Trustworthy AI Tools for the Prediction of Obesity Related Vascular Diseases

HORIZON-HLTH-2022-STAYHLTH-01-04-TWO-STAGE

DELIVERABLE D5.1

ANALYSIS OF GOVERNANCE FRAMEWORKS FOR THE IMPLEMENTATION OF AI-DRIVEN TECHNOLOGIES

Lead beneficiary	KU Leuven
Author(s)	Pascal Borry Eva Van Steijvoort
Dissemination level	Public
Type	Report
Delivery date	30/04/2024

AI-POD is funded by the European Union's Horizon Europe research and innovation programme under grant agreement No 101080302

Funded by the European Union

Table of contents

Exectuvive summary 3
Introduction 4
Key ethical principles for responsible AI implementation 5
Principles of AI Governance 10
Governance models for the application of AI in health care 14
Conclusion 20
References 21

Disclaimer

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

EXECTUVIVE SUMMARY

Technologies based on AI have the potential for strengthening the delivery of public health care and medicine by overcoming the limitations of traditional rules-based clinical decision support systems [1]. However, to achieve a beneficial impact, ethical/safety considerations and human rights should be placed at the centre of the design, development and deployment of AI technologies. If ethical and safety concerns are neglected, there might potentially be serious negative consequences [2, 3].

To limit the risks and maximize the opportunities intrinsic to the use of AI for health, the WHO has proposed some key ethical principles to guide the responsible implementation of AI for health to ensure its great potential [1]. The implementation of AI in healthcare may however present challenges that cannot be adequately addressed by existing ethical principles, laws, or policies. This is due to the fact that the potential risks and opportunities associated with the use of AI in healthcare are not fully understood and may evolve over time. As a result, there is a pressing need for effective governance of AI in healthcare that encompasses various regulatory and decision-making functions carried out by governments and other stakeholders [2].

To ensure consistent solutions and enable countries to support and benefit from each other’s efforts, there is an urgent need for international collaboration and coordination on AI governance for health care [4]. The lack of international coordination for the governance of AI in healthcare may also limit its adoption because of issues of trust, which is seen as a core component for succesful innovation in digital health [4]. A robust and unified governance framework could both enhance trustworthiness and transparancy of AI-systems and mitigate potential challenges associated with the implementation of AI-driven tools [5].

In this report, we present an analysis on the academic literature, guidelines and governance frameworks of AI-based decision-making in healthcare. As a result, we aim to provide valuable guidance for the development and implementation of the AI-POD project and to ensure that the project aligns with established standards and promotes safe and ethically sound AI-based decision-making tools.

Funded by
the European Union

INTRODUCTION

Artificial Intelligence (AI) refers to "the ability of algorithms encoded in technology to learn from data so that they can perform automated tasks without every step in the process having to be programmed explicitly by a human"[2]. Technologies based on AI have the potential for strengthening the delivery of public health care and medicine by overcoming the limitations of traditional rules-based clinical decision support systems [1]. For example, it can be applied to improve prevention, to predict certain diseases, to provide more accurate and faster diagnosis, to optimize treatment decision-making, to avoid human errors or even to allocate resources within health systems [6, 7]. Furthermore it holds great promise to empower patients to take control over their own health. AI could for example assist in health self-management through health monitoring and risk prediction tools that provide specific recommendations to improve health (e.g. nutrition and diet, physical activity, etc.) [2, 8]. However, to achieve a beneficial impact, ethical/safety considerations and human rights should be placed at the centre of the design, development and deployment of AI technologies. If ethical and safety concerns are neglected, there might potentially be serious negative consequences [2, 3]. Safety issues could for example arise after the implementation of AI systems in health care practice because of the unpredictable performance of AI-driven tools in diverse settings, the unknown human-computer interactions, the unclear accountability and liability and the inadequate education or preparedness of the health care workforce [3]. While some concerns are not unique to AI and have been addressed since the introduction of software and computing in healthcare, others are novel and more specific to the use of AI in healthcare [2]. For example, the growing use of mobile health applications and wearables that require near-continuous monitoring and collection of large amounts of data that otherwise would have remain unknown [2].

As AI-driven tools become more prevalent in clinical practice, there is a growing necessity for frameworks to ensure safe adoption and to govern their use effectively [9, 10]. Ensuring robust and effective governance is deemed essential for addressing all existing and potential challenges in the application of AI-driven tools in healthcare [11]. AI governance encompasses a range of procedures concerning the ethical implementation and utilization of AI tools, along with the regulation and accreditation of AI models. It also addresses issues such as liability, accountability, data protection protocols, and educational efforts, among other considerations [10]. To date, there remains a lack of globally acknowledged governance mechanisms for the development and utilization of AI-driven tools in healthcare which has led to important variation among organizations and countries [4]. In the USA, there is a belief that regulation stifles innovation, leading to a preference for minimal governance [12]. In contrast, other countries tends to prioritize stronger regulatory frameworks to facilitate innovation by providing structure and clarity [4]. The European Commission proposed the world's first comprehensive legal framework on AI in April 2021 as part of its digital strategy. In December 2023, the Council and the Parliament reached a political agreement on the EU's new Artificial Intelligence Act (AI Act). This legislation follows a risk-based approach to ensure the safety of AI systems in the EU market while respecting fundamental rights. The AI Act is expected to set a global standard for AI governance, similar to the widespread influence exerted by the General Data Protection Regulation (GDPR) [13].

To ensure consistent solutions and enable countries to support and benefit from each other's efforts, there is an urgent need for international collaboration and coordination on AI governance for health care [4]. The lack of international coordination for the governance of AI in healthcare may also limit its adoption because of issues of trust, which is seen as a core component for succesful innovation in digital health [4]. A robust and unified governance framework could both enhance trustworthiness and transparancy of AI-systems and mitigate potential challenges associated with the implemenation of AI-driven tools [5].

Funded by the European Union

KEY ETHICAL PRINCIPLES FOR RESPONSIBLE AI IMPLEMENTATION

To limit the risks and maximize the opportunities intrinsic to the use of AI for health, the WHO has proposed some key ethical principles to guide the responsible implementation of AI for health to ensure its great potential [2]. These include the following ethical principles:

PROTECT AUTONOMY

In the context of health care, this means that humans should remain in control of health-care systems and medical decisions. Healthcare professionals should be able to override decisions made by AI systems. AI driven technologies should be designed to assist in making informed decisions. This principle also entails the related duty to protect privacy and confidentiality. Finally, patients must give valid informed consent through appropriate legal frameworks for data protection [2].

PROMOTE HUMAN WELL-BEING, HUMAN SAFETY AND THE PUBLIC INTEREST

The designers of AI technologies should satisfy regulatory requirements for safety, accuracy and efficacy for well-defined use cases or indications. Specific measures should be put in place to ensure quality control and quality improvement to ascertain if the AI drive systems are working as designed and to ensure the early identification of any detrimental effect on individual patients or groups [2].

