EU Institutions and Competent Authorities Dealing with the Revolution of Artificial Intelligence for Health

Fidelia Cascini

Abstract The rapid evolution of Artificial Intelligence is transforming healthcare, presenting both significant opportunities and considerable challenges. The European Union has adopted a risk-based regulatory approach with the AI Act (Regulation (EU) 2024/1689) to balance innovation with fundamental rights protection. This law classifies AI systems into four risk levels where high-risk systems, such as those used in medical diagnostics and robotic surgery, are subject to stringent compliance and ex-ante assessments. Despite AI promises improvements in several areas of Medicine (e. g. drug discovery, medical imaging, early diagnosis, disease prevention) potentially enhancing efficiency and reducing medical errors, current critical concerns include automation bias, the lack of long-term prediction accuracy, potential inequalities in technology access, and algorithmic opacity. For these reasons, AI systems can also exacerbate existing forms of human vulnerability and introduce new ones through the interaction between inferences, predictions, and AI-generated content. Competent authorities play a crucial role in ensuring that AI systems for health are governed and implemented effectively, and are prepared with adequate resources to meet market demand for testing and evaluation. Controlled environments (i.e. regulatory sandboxes) facilitate the testing and verification of AI systems through collaboration among developers, healthcare providers, regulators, and patients. These ensure that new technologies are safe, effective, and ethical before widespread clinical adoption, mitigating risks and protecting human vulnerability. The role of Competent Authorities is crucial to guarantee the protection of fundamental rights, health, and safety.

Keywords Artificial intelligence · Regulatory sandboxes · Competent authorities · Healthcare · AI Act

F. Cascini (✉)
University S. Raffaele, Rome, Italy
e-mail: fidelia.cascini@uniroma5.it

© The Author(s) 2025

C. Amodio and A. Diurni (eds.), *Human Vulnerability in Interaction with AI in European Private Law*, https://doi.org/10.1007/978-3-031-93247-2_13

1 The EU Approach Towards Artificial Intelligence for Health

Human vulnerability, while not a consolidated legal concept, is embraced in recent EU regulations and political interventions mainly to protect vulnerable groups or minorities. This means a group-based understanding of vulnerability, i.e. by prohibiting AI systems that aim to exploit individual vulnerabilities due to their age, disability, social or economic conditions. From a broader perspective, AI systems bring the concrete risk to negatively impact on natural person's rights such as those related to privacy protection, freedom of expression and information, health. Consequently, AI systems can exacerbate existing forms of human vulnerability and introduce new ones through the interaction between inferences, predictions, and AI-generated content. This scenario brings additional challenges in the healthcare sector, where the management of potential harm related to AI is characterized by unconsolidated attempts of governance, despite the EU's effort to be the first in the world regulating AI using a risk-based approach and considering vulnerabilities.

On 1 August 2024, Regulation (EU) 2024/1689 (AI Act)[1] came into force, establishing harmonized rules on the use of artificial intelligence systems in the EU territory to protect European citizens from potential negative consequences. The legislator considered it relevant to explore artificial intelligence-based innovations introduced globally, preparing to address their potential risks. Promoting the development of artificial intelligence by strengthening investments and, at the same time, protecting the fundamental rights of individuals and businesses by developing new rules was the strategic approach followed by the European Commission. Regulatory action appeared essential in high-impact sectors, including the environment, the public sector, finance, mobility, home affairs, agriculture and, in our case, healthcare. Through the AI Act, the European Commission proposed a regulatory framework aimed at ensuring that the artificial intelligence systems used are safe, transparent, ethical, impartial and under human control. The scope of the admissibility of artificial intelligence practices has been defined based on four levels of risk: unacceptable, high, limited, and minimal.

Unacceptable risk is that linked to the practice of artificial intelligence for purposes and/or in ways that are contrary to the values of the European Union, for example, because they violate the fundamental rights of natural persons. The prohibitions therefore concern practices that have high potential in terms of manipulating people through subliminal techniques without their awareness or exploiting the vulnerabilities of specific vulnerable groups, such as minors or people with disabilities, to materially distort their behavior to cause psychological or physical harm. The attribution of a social score based on artificial intelligence for general purposes by public

[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance). https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

authorities is also prohibited, as is the use of 'real-time' remote biometric identification systems in spaces accessible to the public for law enforcement purposes, without prejudice to the application of certain limited exceptions.

