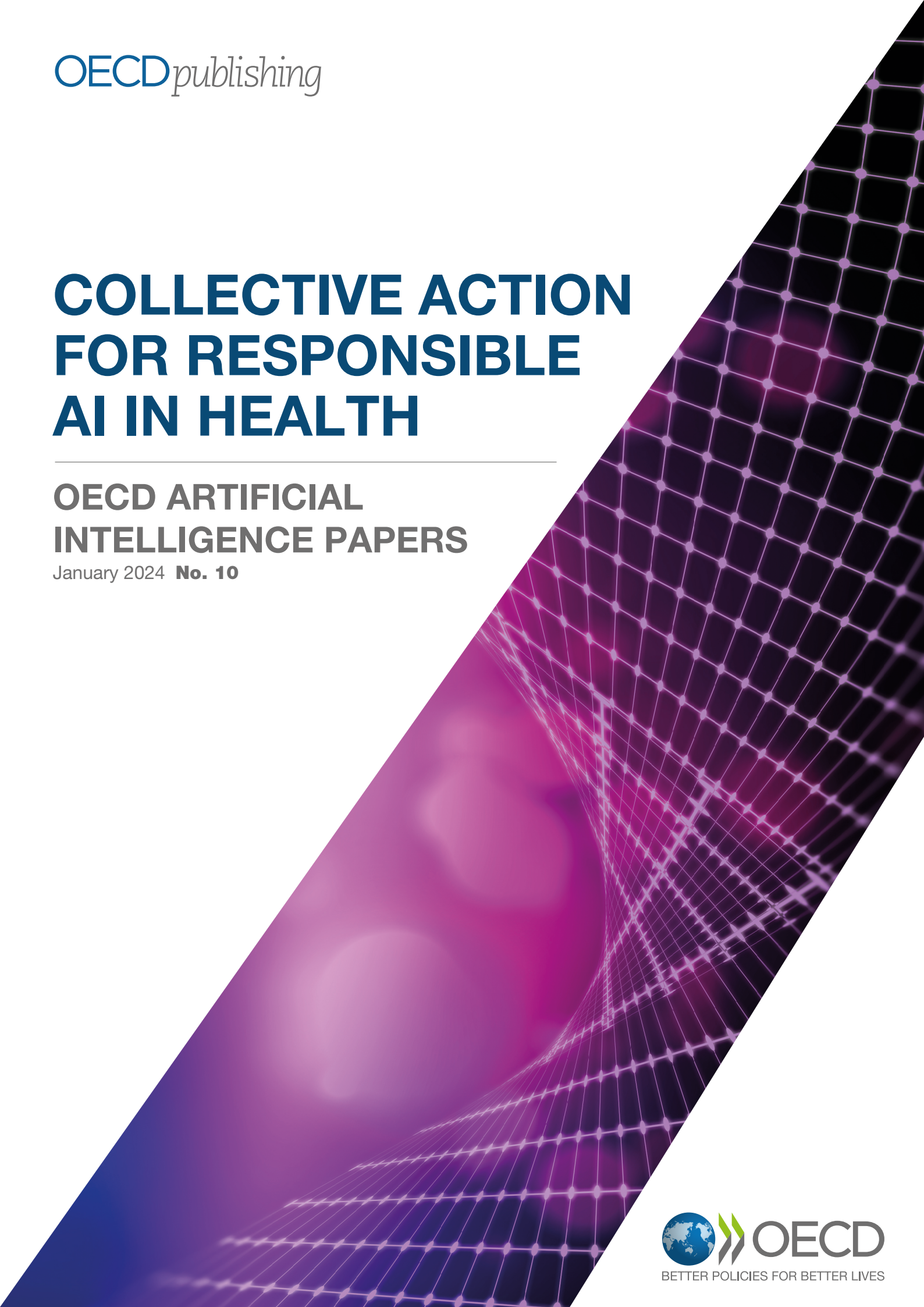
OECD publishing

COLLECTIVE ACTION FOR RESPONSIBLE AI IN HEALTH

OECD ARTIFICIAL INTELLIGENCE PAPERS

January 2024 **No. 10**



OECD

BETTER POLICIES FOR BETTER LIVES

Collective action for responsible AI in health

Brian Anderson, Eric Sutherland

Artificial intelligence will have profound impacts across health systems, transforming health care, public health, and research. Responsible AI can accelerate efforts toward health systems being more resilient, sustainable, equitable, and person-centred. This paper provides an overview of the background and current state of artificial intelligence in health, perspectives on opportunities, risks, and barriers to success. The paper proposes several areas to be explored for policy makers to advance the future of responsible AI in health that is adaptable to change, respects individuals, champions equity, and achieves better health outcomes for all.
The areas to be explored relate to trust, capacity building, evaluation, and collaboration. This recognises that the primary forces that are needed to unlock the value from artificial intelligence are people-based and not technical. The OECD is ready to support efforts for co-operative learning and collective action to advance the use of responsible AI in health.

Acknowledgements

The work presented here was led by Brian Anderson (Chief Digital Health Physician, Mitre Group) and supported by the OECD Secretariat (Eric Sutherland, Rishub Keelara, Sam Eiszele, Margarita Almyranti). An expert group provided invaluable support in the drafting and review of this paper (Laura Adams, Mary Gray, John Halamka, Brenton Hill, Dipak Kalra, Andrew Morris, Laura Rosella, Divya Srivastava, Ariel Dora Stern). The authors would also like to thank Francesca Colombo, Nick Tomlinson, and Mark Pearson at the OECD Directorate for Employment, Labour, and Social Affairs, for their comments. Finally, we thank Nicholson Price, Roger Taylor, Enrico Coiera, Ewan Affleck, and Timo Minssen for their advice and guidance with the analyses.

The views expressed in this paper are those of the authors and not necessarily of the authors' organisations, the governments of OECD member countries, or the European Union. The paper is intended to help inform and stimulate discussion.

Table of contents

Figures

Boxes

Executive summary

Artificial intelligence (AI) is here and here to stay. For health, AI represents both tremendous opportunity and tremendous risk. When managed poorly, AI can widen inequities and cause harms to individuals and society. When managed well, AI can highlight and help to reduce inequalities, save lives, improve work for health providers, and protect society from public health emergencies. AI will be a key contributor to global health by enabling learning from cross-border data sets and populations, helping to identify systemic inequities, and driving health systems to be more resilient and sustainable while evolving healthcare toward integrated and human-centric learning health systems.

While there are significant risks from AI, these can be managed effectively with clear guidance and action toward what is known as “Responsible AI” that is trustworthy, ethical, and minimises risks while respecting human rights and democratic values. Health is well positioned to adopt AI through its experience with risk management and economic evaluation for cost-effectiveness (e.g. *Health Technology Assessments)* and monitoring outcomes, safety, and quality (e.g. *post-market drug surveillance).*

Well-recognised risks with AI include bias, privacy and security risks, discrimination, lack of transparency, lack of oversight, job displacement and de-personalisation, as well as misapplication of context-dependent algorithms. There are cases where the risk of AI is low and the benefit will be high, which represent potential areas of initial focus. One such area is the automation of tasks for health workers to have more time to care, which is estimated to reduce administrative burden by 10-30%. Another area is where AI is used as a “second opinion” in reviewing images. In one recent study, an AI partnering with a radiologist looking at breast cancer images produced exceptional results catching positive cases that would have otherwise been missed. AI can also accelerate drug discovery to address anti-microbial resistance.

Potential advancements with responsible AI in health are promising; however, AI faces human-controllable barriers that hinder its development, implementation, and use. Such barriers include unclear accountability, misaligned policies for timely access to quality data, lack of coherent data standards for data collection, global health system fragmentation, and insufficient skills to optimise digital health (including AI) among health workers, policy makers, and the public. The result of these barriers is that current efforts in AI are, at times, ad hoc with limited ability to scale due to poor data and digital foundations for health systems. Addressing these barriers are critical to ensure that significant financial and resource investments can achieve expected returns and that all people can benefit from the application of responsible AI.

All industries are learning what practical actions to take to enable responsible AI. As the health industry is establishing an approach to AI, it has an opportunity to learn from other industries with similar risk tolerance (e.g. aeronautics) and data sensitivity (e.g. finance). Like the roll-out of rules for health data privacy and access, it is expected that health-specific approaches to AI will be helpful to ensure their consistent adoption within health that also enables co-operation across sectors (e.g. social services, environment) and supports collaboration across borders.

Further, many countries (and broader regions) are learning what practical actions to take to enable responsible AI in health. No country has established a definitive solution for AI in health. Co-ordinated action will achieve optimal outcomes when it is done across countries. AI will leverage more diverse data

sets across populations to discover treatments for rare diseases or help to address health inequities. AI will help provide protection from public health emergencies where an AI solution in health is deployed across health systems to create de facto global datasets that cross country borders to identify patterns of interest to detect, understand, and prevent emerging challenges. Harnessing learning from real-world data is paramount to support the use of real-world evidence in decision-making at all levels of health systems.

Co-operation across countries will also allow joint action that reduces the cost of development of AI solutions, progresses regulatory effectiveness and efficiency, and improves safety of AI solutions should there be poor outcomes or unintended consequences. Working together will unlock economic opportunity from AI as new innovations can scale within countries and across borders. Countries acting independently on AI risks fragmentation of AI solutions that cannot scale across borders which would exacerbate existing digital inequities, fail to discover new treatments for disease, prevent innovation being shared, and not be prepared for future public health emergencies.

Governments can accelerate the adoption of responsible AI in health with focus on improving trustworthiness, building capacity, evaluating and evolving solutions, and accelerating progress together. While any of these areas of focus will help the advancement of responsible AI in general, each of these areas are critical in health to optimise health outcomes for individuals, communities, and society while minimising harms. Action on AI will ensure that local uses of AI are effective and wider uses of AI are able to achieve optimal health benefits for everyone.

Given the complexity, risk, and opportunity of the health sector itself, it is important to develop health-specific guidance for AI. A health-specific approach for AI will help provide guidance in a crucial industry and articulate nuances that are critical for success while ensuring alignment with a whole-of-government approach. Further, understanding the issues of AI in health will help inform approaches in other industries and support cross-industry collaboration.

The OECD is ready to help collective action across countries and regions by providing backbone support for co-learning, joint problem solving, and co-ordination toward agreed outcomes for health purposes. Through the work of the Committee on Digital Economy Policy (CDEP) and its Working Party on Artificial Intelligence Governance (AIGO), the OECD as demonstrated significant leadership in AI, with the publication of the OECD AI Principles in 2019, the OECD.AI Policy Observatory in 2020, the establishment of the OECD AI Network of Experts (ONE.AI), the OECD Framework for the Classification of AI Systems in 2022, and the AI Incident Monitor in 2023.

The OECD would leverage its broad knowledge, convening expertise, and trusted position on AI to work with partners to develop policy guidance and toolkits for developing a health-specific approach to advance responsible AI that is grounded in a multi-sectoral whole-of-government approach. This would identify where health-specific approaches for AI would promote fairness, ethics, people-centredness, transparency, and accountability of responsible AI solutions within and across borders.

Policy guidance and toolkits could be developed by an AI- and health-focussed expert group that is convened and supported by the OECD. This group could align and build upon OECD broader expertise measuring and analysing the policy, economic, and social impacts of AI technologies. The group would partner closely with the WHO/ITU partnership on AI and the WHO-led Global Initiative for Digital Health, among others, to support alignment and coherence across international partners working specifically on health and with wider non-sector-specific processes. The group would identify health specific issues that need to be addressed and align on issues common across industries such on data privacy and bias. The group would review, benchmark, and share practices for AI in health across OECD countries and work with partners to share expertise more broadly while drawing from AI advancements across governments and industries.

1 Guiding the transition to Responsible AI in Health

Artificial intelligence (AI) is transforming society. The pace of change has been accelerating and policy makers are finding it difficult to encourage innovation while establishing protections. Managing this balance is particularly challenging in health where there are significant opportunities and profound risks such as data privacy and security, algorithmic biases, and ethical and social concerns involving AI (Khan et al., 2023[1]).

As health information is increasingly digitalised, AI is already used to aid screening, diagnosis, and health system management. Looking forward, AI has the potential to help solve some of healthcare's most vexing problems, such as cures for cancer and other serious illnesses; improved diagnostic access, speed, and accuracy; detection of public health threats; prevention of chronic illness; protection against superbugs; and providing tools to bolster the health workforce. As AI moves to scale in the healthcare market, the policy response requires a nuanced understanding of the effectiveness and efficiency of AI solutions from an ethical and sustainability perspective.

> "In medical research, to take one example, the use of machine learning and mRNA technology (the same as in the COVID-19 jab) holds tremendous potential. **Vaccines against cancer, cardiovascular and auto-immune diseases could be ready by the end of the decade**." (Geddes, 2023[2])

The use of AI has risks that must be understood and mitigated to not hamper the realisation of its tremendous benefits. These mitigations must be guided by a culture of responsibility and trustworthiness that respects the people who are most impacted by the transition – the public and their care providers.

