

AI IN HEALTH

HUGE POTENTIAL, HUGE RISKS



AI IN HEALTH: HUGE POTENTIAL, HUGE RISKS

The AI Age is here and here to stay

The OECD has been at the frontier in defining comprehensive policy principles for trustworthy development and use of Artificial Intelligence (AI) with its 2019 Principles. These principles seek to mitigate some of AI's most significant risks including worker displacement, expanding inequities, breaches of personal privacy and security, and irresponsible use of AI that is inappropriate for the context or may result in harm. The last quarter of 2023 saw several important events and reports designed to drive the safe implementation of AI from the White House Executive Order on the Safe, Secure and Trustworthy Development and Use of Artificial Intelligence; to the G7 Code of Conduct for AI; the AI Safety Summit at Bletchley Park; the provisional agreement on the European Union AI Act; and, in health, the WHO guidance for the regulation of artificial intelligence (US White House, 2023^[1]; World Health Organization, 2023^[2]; European Commission, 2023^[3]; AI Safety Summit, 2023^[4]; European Parliament, 2023^[5]).

AI has significant potential to save lives, improve health professionals' work, and make health systems more people-centred

AI can help address some of health's largest challenges including a depleted workforce, future threats to public health, ageing populations, and increasing complexity of health due to multiple chronic conditions. It is crucial to make use of this powerful new tool while also mitigating its risks. Oversight and robust governance will be necessary to respond rapidly to emerging issues and opportunities.

Failure to turn principles into action poses significant risk

While there are significant risks from the use of AI in health, there are also significant risks from not taking action to operationalise agreed principles. These risks include exacerbating digital and health inequities, increasing privacy risk, slowing scientific advancement, and hampering trust with the public. At present, AI is being designed, developed, and implemented in health facilities around the world, leveraging local data sets for training and making the results available to their local populations. Bespoke AI applications without the ability or intention to scale (e.g., due to system incompatibility or lack of technical resources), risk a fragmented set of AI innovations that are built and maintained by wealthy health organisations and only available to wealthy segments of the public. Strong and co-ordinated policy, data, and technical foundations, both within and across borders, are necessary to unleash the broad and equitable human value that is possible from AI.

This brief outlines the key opportunities for AI to improve health outcomes, critical risks to be addressed in its deployment in health, and proposes practical policy action to operationalise responsible AI that respects human rights and improves health outcomes within and across borders.

These actions will benefit from common principles and guardrails. Recent policy actions in the EU, US, and at a global level **all confirm the relevance of the OECD AI principles from 2019** (OECD, 2019^[6]).

Moving forward, the OECD AI Principles provide a framework for developing AI policies in health.

AI saves lives, helps the health workforce, and unlocks value

AI is already saving lives and can save more. Empirical evidence suggests that only in 2023 some 163 000 people may have died in Europe due to medical errors. 30% of medical errors are due to communication failures (European Alliance for Access to Safe Medicines, 2022^[7]). AI is ideal for improving communication by surfacing the right information to the right people at the right time for the right context, preventing errors, saving lives, and improving health outcomes. This is in addition to the promise of AI utilising vast amounts of clinical evidence (e.g., imaging, patient histories) to assist health providers in diagnosing and optimising treatment for patients. When used safely and appropriately, this could exponentially expand evidence-based medicine to improve health outcomes and people-centred care.

AI can help health professionals provide more time to care. AI helps health providers to practice their profession and invest their time in engaging with patients rather than spending time on transcribing notes and administrative work. As much as 36% of activities in health and social care could be automated using AI (Chebrolu Kumar, 2020^[8]). These productivity gains would reduce the projected deficit of 3.5 million health professionals required by 2030 across the OECD (OECD, 2023^[9]). AI can help providers integrate leading knowledge and mine health data to find critical signals to prevent patients falling between the cracks and improve adherence to clinical leading practice (Vender and Lynde, 2023^[10]). AI can improve quality of work, quality of human interactions, and quality of outcomes.