High-risk artificial intelligence systems are permitted in the European market and are subject to compliance with certain mandatory requirements and an ex-ante conformity assessment. The classification of a system as high risk is based on its intended purpose in line with the current EU legislation on product safety, which, with respect to medical devices, refers, for example, to the Medical Device Regulation (EU Regulation 2017/745)[2] and Regulation (EU) 2017/746[3]on in vitro diagnostic medical devices. The classification of a product as high risk depends not only on the function performed by the system but also on the specific purposes and ways in which the system is used. In the medical field, for example, artificial intelligence-based medical diagnostic devices and applications in robotic surgery are included, since they represent safety components of products (or even products themselves), which must be carefully assessed by a third-party body that expresses its opinion on the conformity of the product component within the scope of application of a given Union harmonization regulation, both before it is placed on the market and throughout its life cycle.

The extent of the negative impact of artificial intelligence systems on the rights protected by the Charter of Fundamental Rights of the European Union is relevant for classifying a system as high risk. These rights include the right to human dignity, respect for private and family life, and the protection of personal data (Official Journal 2000/C 364/01)[4] In line with the objectives of EU legislation to facilitate the free movement of products in the internal market by ensuring that only safe and compliant products can be placed on the market, it is important that risks that a product may generate because of its digital components (including artificial intelligence) are duly prevented and mitigated. For example, increasingly autonomous robots, both in the context of caring for and assisting people, should be able to operate and perform their functions safely in complex environments. Similarly, increasingly sophisticated artificial intelligence-based systems for diagnostics and for supporting human decisions in healthcare should be reliable and accurate.

Regarding digital components, including artificial intelligent systems and their relationships with medical devices, the Court of Justice of the EU pronounced an interesting decision in 2017[5] as the first instance ruling on standalone software

[2] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng.

[3] Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU. https://eur-lex.europa.eu/eli/reg/2017/746/oj/eng.

[4] Charter of Fundamental Rights of the European Union (2000/C 364/01). https://www.europarl.europa.eu/charter/pdf/text_en.pdf.

[5] Case C-329/16, Syndicat national de l'industrie des technologies médicales (Snitem) and Philips France v Premier ministre and Ministre des Affaires sociales et de la Santé, Judgment of the

within the medical device context. The French Conseil d'État asked whether prescription support software enabling healthcare practitioners to obtain patients' relevant information (e.g., drug interactions, dosage limits) may constitute a medical device according to Article 1(2)(a) of the Medical Device Directive.[6] The Court, agreeing with Advocate-General's Opinion, held that this software can be a medical device notwithstanding the fact that the software does not itself act in or on the human body. A key factor considered by the Court in reaching its decision was that the software in question (IntelliSpace Critical Care Anaesthesia) facilitated the calculation of medication dosage and the proper design of the treatment by providing relevant information to healthcare professionals, therefore suggesting a "medical purpose", i.e., assisting in the prevention, monitoring, treatment or alleviation of the disease. Indeed, the intended purpose of software, alongside the action it performs on data, is central in determining whether the software might be qualified as a medical device.

The AI Act is supported by the work being done on the Data Governance Act, the Open Data Directive and the initiatives under the European strategy of data. In the European strategy of data[7] focused on the relevant deployment of data infrastructures, tools and computing for the European Health Data Space, which will determine the availability of high-quality data essential for training and further developing artificial intelligence systems. This specific space facilitates non-discriminatory access to health data and the training of artificial intelligence algorithms on those datasets in a privacy-preserving, secure, timely, transparent and trustworthy manner supported by institutional governance. In fact, through this regulatory framework, the European Commission is seeking to ensure common normative standards for all high-risk artificial intelligence systems. The health sector's involvement in high-stakes situations, which require the use of sophisticated diagnostic systems and systems supporting human decisions, needs to be reliable and accurate. Building on this aspect is the important element of "human oversight" outlined in Article 14 of the Act, where human oversight will play an important role in preventing or minimizing the risks to health that may emerge because of the implementation of a high-risk system.

For the development of high-risk artificial intelligence systems such as those for health purposes, the EU considers it appropriate to grant entities responsible for digital innovation (such as suppliers, notified bodies and other relevant entities), as well as testing and experimentation facilities and researchers involved in such trials, to access and use high-quality datasets. Common European data spaces available in the EU and facilitating data sharing between companies and with governments and institutions in the public interest are therefore essential to provide reliable, responsible and non-discriminatory access to high-quality data for the training, validation and testing of artificial intelligent systems. The European Health Data Space, for example, will facilitate access to health data and the training of algorithms on

Court (Fourth Chamber) of 7 December 2017. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:62016CJ0329.

[6] Council Directive 93/42/EEC of 14 June 1993 concerning medical devices. https://eur-lex.europa.eu/eli/dir/1993/42/oj/eng.

[7] European Commission (2020a).

such datasets in a secure, timely, transparent, trustworthy, and privacy-preserving manner, as well as with adequate institutional governance. In fact, personal data legally collected for other purposes may be processed for the purposes of developing innovative artificial intelligence systems and carrying out trial tests of these algorithms in the experimental phase.