Governments and health systems are grappling with the transition from a legacy of analogue and paper-based processes to an environment of connected networks generating large volumes of data and powered by digital tools (Srivastava et al., 2023[3]). That transition requires adoption of agile processes with ongoing monitoring and a culture of partnership, knowledge exchange, and continuous learning within and across borders.

Meanwhile, the public expects – and deserves – health services that take full advantage of the digital age. In a world where banking occurs in the palm of our hand and near anything can be bought online, health seems antiquated with its inability to reliably share information across health settings, with health providers, and with patients. Practically, healthcare still sometimes relies on the fax machine with more than 75% of global use for health purposes (Gintux, 2023[4]).

Recent advancements in telemedicine, electronic prescriptions, digital health applications, connected sensors for remote patient monitoring, and patient portals show promise; however, these advancements are at least ten years behind other industries where virtual services and information exchange are common (Beamtree Global Impact Committee, 2023[5]) The promise of AI in health is hampered by weak digital and data foundations where there is uneven health data governance, lack of evidence, inconsistent adoption of technology standards, and limited resources to design, implement, sustain, and monitor novel solutions.

The benefits of AI could help everyone. There are practical actions governments can take immediately to overcome barriers, mitigate risks, and accelerate development, adoption, use, and equitable deployment of artificial intelligence to achieve desired outcomes. That action must be co-ordinated within and across governments where there is alignment toward a future environment where AI is ubiquitous. Legislation, evidence-based policies, and standards must be both robust and agile to evolve with the technology and the understanding of effective governance while enabling responsible innovation. For example:

AI is already saving lives and can save more. AI is 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. Further, AI can reduce medical errors due to miscommunication (estimated at 30% of all errors) (European Alliance for Access to Safe Medicines, 2022[6]). AI is ideal for improving communication by surfacing the right information to the right people at the right time for the right context.

AI is helping health professionals provide more time to care. As much as 36% of activities in health and social care could be automated using AI (Chebrolu Kumar, 2020[7]). These productivity gains would reduce the projected deficit of 3.5 million health professionals required by 2030 across the OECD (OECD, 2023[8]). AI will 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[9]). AI can improve quality of work, quality of human interactions, and quality of outcomes.

AI can protect digital 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[10]). Those cyberattackers are using AI to find and exploit vulnerabilities. Health systems will be able to "fight fire with fire" and use AI to detect threats and co-ordinate action to prevent breaches of health system infrastructure.

AI can unlock value from the 97% of the health data assets that are unused for decision making (Thomason, 2021[11]) (Organization, University and INSEAD., 2019[12]). Design, development, and implementation of AI benefits from timely access to quality data and can do much more with data while ensuring appropriate protections are in place. 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.

When implemented, AI will be a key contributor to global health by enabling learning from cross-border data sets, helping to identify systemic inequities, and driving health systems to be more resilient and sustainable. At the same time healthcare will evolve toward integrated and human-centric learning health systems.

Many countries are taking action to understand how best to enable responsible AI for health within their jurisdiction. There is benefit for collective action and international co-operation to optimise the use of AI for health both within and across borders while effectively mitigating its risks and overcoming barriers.

This paper explores the background of artificial intelligence; highlights predominant barriers to overcome, risks to mitigate, and opportunities to achieve; and proposes evidence-based actions to build trust, grow capacity, evaluate progress, and work together to achieve mutual goals.

It is time for the international community to act collectively to unlock the power of responsible AI in health for mutual benefit of individuals, communities, society, and the planet.

2 The future of AI: Looking back and looking forward

> "Since a major national workshop in 1956, artificial intelligence's potential has been clear. Today that potential has become a reality: **the AI age is here and here to stay**." (Schumer, 2023[13])

Artificial intelligence (AI) has been evolving for over 60 years. From a national workshop in the United States in 1956 to today, there have been significant advancements in AI. These advancements have grown exponentially with computational power and data volume. Novel demonstrations of the possibilities of AI have been demonstrated through mastering games (e.g. chess, go, Jeopardy!). In recent years, AI has been used to help practical problems such as improving aeronautic safety with digital twins (Raj, 2021[14]), automating manufacturing processes to lower cost while improving quality, and detecting fraud in financial transactions (Hemdan, El-Shafai and Sayed, 2023[15]).

The progression of AI in healthcare has been happening at a slower pace than other industries. AI tools were developed by Stanford University to diagnose bacterial infections and recommend antibiotics in the 1970s (Yu, 1979[16]). The first FDA-approved AI-enabled medical device occurred in 1995 and was used to help pathologists in abnormal cell recognition for cervical smears (U.S. Food & Drug Administration, 2022[17]). This year, AI algorithms outperformed the standard clinical risk model for predicting the five-year risk for breast cancer (Arasu, 2023[18]). AI has demonstrably saved a life by detecting a potential brain haemorrhage and informing the care provider who then ensured urgent action was taken. (Siegel-Itzkovich, 2023[19]) Clinicians are increasingly utilising clinical decision aids to treat patients and provide better outcomes.

AI continues to evolve with better creative capability in new and innovative applications. Most recently, a variety of *generative AI* technologies (such as ChatGPT, PaLM, LLaMa, and Gemini) have commoditised the power of AI into people's hands. Envisioning the opportunities in the next 60 years is difficult to conceive; however, the changes brought on by AI will be significant and the time between evolutionary leaps will continue to shrink.

There are aspects of the future of AI in health that are clear. The future will be more digital, more connected, and more personalised. Legislation, policy, and standards will help ensure responsible and equitable health outcomes for all. Immediate actions would shape the health ecosystem to be ready for the future including being adaptable to change and resilient to shocks. As such, action in AI must be guided by principles that are forward-looking for the long-term while aiding immediate and pressing health system issues related to ageing populations, growing inequities, and health workforce deficits. Establishing structures that anticipate, evaluate, and adapt to the dynamic environment while intentionally shaping the future will be the defining success criteria of the AI age.

Box 2.1. What is Artificial Intelligence? What is different about Generative AI?

At its essence, AI uses data and mathematical models to find patterns and use those patterns to make predictions. In health, those predictions can be used to detect the likelihood of a cancer from an image, identify likely antibiotics for superbugs, or populate structured administrative records from unstructured doctor notes.

The OECD defines an AI system as "a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decision that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment." (OECD, 2019[20]) The definition of an AI system in the European Union's Artificial Intelligence Act (EU AI Act) – currently being finalised after a political agreement was reached on 8 December 2023 – is expected to be aligned with the OECD definition.

The potential applications for AI are being explored such as machine learning, natural language processing, computer vision, and robotics. In health, AI is being used to improve medical diagnostics, protect disease outcomes, and develop personal treatment plans.

Generative AI can be understood as a form of AI model specifically intended to produce new digital material as an output (including text, images, audio, video, software code), including when such AI models are used in applications and their user interfaces. These are typically constructed as machine learning systems that have been trained on massive amounts of data. They work by predicting words, pixels, waveforms, data points, etc. that would resemble the models' training data, often in response to prompts (OECD, 2023[21]).

Generative AI itself carries risks including issues of intellectual property and copyright; concerns with proliferation of mis- and dis-information; and bias and discrimination (Lorenz, Perset and Berryhill, 2023[22]). While generative AI is powerful, its creative nature increases uncertainty and thereby risk for the delivery of healthcare. There are low risk use cases for generative AI such as in generating summaries of patient encounters – if sufficient precautions are taken such as preventing the use of identifiable information and keeping health providers in-the-loop.

History has taught us that going too fast with technical innovation may introduce harms and break trust with the public and providers. There have been notable missteps in the evolution of AI that have raised expectations – due to economic, political, or competitive pressures – only to be dashed on the rocks of reality when there was not full consideration for potential risks. IBM invested in its AI solution (Watson) to help fight cancer and was deemed a failure when it got in the way of doctors' work and made them less efficient while failing to provide substantially better outcomes. In 2015, initial versions of facial recognition mistook people with dark skin tones for gorillas. More recently, ChatGPT became notorious for "hallucinating" when it would generate fictitious citations for its advice.

There are also serious concerns that the use of AI may expand inequities due to inherent bias in historical data collection and the availability of AI applications limited to those with access to digital tools. Ensuring that algorithms are trained on data sets that reflect the target populations will be critical. An action plan on AI must acknowledge and address historical bias and injustice. Additionally, AI also requires large amounts of computing power, leading to concerns of its potential climate impact.

Despite these concerns, the potential benefits of AI provide an ethical and moral imperative for its use to improve health outcomes for all. Historically, new innovations that are now ubiquitous have experienced significant concerns when they were introduced. The stethoscope was viewed with scepticism as a means for good diagnosis a century ago and is now considered one of the most important medical inventions

(AMA Journal of Ethics, 2002[23]). The public had to be convinced of the usefulness and safety of electricity through ads in the New York Tribune in 1920 (New York Tribune, 1920[24]). There was resistance to the radio and television over fears of the radio wave transmissions and the impacts it would have on people (The Saturday Evening Post, 2021[25]). During COVID-19, the value of mRNA vaccines was demonstrated with 12 billion shots being given with relatively low frequency of serious side-effects (Dhamanti et al., 2023[26]).

The use of AI in healthcare raises the need for government intervention. There are well-known market failures in the healthcare market including information asymmetries and differentiated (AI) products (Folland, Goodman and Stano, 2016[27]). AI in healthcare displays the properties of a public good. Government intervention is necessary both from an economics and public health perspective to send the appropriate signals to the market but also to ensure the responsible use of AI in healthcare and find an appropriate balance for public / private co-operation and related risks.

A well-designed set of legislation, policy, and regulations by the public sector alongside a set of best practices and standards developed in partnership with the private sector and end-users will help achieve wide-spread and safe adoption and subsequent evolution of AI (see Box 2.2 and Box 2.3). For the health industry – given the highly sensitive nature of health data – a sufficiently specific and coherent set of policies, practices, and standards will be critical to encourage consistency in the development of responsible AI across data collection, access, and protection; to development of AI algorithms; to their use in healthcare settings and processes; and their subsequent evaluation and evolution. Moreover, ensuring that communities have a voice in this process and have transparency into how models are developed and how they perform over time will be critical to set guardrails and foster ongoing trust between the public, governments, and private sector.

Box 2.2. OECD Recommendation of the Council on Artificial Intelligence

In 2019, OECD countries adopted the Recommendation of the Council on Artificial Intelligence (OECD, 2019[20]). The OECD Principles on Artificial Intelligence promote AI that is innovative, trustworthy and respects human rights and democratic values. These principles apply to AI actors and all of government, including implementation of AI in Health.

The OECD AI Principles identify five complementary values-based principles for the responsible stewardship of responsible AI:

AI systems benefit people and the planet by driving inclusive growth, sustainable development, and well-being.

AI should be designed in a way that respects the rule of law, human rights, democratic values, and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society.

AI systems are understood by people such that those systems may be effectively challenged.

AI must function in a robust, secure, and safe way throughout their life cycles and potential risks should be continually assessed and managed.

Organisations and individuals developing, deploying, or operating AI systems should be held accountable for their proper functioning in line with these principles.

Consistent with these principles, the OECD also provides five recommendations to governments:

Facilitate public and private investment in research and development to spur innovation in responsible AI.

Foster accessible AI ecosystems with digital infrastructure and technologies and mechanisms to share data and knowledge.

Ensure a policy environment that will open the way to the deployment of responsible AI systems.

Empower people with the skills for AI and support workers for a fair transition.

Co-operate across borders and sectors to progress on trustworthy stewardship of responsible AI.

Source: (Galindo, Perset and Sheeka, 2021[28]).

Box 2.3. Advancements on policy for artificial intelligence in last quarter of 2023

Responding to the rapid awareness of the importance of artificial intelligence, the last quarter of 2023 has seen significant progress has been made on setting clear guidance for AI.

European Union Artificial Intelligence Act (EU AI Act): Through the EU AI Act, the European Union aims to ensure that fundamental rights, democracy, the rule of law and environmental sustainability are protected from high-risk AI. The Act categorises AI systems based on risk levels, imposing bans for systems that present unacceptable risks, such as social scoring, cognitive manipulation, and real-time and remote biometric identification systems (unless court approved). High-risk AI systems, which impact safety or fundamental human rights, will be subject to pre-market conformity assessment and post-market monitoring obligations. The EU AI Act establishes transparency obligations for the use of AI system that present limited risks (such as chatbots). Additionally, general purpose AI systems (including foundational models and generative AI) must comply to additional requirements, with further obligations for high-impact general purpose AI with systemic risk. The goal is to finalise the AI Act in early 2024, following a provisional agreement between the parliament and EU countries on 8 December 2023 (European Parliament, 2023[29]).

United States Executive Order on AI: On 30 October 2023, President Joe Biden issued an executive order (EO) on *Safe, Secure, and Trustworthy Artificial Intelligence*, outlining a comprehensive federal approach to responsible AI development. The EO aims to enhance AI security and solidify US global leadership. Emphasising a collaborative effort across government, the private sector, academia, and civil society, the EO addresses concerns about fraud, discrimination, bias, disinformation, and national security risks associated with AI. It mandates the safe use of AI, encourages premarket testing, and relies on technical standards and guidance from executive departments and agencies.

