbaz, M., Tabatabaei, S.M. et al. (2023) Artificial intelligence literacy among healthcare professionals and students: a systematic review. *Frontiers in Health Informatics* **12**, 168. <https://doi.org/10.30699/fhi.v12i0.524>
- Kiseleva, A., Kotzinos, D. and De Hert, P. (2022) Transparency of AI in healthcare as a multilayered system of accountabilities: between legal requirements and technical limitations. *Frontiers in Artificial Intelligence* **5**, 879603. <https://doi.org/10.3389/frai.2022.879603>
- Kocak, B., Klontzas, M.E., Stanzione, A. et al. (2025) Evaluation metrics in medical imaging AI: fundamentals, pitfalls, misapplications, and recommendations. *European Journal of Radiology Artificial Intelligence* **3**, 100030. <https://doi.org/10.1016/j.ejrai.2025.100030>
- Lalova-Spinks, T., Meszaros, J. and Huys, I. (2023) The application of data altruism in clinical research through empirical and legal analysis lenses. *Frontiers in Medicine* **10**, 1141685. <https://doi.org/10.3389/fmed.2023.1141685>
- Lekadir, K., Frangi, A.F., Porras, A.R. et al. (2025) FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ* **388**, e081554. <https://doi.org/10.1136/bmj-2024-081554>
- Lekadir, K., Quaglio, G., Tselioudis Garmendia, A. et al. (2022) Artificial intelligence in healthcare: applications, risks, and ethical and societal impacts. Brussels: European Parliamentary Research Service, Scientific Foresight Unit (STOA). PE 729.512. <https://doi.org/10.2861/568473>
- Lian, Z., Wang, W., Zeng, Q. et al. (2026) Prompt-in-content attacks: exploiting uploaded inputs to hijack LLM behavior. In: Chen, J., He, D. and Xie, W. (eds), *Network and System Security: 19th International Conference, NSS 2025, Wuhan, China, December 5–7, 2025, Proceedings*, pp. 146–160. Singapore: Springer. https://doi.org/10.1007/978-981-95-6419-4_9
- Liu, S. and Guo, L.R. (2024) Data ownership in the AI-powered integrative health care landscape. *JMIR Medical Informatics* **12**, e57754. <https://doi.org/10.2196/57754>
- Lopez-Ramos, L.M., Pluktaite, G., Bui, C.K.T. et al. (2025) Operationalizing AI ethics in medicine: a co-creation workshop study. *BMC Medical Ethics* **26**, 150. <https://doi.org/10.1186/s12910-025-01317-y>
- Lundberg, S.M. and Lee, S.-I. (2017) A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pp. 4765–4774. <https://proceedings.neurips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions>
- Ma, Y., Song, Y., Balch, J.A. et al. (2024) Promoting AI competencies for medical students: a scoping review on frameworks, programs, and tools. *arXiv*, arXiv:2407.18939. <https://doi.org/10.48550/arXiv.2407.18939>
- McLennan, S., Fiske, A., Tigard, D. et al. (2022) Embedded ethics: a proposal for integrating ethics into the development of medical AI. *BMC Medical Ethics* **23**, 6. <https://doi.org/10.1186/s12910-022-00746-3>
- MHRA (2024) MHRA's AI regulatory strategy ensures patient safety and industry innovation into 2030. London: Medicines and Healthcare products Regulatory Agency. <https://www.gov.uk/government/news/mhras-ai-regulatory-strategy-ensures-patient-safety-and-industry-innovation-into-2030>. Accessed 11 June 2026.
- Mitchell, M., Wu, S., Zaldivar, A. et al. (2019) Model cards for model reporting. In: *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*, pp. 220–229. New York: Association for Computing Machinery. <https://doi.org/10.1145/3287560.3287596>
- Moore, J., Grabb, D., Agnew, W. et al. (2025) Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers. In: *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*, pp. 599–627. New York: Association for Computing Machinery. <https://doi.org/10.1145/3715275.3732039>
- Muley, A., Muzumdar, P., Kurian, G. and Basyal, G.P. (2023) Risk of AI in healthcare: a comprehensive literature review and study framework. *Asian Journal of Medicine and Health* **21**(10), 276–291. <https://doi.org/10.9734/AJMAH/2023/v21i10903>
- Murdoch, B. (2021) Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Medical Ethics* **22**, 122. <https://doi.org/10.1186/s12910-021-00687-3>
- Musleh, M., Raidou, R.G. and Ceneda, D. (2025) TrustME: a context-aware explainability model to promote user trust in guidance. *IEEE Transactions on Visualization and Computer Graphics* **31**(10), 8040–8056. <https://doi.org/10.1109/TVCG.2025.3562929>
- National Academy of Medicine (2025) An Artificial Intelligence Code of Conduct for Health and Medicine: Essential Guidance for Aligned Action. Adams, L., Fontaine, E., Matheny, M. and Krishnan, S. (eds). *NAM Special Publication*. Washington, DC: National Academies Press. <https://doi.org/10.17226/29087>
- Nemec Zlatolas, L., Welzer, T. and Lhotská, L. (2024) Data breaches in healthcare: security mechanisms for attack mitigation. *Cluster Computing* **27**(7), 8639–8654. <https://doi.org/10.1007/s10586-024-04507-2>