Such systems should be developed to safeguard the public interest, such as for public health purposes, including disease prevention and control. As part of the EU Digital Strategy,[8] the creation of testing and experimentation facilities (TEFs)[9] for the verification and experimentation of artificial intelligence-based technologies and systems started in the EU, providing a common EU platform for the design and manufacturing of artificial intelligence components and systems and reference sites for applications in sectors such as healthcare. TEFs are test beds that have emerged as a remarkably interesting option for testing artificial intelligence in real conditions, implying strong collaboration between startups or companies creating the systems and healthcare facilities.

According to EU legislation, once the artificial intelligence system is on the market, authorities are responsible for market surveillance, and users ensure surveillance and monitoring related to the human use of tools, while providers have a postmarket monitoring system. Providers and users also report serious incidents and malfunctions. Low-risk systems such as chatbots are subject to minimum transparency requirements designed to allow people to make informed decisions (e.g., the user of the chatbot can decide to continue or stop using the application). The vast majority of artificial intelligence systems fall into the low-risk category, except those for health applications, which are not covered by the new rules because they pose only minimal or no risk to the rights or safety of citizens. Artificial intelligence technologies are designed by companies or through public–private partnerships. To strengthen

[8] The EU Digital Strategy (https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age_en) is part of the European Commission political priorities form several years. TheEU's Digital Single Market strategy, launched in 2015, paved the way for closer digital harmonization between the EU Member States with the aim to contribute to economic growth, boosting jobs, competition, investment and innovation in the EU. In March 2021 the Commission proposed a path to the Digital Decade. This policy programme is guided by the 2030 Digital Compass—a plan to achieve digital transformation of the EU economy and society. The Digital Compass (https://eufordigital.eu/discover-eu/eu-digital-strategy/) aims towards a secure human-centred digital ecosystem, where citizens are empowered, and businesses prosper from the digital potential. The Compass indicates four cardinal points for this trajectory: digital skills, secure and performant digital infrastructure, digital transformation of businesses and the digitalization of public services. This EU political agenda aligns with EU norms and standards to strengthen the EU's digital sovereignty. A number of budgetary instruments will support the investments necessary to build Europe's Digital Decade on solid foundations.

[9] TEFs are a combination of physical and virtual facilities, in which technology providers can get primarily technical support to test their latest AI-based software and hardware technologies (including AI-powered robotics) in real-world environments. Thus, they are specialized large-scale reference sites open to all technology providers across Europe to test and experiment at scale state-of-the art AI solutions, including both soft-and hardware products and services, e.g. robots, in real-world environments. More information are available at: https://digital-strategy.ec.europa.eu/en/faqs/testing-and-experimentation-facilities-tefs-questions-and-answers.

EU competitiveness through these partnerships[10] and to engage various stakeholders and investors in the artificial intelligence development market in the EU, the European Commission has invested much money in recent years, with an overall 1.1 billion € spent under the Horizon 2020 research and innovation program, including healthcare. Some of the world's largest technology companies are developing new applications and services, which they either own or invest in.

2 Artificial Intelligence for Healthcare Purposes[11]

Artificial intelligence-based algorithms are transforming medicine, medical research, and public health. The United Nations Secretary-General has stated that safe deployment of new technologies, including artificial intelligence, can help the world achieve health-related objectives under Sustainable Development Goal #3. Artificial intelligence could also be the key to improving global commitments to achieving universal health coverage.[12] On the other hand, a coordinated strategic and regulatory approach from the EU addressing opportunities and challenges of artificial intelligence tools (see paragraph 1) is creating an environment for growth without leaving single Member States lagging. Since 2020, the European Commission has described the characteristics of the policy framework necessary to develop trustworthy and secure artificial intelligence applications for all areas, including healthcare, in the white paper on artificial intelligence.[13] EU willingness is to become an ecosystem of excellence in research and innovation, accelerating the adoption of artificial intelligence-based solutions with the involvement of SMEs and, at the same time, an ecosystem of trust that is compliant with EU rules and regulations, including those protecting fundamental rights and consumer rights, particularly regarding high-risk artificial intelligence systems.

The use of artificial intelligence to improve healthcare systems holds a promising future, with progress already being made in various health-related fields, such as drug discovery, medical imaging, screening, diagnosis, and prevention of diseases. Artificial intelligence could help healthcare providers increase their performance and avoid medical errors. It can allow clinicians to be more focused on providing care to patients and solving complex cases while saving time while achieving better clinical outcomes. One of the main focuses is creating a support system to improve the early diagnosis of various diseases. This is particularly explored in the field of oncology, where artificial intelligence is being evaluated for use in radiological diagnoses, such as in whole-body imaging, colonoscopies, and mammograms. Artificial

[10] Public–Private Partnership on AI, Data and Robotics. https://stip.oecd.org/stip/policy-initiatives/2019%2Fdata%2FpolicyInitiatives%2F26437.