ENSURE TRANSPARENCY, EXPLAINABILITY AND INTELLIGIBILITY

Transparency requires that sufficient information should be published or documented before the design or deployment of an AI technology. This should include information about the assumptions and limitations of the technology, operating costs, the properties of the data and development of the algorithmic model. Such information must be easily accessible and facilitate meaningful public consultation and debate on how the technology is designed and how it should or should not be used. Furthermore information should be tailored, according to the capacity of those to whom the explanation is directed. This could lead to a possible trade-off between full explainability of an AI algorithm (at the cost of accuracy) and improved accuracy (at the costs of explainability) [2].

FOSTER RESPONSIBILITY AND ACCOUNTABILITY

Although AI technologies perform specific tasks, it is the responsibility of stakeholders to ensure that they are used under appropriate conditions and by appropriately trained people. Effective mechanisms should be available for questioning and for redress for individuals and groups that are adversely affected by decisions based on algorithms. To avoid a diffusion of responsibility where 'everybody's problem becomes nobody's responsibility' a collective responsibility has been proposed where all the actors involved in the development and deployment of AI could be held responsible. This collective responsibility could encourage all actors to act with integrity and minimize harm [2].

ENSURE INCLUSIVENESS AND EQUITY

Inclusiveness requires that AI for health should be designed to encourage the widest possible equitable use and access, irrespective of age, sex, gender, income, race, ethnicity, sexual orientation, ability or other characteristics protected under human rights codes. AI technologies should be adaptable to context and needs of different settings and should try to avoid to enlarge the existing 'digital divide'. (Unintended) biases should be avoided or identified and mitigated [2].

Funded by the European Union

PROMOTE ARTIFICIAL INTELLIGENCE THAT IS RESPONSIVE AND SUSTAINABLE

Designers, developers and users should continuously and transparently assess AI applications during actual use to determine whether AI responds adequately and appropriately to expectations and requirements. The use of AI technology should be terminated as soon as possible if necessary. AI systems should also be designed to minimize their environmental consequences and increase energy efficiency. Governments and companies should address anticipated disruptions in the workplace, including training for health-care workers to adapt to the use of AI systems, and potential job losses due to use of automated systems [2].

Funded by
the European Union

ETHICAL CHALLENGES APPLIED TO THE AREA OF AI-BASED DECISION-MAKING IN IMAGING-BASED PREDICTION OF OBESITY-RELATED VASCULAR DISEASES

AUTONOMY & AI PROVIDED DIAGNOSIS

The greatest concerns regarding autonomy when AI technology is used largely lie within the domains of explainability and informed consent. These two domains go hand in hand as explainability of healthcare interventions ensures that true informed consent is obtained. As one of the main principles of health care ethics, protecting the autonomy of patients and their decisions is paramount to creating healthcare that is ethically appropriate and does not lead to greater harm to patients. These issues are compounded when diagnosis or prognosis specifically is provided via AI technology, which AI-POD includes within its Citizen App.

To begin, the main method of protecting patients' autonomy in most healthcare interventions is by gaining informed consent. However, "informed consent within the context of AI poses practical challenges of explainability to patients as the output derived from AI systems can be challenging to interpret", thus explaining to patients exactly how their diagnosis or their treatment was decided upon becomes difficult [14].

Therefore, as WHO guidance states, it is important to keep explainability in mind when creating trustworthy AI tools, as the tools must be intelligible to developers, medical professionals, patients, users and regulators, and making AI technology explainable is one of the approaches to ensuring intelligibility. The WHO goes on to state that "AI technologies should be explainable according to the capacity of those to whom they are explained," so when explaining the use of AI technology in healthcare interventions to patients, a more feasible approach may be that "the right to an explanation requires that the solution is to make artificial intelligence decision-making explainable, not to explain the artificial intelligence model" [2, 15].

To expand, Lehmann [12] posits that with certain AI technology, such as machine learning and deep learning, there "may be limited information on how an AI output is derived for a specific system, in most cases there are important features of AI that can and should be shared with patients". These features include but are not limited to "how sensitive and specific the algorithm is for certain patients, the error rate of an algorithm, how an algorithm's accuracy compares with physicians' decisions, the safeguards put into place to detect and prevent errors, and the consequences for patients' health if the AI algorithm is biased or wrong may be more important to patients than how an algorithm arrived at a decision"[14]. These details can help both patients and doctors, in safeguarding "the autonomy-preserving function of informed consent" for patients without demotivating "healthcare stakeholders from implementing advanced technologies in their daily practice"[15, 16].

When considering the use of AI in diagnostic and prognostic systems specifically, there are several issues that are dependent on the use of data sets for machine learning and deep learning algorithms. Beil et al. [17] discusses how data from cohort studies applied to individuals "carries the risk of false hope, false despair, or continuing uncertainty. A falsely optimistic prognosis based on, for example, an unsuitable dataset for training an AI model could trigger futile, i.e., potentially inappropriate interventions." This dilemma could be alleviated by personalizing probabilities as much as possible, by taking into account more features describing the individual circumstances of the patient [17]. WHO's guidance also corroborates the emotional harm that could come from AI provided diagnosis or warnings, as providing diagnosis to patients that cannot be addressed "because of lack of appropriate, accessible or affordable health care should be carefully managed and balanced against any "duty to warn" that might arise from incidental and other findings" [2]. Liu et al.'s [18] research in AI-aided diagnosis for cardiac diseases suggests that AI diagnosis tools should be used as

Funded by the European Union

auxiliary tools to aid physicians to make "better interpretation decisions and improve diagnosis accuracy and efficiency", but ultimately it should be physicians that are providing the prognosis.

An issue that also arises with autonomy when integrating the use of AI is the question of "whether the use of AI-powered decision aids is compatible with the inherent values of patient-centered care" [16]. The integration of opaque medical AI into healthcare decision-making may lead to a paternalistic healthcare framework, potentially restricting patients' opportunities to communicate their expectations and preferences regarding medical procedures or interventions [16]. Additionally, due to AI's data dependency, there may be the obstacle of depersonalization, where patients are viewed simply as data points fed into the AI learning algorithms rather than holistically, possibly resulting in a decrease in the quantity and quality of patient-provider relations [15]. However, solutions to these issues stemming from AI decision-making systems include creating value-flexible AI systems that take into account differing values and beliefs of patients. This is integral in protecting the autonomy because as Beil et al. expresses, "respect for patients autonomy acknowledges the capacity of individuals for self-determination and the right to make choices based on his/her own values and beliefs" [17].

ALGORITHMIC BIAS

A well-documented issue that has been observed in the use of AI learning algorithms is that of algorithmic bias, where algorithms reproduce flaws present from older data [19, 20]. Such biases can be present when the data sets used for training aren't representative for the target population. AI models are designed to recognize patterns within datasets, which means they may unintentionally perpetuate any biases or unfairness present in the data they're trained on [2, 20]. This is problematic as this produces a downstream impact of bias in data or modelling, "where observed risk is caused by biased care in the underlying data rather than biologically plausible mechanisms for disparate risk"[21]. Thus algorithmic bias creates AI tools that may systematically produce worse outcomes for under-served groups by hardcoding health disparities that result from unequitable access to social determinants of health into clinical practice [21]. To promote better and more equitable health outcomes it's therefore necessary to reduce AI bias [1, 2].