There is a specific focus on removing unnecessary and unhelpful barriers to the use of AI while retaining and strengthening guardrails. President Biden highlighted the EO's goal of earning trust in AI systems and called on Congress to pass bipartisan data privacy legislation addressing AI risks (US White House, 2022[30]).

AI Safety Summit, Bletchley Park: The United Kingdom, as the Chair of the inaugural *AI Safety Summit* at Bletchley Park on 1-2 November 2023, led a global effort to address the opportunities and risks of frontier AI. During the Summit, countries, industry leaders, and civil society representatives agreed to the *Bletchley Declaration on AI safety*, recognising the need for collaborative action. Notably, they committed to state-led testing of next-generation models before release, emphasising the importance of governments and AI developers working together. The UK Prime Minister, Rishi Sunak, highlighted the urgency of addressing the risks associated with AI's unprecedented development,

emphasising the need for global, collaborative action to seize its benefits while minimising potential harm (AI Safety Summit, 2023[31]).

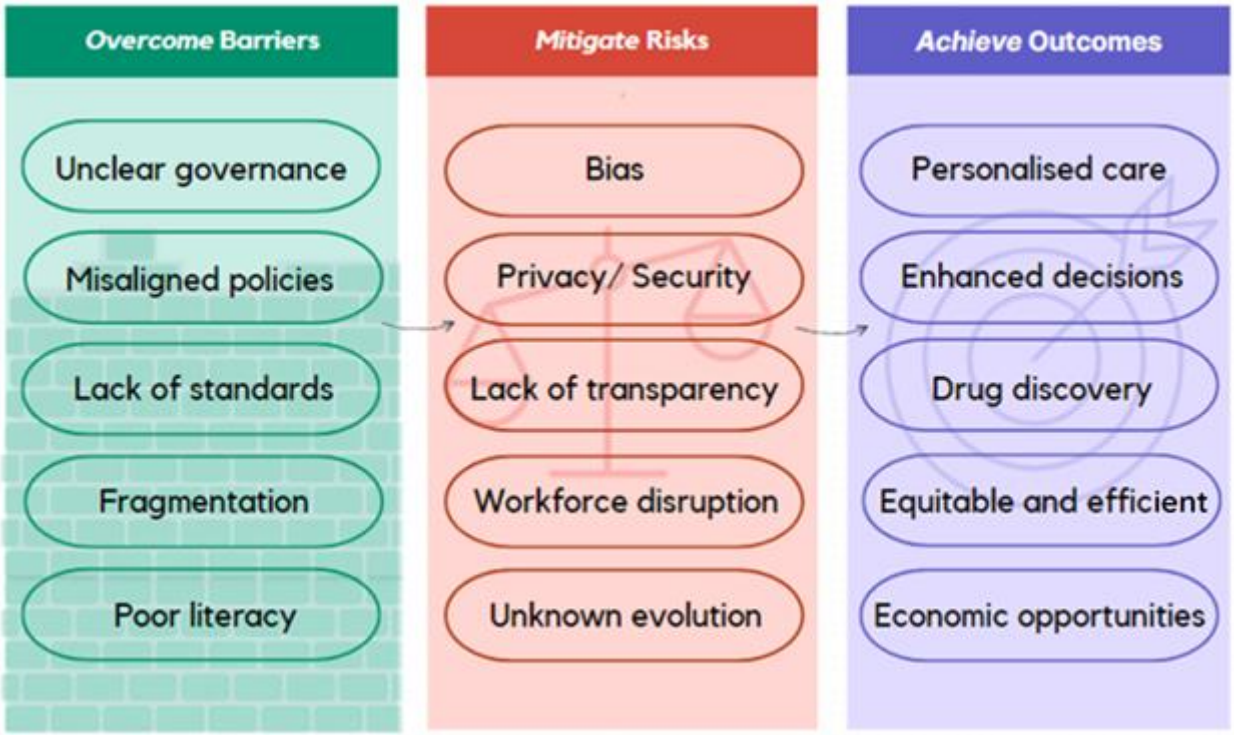
G7 Hiroshima AI Process: Under its 2023 Presidency of the G7, Japan established an ambitious agenda on generative AI, known as the "Hiroshima AI Process". At the G7 Leaders meeting of 6 December 2023, the Hiroshima AI Process concluded with the "Comprehensive Framework on Generative AI", including "International Guiding Principles for All AI Actors and for Organisations Developing Advanced AI Systems", as well as the "Code of Conduct for Organisations Developing Advanced AI Systems". The Code of Conduct provides voluntary guidance for actions by organisations developing advanced AI systems. The living documents build on the existing OECD AI Principles and are meant to help seize the benefits of AI while addressing its risks and challenges (European Council, 2023[32]).

3 Opportunities and risks and barriers for artificial intelligence in health

Artificial Intelligence (AI) is poised to revolutionise the healthcare landscape. This will impact all corners of health systems. AI is saving lives by helping doctors act on anomalies in radiographic images (Israel in August 2023); improving productivity in clinical note taking, allowing emergency room doctors to see 15% more patients (Canada in September 2023); and accelerating discovery of antibiotics for superbugs, reducing processes that could take years to weeks (United Kingdom in May 2023).

As with any innovation, new opportunities come with new risks that must be mitigated to minimise harms while optimising outcomes. While mitigating risks, it will be important to address the human-controllable barriers that prevent progress. When done effectively, it will be possible to achieve optimal health outcomes (see Infographic 3.1). This chapter will elaborate on these areas. Chapter Four will suggest pragmatic action to enable that success.

Figure 3.1. AI in Health: Overcome Barriers. Mitigate Risk. Achieve Outcomes



Source: Authors.

3.1. Barriers for artificial intelligence in health to be overcome

Health systems have been slow adopters of digital health innovations, including AI. While adoption of new telemedicine accelerated during COVID-19, health continues to lag other industries in implementation and scale of innovation (Beamtree Global Impact Committee, 2023[5]). While a primary reason for this slow adoption is concern over safety and unintended harms from innovation, a key contributor is the legacy of fragmented data infrastructure that challenges the ability to scale innovation and realise economies of scale from technology investments – a particular challenge for AI given its dependency on a timely access to quality data for training purposes.

While individual decisions that led to the current environment were well-intended to protect patients from harm, the slow adoption of innovation itself may cause harm – for example in the failure to realise and share benefits from science for individuals, communities, and society (Alberta Virtual Care, 2023[33]).

The tools for the development of the technical aspects of AI solutions already exist and are broadly available. Hence, the primary barriers to AI in health relate to AI not having sufficient quality data to be representative of the people intended to benefit from the solution (e.g. unnecessary conflicts between policies for access and privacy – among others; lack of interoperability standards within countries and/or across borders); or not having sufficient guidance for the implementation and operation of AI solutions (e.g. clarity for liability and accountability; fostering trust with providers and the public). What is notable about these barriers is that they may be overcome with sufficient human effort and co-operation.

Misaligned policies for advancement of responsible AI: Unharmonised policy is frequently a barrier to AI, due to conflicting, inconsistent, or absent policies that encourage desired outcomes. This may be due to the same policy domain with different guidance along “data supply chains” that connect organisations that are working together to deliver collective outcomes (e.g. different levels of data pseudonymisation across organisations). This may also be due to conflicting policies within an organisation, for example when policies for privacy provide guidance criteria for privacy of sensitive data, whereas policies for access provide guidance on their criteria for necessary access in the public interest. The two sets of criteria are often in conflict without a clear path to resolution.

A lack of harmonisation means that individual rights are placed in conflict including the right to (a) benefit from science, (b) privacy and security, (c) health and well-being; and (d) personal autonomy. Without clear governance and policies that provide guidance that recognises the tensions between these rights, advances in AI are often starved for the data necessary to responsibly develop, implement, scale, and sustain solutions.

Policies also need to be agile to be responsive to change and resilient to shocks. For example, policies associated with encryption will need to be updated to recognise technological advances that render old forms of encryption vulnerable to cyberattacks. Similarly, security policies need to be updated for health information infrastructures to move to the next level of (cyber)security. These updates must both equip health institutions to use robust safety techniques (e.g. synthetic datasets and differential privacy) while enabling data access that achieves desired health outcomes.

Those policies would recognise the potential of innovation (such as AI) to be a public good and foster its development, efficiency in implementation, and use for equitable individual, community, and society health outcomes.

Lack of standards as a barrier to responsible AI innovation in health: Standards help accelerate safe AI solutions, by simplifying the interoperability of solutions which accelerates data movement and improves data quality. In this context, standards refer to semantic and technical data standards that help with the movement of data between systems. Standards also refer to the process, quality, ethical, and governance standards that enable legal and operational interoperability. These legal and operational standards help

with predictability and clarity of accountabilities across organisations, which is important for lifecycle management from development through deployment and ongoing maintenance.

Without common data standards (e.g. for the collection of medical diagnosis) the information to train the AI solutions will be unreliable (e.g. using different terms for the same diagnosis or using the same term for different diagnosis), hence the AI solution itself will be unreliable. Without legal and operational standards (e.g. rules for providing protected access to data), AI solutions will have varying quality, timeliness, cost, and unclear accountability, as noted in the example above.

Standards simplify the transparency, monitoring, and regulation of solutions which helps demonstrate trustworthiness and enables economies of scale. Standards that are aligned and reflect ethical and governance considerations will contribute to improved decision-making regarding the effectiveness of responsible AI solutions in health (Srivastava, Scarbrough and Stavropoulou, 2020[34]). For example, the jointly issued European and North American statement on the ethics of AI in radiology calls on the community to develop and adhere to a uniform code of ethics to provide a reliable ethical framework as the technology rapidly advances (Geis et al., 2019[35]).

Policy makers, driving the adoption of clear standards, create more predictability (and consistency) for decision-making and development processes associated with AI, including its ability to scale across healthcare settings. Extending this adoption of standards across borders increases cross-border collaboration for research and trans-national individual care. These standards would be subject to controls through AI governance to encourage their adoption, use, and evolution to improve how data are acquired, managed, and used for AI environments.

Global health system fragmentation as a barrier to responsible AI innovation in health: Within country fragmentation undermines the effectiveness of cross-border health programmes supported by global health actors and threatens the attainment of the health-related Sustainable Development Goals. Five distinct yet interconnected sets of factors causing fragmentation are proliferation of global health actors; problems of global leadership; divergent interests; problems of accountability; and problems of power relations (Spicer et al., 2020[36]).

New global health problems such as COVID-19 heightened the need for global health actors to better co-ordinate their efforts in working towards common goals. Reaping the benefits of AI – for example to collaborate on solutions for rare disease – presents yet another opportunity and incentive to address the issue of global health system fragmentation and lack of standards for cross-border data collaboration.

It is notable that many of the problems of cross-border fragmentation are like the problems of fragmentation within countries – lack of interoperability and misaligned policies. International leaders can work together to determine the most beneficial actions that help advance the use of AI within their borders that also lowers the risk of long-term cross-border fragmentation. Indeed, if leaders choose to solve their local problems first without regard to cross-border co-operation, then it is likely that global fragmentation will persist.

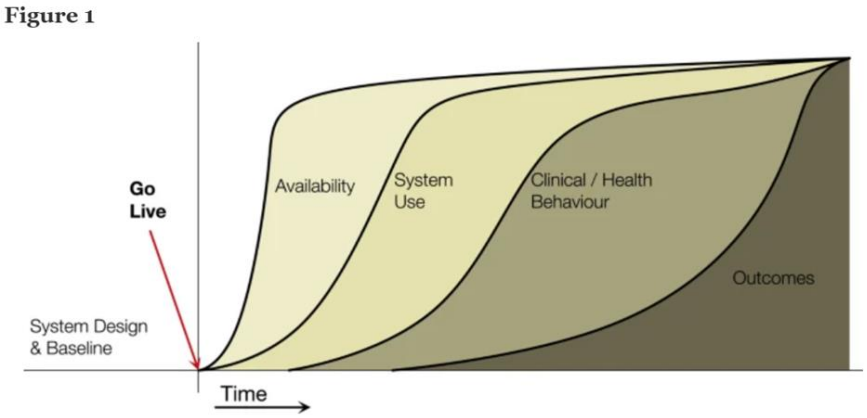
Unclear liability and accountability as a barrier to responsible AI innovation in health: Without clarity around accountability for AI solutions, multiple groups may work on similar issues (e.g. ethics of AI, risk management frameworks) where other critical issues are unaddressed (e.g. health data governance, evidence generation requirements) – resulting in a patchwork of solutions where it is unclear what practices should be followed. Apart from the waste from the duplication of effort, these efforts may cause confusion about necessary priorities and actions, which itself slows down progress.

Further, for health outcomes-related AI solutions, there is an open question of liability for “AI errors”. That is, if the AI solution produces a result that negatively impacts patient outcomes, it is not yet clear who is accountable – the manufacturer that produced the AI solution, the health organisation that acquired the solution, or the health provider that applied the solution to a patient. For AI to scale and thrive, issues of liability must be addressed.