[11] This paragraph is an extract of Cascini et al. (2023).

[12] Dept. of Economic and Social Affairs Sustainable Development, Developmental Goal 3. https://sdgs.un.org/goals/goal3.

[13] European Commission (2020b, 2021).

intelligence can also aid in optimizing radiological treatment dosing, recognizing malignant disease in dermatology or clinical pathology, and guiding RNA and DNA sequencing for immunotherapy.[14] In general, artificial intelligence developments in early diagnosis are being studied in most health-related fields, such as in the early diagnosis of diabetic retinopathy, cardiovascular disease, liver disease and neurological disorders[15] Currently, there are only a handful of prospective clinical trials on the effectiveness of artificial intelligence in early diagnosis, with some showing promise of equivalent detection ability to that of human professionals in specific tasks, with even fewer focusing on the potential benefits of human–machine partnerships. One of the risks of relying excessively on artificial intelligence is the development of automation bias, where medical practitioners might not consider other important aspects of patient care and may overlook errors that should have been spotted by human-guided decision making.

Artificial intelligence can also be used to digitalize and store traditional paper medical records and process large amounts of data from images and other types of inputs or signals (such as motion data or sound data). Steps in image and signal processing algorithms typically include signal feature analysis and data classification via tools such as artificial neural networks, which work via complex layers of decision nodes[16] Medical imaging is one of the most rapidly developing areas of artificial intelligence application in healthcare. While improving automated image interpretation and analysis are a priority, other important aspects of AI application to medical imaging are being explored, such as data security and user privacy solutions for medical image analysis, deep learning algorithms for the restoration/reconstruction and segmentation of complex images and the creation of fuzzy sets or rough sets in medical image analysis.[17]

Furthermore, with health systems, in general, becoming increasingly complex every year, the administration and management of care are becoming increasingly laborious. Artificial intelligence can be used to assist personnel in complex logistical tasks, such as optimization of the medical supply chain, to assume mundane, repetitive tasks or to support complex decision-making[18] This is made possible by a combination of artificial intelligence advancements in the fields of natural language processing, automated scheduling and planning, and expert systems. Many artificial intelligence tools can also be used in specific public health programs or in wide public health approaches to improve wellbeing. Artificial intelligence can be used for health promotion or to identify target populations or locations with "high-risk" behaviour concerning communicable and noncommunicable diseases. Artificial intelligence can improve the effectiveness of communication and messaging specifically directed at certain subpopulations, both in terms of its ability to recognize priority groups and

[14] Bi et al. (2019).

[15] Kamdar et al. (2020).

[16] Wahl et al. (2018).

[17] Artificial Intelligence for Medical Image Analysis, Special Issue. https://www.hindawi.com/journals/bmri/si/617603/.

[18] Schwalbe and Wahl (2020).

its adaptiveness in creating tailored messages to benefit population health (microtargeting).[19] One example of such an application is microtargeting individuals or communities with technological, linguistic, or cultural barriers to better communicate the importance and safety of vaccinations, such as the COVID-19 vaccination[20] AI tools could therefore be adapted to improve access and equity of care, furthering the development of truly personalized medicine. AI can also play a leading role in performing analyses of patterns of data for health surveillance and disease detection[21] AI tools can be used to identify bacterial contamination in water treatment plants, identify foodborne illnesses in restaurants or hospitals, simplify detection and lower costs. Sensors can also be used to improve environmental health, such as by analysing air pollution patterns or using machine learning to make inferences between the physical environment and healthy behaviour.[22]

Another application of AI in public health surveillance is evidence collection and its use to create mathematical models to make decisions. Although many public health institutions are not yet making full use of all possible sources of data, some fields, such as real-time health surveillance, are steadily improving. This has improved the public health outlook on pandemic preparedness and response, although the long-term ramification of such important changes will only be evident in the future[23] The development of public health policy also proves to be fertile ground for artificial intelligence, where attempts at analysing argumentation on food quality in a public health policy were attempted. This action resulted in the creation of models that output new recommendations on the basis of stakeholders' arguments by targeting specific audiences[24] Healthcare has always depended in part on predictions, prognoses, and the use of predictive analytics. AI is just one of the more recent tools for this purpose, and many possible benefits of prediction-based health care rely on the use of this technology. For example, AI can be used to assess an individual's risk of disease, which could be used for the prevention of diseases and major health events. Various studies suggest that artificial intelligence may improve several pathologies, such as heart failure, by utilizing predictive models and telemonitoring systems for clinical support and patient empowerment. For example, given the expected increase in the number of heart failure patients in the future due to the aging of the population, predicting the risk of a patient having heart failure could prevent hospitalizations and readmissions, improving both patient care and hospital management, which would have a high impact on costs and time.[25]

Machine learning is also increasingly being applied to make predictions related to population health: the use of novel big data resources, which are ripe with different data types, may allow for improvements in the prediction algorithms necessary to

[19] Privacy International (2021).