This is relevant to AI-POD and its proposal as there is bias present in the treatment and understanding of both cardiovascular disease and obesity [20, 22-26]. The screening and risk practices of cardiovascular disease have historically been centered around men's presentation of the disease, where "the underrepresentation of women in research partially explains the incomplete understanding of CVD symptomology and presentation in women" [23]. Regarding obesity, Phelan et al.'s narrative review found that there is a "growing body of evidence that physicians and other healthcare professionals hold strong negative opinions about people with obesity", which have been observed impact health outcomes and treatments negatively [24].

Furthermore, in the development of AI-POD's CDSS, the proposal states that the system "will be created by integrating established guidelines from ESC, AHA", however many of the studies used to create the guidelines and treatments of cardiovascular underrepresent women in their research, in addition to not addressing the sex and gender specific effects on cardiovascular risk [20, 23]. There is considerable concern in the likelihood of biased predictions of cardiovascular risk for women and underserved minorities due to their underrepresentation in cardiology and their frequent exclusion in clinical trials for treatment and guideline development [20].

Due to issues like algorithmic bias, it is the duty of funders, developers and users "to measure and monitor the performance of AI algorithms to ensure that AI technologies work as designed and to assess whether they have any detrimental impact on individual patients or groups." Oversight measures such as regular tests and evaluations, alongside human supervision is strongly recommended to prevent biased algorithms. Additional recommendations for addressing bias in AI algorithms encompass ensuring the inclusion and representation

Funded by the European Union

of women and minority populations in data sets, appropriate selection of clear and specific training targets for algorithms being developed, thorough risk assessments before development, and typical ethical oversights such as transparency, validation, and rigorous testing [20]. Consequently, there should be mechanisms for redress and accountability if the AI technology provides wrong or biased predictions [2].

STIGMATIZATION

As touched upon in the previous section, obesity is a commonly and strongly stigmatized characteristic, with bias present in general populations and within the healthcare community, even within providers who specialize in treating obesity [24, 27]. Weight discrimination is one of the most frequent forms of discrimination reported by adults and many healthcare professionals hold negative attitudes about obesity, such as stereotypes that affected patients are "lazy, lack self-control and willpower, are personally to blame for their weight, and are noncompliant with treatment" [25, 26]. When working on health interventions for a stigmatized community, it is imperative that the intervention and treatment process do not reproduce or contain stigmatizing elements. Stigma adversely affects health outcomes and serves as a barrier to care, as patients who experience stigma, especially from healthcare providers, are less likely to seek out healthcare [26].

The use of commercial mobile apps for weight management have become increasingly popular and have the advantages of being widely available, easily accessed, and portable, however "apps that have weight stigmatization unintentionally embedded within them may de-incentivize behavior change and/or cause emotional distress" [28]. This is particularly relevant to the development of the Citizen App for the AI-POD proposal, as the app would have the ability to "'what-if' scenarios informing individual lifestyle decisions with quantitative evidence and prediction" and personalized health education. These notifications should not use stigmatizing elements, such as the use of overweight or obese bodies as negative imagery or grading systems for diet and exercise tracking [28]. As stigmatization plays a role in the healthcare experience of obese persons, underrepresentation of the cardiovascular risk in obese persons in data sets as well as medical bias could result in flawed risk prediction. Therefore, avoiding and preventing stigmatization within the development of AI-POD is of great interest to ensure the successful development of the AI tools and treatments processes.

PRIVACY

With the collection of personal health data, the issue of privacy and how to secure patient-consumer data must be addressed. The AI-POD proposal outlines how it will seek to integrate the AI tools and data into a secure and unique platform that will be "used for data provision and processing within the project, and for data- and algorithm releases to the wider research community in a novel process of data and AI technology sharing across Europe". The platform will also play a role in the management and growth of the data as well as "the validation of algorithms on external data and by external experts". The proposal also notes that the data will be "stored decentralized within the IT-infrastructure of the CDSS user in compliance with the General Data Protection Regulation (GDPR).

Ensuring GDPR compliance for the platform and overall collection of data is the first step in ensuring that the privacy and rights of the participants is prioritized, even when information sharing and continued research remains a goal. From an ethical perspective, "privacy is linked to digital agency, that is, control over access to and use of an individual's personal information" [14]. Although other possibilities exist, consent is a common legal ground for the collection of personal data, and Lehmann [14] discusses that in most instances "individuals must consent to their personal data being collected and used for automated decision-making that significantly or legally affects them, and individuals have the right to obtain human intervention and

Funded by the European Union

contest decisions based on AI." When gaining informed consent, it is important that it is clear what data is collected, why it is collected, and data sharing practices must be disclosed [29]. In this way, it is clear that privacy is linked to autonomy and informed consent, as patients should have the ability to control who has access to their information [29]. Thus, participants must be aware that their data will be shared for the purposes of research and the further development of AI tools.

The use of AI technology does bring about greater concerns regarding privacy as Kerry's research as cited in Zhang et al. [30] discusses "that AI expands the ability to use personal information in ways that can infringe on privacy interests by bringing personal data analysis to new levels of power and speed". It may be difficult to fully gain consent when the use of data in unsupervised AI learning algorithms is not fully explainable. The use of privacy-preserving techniques "including (relative) anonymization , access control (plus encryption), and other models where computation is carried out with fully or partially encrypted input data" should be utilized to provide the greatest degree of protection for participant data [31].

PRINCIPLES OF AI GOVERNANCE

The significant growth of the AI industry has also given rise to a high demand for regulation and normative guidance, what is also referred to as the 'AI ethics boom' [32]. From 2014 onwards, there has been a significant rise in the production of documents related to AI ethics [32, 33]. Corrêa et. al. (2023) performed a meta-analysis of 200 governance policies and ethical guidelines for AI usage published by public bodies, academic institutions, private companies and civil society organizations worldwide (37 countries spread over six continent) to identify the most advocated ethical principles. In addition they assessed if there is a consistent understanding of these principles and how they are distributed globally.

The authors identified 17 principles that encompass the normative discourse within these policies and guidelines. These principles bring together similar and resonant values. Below, you can find an overview of the definitions Corrêa et. al. (2023) have provided for each of these 17 aggregated principles [32]. The five most prevalent principles were: (1) transparency/explainability/auditability, (2) reliability/safety/security/trustworthiness, (3) justice/equity/fairness/non-discrimination, (4) privacy and (5) accountability/liability [32]. These findings are in line with earlier initiatives to map principles and guidelines of ethical AI [34, 35]. Interestingly, the least mentioned principles relate to labor rights, sustainability and truthfulness indicating that the discourse of many of these guidelines might already be outdated due to the significant growth of generative AI. For example, many have raised concerns about the possibility of mass unemployment, yet the proposed measures remain insufficient. Another example is the related to the costs of AI technologies. While the carbon footprint of these processes is well-documented, there is limited understanding of the broader costs beyond CO2 emissions [32].