AI can help to protect digital health infrastructure from security threats. There are ever-increasing cyberattacks on health systems which are projected to cause financial losses of up to USD 10.5 trillion by 2025 (Aiyer et al., 2022^[11]). Cyberattackers are increasingly using AI to find and exploit vulnerabilities. Health systems could learn from the use of AI in other industries to help detect threats and help prevent or address breaches of health system infrastructure. Working together, AI can help build trust in the use of digital tools.

AI can help the health sector to unlock value from the 97% of the health data assets that are not currently used to assist decision making (Thomason, 2021^[12]; Cornell University/INSEAD/WIPO, 2019^[13]). The design, development, and implementation of AI systems in health benefit from timely access to quality data and could do much more with data while ensuring appropriate protections are in place. In health, AI was used to detect early signs of COVID-19 and accelerated the discovery of vaccines that saved millions of lives (Sharma et al., 2022^[14]). AI could be used to help develop treatments for rare diseases, improve safety of health systems by detecting unusual patterns of illness, identify opportunities for prevention of chronic illness, or advance personalised medicine. These innovative applications will be more representative of populations and more impactful if there is co-operation across health organisations, regions, and country borders.

Opportunities of AI in health are significant and warrant urgent action to address barriers and risks. Establishing the policy, data, and technical foundation where responsible and safe AI can be developed, implemented, and scaled would benefit all aspects of society. This must be done in a way that minimises the potential harms from privacy, security, bias, and misinformation while driving positive outcomes for everyone. Pro-active initiatives and strategies are needed to generate positive outcomes while implementing robust solutions to ensure respect for critical rights to privacy, non-discrimination, and safety.

Addressing risks of AI in health

There are risks that must be effectively addressed. Risks from AI include poor outcomes from algorithms (due to bias, low quality data, lack of transparency, or usage in inappropriate contexts); leaks of personal data (due to breaches of privacy and security in data collection or in execution of AI algorithms); solutions imposed on the health workforce stretching already strained workers; and unclear accountability

for the management of AI – and for any possible unfavourable outcome – that prevents collective and equitable impact.

Likewise, there are risks from *not* enabling AI such as expanding digital divides where AI solutions are only available to limited subsets of the public; continued burnout of health professionals from expanding administrative burden; and failure to generate and spread benefits from scientific advancements to improve health outcomes, to name a few.

It is not possible to reduce risks to zero. Action in AI must consider the ways to best remove or mitigate risks of harm in the short-, medium-, and long-term. Effective implementation and oversight of AI would optimally use and leverage data and technology resources while continuously learning and adjusting plans when necessary. This is where having clear policy principles for AI is beneficial in guiding policy and action.

The OECD AI Principles have been adopted by all OECD countries and are reflected in the G20 AI Principles. They provide a common set of guardrails to facilitate the responsible and effective development, deployment, and maintenance of AI solutions. These principles steer implementation of AI to put humans at the centre of design and articulate key requirements for the deployment of AI solutions to foster trust and impact. Principles are a starting point and need to be operationalised through the right tools and processes, policies, and enforcement mechanisms.

It is necessary to operationalise the principles for responsible AI in health. For example, the principle of “*human-centred values and fairness*” would guide actions that include both health providers and the public in development of AI solutions and ensure that those solutions would be equitable, adopted, used, and re-used / leveraged by others as appropriate. This would help mitigate the risk of bespoke AI solutions being deployed where it has not been designed to “fit” into clinical practice – historically, novel technologies have contributed to health provider burnout rather than enabling better health outcomes.