[20] Harris (2021).

[21] Morgenstern et al. (2021).

[22] Roski et al. (2019).

[23] Whitelaw et al. (2020).

[24] Bourguet et al. (2013).

[25] Larburu et al. (2018).

navigate complex health data ecosystems successfully. A good example of this is the integration of data types to better understand complex associations between genetics, the environment, and disease. The Harvard group has used large administrative datasets to untangle the relationship between genetics and the environment in all diseases recorded in health insurance claims data[26] Use of biobanks and their massive datasets allows scientists around the world to discover new genetic variants (e.g., through genome-wide association studies) and novel risk factors associated with disease more efficiently and with higher sensitivity and specificity compared to traditional "one-at-a-time" methods[27] Using electronic medical record data, machine- and deep-learning algorithms have been able to predict many important clinical parameters, including suicide, Alzheimer's disease, dementia, severe sepsis, septic shock, hospital readmission, all-cause mortality, in-hospital mortality, unplanned readmission, prolonged length of stay and final discharge diagnosis.[28]

Overall, predictive models have been used much more widely by clinicians than by public health professionals. However, on closer inspection, any application that improves patient care at any level can be considered relevant to the field of public health. The ability of clinicians and healthcare providers to make better informed decisions on patient health can be improved by context-specific algorithms, which use massive quantities of clinical, physiological, epidemiological, and genetic data. Precision medicine will further benefit from these advanced algorithms, as their accuracy, timeliness and appropriateness in clinical care improve over time, decompressing our reliance on human resources. This advancement, however, still necessitates computer-literate physicians, who are up-to-date with new generation data-driven approaches. The key to the complete incorporation of AI into clinical care will therefore be the integration of human clinical judgement with advanced clinical machine learning algorithms.[29]

To maximize these benefits, however, legal, ethical, economic, and social constraints of artificial intelligence must be addressed first. The potential benefits of these technologies and the economic and commercial potential of artificial intelligence for healthcare warrant ever-greater use worldwide[30] Different organizations throughout the world are building on this important discussion. The Pan American Health Organization, for example, offers its own guiding principles for all applications in healthcare or public health[31]:

1. People-centred: artificial intelligence technologies must respect the rights of the individual.
2. Ethically grounded, progress and discussion must be made in compliance with the principles of human dignity, nonmaleficence, and justice.

[26] Lakhani et al. (2019).

[27] Centers for Disease Control and Prevention (2019).

[28] Topol (2019).

[29] Khemasuwan et al. (2020).

[30] World Health Organization (2021).

[31] Pan American Health Organization (2021).

3. Transparent: When developing artificial intelligence algorithms, having clear objectives and goals is mandatory.
4. Data protection: Data privacy and security are paramount in artificial intelligence development.
5. Scientific integrity: artificial intelligence applications must be reliable, reproducible, fair, honest, and accountable, according to best practices.
6. Open and shareable: Openness and shareability must be the founding principles of every artificial intelligence development process.
7. Nondiscriminatory: artificial intelligence for public health must be based on fairness, equality, and inclusiveness.
8. Human-controlled: All automated decisions must be reviewed by human beings.

These guiding principles are fundamental in developing global cooperation initiatives centred on artificial intelligence use for health while navigating the different regulatory and policy environments worldwide.

3 Critical Aspects of Artificial Intelligence for Healthcare Provision

The application of artificial intelligence for health is very promising. Nevertheless, the current scientific landscape still lacks clinical trials highlighting the impact of artificial intelligence tools in delivering healthcare (in terms of safety, effectiveness, patient outcomes) compared with traditional approaches and their sustainability in terms of economic and ethical aspects. The development of evidence-based guidelines and the implementation of best practices are fundamental to promote the use of artificial intelligence in routine clinical practice, fostering a significant positive transformation in the health sector. Furthermore, artificial intelligence currently poses major technological, ethical, and social challenges, which need competent professionals to address. Critical issues slow the adoption of innovative artificial intelligence-based applications. For example, regarding public health interventions, it is still unknown whether the introduction of artificial intelligence in clinical practice might either improve or worsen health inequities on a large scale.[32] Although the promotion of artificial intelligence addresses personalized recommendations and individual action, this should not threaten the importance of collective action to take care of social and structural determinants of health[33] Moreover, similar to all public health interventions, the use of artificial intelligence to create enduring benefits requires not only a broad coalition of support and partnerships between the public and private sectors but also the trust and enduring support of patients.

In addition, despite their wide testing, artificial intelligence-based algorithms affecting patient care have not yet reached the sufficient level of accuracy needed for

[32] Weiss et al. (2018).