Funded by the European Union

PRINCIPLE	DEFINITION
Accountability/liability	The idea that developers and deployers of AI technologies should be compliant with regulatory bodies. These actors should also be accountable for their actions and the impacts caused by their technologies [32].
Beneficence/non-maleficence	The idea that in AI ethics, human welfare (and harm aversion) should be the goal of AI-empowered technology [32].
Children and adolescents rights	The idea that we must protect the rights of children and adolescents. AI stakeholders should safeguard, respect, and be aware of the fragilities associated with young people [32].
Dignity/human rights	The idea that all individuals deserve proper treatment and respect. In AI ethics, respect for human dignity and human rights (i.e., the Universal Declaration of Human Rights) are used (sometimes) interchangeably [32].
Diversity/inclusion/pluralism/accessibility	The idea that the development and use of AI technologies should be done in an inclusive and accessible way, respecting the different ways that the human entity may come to express itself (gender, ethnicity, race, sexual orientation, disabilities, etc.) [32].
Freedom/autonomy/democratic values/technological sovereignty	The idea that the autonomy of human decision-making must be preserved during human-AI interactions, whether that choice is individual or the freedom to choose together, such as the inviolability of democratic rights and values, also being linked to technological self-sufficiency of nations/states [32].
Human formation/education	The idea that human formation and education must be prioritized in our technological advances. AI technologies require considerable expertise to be produced and operated, and such knowledge should be accessible to all [32].
Human-centeredness/alignment	The idea that AI systems should be centered on and aligned with human values. This principle is also used as a "catch-all" category, many times being defined as a collection of "principles that are valued by humans" (e.g., freedom, privacy, non-discrimination, etc.) [32].

Funded by the European Union

Intellectual property	The idea that underlies this principle is to establish property rights over AI products and their generated outputs [32].
Justice/equity/fairness/non-discrimination	The idea of non-discrimination and bias mitigation (discriminatory algorithmic biases AI systems can be subject to). It defends that, regardless of the different sensitive attributes that may characterize an individual, algorithmic treatment should happen "fairly” [32].
Labor rights	The idea that this principle emphasizes is that labor rights, which are legal and human rights related to the labor relations between workers and employers, should be preserved regardless of whether labor relations are being mediated or augmented by AI technologies [32].
Cooperation/fair competition/open source	This set of principles advocates different means by which joint actions can be established and cultivated between AI stakeholders to achieve common goals. It also relates to the free and open exchange of valuable AI assets (e.g., data, knowledge, patent rights, human resources) [32].
Privacy	The idea of privacy can be defined as the individual’s right to "expose oneself voluntarily, and to the extent desired, to the world." This principle is also related to data protection related-concepts such as data minimization, anonymity, informed consent, and others [32].
Reliability/safety/security/trustworthiness	This set of principles upholds the idea that AI technologies should be reliable, in the sense that their use can be truly attested as safe and robust, promoting user trust and better acceptance of AI technologies [32].
Sustainability	This principle can be interpreted as a manifestation of "intergenerational justice," wherein the welfare of future generations must be considered in AI development. In AI ethics, sustainability pertains to the notion that the advancement of AI technologies should be approached with an understanding of their enduring consequences, encompassing factors such as environmental impact and the preservation and well-being of non-human life [32].
Transparency/explainability/auditability	This set of principles supports the idea that the use and development of AI technologies should be transparent for all interested stakeholders.

	Transparency can be related to "the transparency of an organization" or "the transparency of an algorithm." This set of principles is also related to the idea that such information should be understandable to nonexperts and, when necessary, subject to be audited [32].
Truthfulness	This principle upholds the idea that AI technologies must provide truthful information. It is also related to the idea that people should not be deceived when interacting with AI systems [32].

An in-depth analysis of these 200 governance policies and ethical guidelines highlighted an underrepresentation of certain world regions/countries, with one third of the documents coming from Europe and North America [32]. Yet, as mentioned by the authors, there is also an ongoing discourse present about AI governance in regions/countries that are underrepresented in the sample. The underrepresentation might simply be due to language limitations, lack of representative databases used for the search and/or unfamiliarity with how to find such documents. Furthermore, the authors found that most of the analyzed documents came from private institutions and/or governmental institutions. The equal presence of both groups in the current normative discourse may be related to the recent successes of the AI industry. The AI industry seems to quickly respond to the demands for regulation and accountability from civil society by proposing rules that are supposed to guide its progress [32]. Most of the analyzed documents were also of the normative type (96%), while only 2% of documents provided practical tools to implement ethical principles and norms [32]. Furthermore, most documents offered recommendations to various AI stakeholders (56%), while 24% present self-regulatory or voluntary self-commitment guidelines, and only 20% advocate for regulation administered by countries [32]. The dominance of "soft laws" in these documents, with 98% offering non-binding guidelines, reflects the current lack of convergence towards government-based regulation [32]. Yet, there seems to be a growing understanding that ethical principles will not be enough to govern the AI industry with a growing adoption/proposition of stricter resolutions [32]. We could currently be in a transitioning phase where ethical principles might be more and more translated into actual legally binding forms of regulation [32]. Finally, the analysis revealed that while academic and non-profit organization think more about long term impact of AI-technologies (e.g. AI-related existential risks, super-intelligent AI, etc.) , private corporations are more focused on short-term implications (e.g. legal accountability, algorithmic discrimination, etc.) [32].

Funded by the European Union

GOVERNANCE MODELS FOR THE APPLICATION OF AI IN HEALTH CARE

Various ethical principles and guidance documents have been published to guide the integration of AI tools across industries, including healthcare. These recommendations aim to address that AI in healthcare is safe, effective, and ethically appropriate at every stage. International organizations like the WHO are leading discussions and creating practical tools for using AI in healthcare [2, 4]. Nevertheless, significant tasks remain. The implementation of AI in healthcare presents numerous challenges that cannot be adequately addressed by existing ethical principles, laws, or policies. This is due to the fact that the potential risks and opportunities associated with the use of AI in healthcare are not fully understood and may evolve over time. As a result, there is a pressing need for effective governance of AI in healthcare that encompasses various regulatory and decision-making functions carried out by governments and other stakeholders [2].

As AI becomes more integrated into healthcare systems for tasks like diagnosis, treatment planning, and patient monitoring, establishing robust governance frameworks prioritizing patient safety and well-being will become more and more urgent. The rapid development of AI-driven technology in healthcare has already outpaced the creation of global guidelines to govern its use. International collaboration will be essential to create comprehensive and clear rules for AI in healthcare. This will help countries support and learn from each other [4]. Consensus on best practices, widespread education about AI in healthcare, and the establishment of international standards for AI models will be imperative. By discussing these issues at a more practical level, we can bridge cultural and political gaps and create common ground.

WHO'S CONTRIBUTION TO A FRAMEWORK FOR GOVERNANCE OF AI FOR HEALTH

The WHO [2]has laid out a set of recommendations covering various aspects for the governance of data, control and benefit sharing, governance of the private sector, governance of the public sector, regulatory considerations, policy observatory, model legislation, and global governance of AI in the healthcare sector.

In terms of data governance, the WHO emphasizes the importance of clear data protection laws, meaningful informed consent, and transparency in the use of health data. Governments should also be urged to establish independent data protection authorities to enforce these laws and support community oversight mechanisms. Furthermore, the WHO highlights the need for clear ownership and benefit sharing of AI technologies and algorithms. Research institutions and universities involved in the development of AI technologies should maintain an ownership interest in the outcomes so that the benefits are shared and are widely available and accessible, particularly to populations that contributed their data for AI development. The WHO also recommends that governments should consider alternative “push-and-pull” incentives instead of IP rights, such as prizes or end-to-end push funding, to stimulate appropriate research and development. Furthermore, transparency in regulatory procedures and interoperability should be enhanced and should be fostered by governments as deemed appropriate [2].