Overall, a healthy ecosystem that fosters responsible artificial intelligence would feature strong leadership that support cultural change, enable inter-organisational co-operation, and ensure appropriate financing. Further, governance would clarify who is accountable for doing what, defining clear rules of the road for the development, adoption, use, and scale of AI, and establish effective oversight to provide the appropriate incentives (and dis-incentives) in the interest of achieving agreed outcomes. This is consistent with the OECD Council Recommendation on Health Data Governance (OECD, 2022[37]).

Strong AI leadership would respect the importance of clinical governance in the adoption and use of artificial intelligence. The Clinical Adoption Meta-Model (Price and Lau, 2014[38]) demonstrates that pushing for adoption either too slowly or too quickly will result in sub-optimal outcomes. New innovations with AI will necessarily go through the stages of availability of the innovation, its preliminary use, and incorporation into practice prior to achieving sustained beneficial outcomes.

Figure 3.2. Clinical Adoption Meta-Model



Source: (Price and Lau, 2014[38])

Australia has developed a Clinical Governance Framework for Digital Health that provides principles to better connect across health organisations and agencies. The framework advocates for the active management and monitoring of risks as part of an overall governance framework. This would help ensure clinical safety while effectively managing security and privacy. (Australian Digital Health Agency, 2023[39]). The United States, Canada, and the United Kingdom jointly developed an AI in health safety framework with ten principles for good machine learning (U.S. Food and Drug Administration (FDA), Health Canada and Medicines and Healthcare products Regulatory Agency (MHRA), 2021[40]). The principles are designed to promote safe, effective, and high-quality medical devices that use artificial intelligence and machine learning.

Insufficient trust in and understanding of digital health (including AI) as a barrier to responsible AI innovation in health: Lack of trust for the use of digital tools among providers, patients, and the public means that when AI innovations are developed, that they may not be adopted, used, or scaled, negating their value. A survey in early 2023 showed that 70% of the public in the United States. did not want AI to be used for their diagnosis (Tyson et al., 2023[41]).

Further, without a good understanding of how to make use of the innovation, the health workforce or the public may not have sufficient knowledge to make use of new tools, even when they are accepted as being beneficial. A comprehensive study carried out by the International Labour Association revealed that basic digital health literacy in multiple high-income countries, including Germany, Hungary and Italy, is only present in approximately 50% of the population, making it difficult to take advantage of innovation (International Labour Association, 2023[42]).

Policy makers also require good understanding about digital health and artificial intelligence to ensure that evidence-based decisions are made that achieve appropriate outcomes while mitigating risks. Decisions would leverage and harness real-world data and real-world evidence arising from AI solutions. For example, once deployed, developers who continue to evidence the effectiveness of their technology in the real-world will help policy makers to better understand their functionality. This puts emphasis on the need for a robust regulatory environment to monitor and benchmark these technologies as they continually evolve. This could leverage similar processes to post-market drug surveillance and the notion of digital health technology vigilance as part of the regulatory response (Srivastava et al., 2023[3]).

Significant effort is needed to demonstrate trustworthiness with new digital health and AI solutions in health among providers and the public. While some efforts could be made to improve digital literacy in those groups, equal (if not more) efforts are required to simplify the message of digitalisation in health – in its benefits, risks, and mitigations – so that solutions are designed for an optimal user experience that simplifies its use, articulates the value of novel tools, and visibly applies necessary controls for protection. Applying Classification frameworks will identify relatively low-risk high-value activities – such as automating clinical administrative tasks (estimated at up to 10-30% of time of health providers). This will provide immediate value to health providers and minimise their administrative time so they have more time to care with their patients.

3.2. Risks of artificial intelligence in health

AI in healthcare has the potential to revolutionise patient care and improve outcomes. However, the risks associated with AI must be addressed to ensure that its benefits are generated and shared safely and equitably. By adopting a proactive approach to mitigating these risks, the healthcare sector can harness the full potential of AI while minimising potential harms.

While advancing the use of responsible AI, it must be complemented with a robust risk management programme including risk mitigation, risk analysis, and governance. This would include risk prioritisation and differential controls based on level of risk. Risk programmes will benefit from international co-operation and collective action. There are five major risks to be managed and mitigated.

AI solutions may be built on biased data, leading to biased and harmful outcomes or proliferating misinformation: Large patient data sets are necessary to train AI applications. Data used in these applications can unintentionally lead to biases in the results such as lack of applicability to certain population-sub-groups, and thereby raising questions of whether the results are appropriate (Buolamwini and Gebru, 2018[43]).

When designing and deploying AI solutions, efforts should be made to ensure that data used to build the solution are representative of the target population that will benefit from the AI solution. Several countries (e.g. Australia, Canada, France, Korea, the Netherlands, the United Kingdom, the United States, among others) are developing patient cohorts to support research based on large-scale population data. These cohorts compare the demographics of their patient population to that of the whole population to understand where biases may exist.

At minimum the AI solution must be transparent about the nature of the data used in its construction and thereby where it is most fit-for-purpose. Models should also be tested across important subgroups as part of the validation. (Abràmoff et al., 2023[44]).

AI solutions leverage significant amounts of data, leading to privacy and security concerns: AI requires vast amounts of data for training. While individually identifiable information will be removed (such as name, address, phone number), there is concern that when information about the person is linked, that it may provide enough clues for an individual to be identifiable. When bad actors (or incompetent users) get access to that data, it could result in a privacy breach and potential harm from data misuse.

A related risk is that training data sets for AI may be a target of cyberattacks given their potential value. Such cyberattacks may cause a privacy breach, render AI solutions unusable, or change the training data sets to build bias (and thereby distrust) into AI algorithms.

To mitigate privacy and security concerns, there are various controls that could be implemented. These include privacy-enhancing technologies (such as encryption, pseudonymisation, or synthetic data), establishing a culture of privacy and security through training programmes, applying a clear code-of-conduct (with consequences), and updating guidance for access. Together, these will ensure the right person can access data in the right place at the right time with the right level of granularity and protections.

With the arrival of GDPR in the EU, some of the data and privacy concerns may in part be mitigated in this region but the challenge remains. For example, the trend among most OECD member states is that decision-making with regards to AI solutions – and more broadly digital health technologies – occur in silos with data authorities concerned about privacy while health authorities are concerned with quality, safety, and outcomes. This tension exists with poor co-ordination or minimal incentives to collaborate.

Regulation and policy frameworks for health data privacy and security should be modernised to recognise the value of national, regional, and global data collaboration with data and AI while continuing to provide protections on sensitive health data. A regulatory approach that supports innovation while also monitoring AI solution development is necessary. Explicitly, this does not require new regulators – rather authorities that exist should be encouraged to work together across borders toward common objectives that optimise outcomes while minimising harm.

This modernisation should recognise the variation in risks and rewards of AI solutions by providing clear guidance for classification of AI solutions with controls calibrated to the level and nature of risk. This is especially true where AI can meaningfully improve productivity of health providers without introducing risk to providers or their patients.

AI solutions may not be explainable, impacting accountable evidence-based decision-making: The "black box" nature of AI algorithms can lead to difficulty in understanding the rationale behind specific AI-driven outputs. This difficulty in understanding can grow to a lack of trust in solutions when coupled with the risk of AI solutions being trained on biased data.

While it is difficult to fully articulate the underlying mathematics in an easy-to-consume manner, it is important to develop guidance for explainability of AI solutions to ensure that sufficient information is provided to establish trust in outcomes.

Where appropriate, sufficient transparency should be provided to both the users of AI algorithms (e.g. health providers) as well as those impacted by its outcomes (e.g. patients). This should be communicated in language that is appropriate and consumable for the target audience while respecting intellectual property and preventing breaches of privacy. Transparency into the demographic data used in AI models will allow AI users to evaluate the appropriateness of the model in a given clinical context.

AI solutions may displace jobs, result in de-personalised care, or introduce new issues of liability: There is significant concern over the ability of AI to replace health workers through automation. Overdependence on AI and anthropomorphism (overly humanizing AI) can lead to a loss of human connection in healthcare. At an extreme, this would result in health providers being out of the loop with patients, only interacting with patients in more complicated situations where a health professional can provide assurance and guidance to the patient. The relationship between patient and provider is among the most trusted among government services. AI should serve to complement this relationship rather than displace it.

This risk can be partially mitigated through clear guidance on keeping "health providers in the loop" so that health providers are providing health advice – where that health advice may be informed by AI. Further, given the health workforce crisis that is being experienced by many countries, the adoption of AI could

contribute to improved workforce retention and productivity when solutions are designed around clinical workflows. There is an opportunity with AI to extend the capabilities of the existing workforce and improving their productivity while automating low value-added tasks. There is emerging evidence that leveraging AI in that way improves satisfaction for providers, gives them more time to care, and improves the quality of data that are produced. Using AI for low-risk tasks represents a win-win-win situation.

However, this will not always be an effective solution, as it requires that health practitioners are sufficiently digitally literate to use AI tools and that they have the health expertise to make judgements about the utility of the information generated by the AI solution. Further there may be reluctance to adopt new tools due to prior missteps with deployment of digital tool that added administrative burden without achieving promised benefits. At present, burnout and lack of health workforce capacity pose a significant risk to health systems and making use of tools such as AI can help relieve some of this burden. While not a solution to the health workforce crisis, AI can meaningfully contribute to solutions (Thomas Craig et al., 2021[45]). These are known risks that require a proactive regulatory environment.

AI solutions are context-dependent and emergent by nature: The performance of AI solutions can degrade precipitously after the transition from training and test data to production data; they can also initially perform well in a specific setting, only to decline over time; and their performance can vary significantly across settings. In addition, unlike a new drug or a new device, adaptive AI solutions are not static, rather they are emergent in nature, as they learn over time and new system properties can unpredictably appear.

The risk can be mitigated by requiring continual post-deployment monitoring and improvement. This could include demonstrating evidence, which may include public reporting of the explainability of outcomes, model performance as it relates to equity, effectiveness, safety, and accuracy. Tracking AI incidents (as has been recently established by the OECD (OECD.AI expert group, 2023[46])will enable wide-spread learning to ensure appropriate oversight and necessary response when issues occur.

A side-effect of this emergent nature may lead to inequities. When quality data about equity deserving groups are not available, then any AI solutions may not be fit-for-purpose for those populations. Being aware of this risk would lead to guidance on data required for training AI and for its ongoing surveillance.

3.3. Opportunities for artificial intelligence in health

Once barriers are removed, then risk controls can operate effectively for artificial intelligence to create benefits for individuals, health providers, populations, and health systems.

These opportunities rely on timely access to quality data that are fit-for-purpose for the intended use that would be achieved through effective health data governance. Data from AI solutions would inform providers and patients, feed back into decision-making like health technology assessments (HTA) to inform regulation or set evidence thresholds. It could monitor the technology once it is on the market like post-market drug surveillance. Insights from this data gathering would both improve safety and be used to inform upstream design of future AI solutions as well as to understand its degree of generalisability (Srivastava et al., 2023[3]).

Note that examples provided in this section are based on ad hoc and small-scale AI solutions. Many countries are exploring actions to enable the ability to utilise AI solutions for broader populations.

Empowering individuals and personalising care: Health AI can support individuals by providing them with greater insights based on their personal health data (Price and Cohen, 2019[47]). AI-powered tools can help patients better understand their medical records as well as aiding navigation of complex healthcare systems (Krumholz, 2014[48]). Generative AI models can also serve as “co-pilots” for patients, helping them make informed decisions regarding their care (Topol, 2019[49]). These AI-driven opportunities can lead to

more personalised and effective healthcare, as patients become more engaged in their care and make informed decisions that align with their health goals.

For example, a person with diabetes can use a continuous blood glucose monitor to track their blood sugar. When there is a change in blood sugar level, the AI could suggest specific interventions for the treatment of their chronic disease appropriate for their personalised characteristics. Further, the AI can learn patterns over time when the person frequently experiences change in their blood sugar and make proactive recommendations to prevent negative incidents. As such, personalised healthcare can contribute to improved health outcomes, reduced healthcare costs, and increased patient satisfaction (Hibbard and Greene, 2013[50]).

Overall, the role of the patient may change dramatically following the broad adoption of AI in healthcare. There is a large unmet need for people suffering from chronic diseases, where they only interact with specialists every 3-6 months (for example, with auto-immune diseases). Under the guidance of a health provider, advancements with AI can leverage innovations such as cell phone cameras can take very high-resolution photos of the eye and other bodily areas that are rich in diagnostic information and avoid expensive and time-consuming lab tests. Commoditising AI tools to be available on cell phones to give medically approved advice on interventions on an hourly basis can help with management of chronic conditions (e.g. chronic pain). Given the growth in patients with multiple chronic illnesses – and its related complexity – this will be a revolutionary and necessary step in the right direction for improved quality of life, better health outcomes, lower risk of disease, and the potential to lower healthcare costs.