[33] Panch et al. (2019).

precise long-term predictions. This poses a serious challenge for healthcare workers, as long-term predictions of limited reliability could impact an individual's life for years in the future (e.g., both false-positive/negative predictions on an essential diagnosis could affect the level of risk that clinicians are willing to undertake to treat a disease, thus heavily impacting patient health outcomes). Prediction algorithms could also be biased towards or against certain population subgroups (e.g., ethnic groups, religious groups), both in terms of potentially discriminatory health practices suggested and issues related to individual autonomy in personal data use and informed consent. These potential pitfalls of artificial intelligence and its long-term health implications raise essential ethical concerns that must be addressed by all the stakeholders involved (i.e., governmental institutions, healthcare authorities, medical and patient associations, healthcare providers and industries).[34] From a public health perspective, one of the main and most obvious implications of artificial intelligence use for healthcare is the risk of inequalities in accessing technologies and in the financial burdens that they can generate.[35] This gap could emerge among patients population subgroups and between researchers and clinicians operating in different contexts or areas and between the public and private sectors as well as in the health systems of different regions and countries.[36]

As another issue, artificial intelligence systems might be programmed according to certain values and judgments that could create or worsen health inequities, such as those related to gender, race or belonging to minority groups. Moreover, there could be disagreement about the system of values that informs artificial intelligence systems.[37] At the research stage, proper development of a precautionary supervised system of artificial intelligence trials would ensure that they are ethical, legal, and inclusive in scope. The greatest challenge related to artificial intelligence tools for health remains their adoption in everyday clinical practice following the consensus of the scientific community in different medical areas. In fact, for widespread adoption to take place, artificial intelligence systems must be approved by scientists and regulators, integrated with digital health platforms through health data pipelines, standardized at a level such that similar products perform similarly, taught to clinicians, accepted by patients, paid for by public or private payer organizations, and updated over time. The dysregulated use of artificial intelligence can, in fact, negatively affect patients, professional and health systems because of some characteristics of this family of technologies, possibly generating critical issues regarding the absence of transparency and comprehensibility of their way of functioning, the generation of biases related to the programming and/or training of machine learning systems, the lack of robust data protection and cybersecurity requirements, the difficult definition of responsibility and accountability concerning their use and the reliability of their performance. Biased algorithms are, for example, those trained mainly on healthy men but are also used in clinical practice with women and children. If errors in

[34] World Health Organization (2019).

[35] IEEE (2017).

[36] Smith et al. (2020).

[37] Caliskan et al. (2017).

an algorithm used for the diagnosis of serious and potentially fatal diseases such as cancer are undetectable by the physician, the physician can make incorrect diagnoses with serious consequences for the patient's outcome. Furthermore, the standards of care in the artificial intelligence-based decision support system medicine era and who will assume risks and liability for medical malpractice caused by errors in AI systems are still unknown. Common issues related to the uncontrolled use of artificial intelligence tools include transparency and explainability,[38] bias and discrimination,[39] responsibility and accountability,[40] data privacy and security,[41] reliability and performance.[42]

4 Regulatory Sandboxes in Healthcare Are Needed to Solve Some Critical Problems

Article 57 (1) of the EU AI Act requires that Member States ensure that their competent authorities establish at least one AI regulatory sandbox at the national level to facilitate the development and testing of innovative AI systems under strict regulatory oversight before these systems are placed on the market. Regulatory sandboxes can be considered controlled environments to foster innovation through the cooperation of multiple stakeholders (producers and innovators, providers, end-users, and authorities). The sandbox participants collaborate to design and drive innovation without the associated risks. Competent authorities provide, as appropriate, guidance, supervision and support with a view to identifying risks, particularly fundamental rights, health and safety.

Stakeholders' ability to collaborate effectively is one of the key measures identified for the success of sandboxes in general. This collaboration can be assisted via directed consortia, a partnership framework that provides collaborative environments

[38] The lack of transparency and explainability means that the healthcare professional may not understand the machine's decision-making process, with impact on trust (risk of both excessive distrust and unjustified trust), impacting on patient's outcome.

[39] The artificial intelligence system can perpetuate or amplify existing biases in the health dataset used for its design and training. This can lead to wrong decisions of the professional in providing the right healthcare services or treatments to the right patients.

[40] The question of responsibility is crucial: it is not clear who should take responsibility for the wrong decisions made by artificial intelligence systems. This raises ethical and legal questions, especially when such decisions have significant impacts on people's health and lives.

[41] The use of artificial intelligence tools in the healthcare sector involves the collection and analysis of large amounts of personal health data, raising concerns about privacy. In all the cases where rigorous measures to protect patient's data and ensure their ethical use are not implemented, there is a concrete risk of violations of natural persons rights.