Regulatory considerations for the use of AI in healthcare are believed to be crucial to ensure responsible innovation and to safeguard patient safety. According to the WHO, governments should introduce and enforce regulatory standards for new AI technologies to prevent the use of harmful or insecure systems. Transparency of AI technologies is considered to be essential, including the disclosure of source code, data inputs, and analytical approaches. Governments should also mandate prospective testing in randomized trials to assess AI system performance accurately. To address safety and human rights concerns, regulators should provide

Funded by the European Union

incentives to developers and integrate relevant guidelines into precertification programs. They should also conduct robust marketing surveillance to identify biases [2].

In addition, the WHO also advocates for a coordinated approach with intergovernmental organizations to formulate ethical laws, policies, and best practices for AI technologies in healthcare. Global governance of AI for health is seen as essential by the WHO to ensure adherence to ethical norms, human rights, and legal obligations. Global health bodies should commit to uphold human rights obligations, legal safeguards, and ethical standards, while international agencies, such as the Council of Europe, OECD, UNESCO should develop a common plan to address ethical challenges and opportunities in AI for health, providing legal and technical support to governments to comply with international ethical guidelines and principles [2].

Below we also provide two more specific governance frameworks that have been proposed to govern trustworthy AI in health care.

GOVERNANCE MODEL FOR AI IN HEALTHCARE (GMAIH)

To address ethical, regulatory and safety and quality concerns, Reddy et al. (2020) have proposed a governance model for AI application in health care. The proposed governance model 'Governance Model for AI in Healthcare (GMAIH) has four main components which include fairness, transparency, trustworthiness and accountability (see Figure 1). By incorporating basic elements essential to the safe and ethically responsive use of AI in health care, it is designed to be flexible enough to accommodate changes in AI technology. Below we will discuss the four components of the proposed model in more detail.

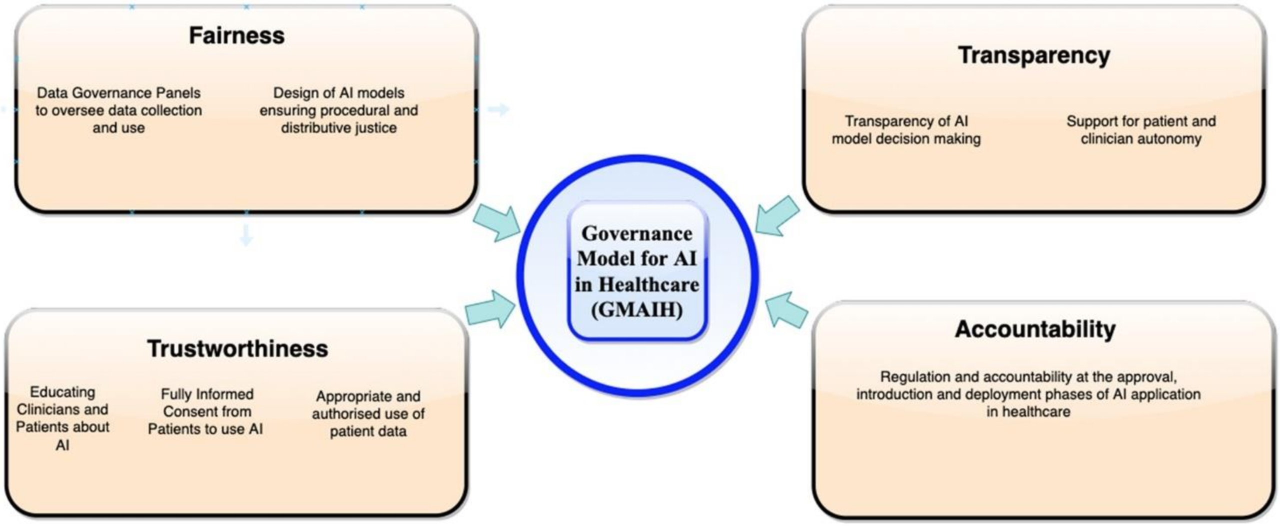
Fairness
Data Governance Panels to oversee data collection and use
Design of AI models ensuring procedural and distributive justice
Transparency
Transparency of AI model decision making
Support for patient and clinician autonomy
Governance Model for AI in Healthcare (GMAIH)
Trustworthiness
Educating Clinicians and Patients about AI
Fully Informed Consent from Patients to use AI
Appropriate and authorised use of patient data
Accountability
Regulation and accountability at the approval, introduction and deployment phases of AI application in healthcare

Figure 1: Outline of the Governance Model for Artificial Intelligence (AI) in Health Care. (Sandeep Reddy, Sonia Allan, Simon Coghlan, Paul Cooper, A governance model for the application of AI in health care, Journal of the American Medical Informatics Association, Volume 27, Issue 3, March 2020, Pages 491–497)

1. Fairness

The first component of the GMAIH model centers around the principle of fairness. Reddy et al. (2020) underline the importance of establishing a data governance panel led by AI developers, that also includes representatives from patient and target groups, clinical experts, and individuals with relevant AI, ethical, and legal backgrounds[1]. This panel has the important task to review datasets used for training AI to ensure that data are representative and sufficient to inform model outcomes and to develop a clear data collection strategy, guiding documentation, workflow, and monitoring standards. Additionally, the panel

Funded by the European Union

should also review algorithms, acknowledging their integral role in AI model development [1]. Furthermore the authors underline that the design of AI models should ensure procedural and distributive justice. Normative standards for AI application in healthcare should be developed by governmental bodies and healthcare institutions, emphasizing principles of justice to ensure fairness in access to healthcare. To protect against adversarial attack or the introduction of biases or errors through self-learning or malicious intent, it's important to ensure both procedural (fair process) and distributive justice (fair allocation of resources).

2. *Transparency*

While the performance of AI algorithms in healthcare are promising, they are also often hard to interpret and explain. This poses an important issue, especially in medicine where transparency and explainability of clinical decision making are considered to be of utmost importance. Sufficient transparency and explainability are requested by the ethical principle of autonomy. A lack thereof can hinder trust in AI models and makes it difficult to validate clinical recommendations of AI models or the detection of errors or biases. To tackle this challenge, the concept of explainable AI (XAI) has emerged, aiming to provide techniques for maintaining performance while enabling explainability. In medicine, XAI techniques focus on understanding the functional logic of models rather than low-level details. Although explainable algorithms may be less accurate and have a lower predictive performance, they are seen as essential for ensuring transparency in medical decision-making. Additionally, AI agents must adhere to principles of respect for autonomy, supporting patients' freedom to make decisions based on clear understanding without coercion or undue pressure. Therefore, the governance model emphasizes ongoing explainability, promoting the use of interpretable frameworks alongside deep learning models to enhance decision-making in healthcare. Recent medical studies have demonstrated the feasibility of this approach through various explainable tools, ranging from visual aids to direct measurement tools [1].