There are clear signals that the public are looking to benefit from the opportunities from better use of health data through artificial intelligence. A study from Canada identified that patients expect their data to be used to (a) improve their health; (b) improve safety of the health system; and (c) improve understanding of treatment and prevention of disease (Burt et al., 2022[51]). The European Patients Forum (EPF) recognised that AI would enable providers to have more time to care, enhance the quality and efficiency of research and development, improve the quality of diagnosis, and delivery personalised care. In achieving these benefits, the EPF had nine recommendations including respecting human dignity, engaging patients and providers in design, implementation, and operation of AI solutions, and keeping humans in control. (European Patients Forum, 2023[52])

Streamlining processes and enhancing decision-making: Health AI can support healthcare professionals by serving as a documentation assistant, reducing the administrative burden on clinicians (Shanafelt et al., 2016[53]), which may free up between 10-30% of their time for care (Beamtree Global Impact Committee, 2023[5]). AI algorithms can also scan and analyse longitudinal patient records alongside the most current scientific literature to identify critical insights for patient care (Hinton, 2018[54]) In radiology, AI can assist in scanning medical images, enabling faster and more accurate diagnoses (Lakhani and Sundaram, 2017[55]). AI can also scan academic literature to surface the most relevant articles for consideration to close the 17-year gap of publication-to-practice (Morris, Wooding and Grant, 2011[56]) – a problem that will only become more difficult for health providers when there are more than 3 000 publications to read each year (Mimouni et al., 2010[57]) (Alper et al., 2004[58])

By streamlining processes and enhancing decision-making, AI can help health professionals focus more on direct patient care, leading to improved outcomes and increased job satisfaction. Additionally, AI can support the ongoing professional development of healthcare workers by identifying knowledge gaps and facilitating access to relevant resources (Davenport and Kalakota, 2019[59]). When involved in a patient consultation, AI can also prompt the provider to ask questions based on the patients' profile.

Advancing drug discovery: Health AI has the potential to significantly impact public health by aiding in drug discovery for superbugs (Liu et al., 2023[60]) and combating antimicrobial resistance (Vamathevan et al., 2019[61]). Additionally, AI can track the continuum of care and wraparound services as part of community health efforts, predicting the best preventive and chronic care management opportunities.

As reported in May 2023, a short list of possible antibiotics for a superbug was discovered in 2 hours – identifying 240 out of over 7 000 candidates and ultimately accelerating the time to start clinical trials and overall required fewer resources. This demonstrated the potential of AI as a tool to help with addressing increasing anti-microbial resistance (Liu et al., 2023[60]).

Looking forward to personalised medicine, AI will help unlock insights from genomics and synthetic biology while being sure to continue to protect privacy and security.

Optimising efficiency and addressing inequities: Health AI can help find inequities of healthcare systems (such as underserved populations), identify geographic areas requiring assistance, and support workforce retention (Chen et al., 2023[62]).The AI can then identify the levers that would improve those situations for policy makers to make evidence-based decisions.

By automating non-clinical activities for providers, particularly in rural and under-resourced regions, AI can contribute to improved healthcare delivery and outcomes while also improving provider satisfaction and performance (Chen et al., 2023[62]; Jha et al., 2013[63]).

The optimisation of health systems through AI can lead to more efficient resource allocation to increase preparedness and resilience to public health threats (Braithwaite et al., 2018[64]). AI can also assist in predicting and managing healthcare demand, enabling health systems to adapt to changing needs and enhancing the overall quality of care (Song and Veeraraghavan, 2018[65]).

Economic opportunities: In addition to the significant human benefits from AI as noted above, it is projected that the economic value of the AI market will grow USD 183B in 2023 to USD 995B in 2028, which represents more than 450% growth (Lawson, 2023[66]).

This growth could be disproportionately larger in health given both the size of data assets and the breadth of opportunity. These gains will come from improving productivity of front-line providers (e.g. automating administrative functions); better leveraging existing data (e.g. diagnostic support for health providers by leveraging comprehensive patient health records, coupled with leading literature); and discovery of new innovations for improved treatment or prevention. These opportunities will only be realised if there is timely access to quality health data that are used to create responsible AI.

4 Practical actions to enable AI for the long-term

There are practical actions that can be taken immediately to advance the responsible use of artificial intelligence in health while also initiating action to prepare governments to be ready to meet future challenges – such as public health emergencies, global ageing, and climate impacts on health. When designed in the broader context of health in the digital age, these actions will also benefit the adoption of digital tools and the use of quality data to improve individual, population, and public health.

As such, AI will be a key contributor to global health, enabling learning from cross-border data sets, helping to identify systemic inequities, and driving health systems to be more resilient and sustainable while evolving healthcare toward integrated and person-centric learning health systems.

Actions fall into four categories – improving trustworthiness, building capacity, evaluating, and evolving solutions, and accelerating progress together.

4.1. Actions to improve trust in the use of AI solutions by health systems

While artificial intelligence has been maturing for decades – with increasing accuracy, usability, versality, and protections – it has only recently risen into the public's awareness by being more widely available to individual consumers. Many references to AI in popular culture have presented AI as villainous and to be feared. Several attempts to leverage AI have had embarrassing results such as IBM's Watson being ineffective in managing cancer patients, skin lesion scans not being effective for people with darker skin, and recent "hallucinations" produced by generative AI systems. As of February 2023, more than 60% of US patients have indicated they would be uncomfortable with the use of AI in their care (Tyson et al., 2023[41]). Advancing AI without the trust of the people who will be impacted most by its use is likely to further damage trust or to have no return on investment.

Democratic values and respect for human rights and dignity are core to health innovation. Bioethics has long argued that foundational consent – the right of individuals to knowingly contribute their health experiences to inform future medical practices and interventions – is the cornerstone of the public's trust in medical sciences. This is not an abstract tenet or bureaucratic formality. Meaningful consent and a focus on patients' rights are explicit efforts to tend to the fragile state of the public's trust in medical sciences. Advancing AI in biomedicine and health practices will depend on even deeper learning from people and the data that they generate to innovate. AI innovation will hinge on developing the capacity to engage people in their health journeys, inviting them to share intimate details about the social circumstances that drive their individual and community health. AI in health will require more trust than we have ever asked of patients.

Hence, efforts should be made to understand the requirements for AI to be considered trustworthy in its use. Those requirements will vary by the context for the use of AI as well as the culture of the environment in which the AI is to be used. Once the requirements are understood, then rules can be implemented so that the trustworthiness of AI solutions may be demonstrated, and feedback sought from impacted

stakeholders. This process can be iterated to continuously improve the efficacy, effectiveness, and efficiency of AI solutions.

Some specific areas to explore in establishing the trustworthiness of AI:

- **Engage and listen to public and provider concerns and involve stakeholders with respect to opportunities, risks, and controls:** Impacted stakeholders (e.g. providers and the public) will have concerns about AI solutions in terms of its development, adoption, and use. Trust will be fostered by listening to those concerns, addressing them directly, mitigating risks, and communicating clearly and directly. Involvement could include co-design of solutions to ensure they are effectively incorporated into clinical workflow and practice. Communication could include rationale for the controls, residual risks, and how they will be managed.
- **Publicly report and oversee AI performance:** Establishing transparent reporting for AI performance can be helpful in demonstrating trustworthiness. This may include comparing the benefits of AI performance over traditional methods as well as their effectiveness over time and across populations (e.g. for groups experiencing inequities). Public reporting will be part of human oversight of implementation of AI solutions. This may involve creating guidelines for AI safety and accountability, evidence generation to prove its efficacy, effectiveness, and quality, fostering collaboration between AI developers, healthcare professionals, and regulatory authorities, and ensuring that AI systems are grounded in human values and ethics. Transparent reporting about AI incidents – including impacts, lessons learned, and adjustments – will be a key component of trustworthiness.
- **Establish rules about data control:** Governments should work with the public, communities, providers, and other key stakeholders to establish methods of data control that are appropriate for the digital age. Controls would recognise the transformative value of data and AI while also understanding its risks. One way to build capacity to collect valuable, contextualised data to develop AI models would be to engage patient communities in the governance and stewardship of the uses of their data, both as individuals and as members of communities. The alternative – taking patient data without meaningful engagement – will only make people more protective of their health journeys and less trusting of entities that need to collect data from them. This would establish new approaches to access, privacy and consent of personal health data that are suitable for the digital age.
- **Incentivise and oversee adherence to responsible AI practices:** Codes of conduct are standard practice in healthcare. This concept could be extended to define a code of conduct for responsible AI in health (See Box 4.1). Once that has been developed and implemented, rules may be put into place that encourages adherence to the code of conduct. For example, if the public have concerns about their data being re-identified and then breaching their privacy an action could be to have a policy where there are significant fines or legal action should there be an unauthorised intentional re-identification of data. Such policies are already embedded in privacy laws such as the EU's General Data Protection Regulations.

Box 4.1. Spotlight: AI Codes of Conduct for health

Codes of Conduct provide frameworks to clarify responsible behaviour within an industry and may be used as the basis for incentives that encourage adherence to desired behaviours.

The acceleration of science in the past century and the rapid advancement of computing power and the development of AI have identified both new opportunities and risks alongside increasing complexity. This is driving the development of a forward-thinking code of conduct for AI.

AI Codes of Conduct can serve as a moral compass for organisations to navigate situations where the legal framework might be unclear, inadequate, or evolving as well as establishing a standards framework that then can be used to validate applications and assess AI's effects. These codes of conduct are developed with multistakeholder collaboration with input from diverse representation such as law, medicine, governance, science, and ethics; as well as individuals and communities representing those affected by the use of AI.

The G7 published guiding principles for all AI Actors and for organisations developing advanced AI systems and a code of conduct for organisations developing advanced AI systems in October 2023 (European Commission, 2023[67]). In parallel, the National Academy of Medicine has been undertaking an AI Code of Conduct specific for health, healthcare, and biomedical science. These AI Codes of Conduct are designed to be comprehensive and flexible to evolve with changing environments. They are designed to ensure the safe and responsible use of AI by providing a holistic approach, combining ethical values, technical best practices, and legal compliance. The objectives of these codes are to prevent harm stemming from AI systems while promoting innovation to address complex and challenging problems.

Contextualising an AI Code of Conduct specific to health would help address the complex balance of priorities in healthcare involving highly sensitive data, significantly large data volumes, and the ability to drastically impact human lives – both for good and bad. Such a code of conduct must be aligned to whole-of-government approaches for AI while addressing health-specific concerns.

Common components in AI Codes of Conduct for health could include:

- respect for human dignity, autonomy, and rights;
- equity;
- the avoidance of harm through the full AI life cycle; and
- the adherence to the rule of law,
- with an emphasis on transparency, safety, data protection, and environmental impact.

In addition to the high level ethical guiding principles, codes of conduct include practical guidelines for deployment consistent with the principles. Examples include:

- risk management,
- data integrity,
- confidentiality,
- traceability, and
- respect for intellectual property.

As has been experienced with other digital health advances, expectations, norms, and values change after experience with new technology and the outcomes of societal debate on the issues. Therefore, ongoing testing, validation, and improvement of the codes are necessary to keep pace with change.

Source: (National Academy of Medicine, 2023[68]; ECP, 2018[69]).

4.2. Actions to build capacity for responsible AI

Humans are at the centre of AI. Humans decide on the methods and data to be used in AI design and development; how the AI solution is deployed and adopted; where the AI solutions are used; how it is monitored over time; and how the AI may be scaled to other use cases, populations, or geographies. Just as it is essential for AI to keep humans in the loop, it is essential for humans to understand AI to leverage it for the public good.

For humans to be at the centre of AI, there is a need to build capacity among the health workforce to be involved in the design and use of AI, for policy makers to guide and oversee their implementation, and for professionals to develop, deploy, operate, and evolve AI solutions. Building that capacity may be multi-faceted, starting with targeted continuous education for existing professionals, embedding AI in medical programmes, and eventually including AI and digital health as part of pre-university education.

Although job displacement is a concern in many industries, the healthcare sector faces a workforce crisis that AI could help address. Studies have shown that AI can automate at least 15% of administrative work in healthcare in addition to reducing errors and wasteful redundant testing. Ensuring that AI is used to complement human workers, rather than replace them, is vital. That involves demystifying AI and engaging health workforces and labour unions when possible (Beamtree Global Impact Committee, 2023[5])

Some specific areas to explore in building capacity:

Grow AI and data-driven culture among all stakeholders: The future of artificial intelligence involves buy-in from the **public** and their **providers** as well as enabling action from **health system leaders** and managers. Efforts should be taken to improve awareness, transparency, and literacy for all aspects of digital health (including AI). Where appropriate, incentives (or dis-incentives) could be used to help encourage action. Includes working with regulators on practical solutions that foster trust and developing an evidence-base to understand the economic and health system impact of AI, while working with the media to provide balanced perspectives on adoption of AI. This should be approached as a long-term exercise to both educate and involve stakeholders.

Workforce (re-)training: In general, health workers are required to engage in continuous education to learn and apply new health discoveries and innovations. The use of AI is like the discovery of a new protocol for treating diabetes – the health worker needs to know why the innovation matters, what they must know, and how to make most effective use of the innovation. While AI is an exceptionally broad topic, continuous education can involve specific applications of AI while providing background on how the solution was developed, the value of it being adopted, how it may be used, and how it is monitored. The training would involve understanding the risks and benefits of using AI solutions, how to communicate its use with the public, their role working effectively alongside AI systems, and fostering a culture that draws on AI to create more opportunities for empathy and human connection.