GENERATE DESIRED OUTCOMES	ADDRESS KEY RISKS of AI in HEALTH	OECD AI PRINCIPLES
Adoption, use and evolution of AI in health is effectively governed with appropriate enforcement and transparent reporting of positive and negative AI incidents	Unclear accountability for management of AI currently and through its evolution in health and across sectors	Accountability
Capacity for health workforce to use AI to improve outcomes for individuals and populations	Disruption of the health workforce stretching already strained workers	Human-centred values and fairness
AI in health solutions are designed to be broadly and equitably accessible for the public and their health providers	Human and technical resources invested in bespoke health solutions that only benefit a subset of the population	Inclusive growth, sustainable development, and well-being
AI in health solutions are clear, trusted, and understood by providers and their patients	Algorithms that are biased or not transparent	Transparency and explainability
AI in health solutions responsibly and safely use sensitive health data	Leaks of sensitive personal data due to breaches of privacy and security / cyberthreats	Robustness, security, and safety



The adoption, use and evolution of AI should be effectively governed with appropriate enforcement and reported transparently

Risk: Without oversight of the implementation and evolution of AI solutions, there is a risk that AI is designed, developed, or implemented in a way that causes harm to patients and health systems and possible even without repercussions for those causing the harm. Further, as capabilities in AI evolve oversight, enforcement and redress mechanisms are needed to prevent its harms and benefit from the opportunities it presents.

Gaps: Some countries are establishing AI oversight bodies and mechanisms; however, relatively few have a focused approach to AI in health. As a result, it is not clear how much AI is being adopted in health systems, what are the primary barriers challenging AI adoption, the benefits of AI on health workforces and health outcomes, and the frequency and quantity of negative AI incidents. Governance should be integrated with financial incentives to support the policy, data, and technical foundation (discussed above).

Further, the trust for the use of AI in health is mixed among the public – ranging from the belief that AI can accelerate the promise of personalised medicine to refusing that AI technologies are involved in any part of health care. The public has voiced concerns on the advancement of AI in the public sector (e.g. government having access to personal health data) and in the private sector (e.g. generating profits without sharing that value with the public).

Importantly, as in other sectors, it is not yet clear who is liable when health AI solutions cause harms – the developer of the solution, the organisation that purchased the solution, or the user that applies the AI solution. Without that clarity, potentially significant personal, organisational, and corporate risks could incite inaction by some or all parties.

Possible policy directions to maximise the benefit of AI and minimise the risk

- Enabling oversight within and across countries to advance AI in health that is aligned with key policy principles of trustworthy AI and leverages / contributes to the policy, data, and technical foundation required for AI and promotes fairness, ethics, people-centredness, transparency, and accountability.
- Establishing country-level and cross-country measurements and public reporting on the progression of the adoption of AI in health, its benefits, and its harms.
- Defining requirements and liability for public and private sector development, implementation, and evolution of AI solutions in health.



Capacity for health workforce to use AI to improve health outcomes

Risks: When AI in health is implemented, job functions will change with the result that some roles may no longer be necessary or may have very different skills requirements. At the aggregate level, it is estimated that 27% of jobs are at high risk of automation (OECD, 2023^[15]; European Patients Forum, 2023^[16]). These jobs will not disappear, but their tasks will be profoundly modified. However, while this was a concern at the start of the use of AI in health, recent work has explored the use of AI as an assistant to improve the productivity of health professionals. Studies repeatedly show that humans working with AIs outperform either humans or AIs working alone.

Gaps: Numerous health professionals have not yet been equipped with the skills to understand the value of AI tools in their practice to save time and improve health outcomes. Further, the development of health-related AI tools does not always involve health professionals and resulting solutions can add to the burden of already stretched workloads, contributing to worker burnout.

Health professionals tend to be risk averse. The introduction of AI needs to adapt to health professionals, respect their integrity, and address their concerns, while encouraging adoption and enriching skills.

Possible policy directions to maximise the benefit of AI and minimise the risk

- Improving training and capacity building for health and IT professionals (workforce size, upskilling, culture) for effective use of AI to free time for high-quality care.
- Developing requirements for the inclusion of health professionals in creation of AI solutions throughout the AI lifecycle – design, development, implementation, operation, and evolution.
- Establishing programmes to encourage the adoption and responsible use of authorised AI solutions into clinical practice, including identifying good AI solutions in an authoritative manner.