3. *Trustworthiness*

Clinicians' understanding of the methods employed by AI to aid decision-making is crucial. However, issues like explainability and potential autonomous functioning of AI can hinder their acceptance. To address this, Reddy et al. (2020) have proposed a multi-faceted approach, including technical education, health literacy, informed consent, and clinical audits [1]. Educating healthcare professionals about AI basics could build trust in AI-driven tools. This approach can also enable health care professionals to become partners in the control of AI-driven tools instead ofremaining more passive recipients of the outputs of these tools. Extending education to patients could ensure patients receive the information they need to make autonomous and informed choices. This also entails that patients should also be informed when the clinical decision making of health care professionals is supported by AI-driven tools, what the limitations of these AI-driven tools are and that they have the right to refuse treatment involving AI. Fully informed consent should also be sought from patients to be able to share data with developers of AI software. When data are shared a high standard of data anonymization should be the aim to protect the privacy from patients and to avoid patient reidentification. When possible, public datasets should be prioritized, to minimize privacy breaches [1].

4. *Accountability*

Accountability is seen as integral from AI model development to clinical application and eventual retirement. This complex process involves stakeholders like software developers, regulatory agencies, healthcare providers, professional bodies, and patient advocacy groups. To address this complexity, the

Funded by
the European Union

GMAIH proposes a structured approach focusing on monitoring and evaluation at key stages namely approval, introduction, and deployment. In the approval stage (includes both permission for the marketing and use of AI in healthcare delivery), regulatory bodies or governmental bodies have an important role to play to organize risk review and pre-market approval of AI-based software as a medical device. In the future, the challenge will be to find the right balance to ensure safety and quality of AI-based software as a medical device while avoiding the creation of undue obstacles for AI developers to bring their software to market. During the introduction stage, healthcare services should evaluate AI products present in the market, assess them for suitability and establish relevant policies and procedures to allow for the incorporation of AI-based software as a medical device. Due to the rapid progression in AI technology, it has been suggested that a benchmarking system could aid health services to assess the performance and robustness of AI-driven tools and to compare different AI models through a dashboard of performance metrics.

In the deployment stage, accountability includes liability, monitoring, and reporting. Determining responsibility for safety and quality issues arising from AI software usage requires appropriate legal guidance, as current laws may not adequately cover autonomous or semi-autonomous medical software scenarios. A balanced regulatory approach prioritizing patient safety, clinician autonomy, and AI-driven decision support is necessary. The models therefore foresees the need for responsive regulation with ongoing safety monitoring through audits and reporting, including assessments of bias, accuracy, predictability, and decision transparency [1].

CONCEPTUAL AI GOVERNANCE FRAMEWORK

Stogiannos et al. (2023) performed a scoping review to map out available literature on AI governance in the UK, focusing on medical imaging and radiotherapy. Based on their findings they have proposed a comprehensive AI governance framework based on 7 pillars of AI governance [10]. The proposed conceptual framework includes validation and evaluation procedures of AI-drive tools, monitoring of the safety and clinical effectiveness of AI models, compliance with appropriate accreditation bodies and regulatory standards, fundamental ethical principles that should be followed, appropriate staff training, innovation and growth and effective leadership and staff management [10]. Below we will discuss every pillar of the proposed model in more detail.

Funded by the European Union

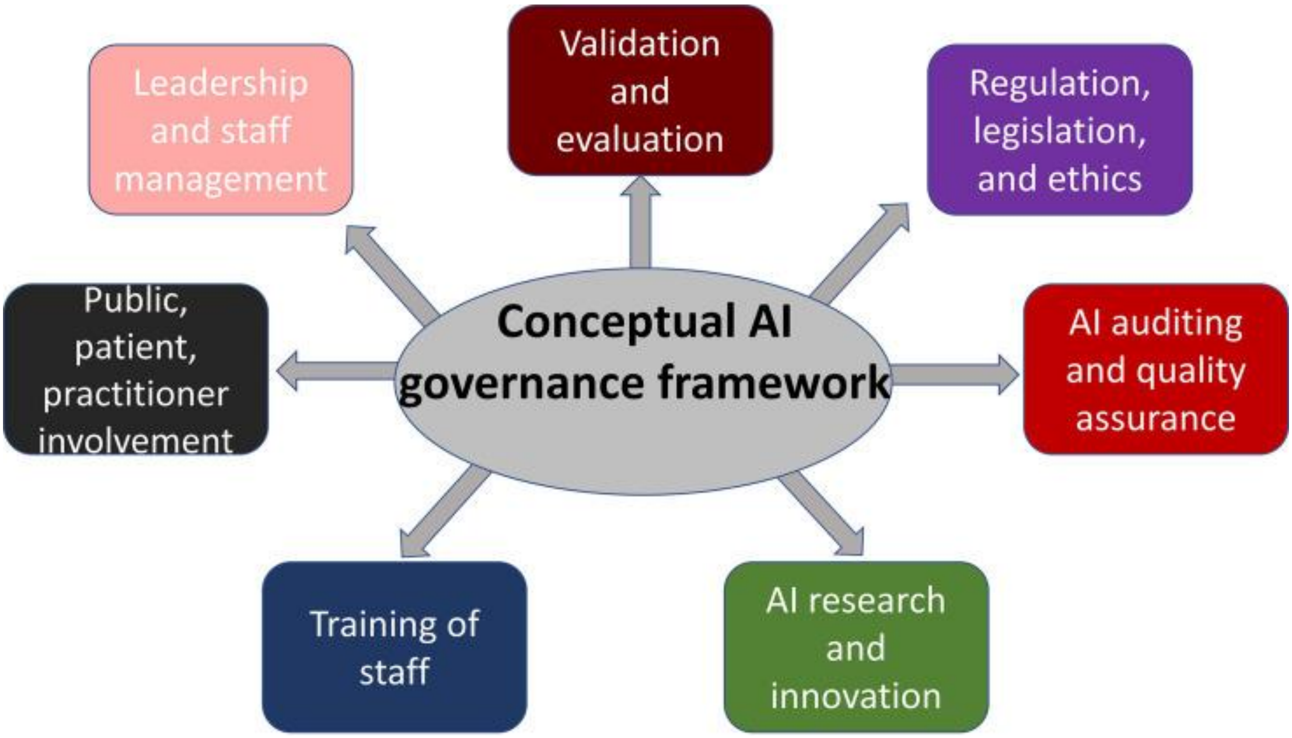
Conceptual AI governance framework
Leadership and staff management
Validation and evaluation
Regulation, legislation, and ethics
Public, patient, practitioner involvement
AI auditing and quality assurance
Training of staff
AI research and innovation

Figure 2: Suggested AI governance framework. (Stogiannos N, Malik R, Kumar A, Barnes A, Pogose M, Harvey H, McEntee MF, Malamateniou C. Black box no more: a scoping review of AI governance frameworks to guide procurement and adoption of AI in medical imaging and radiotherapy in the UK. Br J Radiol. 2023 Dec;96(1152):20221157. doi: 10.1259/bjr.20221157. Epub 2023 Oct 3. PMID: 37747285; PMCID: PMC10646619.)