Build sufficient technical capability and capacity to support responsible AI solutions: The health workforce is dominated by front-line health providers who provide direct care. In health, approximately 5% of budgets are allocated to the expense of building, implementing, and operating the technology solutions that support those front-line workers. Other industries typically allocate approximately 8% of their budgets to technology. Notably, the financial industry – managing large volumes of sensitive personal data – spend around 10% on technology. As health seeks to implement AI at scale across health systems, it is likely that expanding the size of the technical workforce supporting those systems will be necessary. Establishing communities of practice for technical health professionals – both within countries and across borders – can help build capabilities and lower barriers to co-operation and collaboration (Weins, 2020[70]).

Box 4.2. Spotlight: Building capacity for responsible AI

Finland: In December 2019, Finland decided to invest in the future of skills of Europeans through the development of a free online course that came to be known as "Elements of AI". With its initial roll-out, the aspiration was to de-mystify artificial intelligence for the public and policy makers. Common understanding would be a foundation from which trust could be established in the creation, implementation, and use of AI across the country. While not a health-specific course, course materials often considered the benefits of AI for cancer treatment and other aspects of healthy lives.

The original objective for the course was to train 1% of people in Finland, or around 60 000 people. Since that time, the aspiration grew to train 1% of people in Europe. As of September 2023, more than 1 million people have successfully completed the online course.

Canada: Canada is on the verge of an AI revolution with the potential to impact every sector in the coming decade. AI's predictive, analytical, and automation capabilities offer solutions to complex challenges, from healthcare improvements to assisting in climate change mitigation. To engage and inform the Canadian public, the Advisory Council on Artificial Intelligence established the Public Awareness Working Group. Through a national survey and online workshops involving over one thousand Canadians, the group assessed public awareness, perceptions, and educational needs regarding AI. The consensus was clear: enhancing AI education, focusing on responsible use and critical assessment, is essential. Canada's approach to education and responsible use of AI includes:

- Creating a national AI community of practice
- Utilising existing initiatives for AI awareness and offering a free online AI literacy course
- Encouraging public dialogues to align AI use with public interests
- Launching government-led information campaigns
- Funding inclusivity strategies such as ensuring high-speed internet access

Australia: As part of the work of Australia's Digital Transformation Agency, they established a toolkit to encourage the use of responsible AI in the public service. The toolkit covered areas such as:

- Decision-making and accountability hierarchies
- Risk management frameworks
- Ethical adoption of AI technologies
- Enabling human, societal, and environmental well-being
- Human-centred values and Fairness
- Privacy protection and security
- Reliability and safety
- Transparency and explainability
- Contestability and accountability

As a complement to this work, the Australian Public Service Academy offers courses on data literacy as well as data leadership for senior executives in the public service.

Source: (Australian Government, 2023[71]; Government of Canada, 2022[72]; University of Helsinki, 2021[73]).

4.3. Actions to evaluate, validate, and evolve AI solutions

Health has processes in place to measure and understand the risk of novel tools used to provide care. Health Technology Assessments are used to evaluate and validate new technologies to determine if they are safe, cost-effective, and valuable for the public. Post-market drug surveillance is used to ensure the ongoing efficacy of drugs and to capture any unintended side-effects. This was crucial in the vaccination roll-out of COVID-19 to validate its safety while also checking its effectiveness. Note that AI will likely not require new regulators for health AI. Rather, there is opportunity to extend already existing regulatory capabilities that already have significant economic evaluation and risk management expertise.

There will be value for patients, clinicians, health policy makers, and regulators in understanding how to best evaluate and update AI solutions based on new information and technical advances. There is value in co-ordinating these evaluation activities across borders to share information about incidents as well as methods to scale responsible AI innovations.

- **Evaluate impact of AI/digital investments:** Many countries are advancing evidence-based policy as a foundational approach to policy development. At present, metrics associated with digital health programmes tend to be focussed on volume rather than outcomes. Solutions should be increasingly designed to be able to answer questions of the impact of investments. The collection of high-quality data – much of which is a natural by-product of the use of AI tools – will facilitate policy for the value-based use and reimbursement of digital tools (including, but not limited to, AI). For example, evaluations could include measuring the availability of quality data for AI innovations or scale of AI deployments. These statistics can be correlated with health system care indicators to determine which investments in AI solutions generate value-added outcomes. That would be used in financial decision-making to encourage health data re-use, adoption of digital tools, and AI enablement. These evaluations will also help with ongoing validation and regulation of responsible AI solutions. Economic evaluation will support transparency in methods of AI research, demonstrate their effectiveness and the health benefits they bring. They will lead to knowledge generation and increase policy makers capacity and understanding of AI solutions, supporting digital health technology vigilance (Srivastava et al., 2023[3]). In terms of impacts of AI investments, solutions could measure reduction of medical errors, reduction in use of unnecessary health services, improvements in population health, reduction of chronic illness, reduction of non-value-added administrative tasks, improvements in accuracy of diagnosis and prognosis, and improvements in the timeliness of health service delivery.
- **Evaluate AI/Digital capacity:** Much of recent focus in health workforce has been centred on the deficit of front-line health providers. In the digital age, it is possible that a digital resource can have an outsized benefit on health outcomes. If the implementation of an AI algorithm can save 5 minutes per health provider per day, it would represent a productivity improvement of 1%. While seemingly small, across the health workforce, this could translate to be the effective equivalent of avoiding the need to add 1% new health providers – which on a workforce of 100 000 employees would be the equivalent of 1 000 additional staff. At present, few countries are measuring the size of the digital workforce, which would include the workforce for AI, data science, data stewards, data protection officers, and IT functions. That would provide insight into the value of building capacity programmes to grow digital and AI capabilities to help the health workforce of the future.
- **Evolve AI solutions for new populations and environmental factors:** AI solutions are trained using data available at that time. As considerations are made to scale AI solutions to new environments, an evaluation of the appropriateness of the solutions should be made to ensure it is fit-for-purpose given the characteristics of the new community. For example, chest X-rays before 2020 didn't have COVID-19 infections present, and now many of them do. Governance of AI needs to ensure that the architecture of AI solutions can be updated to evolve with changes in the environment – including changing socio-demographics (e.g. growth of people with multiple chronic

conditions) – and that the use of AI solutions is fit-for-purpose given the context of how the AI solution has been developed.

Evolve (Dynamic) Health Technology Assessments for the Digital Era: Many countries use Health Technology Assessments (HTAs) to assess the value of new medical technologies (drugs, devices, procedures, and increasingly, digital health tools) and determine their reimbursement –both whether and how much to pay. Such a system typically involves a clinical benefit assessment followed by price negotiations that are informed by the data from the assessment to better align the prices paid with the value created by new medical products (Lauenroth et al., 2020[74]). Countries, such as Austria, use a system of "reference pricing" to determine prices for new medicines based on what is paid in comparable markets. Germany has recently introduced a fast-track procedure for digital health applications (Leitfaden Für Hersteller and Und Anwender, 2023[75]). Other countries, such as England, conduct cost-benefit assessments that are based on the calculation of a cost-effectiveness ratio of a new product vis-à-vis a pre-determined willingness-to-pay threshold, leading to a yes/no coverage decision for a new product. In the context of digital products, health economists have argued for the need for a *dynamic* approach ("Dynamic HTA"), which would incorporate "the dynamic nature of both product development and product evaluation" (Brönneke et al., Forthcoming[76]). Such an approach is highly relevant for AI tools, which are a special class of digital products. This approach would need to be led by government and involve public and private sector in the implementation and evolution of this innovation.

Box 4.3. Spotlight: Emerging regulatory tools for AI

United Kingdom: The National Institute for Health and Care Excellence (NICE) in the United Kingdom updated their evidence standards framework (ESF) for digital health technologies to include evidence requirements for artificial intelligence (AI) and data-driven technologies with adaptive algorithms (National Institute for Health and Care Excellence, 2018[77]). The ESF consists of twenty-one standards, mapped to the phases of the digital health technology product lifecycle: design factors, describe value, demonstrating performance, delivering value and deployment considerations. A consistent framework provides structure and predictability for innovators to be able to develop solutions that will be interoperable with the UK's digital health ecosystem.

OECD: The OECD published an AI Classification Framework in 2022. This framework proposed that classification of AI should occur along five dimensions: People & Planet, Economic Context, Data & Input, AI model, Task & Output (OECD, 2022[78]). Different types of AI systems raise different policy opportunities and challenges. The framework allows users to zoom in on specific risks that are typical of AI, such as bias, explainability and robustness, yet generic in nature. It facilitates nuanced and precise policy debate. The framework can also help develop policies and regulations, since AI system characteristics influence the technical and procedural measures they need for implementation. The dimensions of the OECD Framework for the Classification of AI Systems can be associated with stages of the AI system lifecycle to identify a dimension's relevant AI actors, which is relevant to accountability.

HealthAI (formerly I-DAIR – International Digital Health and AI Research Collaborative): HealthAI is a transnational not-for-profit organisation that is seeking to advance the governance and regulation of AI to build trust, advance equity, and deliver on the potential of emerging technologies. HealthAI will establish a common methodology that evaluates, validates, and learns from experience. When implemented, this will build trust in the use of AI by supporting validation that certifies that attributes of the AI algorithm meet the standard to be considered responsible AI. HealthAI will also champion networks to share practices that enable the scale of AI across borders. This will encourage innovation, economic growth, and help to reduce inequities (Health AI, 2023[79]).

4.4. Actions to accelerate action on AI in health together

To counter the potential risks of an AI arms race, the establishment of a pre-competitive space is essential so that advancement with AI solutions in health can be made together. This would create common guardrails for all countries and innovators for responsible AI development and deployment for health – building on policies such as the EU AI Act and declaration from the AI Safety Summit at Bletchley Park (see Box 2.3). Collective action would also mitigate the risk of expanding digital divides from bespoke AI solutions developed in wealthy health facilities and only available to wealthy patrons with no plan or capability to scaling to broader populations within and across borders.

Recognising that there will be bad actors advancing simultaneously, it is crucial to be prepared to meet them on the field of battle (e.g. security) and collectively guard against them (by prevention of the "weak link"). Having said that, countries are different in places economically and culturally to be ready to adopt AI.

Further, the lack of alignment on core principles for responsible AI can pose risks to economic development and global competitiveness. Harmonisation (or convergence) of public sector AI-related legislation and regulation would enable global companies to navigate fragmented AI governance structures and provide smaller entrepreneurs with a level playing field to compete with larger corporations. That harmonisation will also enable countries to learn from each other so that regulatory processes are effective and scaled.

The significant potential harms from AI will not stop at borders. These are some potential actions to help accelerate collective action within and across borders to optimise the beneficial impact of AI in health while mitigating its risks.

- **Adopt nimble approach to health AI regulation and oversight:** Legal and regulatory frameworks are often designed for analogue paper processes and may not be suited to address the unique challenges posed by digital solutions such as AI as well as the intersection of AI and health. A nimble approach to policy and regulation should be adopted to achieve desired (real-time) outcomes while mitigating potential harms. That approach would embrace continuous review-recommend-approve-update cycles. This may involve updating existing laws and regulations to provide structure to govern AI solutions and flexibility to adapt laws to evolving capabilities. Such laws would create new standards for AI safety, quality, transparency, economic evaluation, liability, and accountability and foster international collaboration to harmonise AI in health regulatory frameworks while supporting innovation. That approach would actively rationalise regulations and incentives to ensure they were appropriate, and implementation was aligned with intention. This approach would adopt and operationalise principles for responsible AI in health.
- **Establish public-private-person-provider-partnerships:** In health, data are generally managed by public sector resources whereas the private sector has capability and capacity to innovate (e.g. new medical devices, pharmaceuticals, personalised digital therapeutics) to make use of that data for the public good. There are opportunities to encourage relationships between the public and private sector. It is important to extend those partnerships so that health providers and the people that are impacted by solutions are directly involved in the design of new innovations. This may include co-developing systems with patient communities, updating procurement guidance, good practices, and education for decision makers in how to determine if, when, and what AI solutions to purchase based on economic evaluation, and how to incorporate appropriate risk assessment, risk management and governance into contracts.
- **Elevate health data governance:** OECD countries are implementing the Recommendation of the Council on Health Data Governance (OECD, 2022[37]). Efforts should be accelerated to ensure that quality data is available for the innovations in AI and its subsequent adoption, use, and scale – both within countries and across borders. This would define the "pre-competitive space" that establishes the policy and technical environment for AI to flourish with trust, respect, and equity.