- Nikhilesh, N. (2023) Informed consent for clinical drug trials: introspection to the human right issues. *International Journal of Law Management & Humanities* **6**(6), 320–343. <https://ijlmh.com/paper/informed-consent-for-clinical-drug-trials-introspection-to-the-human-right-issues/>
- Nyariro, M., Emami, E., Caidor, P. et al. (2023) Integrating equity, diversity and inclusion throughout the lifecycle of AI within healthcare: a scoping review protocol. *BMJ Open* **13**(9), e072069. <https://doi.org/10.1136/bmjopen-2023-072069>
- Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S. (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
- OECD (2025) AI compute. Paris: Organisation for Economic Co-operation and Development. <https://www.oecd.org/en/topics/ai-compute.html>. Accessed 11 June 2026.
- Olawade, D.B., Wada, O.J., David-Olawade, A.C. et al. (2023) Using artificial intelligence to improve public health: a narrative review. *Frontiers in Public Health* **11**, 1196397. <https://doi.org/10.3389/fpubh.2023.1196397>
- Omar, M., Soffer, S., Agbareia, R. et al. (2025a) Sociodemographic biases in medical decision making by large language models. *Nature Medicine* **31**, 1873–1881. <https://doi.org/10.1038/s41591-025-03626-6>
- Omar, M., Sorin, V., Collins, J.D. et al. (2025b) Multi-model assurance analysis showing large language models are highly vulnerable to adversarial hallucination attacks during clinical decision support. *Communications Medicine* **5**, 330. <https://doi.org/10.1038/s43856-025-01021-3>
- Omiye, J.A., Lester, J.C., Spichak, S. et al. (2023) Large language models propagate race-based medicine. *npj Digital Medicine* **6**, 195. <https://doi.org/10.1038/s41746-023-00939-z>
- Oviedo-Trespalcacios, O., Peden, A.E., Cole-Hunter, T. et al. (2023) The risks of using ChatGPT to obtain common safety-related information and advice. *Safety Science* **167**, 106244. <https://doi.org/10.1016/j.ssci.2023.106244>
- Palmiotto, F. (2024) When is a decision automated? A taxonomy for a fundamental rights analysis. *German Law Journal* **25**(2), 210–236. <https://doi.org/10.1017/glj.2023.112>
- Peng, C., Yang, X., Chen, A. et al. (2023) A study of generative large language model for medical research and healthcare. *npj Digital Medicine* **6**, 210. <https://doi.org/10.1038/s41746-023-00958-w>
- Pennestri, F., Cabitza, F., Picerno, N. and Banfi, G. (2025) Sharing reliable information worldwide: healthcare strategies based on artificial intelligence need external validation. Position paper. *BMC Medical Informatics and Decision Making* **25**, 56. <https://doi.org/10.1186/s12911-025-02883-2>
- Ramspek, C.L., Jager, K.J., Dekker, F.W. et al. (2021) External validation of prognostic models: what, why, how, when and where? *Clinical Kidney Journal* **14**(1), 49–58. <https://doi.org/10.1093/ckj/sfaa188>
- Ratti, E., Morrison, M. and Jakab, I. (2025) Ethical and social considerations of applying artificial intelligence in healthcare: a two-pronged scoping review. *BMC Medical Ethics* **26**, 68. <https://doi.org/10.1186/s12910-025-01198-1>
- RDA FAIR Data Maturity Model Working Group (2020) FAIR Data Maturity Model: specification and guidelines. *Research Data Alliance*. <https://doi.org/10.15497/RDA00050>
- Responsible AI Collaborative (2024) AI Incident Database. Responsible AI Collaborative. https://incidentdatabase.ai/apps/discover/?hideDuplicates=1&is_incident_report=true&s=medicine. Accessed 11 June 2026.
- Rincón, E.H.H., Jimenez, D., Chavarro Aguilar, L.A. et al. (2025) Mapping the use of artificial intelligence in medical education: a scoping review. *BMC Medical Education* **25**, 526. <https://doi.org/10.1186/s12909-025-07089-8>
- Ríos Insua, D., Naveiro, R., Gallego, V. and Poulos, J. (2023) Adversarial machine learning: Bayesian perspectives. *Journal of the American Statistical Association* **118**(543), 2195–2206. <https://doi.org/10.1080/01621459.2023.2183129>