1. *Validation & Evaluation*

The first pillar refers to the need for validation and evaluation of AI models to assess their technical performance and clinical effectiveness before clinical implementation and to ensure that the proposed AI model is aligned with its intended purpose [10]. Using AI-driven tools lacking sufficient validation and/or evaluation could potentially harm patients involved. Morley et al. (2021) have proposed a three-step validation process that consists of internal validation against the test data set, external validation against unseen data and validation against diverse datasets from multiple centres [36]. Yet, this approach does face some challenges as large, diverse, and multi-site datasets may be inaccessible to AI developers due to implemented data privacy and security measures [36]. Furthermore this also requires a common interoperable software framework to enable good data flows across different sites [37]. An additional step that shouldn't be neglected is the evaluation of expected costs compared to standard practices that are in place and their impact on the use of available resources[10].

2. *Regulation, legislation & ethics*

The second pillar in the model of Stogiannos et al. (2023) focuses on regulation, legislation & ethics. To be able to safely use AI-driven tools there is a need for regulations applying to data protection, safety and ethical use of AI models which haven't been standardized in many countries [10]. For example, patients have the right to be informed about when and for what purpose their data will be used through a dynamic process [38]. Furthermore, steps should be taken to ensure data safety, which might be compromised because of cybersecurity issues or updates that impact the algorithm's performance [38, 39]. Accountability & liability issues should also be taken into account. This can include both liability of developers (= product design liability) and/ or health care professionals (=medical malpractice or negligence) [40]. Regulatory standards should also be put in place to eliminate discrimination,

Funded by the European Union

stigmatization and unfair bias [10]. To build trust, AI models should be transparent (or available to interrogation), inclusive and easy to evaluate [10]. Standardized documentation of all development processes should therefore be made available [41]. In addition, the reasoning behind the decision-making process of an AI-model should be explainable to end users (e.g. patients, health care professionals). Explainability should not be confused with interpretability (=the ability of a model to make correct causal associations)[10].

3. *Auditing and quality assurance*

The third pilar focusing on auditing and quality assurance of AI-drive tools. Ongoing procedures to test the performance of an AI algorithm throughout it's life cycle should be put in place to assess any deviation of the intended purpose over time. These procedures should focus specifically on safety, accuracy, bias present in the model and clinical/technical performance. Particular caution should be also present after model updates to mitigate an risks from these updates [10, 42].

4. *Research and innovation*

More prospective research studies are deemed essential to be able to assess and evaluate the real added value of AI in healthcare and to be able to continuously improve [43]. Therefore research and innovation has been included as the fourth pillar of AI governance. It will be essential to strengthen partnerships between academia, clinicians and the industry, while maintaining the impartiality of researchers [44]. Some specific guidelines and checklists have already been developed to increase the quality of conduct and reporting of research studies from the algorithm development to the clinical trial stage [43].

5. *Training of staff*

Training of health care staff on AI principles is also considered to be an essential pillar for responsible and safe AI adoption in the governance model proposed by Stogiannos et al. (2023)[10]. For health care professionals of the future, digital competencies will be paramount and therefore it should be included as a core competence in health care education/training programs. Health care professionals should have knowledge about basis AI principles, validation and evaluation, clinical applications, governance and ethics, regulation, technology implementation and the limitations of AI models so they are confident to use AI-driven tools in a safe and effective way [45]. Providing appropriate education/training to healthcare professionals could in it's turn also increase trustworthiness [46].

6. *Public, patient & practitioner involvement*

The proposed AI governance model also foresees that prospective user, patient and public involvement should be included throughout the entire life cycle [47]. Key stakeholders (e.g. clinicians, patients, hospital administrators, regulatory agencies, etc.) should be asked to provide feedback on the user-friendlyness of interfaces and the accessibility/inclusiveness of specific applications to facilitate effective AI adoption [10].

7. *Effective leadership & staff management*

Finally, effective leadership will be essential to support responsible use of AI-driven tools. Informed and agile senior leadership will play a vital role in identifying and backing AI champions, fostering cultural change, and facilitating knowledge transfer in key practice areas. Additionally, it will be crucial to empower diverse and multidisciplinary decision-making teams to drive progress, rather than relying

Funded by the European Union

solely on specific professionals. This approach ensures a robust foundation for AI implementation, promoting innovation, collaboration, and sustainable growth across healthcare settings [10, 45].

CONCLUSION

Artificial Intelligence is revolutionizing healthcare, offering unprecedented opportunities for improved diagnosis, treatment, and patient care. However, as AI-driven technologies increases rapidly within healthcare systems worldwide, the need for effective governance framework becomes more urgent. While the potential benefits of AI in healthcare are significant, so are the associated risks. Without appropriate governance frameworks in place, these risks could undermine patient safety, data privacy, and the overall trust in AI systems. Recognizing the critical need for AI governance in healthcare, the World Health Organization (WHO) has taken significant steps in formulating recommendations. Within this report we have presented recommendations with regard to several areas of governance made by the WHO and two more specific governance frameworks that have been proposed to govern trustworthy AI in health care. These recommendations cover a wide range of aspects, including data governance, benefit sharing, transparency, accountability, etc. Future oriented, international collaboration and coordination on AI governance for healthcare will be essential to ensure coherent solutions and enable countries to support and learn from each other's experiences. By working together, countries can address the challenges and seize the opportunities presented by AI in healthcare more effectively. Additionally, international collaboration can help ensure that AI governance frameworks are adaptable and scalable to meet the evolving needs of healthcare systems worldwide. In conclusion, this report provides a comprehensive analysis of the academic literature, guidelines, and governance frameworks surrounding AI-based decision-making in healthcare. These insights aim to offer valuable guidance for the development and implementation of the AI-POD project. By aligning the project with established standards, we strive to promote the creation of safe and ethically sound AI-based decision-making tools in healthcare.

Funded by
the European Union

REFERENCES

1. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc. 2020;27(3):491-7.

2. Organization WH. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva2021.

3. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med. 2019;25(1):30-6.

4. Morley J, Murphy L, Mishra A, Joshi I, Karpathakis K. Governing Data and Artificial Intelligence for Health Care: Developing an International Understanding. JMIR Form Res. 2022;6(1):e31623.

5. Guan J. Artificial Intelligence in Healthcare and Medicine: Promises, Ethical Challenges and Governance. Chin Med Sci J. 2019;34(2):76-83.

6. Fan R, Zhang N, Yang L, Ke J, Zhao D, Cui Q. AI-based prediction for the risk of coronary heart disease among patients with type 2 diabetes mellitus. Sci Rep. 2020;10(1):14457.

7. Yan Y, Zhang JW, Zang GY, Pu J. The primary use of artificial intelligence in cardiovascular diseases: what kind of potential role does artificial intelligence play in future medicine? J Geriatr Cardiol. 2019;16(8):585-91.

8. The Topol Review: Preparing the healthcare workforce to deliver the digital future. 2019, https://topol.hee.nhs.uk/.

9. Baig MA, Almuhaizea MA, Alshehri J, Bazarbashi MS, Al-Shagathrh F. Urgent Need for Developing a Framework for the Governance of AI in Healthcare. Stud Health Technol Inform. 2020;272:253-6.

10. Stogiannos N, Malik R, Kumar A, Barnes A, Pogose M, Harvey H, et al. Black box no more: a scoping review of AI governance frameworks to guide procurement and adoption of AI in medical imaging and radiotherapy in the UK. Br J Radiol. 2023;96(1152):20221157.