Health AI solutions should be designed to be broadly accessible

Risks: When health AI is implemented, the benefits of AI may only be available to a subset of the population, leading to inequities. Inequities may be related to culture, gender, income, or geographies, among others. Further, the development of bespoke AI solutions that are unable to scale will result in duplication of effort and technical infrastructure, wasting limited public resources.

Gaps: There are few systemic measurements of the scope of implementation and scale of impact of AI solutions in the health sector that incorporate population-level benefits segmented by gender, equity-deserving groups, or geographies. Hence, policy makers are not able to understand how the benefits of AI are impacting people and populations.

Initial investments in digital health have led to a severely fragmented policy, data, and technical foundation in the health sector where it is difficult to scale AI solutions within countries and across borders. This fragmentation is impairing the equitable use of AI solutions and prevents the ability to leverage innovations across organisations which may result in duplicated efforts and lower their impact.

In the long-term, individuals who have access to their personal health data could directly benefit from certified AI solutions to be empowered in their health outcomes; however, these benefits will be diminished when the data and technical foundation are not interoperable with AI solutions.

Possible policy directions to maximise the benefit of AI and minimise the risk

- Developing measures for the availability, use, and impact of AI on health (e.g., integrating AI with personal health records) that include analysis of outcomes across geographies, genders, and other relevant groups.
- Harmonising the policy, data, and technical foundation – aligning with the OECD Council Recommendation on Health Data Governance (OECD, 2022^[17]) – within countries (and where appropriate across borders) that enable scale of AI solutions – for population, clinical, and individual use.



AI training data are representative and AI solutions are transparent and explainable

Risks: Machine learning AI uses large amounts of data to discover patterns and make predictions. If the data that are used by AI are not representative of the human population that the solution is meant to serve and/or of low quality, then it could be biased and produce poor, harmful, or discriminatory outcomes (e.g., when AI is trained based on outcomes data from men and solutions are then applied to women). Further, the AI solution may not be transparent or explainable, leading to a lack of trust in recommendations and resulting in the AI solution being ignored.

Gaps: Several of the principles for AI incorporate the need for transparency and explainability of AI solutions. Additional national and international guidance and harmonisation is needed on what information

– such as the training data used – is required about AI systems used in healthcare for them to be trusted by stakeholders including health providers and the public.

Guidance should be extended to criteria for what is considered a “responsible AI” solution for health. Consistency and practicality will be necessary for effective regulation in the use of sensitive personal health data and to minimise bias. Where appropriate, criteria would include whether the AI in health solutions is providing qualified clinical guidance.

The worldwide AI health care market is projected to grow by 16x by 2030 from USD 11 billion (in 2021) to USD 188 billion (Conor Stewart, 2023^[18]). To date, this growth has been largely without any regulation or oversight. Adopting methods to certify and regulate AI solutions in health where necessary will help protect the public and help build trust. This may learn from existing health processes for health technology assessments or drug approvals.

Possible policy directions to maximise the benefit of AI and minimise the risk

- Developing guidance on criteria for “responsible AI” solutions that incorporates transparency, explainability, and clinical appropriateness.
- Establishing methods for the certification and regulation of AI solutions in health.



AI solutions protect personal data while improving health outcomes

Risks: Using vast amounts of detailed personal health information in the development and operation of AI solutions risks breaching an individuals’ right to privacy of their health information. Further, the large set of data for AI represents a significant target for cyberattacks.

Gaps: The development of AI solutions must minimise harms. Harms may arise from data use, misuse, and non-use, such as breaches of privacy, discrimination, or the failure to provide effective care to individuals. A comprehensive approach to privacy and security in the AI age would recognise that privacy risk will always exist if health systems are to improve individual health outcomes, protect public health, and enable equitable outcomes for all.