- Rissetto, M.A. (2021) European Health Data Space and the Artificial Intelligence Act proposal: are they tangled? University of Vienna Blog. <https://blog-idunivienne.univie.ac.at/all-blog-entries/european-health-data-space-and-the-artificial-intelligence-act-proposal-are-they-tangled/>. Accessed 11 June 2026.
- Sarra, C. (2025) Artificial intelligence in decision-making: a test of consistency between the “EU AI Act” and the “General Data Protection Regulation”. *Athens Journal of Law* **11**(1), 45–62. <https://doi.org/10.30958/ajl.11-1-3>
- Semigran, H.L., Linder, J.A., Gidengil, C. and Mehrotra, A. (2015) Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ* **351**, h3480. <https://doi.org/10.1136/bmj.h3480>
- Shafik, W., Hidayatullah, A.F., Kalinaki, K. et al. (2026) A systematic literature review on transparency and interpretability of AI models in healthcare: taxonomies, tools, techniques, datasets, open research challenges, and future trends. *Health and Technology* **16**, 209–230. <https://doi.org/10.1007/s12553-025-01051-w>
- Soenksen, L.R., Ma, Y., Zeng, C. et al. (2022) Integrated multimodal artificial intelligence framework for healthcare applications. *npj Digital Medicine* **5**, 149. <https://doi.org/10.1038/s41746-022-00689-4>
- Stern, A.D., Goldfarb, A., Minssen, T. et al. (2022) AI insurance: how liability insurance can drive the responsible adoption of artificial intelligence in health care. *NEJM Catalyst Innovations in Care Delivery* **3**(4). <https://doi.org/10.1056/CAT.21.0242>
- Subramanian, H.V., Canfield, C.I. and Shank, D.B. (2024) Designing explainable AI to improve human-AI team performance: a medical stakeholder-driven scoping review. *Artificial Intelligence in Medicine* **149**, 102780. <https://doi.org/10.1016/j.artmed.2024.102780>
- Tao, C., Gao, J. and Wang, T. (2019) Testing and quality validation for AI software—perspectives, issues, and practices. *IEEE Access* **7**, 120164–120175. <https://doi.org/10.1109/ACCESS.2019.2937107>
- TEHDAS (2023a) Report on lessons learned to be applied and recommendations for data altruism practices in the implementation of construction of national and European health data spaces (including broad consent). Helsinki: TEHDAS. <https://tehdas.eu/app/uploads/2023/09/tehdas-recommendations-for-data-altruism-practices.pdf>
- TEHDAS (2023b) TEHDAS’ proposals for data quality and utility in EHDS. Helsinki: TEHDAS. <https://tehdas.eu/tehdas1/results/tehdas-proposals-for-data-quality-and-utility-in-ehds/>
- TEHDAS (2023c) Recommendations on a Data Quality Framework for the European Health Data Space for secondary use. Deliverable 6.3. Helsinki: TEHDAS. <https://tehdas.eu/app/uploads/2023/09/tehdas-recommendations-on-a-data-quality-framework.pdf>
- Templin, T., Perez, M.W., Sylvia, S., Leek, J. and Sinnott-Armstrong, N. (2024) Addressing 6 challenges in generative AI for digital health: a scoping review. *PLOS Digital Health* **3**(5), e0000503. <https://doi.org/10.1371/journal.pdig.0000503>
- Topol, E.J. (2019) High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine* **25**(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
- Touw, S., McCormack, G., Himmelstein, D.U., Woolhandler, S. and Zallman, L. (2021) Immigrant essential workers likely avoided Medicaid and SNAP because of a change to the public charge rule. *Health Affairs* **40**(7), 1090–1098. <https://doi.org/10.1377/hlthaff.2021.00059>
- Trocin, C., Mikalef, P., Papamitsiou, Z. and Conboy, K. (2023) Responsible AI for digital health: a synthesis and a research agenda. *Information Systems Frontiers* **25**(6), 2139–2157. <https://doi.org/10.1007/s10796-021-10146-4>
- Turbé, H., Bjelogrić, M., Lovis, C. and Mengaldo, G. (2023) Evaluation of post-hoc interpretability methods in time-series classification. *Nature Machine Intelligence* **5**(3), 250–260. <https://doi.org/10.1038/s42256-023-00620-w>
- UK Government (2024) Equity in medical devices (ed. Whitehead, M.). London, UK: Department of Health and Social Care. <https://assets.publishing.service.gov.uk/media/65e89f255b65240011f21b03/equity-in-medical-devices-independent-review-report-print-ready.pdf>