11. Ho CWL, Soon D, Caals K, Kapur J. Governance of automated image analysis and artificial intelligence analytics in healthcare. Clin Radiol. 2019;74(5):329-37.

12. Allen B. The Role of the FDA in Ensuring the Safety and Efficacy of Artificial Intelligence Software and Devices. J Am Coll Radiol. 2019;16(2):208-10.

13. Pehlivan CN. The EU Artificial Intelligence (AI) Act: An Introduction Global Privacy Law Review. 2024, https://ssrn.com/abstract=4746840.

14. Lehmann LS. Ethical Challenges of Integrating AI into Healthcare. In: Lidströmer N, Ashrafian H, editors. Artificial Intelligence in Medicine;10.1007/978-3-030-58080-3_337-2. Cham: Springer International Publishing; 2020. p. 1-6.

15. Astromskė K, Peičius E, Astromskis P. Ethical and legal challenges of informed consent applying artificial intelligence in medical diagnostic consultations. AI & SOCIETY. 2021;36(2):509-20.

16. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310.

17. Beil M, Proft I, van Heerden D, Sviri S, van Heerden PV. Ethical considerations about artificial intelligence for prognostication in intensive care. Intensive Care Med Exp. 2019;7(1):70.

18. Liu Y, Qin C, Liu C, Liu J, Jin Y, Li Z, et al. Multiple high-regional-incidence cardiac disease diagnosis with deep learning and its potential to elevate cardiologist performance. iScience. 2022;25(11):105434.

19. Mathur P, Srivastava S, Xu X, Mehta JL. Artificial Intelligence, Machine Learning, and Cardiovascular Disease. Clin Med Insights Cardiol. 2020;14:1179546820927404.

20. Adedinsewo DA, Pollak AW, Phillips SD, Smith TL, Svatikova A, Hayes SN, et al. Cardiovascular Disease Screening in Women: Leveraging Artificial Intelligence and Digital Tools. Circ Res. 2022;130(4):673-90.

21. O'Reilly-Shah VN, Gentry KR, Walters AM, Zivot J, Anderson CT, Tighe PJ. Bias and ethical considerations in machine learning and the automation of perioperative risk assessment. Br J Anaesth. 2020;125(6):843-6.

22. Organization WH. WHO European Regional Obesity Report 2022. 2022.

23. Gauci S, Cartledge S, Redfern J, Gallagher R, Huxley R, Lee CMY, et al. Biology, Bias, or Both? The Contribution of Sex and Gender to the Disparity in Cardiovascular Outcomes Between Women and Men. Curr Atheroscler Rep. 2022;24(9):701-8.

Funded by the European Union

24. Phelan SM, Burgess DJ, Yeazel MW, Hellerstedt WL, Griffin JM, van Ryn M. Impact of weight bias and stigma on quality of care and outcomes for patients with obesity. Obes Rev. 2015;16(4):319-26.

25. Puhl R, Suh Y. Health Consequences of Weight Stigma: Implications for Obesity Prevention and Treatment. Curr Obes Rep. 2015;4(2):182-90.

26. Rubino F, Puhl RM, Cummings DE, Eckel RH, Ryan DH, Mechanick JI, et al. Joint international consensus statement for ending stigma of obesity. Nature Medicine. 2020;26(4):485-97.

27. Schwartz MB, Chambliss HO, Brownell KD, Blair SN, Billington C. Weight bias among health professionals specializing in obesity. Obes Res. 2003;11(9):1033-9.

28. Olson KL, Goldstein SP, Lillis J, Panza E. Weight stigma is overlooked in commercial-grade mobile applications for weight loss and weight-related behaviors. Obes Sci Pract. 2021;7(2):244-8.

29. Nebeker C, Torous J, Bartlett Ellis RJ. Building the case for actionable ethics in digital health research supported by artificial intelligence. BMC Medicine. 2019;17(1):137.

30. Zhang Y, Wu M, Tian GY, Zhang G, Lu J. Ethics and privacy of artificial intelligence: Understandings from bibliometrics. Knowledge-Based Systems. 2021;222:106994.

31. Müller VC. Ethics of Artificial Intelligence and Robotics. In: Adamson P, editor. Stanford Encyclopedia of Philosophy: Stanford Encyclopedia of Philosophy; 2012. p. 1-70.

32. Corrêa NK, Galvão C, Santos JW, Del Pino C, Pinto EP, Barbosa C, et al. Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. Patterns. 2023;4(10).

33. Zhang DT, Mishra S, Brynjolfsson E, Etchemendy J, Ganguli D, Grosz B, et al. The AI Index 2021 Annual Report. ArXiv. 2021;abs/2103.06312.

34. Hagendorff T. The Ethics of AI Ethics: An Evaluation of Guidelines. Minds and Machines. 2020;30(1):99-120.

35. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nature Machine Intelligence. 2019;1(9):389-99.

36. Morley J, Morton CE, Karpathakis K, Taddeo M, Floridi L. Towards a framework for evaluating the safety, acceptability and efficacy of AI systems for health: an initial synthesis. ArXiv. 2021;abs/2104.06910.

37. Tang A, Tam R, Cadrin-Chênevert A, Guest W, Chong J, Barfett J, et al. Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology. Can Assoc Radiol J. 2018;69(2):120-35.

38. Akinci D'Antonoli T. Ethical considerations for artificial intelligence: an overview of the current radiology landscape. Diagn Interv Radiol. 2020;26(5):504-11.

39. Cohen IG, Evgeniou T, Gerke S, Minssen T. The European artificial intelligence strategy: implications and challenges for digital health. Lancet Digit Health. 2020;2(7):e376-e9.

40. Maliha G, Gerke S, Cohen IG, Parikh RB. Artificial Intelligence and Liability in Medicine: Balancing Safety and Innovation. Milbank Q. 2021;99(3):629-47.

41. Lekadir K, Osuala R, Gallin CS, Lazrak N, Kushibar K, Tsakou G, et al., editors. FUTURE-AI: Guiding Principles and Consensus Recommendations for Trustworthy Artificial Intelligence in Future Medical Imaging2021.

42. Allen B, Dreyer K, Stibolt R, Agarwal S, Coombs L, Treml C, et al. Evaluation and Real-World Performance Monitoring of Artificial Intelligence Models in Clinical Practice: Try It, Buy It, Check It. Journal of the American College of Radiology. 2021;18(11):1489-96.

43. Excellence NIfHaC. Evidence standards framework (ESF) for digital health technologies [Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies.

44. innovation Cfdea. The roadmap to an effective AI assurance ecosystem 2021 [Available from: https://www.gov.uk/government/publications/the-roadmap-to-an-effective-ai-assurance-ecosystem.

45. NHS. Understanding healthcare workers' confidence in AI. 2022, https://digital-transformation.hee.nhs.uk/binaries/content/assets/digital-transformation/dart-ed/understandingconfidenceinai-may22.pdf.

46. Mirbabaie M, Hofeditz L, Frick NRJ, Stieglitz S. Artificial intelligence in hospitals: providing a status quo of ethical considerations in academia to guide future research. AI Soc. 2022;37(4):1361-82.

Funded by the European Union

47. (UK) DoHaSC. A guide to good practice for digital and data-driven health technologies. 2021, https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology.

Funded by
the European Union