As we move to new data governance infrastructure, prioritise interoperability, and the rights of patients to be key partners in data governance. Strengthening health data governance is one of the most impactful ways to mitigate risk of cyberthreats while lowering the cost and complexity of innovation. Greater maturity with health data governance would strengthen the approach to digital security – including partnership across industries and borders – to understand threats and develop co-ordinating approaches to respond to cyberattacks. Health data governance would also simplify interoperability for timely access to quality data for AI solutions – by harmonising policies and standards in the areas that matter most to foster innovation and remove unnecessary and unhelpful barriers to responsible AI in health.

- **Leadership and collaboration for collective impact**: Most governance rules, policies, good practices, and evidence of value should not be tightly siloed within (or across) countries. Collaborating with other countries and international organisations, reusing approaches, and sharing knowledge and learnings will be an effective use of resources, accelerate progress, and lead to better and safer AI benefits realisation. Collaboration will also help development, spread, and scale of innovation. Effective collaboration would create space for the development and testing of new innovations that demonstrate the potential of the innovation at scale while minimising harms and mitigating risks together. This provides inspiration for co-problem solving across stakeholders (including across borders) and identify the key factors that contribute to the development, adoption, use, and scalability of new innovations. For example, this could help share insights on facilitating culture change, enabling inter-organisational collaboration, and developing financing models / incentives that advance digital transformation of health and care. Leadership would leverage classification and risk frameworks to identify its tolerances to minimise risk of AI implementations in health. For example, leadership may prioritise initial work that is of low risk and high value (e.g. automating administrative tasks for health providers).

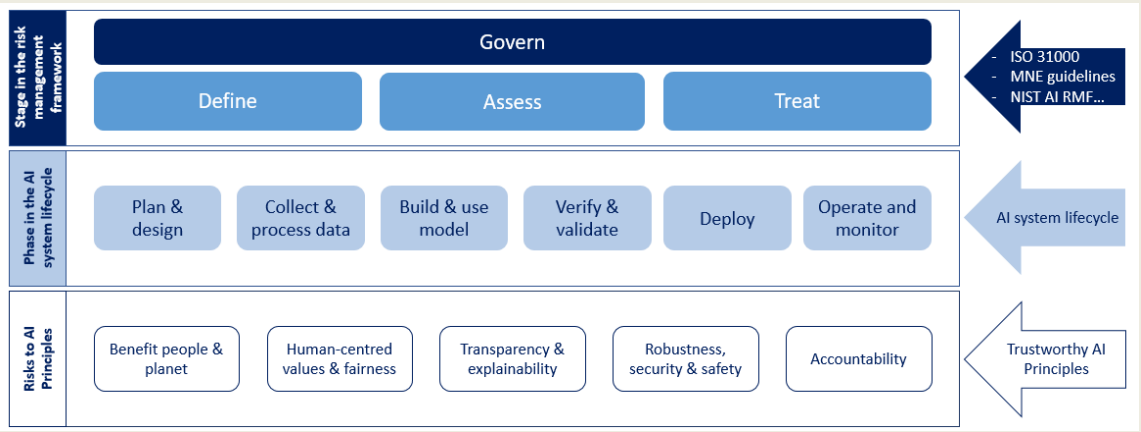
Box 4.4. Spotlight: Regulatory frameworks for Artificial Intelligence

The OECD AI Policy Observatory has been tracking the development of AI-specific legislation and regulatory frameworks around the world. For example, **Canada** proposed the Artificial Intelligence and Data Act (AIDA) which aims at establishing guardrails for AI and the responsible use of data for commercial applications; the **European Union** proposed regulation for harmonised rules on AI (EU AI Act); the Council of Europe is working on a Framework Convention on AI, human rights, democracy, and the rule of law. The **National Institute of Standards and Technology** in the **United States** established an AI Risk Management Framework. AI standardisation initiatives are being led by international organisations including the **International Standards Organisation (ISO);** the **Institute for Electric and Electrical Engineers (IEEE),** the **International Telecommunication Union (ITU)**, and the **Internet Engineering Task Force (IETF)**.

The OECD AI Expert Group on Risks and Accountability compiled information about these frameworks. When synthesised, these frameworks all provide common guidance:

- **DEFINE**: Clearly establish scope, context, criteria, relevant AI principles, stakeholders, and actors for each AI system lifecycle phase.
- **ASSESS**: Identify and analyse risks at individual, aggregate, and societal levels, evaluating their likelihood and potential harm, considering also that small risks can accumulate.
- **TREAT**: Mitigate risks to prevent adverse impacts, proportionate to their likelihood and severity.
- **GOVERN**: Embed a risk management culture within organisations, monitor and review the process continually, and document, communicate, and consult on the process and outcomes.

This synthesis continued to identify common steps in the AI lifecycle as well as common principles to guide the development of responsible AI.



To date, there is no identified legislation that is specific to AI in Health, although some AI systems used in healthcare may be considered as high-risk within the EU AI Act ("access to essential private and public services and benefits"). Furthermore, existing legislation (such as data protection legislation and other health legislation) already include provisions relevant for AI use in health.

In the last quarter of 2023, draft frameworks for AI and health have been published for discussion including from PATH (PATH, 2023[80]), UK Law Society (Dejeu-Castang, 2023[81]), and Global Partnership on AI (GPAI, 2023[82]), among others.

Source: (OECD, 2023[83]).

5 Accelerating Artificial Intelligence in Health

AI presents tremendous opportunities and tremendous risks. The risks are manageable and not proceeding with urgency introduces additional risks of expanding digital divides and continued fragmentation. The promise of AI is in the opportunity to co-ordinate national efforts to learn from each other and unlock the opportunities presented by responsible AI in health. To develop AI in the public interest, the public sector must have its hands on the proverbial steering wheel.

There are real, practical, and low-cost actions that can be taken immediately. Failure to act risks the development of ineffective and/or irresponsible AI which causes real harm and sets back the willingness of providers and the public to adopt these advanced technologies while expanding the digital divide. At an extreme, such harms may prompt significant resistance toward AI that prevents the benefit from being realised. Leadership is required to define the path forward.

For health to achieve the benefits that other industries are already enjoying, health leaders must start to take considered action that manages risks while being focused on beneficial outcomes. AI is already in use in industries that are high-risk (aeronautics), high sensitivity (banking), and high-volume (consumer goods). While health is an important case, it is not unique. There is opportunity for learning.

Simply applying innovative technology to broken systems will not achieve the desired transformation or benefits. As countries are recovering their health systems in the wake of COVID-19, many are looking to address challenges with the health workforce while increasing resilience to future pandemics and incorporating person-centric and integrated models of care. As solutions are developed, recognising that digital health and AI could contribute to these areas, it is important to formulate plans that challenge the digital divide to improve health equity, so everyone has access to quality health services when and where they are needed.

There are tactical actions that can be initiated together – to elevate health data governance, harmonise policy, collaborate across governments, engage with industry to set guidance for procurement, and building AI and digital capacity in the health workforce. There are also strategic actions that are equally important, but will take more time, creativity, and will: (1) Evaluate the impact of AI and data investments; (2) Examine opportunities to build agility and resilience into legislation, policy, and processes; (3) Drive an AI and data-aware culture among all stakeholders; and (4) Establish regulatory sandboxes to encourage responsible experimentation and innovation.

The OECD is ready to help collective action across countries and regions by mobilising action to move from principle to action. The OECD would focus its work in three areas.

- First will be to help communicate the opportunities and risks of AI on health outcomes to articulate how AI will save lives and how individuals will be treated with dignity. This would involve leveraging the OECD AI Incident Monitor, mining academic literature, and identifying tangible cases. This work would evolve economic modelling of the costs and benefits of AI in health over time to support appropriate investments in foundational infrastructure to drive innovation and better health outcomes for all.

- Second, the OECD would focus on activities to operationalise policies and codes of conduct that remove the unnecessary and unhelpful barriers to responsible AI while ensuring appropriate risk classification frameworks, mitigations, and oversight are in place. Immediate action would provide a deeper analysis of the specific risks for AI in health related to key use cases (e.g. drug discovery, diagnostic support, administrative automation) along AI systems' lifecycles. This would help develop guidance for effective risk mitigation and accountability for each AI actor along with the approach to liability. In parallel, policy guidance for responsible AI in health would be created, along with guidance for harmonisation of policies that support cross-border collaboration on AI that optimises health outcomes for all, drives economic growth, and protects individuals, communities, and populations. This work would be grounded in the OECD AI Principles and emergent AI in health frameworks.
- Third, the OECD would track progress with development of AI policy and foster knowledge sharing for collective learning and response. This would enable co-learning, joint problem solving, and co-ordination toward mutual beneficial outcomes.

In doing this, the OECD would leverage its broad knowledge, convening expertise, and trusted position to work with partners to develop policy guidance and toolkits for developing a health-specific approach to advance responsible AI that is grounded in a multi-sectoral whole-of-government approach. This would identify where health-specific approaches for AI may be needed and/or beneficial while promoting fairness, ethics, people-centredness, transparency, and accountability of responsible AI solutions. Understanding the issues of AI in health will help inform approaches in other industries and support cross-industry action.

Policy guidance and toolkits could be developed by an AI- and health-focussed expert group that is convened and supported by the OECD. This group would partner closely with the WHO/ITU partnership on AI, Global Digital Health Partnership, and the WHO-led Global Initiative for Digital Health, among others, to support alignment and coherence across international partners while also being aligned with the directives to the OECD from the G7 Hiroshima agreement and other similar directives. The group would also review, benchmark, and share practices for AI in health across OECD countries.

While it will take time, leadership, will, effort, and investment to achieve and sustain benefits from AI in health, **urgent action is necessary**. Time is running out for policy makers to stay ahead of the curve and take control of the evolution of AI in health systems before technology dictates its own future.

References

Abràmoff, M. et al. (2023), "Considerations for addressing bias in artificial intelligence for health equity", *npj Digital Medicine*, Vol. 6/1, p. 170, https://doi.org/10.1038/s41746-023-00913-9. [44]

AI Safety Summit (2023), "AI Safety Summit 2023: Roundtable Chairs' Summaries", https://www.gov.uk/government/publications/ai-safety-summit-1-november-roundtable-chairs-summaries/ai-safety-summit-2023-roundtable-chairs-summaries-1-november--2 (accessed on 4 November 2023). [31]

Aiyer, B. et al. (2022), "New survey reveals $2 trillion market opportunity for cybersecurity technology and service providers", *McKinsey & Company*. [10]

Alberta Virtual Care (2023), *Interoperability Saves Lives*, https://open.alberta.ca/dataset/1394ebca-9869-40d6-b5af-3c6870557f21/resource/d9558cbb-220e-4b28-a05e-. [33]

Alper, B. et al. (2004), *How much effort is needed to keep up with the literature relevant for primary care?*. [58]

AMA Journal of Ethics (2002), "Learning to Listen", *AMA Journal of Ethics*, Vol. 4/3, https://doi.org/10.1001/virtualmentor.2002.4.3.mhst1-0203. [23]

Arasu, V. (2023), *AI Outperformed Standard Risk Model for Predicting Breast Cancer*, Radiology Society of North America. [18]

Australian Digital Health Agency (2023), "Clinical Governance Framework for Digital Health". [39]

Australian Government (2023), *Adoption of Artificial Intelligence in the Public Sector*, Australian Government Digital Transformation Agency. [71]

Beamtree Global Impact Committee (2023), *More Time to Care - Automation, Digitisation and the Workforce*. [5]

Braithwaite, J. et al. (2018), "The future of health systems to 2030: a roadmap for global progress and sustainability", *International Journal for Quality in Health Care*, Vol. 30/10, pp. 823-831, https://doi.org/10.1093/intqhc/mzy242. [64]

Brönneke, J. et al. (Forthcoming), "Dynamic HTA for Digital Health Solutions: Opportunities and Challenges for Patient-Centered Evaluation", *International Journal of Technology Assessment in Health Care*. [76]

Burt, J. et al. (2022), *SOCIAL LICENCE FOR USES OF HEALTH DATA: A REPORT ON* [51]

PUBLIC PERSPECTIVES.