- Urbina, F., Lentzos, F., Invernizzi, C. and Ekins, S. (2022) A teachable moment for dual-use. *Nature Machine Intelligence* **4**(7), 607. <https://doi.org/10.1038/s42256-022-00511-6>
- Vorisek, C.N., Lehne, M., Klopfenstein, S.A.I. et al. (2022) Fast Healthcare Interoperability Resources (FHIR) for interoperability in health research: systematic review. *JMIR Medical Informatics* **10**(7), e35724. <https://doi.org/10.2196/35724>
- Wasserman-Rozen, H., Gilad-Bachrach, R. and Elkin-Koren, N. (2024) Lost in translation: the limits of explainability in AI. *Cardozo Arts & Entertainment Law Journal* **42**(2), 391–438. <https://heinonline.org/HOL/Page?handle=hein.journals/caelj42&div=16>
- WHO (2023) Regulatory considerations on artificial intelligence for health. Geneva: World Health Organization. <https://www.who.int/publications/i/item/9789240078871>
- WHO (2025) Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. Geneva: World Health Organization. <https://www.who.int/publications/i/item/9789240084759>
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Wilkinson, M.D., Dumontier, M., Sansone, S.-A. et al. (2019) Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Scientific Data* **6**, 174. <https://doi.org/10.1038/s41597-019-0184-5>
- Willem, T., Fritzsche, M.-C., Zimmermann, B.M. et al. (2025) Embedded ethics in practice: a toolbox for integrating the analysis of ethical and social issues into healthcare AI research. *Science and Engineering Ethics* **31**(1), 3. <https://doi.org/10.1007/s11948-024-00523-y>
- Williamson, S.M. and Prybutok, V. (2024) Balancing privacy and progress: a review of privacy challenges, systemic oversight, and patient perceptions in AI-driven healthcare. *Applied Sciences* **14**(2), 675. <https://doi.org/10.3390/app14020675>
- Xia, Z., Xu, J., Zhang, Y. and Liu, H. (2025) A survey of uncertainty estimation methods on large language models. In: *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 21381–21396. Association for Computational Linguistics. <https://aclanthology.org/2025.findings-acl.1101/>
- Xu, H.-L., Gong, T.-T., Song, X.-J. et al. (2025) Artificial intelligence performance in image-based cancer identification: umbrella review of systematic reviews. *Journal of Medical Internet Research* **27**, e53567. <https://doi.org/10.2196/53567>
- You, J.G., Hernandez-Boussard, T., Pfeffer, M.A. et al. (2025) Clinical trials informed framework for real-world clinical implementation and deployment of artificial intelligence applications. *npj Digital Medicine* **8**, 107. <https://doi.org/10.1038/s41746-025-01506-4>
- Young, M.-E.D.T., Payan, D.D. and Guzman-Ruiz, I.Y. (2022) The structural impacts of enforcement policy on Latino immigrant health. *Frontiers in Public Health* **10**, 928435. <https://doi.org/10.3389/fpubh.2022.928435>
- Young, M.-E.D.T., Tafolla, S., Saadi, A. et al. (2023) Beyond “chilling effects”: Latinx and Asian immigrants’ experiences with enforcement and barriers to health care. *Medical Care* **61**(5), 306–313. <https://doi.org/10.1097/MLR.0000000000001839>
- Zhang, C., Xie, Y., Bai, H. et al. (2021) A survey on federated learning. *Knowledge-Based Systems* **216**, 106775. <https://doi.org/10.1016/j.knosys.2021.106775>
- Zicari, R.V., Ahmed, S., Amann, J. et al. (2021) Co-design of a trustworthy AI system in healthcare: deep learning based skin lesion classifier. *Frontiers in Human Dynamics* **3**, 688152. <https://doi.org/10.3389/fhumd.2021.688152>