[42] When artificial intelligence systems are not properly tested and evaluated through clinical trials, the risk they generate confusing or inaccurate results in clinical practice is high. Additionally, the reliability of artificial intelligence systems for healthcare applications requires constant monitoring of their performance allowing professionals to be aware about their limitations.

for multiple organisations with differing priorities. The use of clear mission statements, targeted deliverables and timeframes, workstreams defined by the participants involved, and expected milestones are key success factors.[43] In the healthcare sector, the engagement of clinicians and patients and public and private healthcare providers working closely with regulators, technology developers, and academics is essential in developing artificial intelligence products for health and ensuring the uptake and diffusion of innovation within the healthcare system.

The sandboxes allow innovators to trial products, services, and business models in a safe space to confirm their compliance with existing regulations before wide implementation in the sector. Within these spaces, it is possible to test new technologies in real conditions without immediately incurring all the normal regulatory consequences.[44] Artificial intelligence-based technologies and advanced therapy medicinal products can be, for instance, valuably assessed in these environments before their adoption and implementation for health. In fact, regulatory sandboxes enable the testing of late-stage products, the verification of compliance with existing regulations, and the exploration of models that may violate current regulations while providing substantial benefits if implemented. They can be seen as *ex ante* instruments to certify the safety and quality of artificial intelligence-based products and services in an efficient manner: testing an artificial intelligence product in a regulatory sandbox helps an agency or authority decide to certify its quality or safety.[45] Using an anticipatory sandbox approach such as facilitating the development of appraisal methods, policies, and processes and adapting to the unique features of disruptive innovations would enable all stakeholders involved to have a constructive discussion that would facilitate faster uptake and patient access to these technologies.

Healthcare regulatory sandboxes have been classified as outcome-focused sandboxes or data-focused sandboxes according to their aims and outputs.[46] The outcome-focused methods are designed to improve clinical outcomes through the adoption of new technologies such as digital triage tools, the use of machine learning applications within diagnostics, interventions related to telemedicine and mobile health, and the use of blockchain technology. Other uses include testing service models such as those focused on restructuring social care services and determining the appropriate staff skill mix to support care home services. The data-focused sandboxes are designed to use electronic health data to facilitate the testing of new technologies. The access to and use of health data, such as electronic health records, is thus facilitated for innovative product testing.

There are consistent benefits related to regulatory sandboxes that concern the producer of new technologies, the user (as in the healthcare sector patients, clinicians, providers), and the policymaker. The benefits for artificial intelligence producers

[43] Lim (2014).

[44] The sandbox approach is an example of anticipatory regulation where regulation is seen as a support tool rather than a barrier to innovation. Anticipatory regulation is designed to anticipate future challenges, and support innovators when faced with them.

[45] Buocz et al. (2023).

[46] Leckenby et al. (2021).

include less time to market (shorter authorization process), less regulatory uncertainty (e.g., the prohibition of new technologies), and more feedback on regulatory requirements and risks. With respect to the user, the most relevant benefit is the greater ability to choose new safer AI-based products and services for health. For policy makers, benefits include the opportunity to be informed through learning and experimentation, being supported to update regulations, and being connected and engaged with market stakeholders. Despite all these benefits, stakeholders' engagement can be characterized by some issues related to cultural and organizational resistance that can limit the adoption of innovations. On the other hand, competent authorities are asked to be prepared (with adequate implementation resources) to meet the increasing market demand regarding artificial intelligence system testing and evaluation through regulatory sandboxes, especially in the healthcare sector, and to meet the need for the integration of the EU and national regulations involving artificial intelligence for health and health data (governance, use, protection, security) used to train artificial intelligence algorithms. Standardized approaches and methods to validate artificial intelligence algorithms before marketing should also be prepared, taking into account scientific and legal principles concerning biomedical research for clinical practice.

References

Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A et al (2019) Artificial intelligence in cancer imaging: clinical challenges and applications. CA Cancer J Clin 69(2):127–157

Bourguet JR, Thomopoulos R, Mugnier ML, Abécassis J (2013) An artificial intelligence-based approach to deal with argumentation applied to food quality in a public health policy. Expert Syst Appl 40(11):4539–4546

Buocz T, Pfotenhauer S, Eisenberger I (2023) Regulatory sandboxes in the AI Act: reconciling innovation and safety? Law Innov Technol 15(2):357–389

Caliskan A, Bryson JJ, Narayanan A (2017) Semantics derived automatically from language corpora contain human-like biases. Science 356:183–186

Cascini F, Buttigieg S, Pastorino R, Ricciardi W, Boccia S (2023) Personalized medicine through artificial intelligence: a public health perspective. In: Cesario A, D'Oria M, Auffray C, Scambia G (eds) Personalized medicine meets artificial intelligence. Springer, Cham. https://doi.org/10.1007/978-3-031-32614-1_1