Chebrolu Kumar, R. (2020), *Smart use of artificial intelligence in health care*, Deloitte. [7]

Chen, Y. et al. (2023), "Human-Centered Design to Address Biases in Artificial Intelligence", *Journal of Medical Internet Research*, Vol. 25, p. e43251, https://doi.org/10.2196/43251. [62]

Davenport, T. and R. Kalakota (2019), "The potential for artificial intelligence in healthcare", *Future Healthcare Journal*, Vol. 6/2, pp. 94-98, https://doi.org/10.7861/futurehosp.6-2-94. [59]

Dejeu-Castang, L. (2023), *'Not OK Computer: A Proposed AI Transparency Framework for the UK'*, Counsel. [81]

Dhamanti, I. et al. (2023), "Adverse Reactions of COVID-19 Vaccines: A Scoping Review of Observational Studies", *International Journal of General Medicine*, Vol. Volume 16, pp. 609-618, https://doi.org/10.2147/ijgm.s400458. [26]

ECP (2018), *Artificial Intelligence Impact Assessment*. [69]

European Alliance for Access to Safe Medicines (2022), "Medication Errors – the Most Common Adverse Event in Hospitals Threatens Patient Safety and Causes 160,000 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). [6]

European Commission (2023), *Hiroshima Process International Code of Conduct for Advanced AI Systems*. [67]

European Council (2023), *G7 Leaders' Statement*. [32]

European Parliament (2023), *Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI*, European Parliament. [29]

European Patients Forum (2023), *ARTIFICIAL INTELLIGENCE IN HEALTHCARE Advancing Patient-Centric Care through Co-design and Responsible Implementation*. [52]

Folland, S., A. Goodman and M. Stano (2016), *The Economics of Health and Health Care*, Routledge. [27]

Friedler, S. and C. Wilson (eds.) (2018), *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, PMLR, https://proceedings.mlr.press/v81/buolamwini18a.html. [43]

Galindo, L., K. Perset and F. Sheeka (2021), "An overview of national AI strategies and policies", *OECD Going Digital Toolkit Notes*, No. 14, OECD Publishing, Paris, https://doi.org/10.1787/c05140d9-en. [28]

Geddes, L. (2023), *Cancer and heart disease vaccines 'ready by end of the decade'*, The Guardian, https://www.theguardian.com/society/2023/apr/07/cancer-and-heart-disease-vaccines-ready-by-end-of-the-decade (accessed on 2 October 2023). [2]

Geis, J. et al. (2019), "Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement", *Radiology*, Vol. 293/2, pp. 436-440, https://doi.org/10.1148/radiol.2019191586. [35]

Gintux (2023), *The Most Surprising Fax Machine Usage Statistics And Trends in 2023*, https://blog.gitnux.com/fax-machine-usage-statistics/. [4]

Government of Canada (2022), *Learning together for responsible artificial intelligence*. [72]

GPAI (2023), *The Global Partnership on Artificial Intelligence*, GPAI, https://gpai.ai/projects/responsible-ai/ (accessed on 21 December 2023). [82]

Health AI (2023), *Advancing the development and adoption of Responsible AI solutions in health*, Health AI, https://www.healthai.agency/. [79]

Health.AI (Forthcoming), *The International Digital Health and AI Research Collaborative*, https://www.i-dair.org/. [86]

Hemdan, E., W. El-Shafai and A. Sayed (2023), "Integrating Digital Twins with IoT-Based Blockchain: Concept, Architecture, Challenges, and Future Scope", *Wireless Personal Communications*, Vol. 131/3, pp. 2193-2216, https://doi.org/10.1007/s11277-023-10538-6. [15]

Hibbard, J. and J. Greene (2013), "What The Evidence Shows About Patient Activation: Better Health Outcomes And Care Experiences; Fewer Data On Costs", *Health Affairs*, Vol. 32/2, pp. 207-214, https://doi.org/10.1377/hlthaff.2012.1061. [50]

Hinton, G. (2018), "Deep Learning—A Technology With the Potential to Transform Health Care", *JAMA*, Vol. 320/11, pp. 1101-1102, https://doi.org/10.1001/jama.2018.11100. [54]

International Labour Association (2023), *Country Reports: Digital Health Literacy to Increase the Resilience of the Disadvantaged Group*. [42]

Jha, A. et al. (2013), "The global burden of unsafe medical care: analytic modelling of observational studies", *BMJ Quality & Safety*, Vol. 22/10, pp. 809-815, https://doi.org/10.1136/bmjqs-2012-001748. [63]

Khan, B. et al. (2023), "Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector.", *Biomedical materials & devices (New York, N.Y.)*, pp. 1-8, https://doi.org/10.1007/s44174-023-00063-2. [1]

Krumholz, H. (2014), "Big Data And New Knowledge In Medicine: The Thinking, Training, And Tools Needed For A Learning Health System", *Health Affairs*, Vol. 33/7, pp. 1163-1170, https://doi.org/10.1377/hlthaff.2014.0053. [48]

Lakhani, P. and B. Sundaram (2017), "Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks", *Radiology*, Vol. 284/2, pp. 574-582, https://doi.org/10.1148/radiol.2017162326. [55]

Lauenroth, V. et al. (2020), "Lessons From The Impact Of Price Regulation On The Pricing Of Anticancer Drugs In Germany", *Health Affairs*, Vol. 39/7, pp. 1185-1193, https://doi.org/10.1377/hlthaff.2019.01122. [74]

Lawson, A. (2023), *AI vs. Responsible AI: Why is it Important?*, Responsible Artificial Intelligence Institute, https://www.responsible.ai/post/ai-vs-responsible-ai-why-is-it-important (accessed on 2 October 2023). [66]

Leitfaden Für Hersteller, E. and L. Und Anwender (2023), *The Fast-Track Process for Digital Health Applications (DiGA) according to Section 139e SGB V. A Guide for Manufacturers,* [75]

Service Providers and Users.

Liu, G. et al. (2023), "Deep learning-guided discovery of an antibiotic targeting Acinetobacter baumannii", *Nature Chemical Biology*, https://doi.org/10.1038/s41589-023-01349-8. [60]

Lorenz, P., K. Perset and J. Berryhill (2023), "Initial policy considerations for generative artificial intelligence", *OECD Artificial Intelligence Papers*, No. 1, OECD Publishing, Paris, https://doi.org/10.1787/fae2d1e6-en. [22]

Mimouni, D. et al. (2010), "Trends in Dermatology Publications over the Past 15 Years", *American Journal of Clinical Dermatology*, Vol. 11/1, pp. 55-58, https://doi.org/10.2165/11530190-000000000-00000. [57]

Morris, Z., S. Wooding and J. Grant (2011), "The answer is 17 years, what is the question: understanding time lags in translational research", *Journal of the Royal Society of Medicine*, Vol. 104/12, pp. 510-520, https://doi.org/10.1258/jrsm.2011.110180. [56]

National Academy of Medicine (2023), *Toward a Code of Conduct for Artificial Intelligence Used in Health, Health Care, and Biomedical Science*, National Academy of Medicine, https://nam.edu/programs/value-science-driven-health-care/health-care-artificial-intelligence-code-of-conduct/ (accessed on 2 October 2023). [68]

National Institute for Health and Care Excellence (2018), *Evidence standards framework for digital health technologies*, http://www.nice.org.uk/corporate/ecd7. [77]

New York Tribune (1920), "Electricity in Industry", *New York Edison Company*, https://blogs.loc.gov/inside_adams/2013/07/featured-advertisement-electricity/ (accessed on 23 November 2023). [24]

OECD (2023), "Common guideposts to promote interoperability in AI risk management", *OECD Artificial Intelligence Papers*, Vol. 43/5. [83]

OECD (2023), *G7 Hiroshima Process on Generative Artificial Intelligence (AI): Towards a G7 Common Understanding on Generative AI*, OECD Publishing, Paris, https://doi.org/10.1787/bf3c0c60-en. [21]

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. [8]

OECD (2022), "OECD Framework for the Classification of AI systems", *OECD Digital Economy Papers*, No. 323, OECD Publishing, Paris, https://doi.org/10.1787/cb6d9eca-en. [78]

OECD (2022), *Recommendation of the Council on Health Data Governance*, http://legalinstruments.oecd.org. [37]

OECD (2019), *Recommendation of the Council on Artificial Intelligence*, http://legalinstruments.oecd.org. [20]

OECD AI Policy Observatory (2023), *OECD AI Principles overview*, OECD AI Policy Observatory, https://oecd.ai/en/ai-principles (accessed on 18 November 2023). [87]

OECD AI Policy Observatory (2023), *Policies, data and analysis for trustworthy artificial intelligence*, OECD AI Policy Observatory. [85]

OECD.AI expert group (2023), *OECD AI Incidents Monitor*, OECD. [46]

Organization, W., C. University and INSEAD. (2019), *The Global Innovation Index 2019 : Creating Healthy Lives - The Future of Medical Innovation.*. [12]

Paris, V. and A. Belloni (2013), "Value in Pharmaceutical Pricing", *OECD Health Working Papers*, No. 63, OECD Publishing, Paris, https://doi.org/10.1787/5k43jc9v6knx-en. [84]

PATH (2023), *The power of AI for health equity*, PATH, https://www.path.org/our-impact/articles/the-power-of-ai-for-health-equity/ (accessed on 21 December 2023). [80]

Price, M. and F. Lau (2014), "The clinical adoption meta-model: a temporal meta-model describing the clinical adoption of health information systems", *BMC Medical Informatics and Decision Making*, Vol. 14/1, https://doi.org/10.1186/1472-6947-14-43. [38]

Price, W. and I. Cohen (2019), "Privacy in the age of medical big data", *Nature Medicine*, Vol. 25/1, pp. 37-43, https://doi.org/10.1038/s41591-018-0272-7. [47]

Raj, P. (2021), "Empowering digital twins with blockchain", in *Advances in Computers, The Blockchain Technology for Secure and Smart Applications across Industry Verticals*, Elsevier, https://doi.org/10.1016/bs.adcom.2020.08.013. [14]

Schumer, C. (2023), *SAFE Innovation Framework*, https://www.democrats.senate.gov/imo/media/doc/schumer_ai_framework.pdf (accessed on 2 October 2023). [13]

Shanafelt, T. et al. (2016), "Potential Impact of Burnout on the US Physician Workforce", *Mayo Clinic Proceedings*, Vol. 91/11, pp. 1667-1668, https://doi.org/10.1016/j.mayocp.2016.08.016. [53]

Siegel-Itzkovich, J. (2023), "Patient saved by AI app at Galilee Medical Center", *The Jerusalem Post*. [19]

Song, H. and S. Veeraraghavan (2018), "Quality of Care", in *Handbook of Healthcare Analytics*, John Wiley & Sons, Ltd, https://doi.org/10.1002/9781119300977.ch5. [65]

Spicer, N. et al. (2020), "'It's far too complicated': why fragmentation persists in global health", *Globalization and Health*, Vol. 16/1, https://doi.org/10.1186/s12992-020-00592-1. [36]

Srivastava, D. et al. (2023), "Promoting the systematic use of real-world data and real-world evidence for digital health technologies across Europe: a consensus framework", *Health Economics, Policy and Law*, pp. 1-16, https://doi.org/10.1017/s1744133123000208. [3]

Srivastava, D., H. Scarbrough and C. Stavropoulou (2020), *Current regulatory challenges to support the spread of digital health technologies*, https://www.prnewswire.com/news-releases/worldwide-digital-health-market-to-hit-. [34]

The Saturday Evening Post (2021), "The Perceived Threat of Electromagnetic Radiation, Now and Then", https://www.saturdayeveningpost.com/2021/11/the-perceived-threat-of-electromagnetic-radiation-now-and-then/ (accessed on 23 November 2023). [25]

Thomas Craig, K. et al. (2021), "The burden of the digital environment: a systematic review on organization-directed workplace interventions to mitigate physician burnout", *Journal of the American Medical Informatics Association*, Vol. 28/5, pp. 985-997, https://doi.org/10.1093/jamia/ocaa301. [45]

Thomason, J. (2021), "Big tech, big data and the new world of digital health", *Global Health* [11]

Journal, Vol. 5/4, pp. 165-168, https://doi.org/10.1016/j.glohj.2021.11.003.

Topol, E. (2019), "High-performance medicine: the convergence of human and artificial intelligence", *Nature Medicine*, Vol. 25/1, pp. 44-56, https://doi.org/10.1038/s41591-018-0300-7. [49]

Tyson, A. et al. (2023), *60% of Americans Would Be Uncomfortable With Provider Relying on AI in Their Own Health Care*, http://www.pewresearch.org. [41]

U.S. Food & Drug Administration (2022), *Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices*, U.S. Food & Drug Administration, https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices (accessed on 2 October 2023). [17]

U.S. Food and Drug Administration (FDA), Health Canada and Medicines and Healthcare products Regulatory Agency (MHRA) (2021), *Good Machine Learning Practice for Medical Device Development: Guiding Principles*, FDA. [40]

University of Helsinki (2021), *Elements of AI online course now available in 22 EU countries and languages*, https://www.helsinki.fi/en/news/teaching/elements-ai-online-course-now-available-22-eu-countries-and-languages (accessed on 2 October 2023). [73]

US White House (2022), *BLUEPRINT FOR AN AI BILL OF RIGHTS*, https://www.whitehouse.gov/ostp/ai-bill-of-rights. [30]

Vamathevan, J. et al. (2019), "Applications of machine learning in drug discovery and development", *Nature Reviews Drug Discovery*, Vol. 18/6, pp. 463-477, https://doi.org/10.1038/s41573-019-0024-5. [61]

Vender, R. and C. Lynde (2023), "AI-Powered Patient Identification to Optimize Care", https://pentavere.ai/wp-content/uploads/2023/03/CDA-Poster_AI-Powered-Patient-Identification-to-Optimize-Care.pdf (accessed on 20 December 2023). [9]

Weins, K. (2020), *IT Spending by Industry*, Flexera, https://www.flexera.com/blog/technology-value-optimization/it-spending-by-industry/ (accessed on 18 November 2023). [70]

Yu, V. (1979), "Antimicrobial Selection by a Computer", *JAMA*, Vol. 242/12, p. 1279, https://doi.org/10.1001/jama.1979.03300120033020. [16]