EASAC is a network of the following European national academies and academic bodies. All EASAC member academies have endorsed this report in accordance with EASAC procedures. These include that all efforts have been made — with the nominations for the working group, during the writing phase of the report and through the independent review procedure — to ensure that this report reflects the best available scientific evidence. EASAC focuses with its recommendations on addressing topics and challenges for Europe at the transnational scale, and recognises that some of its member academies may need to weigh in on national issues in the advice to their governments.

The Austrian Academy of Sciences
The Royal Academies for Science and the Arts of Belgium
The Bulgarian Academy of Sciences
The Croatian Academy of Sciences and Arts
The Cyprus Academy of Sciences, Letters and Arts
The Czech Academy of Sciences
The Royal Danish Academy of Sciences and Letters
The Estonian Academy of Sciences
The Council of Finnish Academies
The Académie des sciences (France)
The German National Academy of Sciences Leopoldina
The Academy of Athens
The Hungarian Academy of Sciences
The Royal Irish Academy
The Accademia Nazionale dei Lincei (Italy)
The Latvian Academy of Sciences
The Lithuanian Academy of Sciences
The Royal Netherlands Academy of Arts and Sciences
The Norwegian Academy of Science and Letters
The Polish Academy of Sciences
The Academy of Sciences of Lisbon
The Romanian Academy
The Slovak Academy of Sciences
The Slovenian Academy of Sciences and Arts
The Spanish Royal Academy of Sciences
The Swiss Academies of Arts and Sciences
The Royal Swedish Academy of Sciences
The Royal Society (United Kingdom)
Academia Europaea

The affiliated network for Europe of



EASAC contributes to the implementation of the



EASAC Secretariat
Austrian Academy of Sciences
Dr. Ignaz Seipel-Platz 2
1010 Vienna
Austria

tel: +43 1-51581-1208
secretariat@easac.eu



FEAM

Federation of European
Academies of Medicine

FEAM is a network of the following European national academies of medicine, pharmacy and veterinary science. This report has been formally endorsed by the FEAM Council. FEAM focuses with its recommendations on addressing topics and challenges for Europe at the transnational scale, and recognises that some of its member academies may need to weigh in on national issues in the advice to their governments.

MEMBER ACADEMIES

- | | |
|--|--|
| <ul style="list-style-type: none">• The Austrian Academy of Sciences• The Croatian Academy of Medical Sciences• The French Academy of Veterinary Sciences• The Swiss Academy of Medical Sciences• The Royal Academy of Veterinary Sciences of Spain• The French Academy of Medicine• The Belgian Royal Academy of Medicine (KAGB)• The Belgian Royal Academy of Medicine (ARMB)• The French Academy of Pharmacy• The Portuguese Academy of Medicine• The Israeli National Academy of Science in Medicine• The Hungarian Academy of Sciences• The Irish Academy of Medical Sciences | <ul style="list-style-type: none">• The Academy of Athens• The Royal Netherlands Academy of Arts and Sciences (KNAW)• The Lithuanian Academy of Sciences• The Spanish Royal National Academy of Medicine• The Academy of Medical Sciences of the Serbian Medical Society• The Italian National Academy of Medicine• The Czech Medical Academy• The Romanian Academy of Medical Sciences• The National German Academy of Sciences 'Leopoldina'• The UK Academy of Medical Sciences• The National Academy of Medical Sciences of Ukraine• The Slovenian Academy of Medicine |
|--|--|

FEAM is affiliated to the InterAcademy Partnership (IAP)



FEAM Secretariat

Palais des Académies, Rue Ducale 1, 1000 Brussels, Belgium
info@feam.eu www.feam.eu