Most countries have adopted legislation to regulate the protection of personal health information. In the AI age, the practical application of legislation has been in tension with those developing AI solutions in health – in particular, where incentives are to have zero risk of privacy breaches at the expense of potential positive health outcomes. Other innovations such as privacy enhancing technologies would support both protection of personal health information and the advancement of AI for the public good; however, there would be a non-zero level of privacy risk to optimise the value of AI solutions.

Possible policy directions to maximise the benefit of AI and minimise the risk

- Modernising and harmonising “codes of conduct” for AI that addresses risks from AI that includes considerations both of data use and of data non-use, in alignment with the OECD Council Recommendation on Health Data Governance (OECD, 2022^[17]).
- Operationalising “codes of conduct” with disincentives (e.g., penalties for unauthorised re-identification) that provide robust safeguards to the public while enabling beneficial AI solutions.
- Strengthening cross-border and cross-industry collaboration for resilience to cyberattacks in health.



Operationalising AI principles into healthcare policy and practice

When future generations look back at the 2020's, they may wonder why society was not more rapidly and broadly leveraging a massively abundant asset – health data – to save and improve lives. As solutions to protect data become more widespread, the harms of inaction are becoming apparent. Further, collective action on both *using* AI in health and *protecting* personal data from privacy and security risks is needed to improve health outcomes.

Sufficient international consensus on AI principles has already been reached to guide the development of policies for responsible AI in health. When developing health policy, it is critical to maximise policy interoperability across borders, learn from best practices in other sectors – and encourage policy consistency across sectors – to support both the scaling of beneficial solutions and consistent approaches to oversight and enforcement. The time to operationalise principles and codes of conduct into health policy and practice for AI is now.

Cross-border and cross-industry co-ordination is necessary to optimise the benefits of AI in health while mitigating its risks. Support is needed to harmonise policy and adapt legislation for AI, such as enabling consistent rules for access to personal health data and its pseudonymisation that protects individuals and enables the use of protected data for purposes of research into rare disease or monitoring for public health events. Cross-industry co-ordination will support broader insights into the interaction between health and environment, finance, or social programs, among others.

Urgent action is necessary. It will take time, leadership, will, effort, and investment to achieve and sustain benefits from AI in health. Policymakers should pro-actively shape the evolution of AI in health systems such that it generates beneficial equitable health outcomes while respecting rights. If not, there is a risk that fragmented AI solutions become entrenched leading to inequity, excess cost, and sub-optimal benefits.

The OECD can help collective action across countries and regions in three areas.

1. Assessing and quantifying the opportunities and risks of AI on health outcomes to articulate how responsible AI can save lives and how individuals can be treated equitably and with dignity.
2. Supporting countries in shaping and operationalising health policies and codes of conduct that remove unnecessary barriers to responsible AI while ensuring appropriate risk frameworks, mitigations, and oversight are in place, including through learning from good practices across countries.
3. Benchmarking the development of AI policy in health, foster knowledge sharing, and track AI incidents and good practices for collective learning and response. This enables co-learning, joint problem solving, and co-ordination toward mutual beneficial outcomes.

The OECD is ideally positioned in this work given the in-depth of understanding in health and artificial intelligence as well as in other key industries that enables mutual learning. The OECD can drive health-specific action within the context of whole-of-government solutions. The OECD is already tracking health-related AI incidents; gathering an inventory of health-related AI tools; and observing international progress with AI-related policies that will have an impact on health. OECD knowledge, convening expertise, and trusted position to work with partners to develop policy guidance and toolkits has a long track record of sustained success.

The OECD works with critical international partners – such as the World Health Organization, World Bank, and Global Digital Health Partnership – while contributing efforts toward the broader advancement of the Global Initiative for Digital Health and achievement of the UN Sustainable Development Goals.