Centers for Disease Control and Prevention (2019) Data science and machine learning in public health: promises and challenges. https://blogs.cdc.gov/genomics/2019/09/20/data-science/

European Commission (2020a) A European strategy for data. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN

European Commission (2020b) White Paper on Artificial Intelligence—a European approach to excellence and trust. https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

European Commission (2021) Regulatory framework on AI. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Harris B (2021) As vaccine demand slows, doctors shift to 'microtargeting' vulnerable communities. NBC News, 1 May 2021

IEEE (2017) Ethically aligned design, version 2: ethics in action for the development of autonomous and intelligent systems. https://standards.ieee.org/news/2017/ead_v2.html
Kamdar JH, Jeba Praba J, Georrge JJ (2020) Artificial intelligence in medical diagnosis: methods, algorithms and applications. In: Jain V, Chatterjee J (eds) Machine learning with health care perspective. Learning and Analytics in Intelligent Systems, vol 13. Springer, Cham
Khemasuwan D, Sorensen JS, Colt HG (2020) Artificial intelligence in pulmonary medicine: computer vision, predictive model and COVID-19. Eur Respir Rev 29(157):200181
Lakhani CM, Tierney BT, Manrai AK, Yang J, Visscher PM, Patel CJ (2019) Repurposing large health insurance claims data to estimate genetic and environmental contributions in 560 phenotypes. Nat Genet 51(2):327–334
Larburu N, Artetxe A, Escolar V, Lozano A, Kerexeta J (2018) Artificial intelligence to prevent mobile heart failure patients decompensation in real time: monitoring-based predictive model. Mob Inf Syst, Article ID 1546210
Leckenby E, Dawoud D, Bouvy J, Jónsson P (2021) The sandbox approach and its potential for use in health technology assessment: a literature review. Appl Health Econ Health Policy 19(6):857–869
Lim MD (2014) Consortium sandbox: building and sharing resources. Sci Transl Med 6:242cm6
Morgenstern JD, Rosella LC, Daley MJ, Goel V, Schünemann HJ, Piggott T (2021) "AI's gonna have an impact on everything in society, so it has to have an impact on public health": a fundamental qualitative descriptive study of the implications of artificial intelligence for public health. BMC Public Health 21
Pan American Health Organization (2021) Artificial intelligence in public health. Digital transformation knowledge capsules. https://www.paho.org/en/documents/digital-transformation-knowledge-capsules-artificial-intelligence-public-health
Panch T, Pearson-Stuttard J, Greaves F, Atun R (2019) Artificial intelligence: opportunities and risks for public health. Lancet Digit Health 1:e13–e14
Privacy International (2021) Microtargeting. https://privacyinternational.org/learn/microtargeting
Roski J, Chapman W, Heffner J, Trivedi R, Del Fiol G, Kukafka R et al (2019) How artificial intelligence is changing health and health care. In: Matheny M, Thadaney Israni S, Ahmed M, Whicher D (eds) Artificial intelligence in health care: hope, hype, promise, and peril. National Academy of Medicine, Washington DC. https://nam.edu/artificial-intelligence-special-publication/
Schwalbe N, Wahl B (2020) Artificial intelligence and the future of global health. Lancet 395:1579–1586
Smith MJ, Axler R, Bean S, Rudzicz F, Shaw J (2020) Four equity considerations for the use of artificial intelligence in public health. Bull World Health Organ 98:290–292
Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. Nat Med 25:44–56
Wahl B, Cossy-Gantner A, Germann S et al (2018) Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob Health 3:e000798
Weiss D, Rydland HT, Øversveen E, Jensen MR, Solhaug S, Krokstad S (2018) Innovative technologies and social inequalities in health: a scoping review of the literature. PLoS ONE 13:e019544
Whitelaw S, Mamas MA, Topol E, Van Spall GC (2020) Applications of digital technology in COVID-19 planning and response. Lancet Digit Health 2:e435–e440
World Health Organization (2019) WHO consultation towards the development of guidance on ethics and governance of artificial intelligence for health: meeting report, Geneva, Switzerland, 2–4 October 2019. https://iris.who.int/handle/10665/340089
World Health Organization (2021) Ethics and governance of artificial intelligence for health. https://www.who.int/publications/i/item/9789240029200

Fidelia Cascini, M.D., Ph.D., Associate Professor of Hygiene and Public Health at University S. Raffaele, Rome, Italy. Head of the Scientific Secretariat of the Presidency, National Institute of Health. Digital health advisor, Ministry of Health. Italian representative at the Committee on the European Health Data Space, European Commission, and Co-Chair at the Global Digital Health Partnership. Author of over 90 papers and 4 monographs.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