References

- AI Safety Summit (2023), “AI Safety Summit 2023: Roundtable Chair’s Summaries”, [4]
www.gov.uk/government/publications/ai-safety-summit-1-november-roundtable-chairs-summaries (accessed on 8 January 2024).
- Aiyer, B. et al. (2022), “New survey reveals \$2 trillion market opportunity for cybersecurity technology and service providers”, *McKinsey & Company*. [11]
- Chebrolu Kumar, R. (2020), *Smart use of artificial intelligence in health care*, Deloitte. [8]
- Conor Stewart (2023), *AI in healthcare market size worldwide 2021-30*, STATISTA. [18]
- Cornell University/INSEAD/WIPO (2019), “The Global Innovation Index 2019: Creating Healthy Lives—The Future of Medical Innovation”. [13]
- European Alliance for Access to Safe Medicines (2022), “Medication Errors – the Most Common Adverse Event in Hospitals Threatens Patient Safety and Causes 16/1/2 Deaths per Year”, *European Alliance for Access to Safe Medicines*, <https://eaasm.eu/en-gb/2022/09/13/press-release-medication-errors-the-most-common-adverse-event-in-hospitals-threatens-patient-safety-and-causes-160000-deaths-per-year/> (accessed on 9 December 2023). [7]
- European Commission (2023), *Hiroshima Process International Code of Conduct for Advanced AI Systems*, <https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-code-conduct-advanced-ai-systems> (accessed on 8 January 2024). [3]
- European Parliament (2023), *Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI*, European Parliament. [5]
- European Patients Forum (2023), “Artificial Intelligence in Healthcare: Advancing Patient-Centric Care through Co-design and Responsible Implementation”. [16]
- OECD (2023), *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market*, OECD Publishing, Paris, <https://doi.org/10.1787/08785bba-en>. [15]
- OECD (2023), *Ready for the Next Crisis? Investing in Health System Resilience*, OECD Health Policy Studies, OECD Publishing, Paris, <https://doi.org/10.1787/1e53cf80-en>. [9]
- OECD (2022), *Health Data Governance for the Digital Age: Implementing the OECD Recommendation on Health Data Governance*, OECD Publishing, Paris, <https://doi.org/10.1787/68b60796-en>. [17]
- OECD (2019), *Recommendation of the Council on Artificial Intelligence*, <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>. [6]
- Sharma, A. et al. (2022), “Artificial Intelligence-Based Data-Driven Strategy to Accelerate Research, Development, and Clinical Trials of COVID Vaccine.”, *BioMed research international*, Vol. 2022, p. 1/1/2 241, <https://doi.org/10.1155/2022/7205241>. [14]
- Thomason, J. (2021), “Big tech, big data and the new world of digital health”, *Global Health Journal*, Vol. 5/4, pp. 165-168, <https://doi.org/10.1016/j.glohj.2021.11.003>. [12]

- US White House (2023), *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/ (accessed on 8 January 2024). [1]
- Vender, R. and C. Lynde (2023), “AI-Powered Patient Identification to Optimise Care”, https://pentavere.ai/wp-content/uploads/2023/03/CDA-Poster_AI-Powered-Patient-Identification-to-Optimise-Care.pdf (accessed on 20 December 2023). [10]
- World Health Organization (2023), *Regulatory considerations on artificial intelligence for health*, <https://iris.who.int/bitstream/handle/10665/373421/9789240078871-eng.pdf?sequence=1> (accessed on 8 January 2024). [2]

Contact

Eric SUTHERLAND (✉ eric.sutherland@oecd.org)

Rishub KEELARA (✉ rishub.keelara@oecd.org)

Read more

Anderson, B. and Sutherland, E. (2024), “Collective action for responsible AI in health”, *OECD Artificial Intelligence Papers*, No. 10, OECD Publishing, Paris, <https://doi.org/10.1787/f2050177-en>

This document should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the authors.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at www.oecd.org/termsandconditions.