Ethical Dilemmas and Regulatory Landscape in Healthcare AI

Yasaman Yousefi[1], Daniel Karpati[2], Christoph Schommer[3]

Abstract

The integration of artificial intelligence (AI) into the healthcare sector presents transformative opportunities amidst significant challenges faced by global healthcare systems, including an ageing population, a shortage of healthcare personnel, and pervasive inefficiencies. AI technologies have demonstrated potential in various domains, such as diagnostics and prognostics, personalised medicine, and increasing operational efficiency. However, the deployment of AI in healthcare also introduces substantial ethical, societal, and regulatory challenges. Issues such as data privacy, algorithmic bias, transparency, and accountability are paramount. In the European Union (EU), the regulatory landscape is evolving, with the Artificial Intelligence Act (AI Act) and the Medical Devices Regulation (MDR) creating a complex environment for developers and deployers of AI in healthcare. This work, through the lens of different levels of abstraction, examines the opportunities and challenges of the integration of AI in healthcare, as well as the ethical and regulatory challenges we face in the EU.

Keywords: Healthcare, Ethics of AI, AI Act, Medical Device Regulation

Introduction: Opportunities and Challenges of Healthcare AI

Healthcare systems worldwide are grappling with a multitude of pressing challenges that threaten their efficiency and effectiveness. Among these are the rapid increase in the ageing population (Cristea, M. et al, 2020), a critical shortage of healthcare personnel (World Health Organization, 2022), pervasive inefficiencies within healthcare delivery systems (Behr, A. & Theune, K., 2017), and systemic inequities in access to care (Mackenbach, J. P., 2006; Baeten, R. et al, 2018). Each of these issues compounds the others, creating a complex web of obstacles that impede the provision of optimal healthcare.

This is where artificial intelligence (AI) comes into play, a technology that is at a critical juncture and holds great promise to transform the efficiency, cost, and delivery of healthcare services globally.

[1] University of Bologna, University of Luxembourg Yasaman.yousefi3@unibo.it
[2] University of Luxembourg. Daniel.karpati@uni.lu
[3] University of Luxembourg. Christoph.schommer@uni.lu

The COVID-19 pandemic has underscored the potential of AI in healthcare. AI algorithms have been instrumental in quickly diagnosing COVID-19 by synthesising chest CT images and laboratory data (Mei et al., 2020). Furthermore, AI-driven screening processes have accelerated the development of antiviral drugs and predicted potential future viruses, showcasing AI's flexibility and rapid responsiveness in crisis situations (Ke et al., 2020). AI's utility extended to public spaces such as airports and schools, where thermal scanners equipped with body and facial recognition technologies identified individuals with elevated temperatures. Additionally, AI-generated *health rapid response codes* considered individuals' health status, travel history, and contact patterns to categorise populations during the pandemic (Sharara, S. & Radia, S., 2022). Such systems were pivotal in formulating targeted anti-epidemic measures and significantly curtailed the virus's spread.

Integrating genomic and health data can further enhance AI's precision in managing public health crises. Personalised risk assessments based on genetic predispositions and pre-existing conditions can refine AI's predictive capabilities, enabling more effective interventions and resource allocation during pandemics. This holistic approach not only improves the efficiency of monitoring systems but also aligns with the broader goal of individualised medicine (Johnson et al., 2021).

On a broader scale, AI facilitates personalised medicine by analysing extensive patient data to tailor treatments to individual needs. Algorithms can predict patients' responses to different treatments, allowing for customised therapy plans that maximise efficacy and minimise side effects. The use of AI with genomics enables a deeper understanding of genetic variations and their impacts on diseases, fostering more personalised healthcare strategies (Collins & Varmus, 2015).

Operational efficiencies are also enhanced through AI applications. Systems like IBM Watson streamline administrative tasks such as scheduling, billing, and patient data management. This not only reduces costs but also allows medical staff to devote more time to direct patient care (Fichman et al., 2011).

In diagnostics, AI has demonstrated remarkable capabilities, particularly through deep learning algorithms in image-based diagnosis. Google DeepMind's algorithm for diagnosing eye diseases, for example, has matched the accuracy of top specialists in identifying conditions such as diabetic retinopathy and age-related macular degeneration (De Fauw et al., 2018). These advancements enhance diagnostic accuracy and expedite the diagnostic process, allowing for quicker intervention and treatment.

However, AI-driven diagnoses are not always entirely accurate and error-free. The quality of training data and their representativeness play crucial roles in the performance of AI models (Kilkenny & Robinson, 2018). If the training data are biased or lack diversity, the AI system may exhibit inaccuracies and fail to generalise well across different population groups, leading to disparities in healthcare outcomes (Gianfrancesco et al., 2018; Parikh et al., 2019).

The implementation of AI in healthcare is fraught with complexities, including potential algorithmic challenges and issues of generalisation. The intricacy of medical conditions and the variability in their manifestation across different patients further compound these challenges. Kelly et al. (2019) underscore the need for robust clinical evaluation, comprehensible performance metrics, effective regulatory implementation, and vigilant post-market surveillance. They caution about the risks of algorithmic bias and unfairness and advocate for efforts to identify and mitigate these issues. Enhancing the generalizability and interpretability of machine learning predictions is crucial. Continuous monitoring, validation, and updating of AI models with new and diverse data are recommended to bolster the robustness of AI diagnostics.

Acknowledging that the integration of AI in healthcare brings transformative benefits, such as minimising errors and alleviating the administrative burden on healthcare providers, it should also be noted that it introduces several societal, legal, and ethical challenges. One of the most pressing concerns is data privacy, security, and safety. AI systems require access to extensive personal and medical data, making them potential targets for cyberattacks. Robust discussions and measures are essential to ensure the protection of this sensitive information (Mennella et al., 2024).

Moreover, the ethical implications of AI in healthcare are significant and multifaceted. Questions arise regarding the autonomy of patients and medical professionals when AI-driven decisions are involved (Laitinen & Sahlgren, 2021). Concerns also exist about accountability and informed consent (Luxton, 2014). These systems, while beneficial, must be designed to operate transparently and with explainability, ensuring that patients and practitioners understand how and why decisions are made (Shin, 2020). Additionally, there is a risk of AI perpetuating existing biases found in the training data. Studies have highlighted instances where AI systems exhibited racial bias in treatment recommendations (Obermeyer et al., 2019), underscoring the necessity for diverse datasets and algorithms that can detect and correct biases to ensure fairness and accessibility.

These challenges are not merely technical but also ethical and legal, relating to broader concerns about patient privacy, data protection, fairness, accessibility, and liability. In response to these challenges, there is a growing discourse around AI ethics, which aims to guide the development and implementation of AI technologies in a manner that upholds ethical standards. This movement reflects a collective endeavour to harness the benefits of AI in medicine while safeguarding against its potential harms, ensuring that AI serves as a tool for enhancing healthcare equity and quality without compromising ethical values.

This work will explore, in its first section, the ethical landscape of healthcare AI by raising critical and provocative questions through different hypothetical scenarios. In the same section, different levels of abstraction to study such ethical dilemmas, namely the individual, systemic, and inter-systemic levels, will be explored.

In the second section, the transition from bioethical principles to the ethics of AI, and the further transition of those ethical principles to the existing and forthcoming regulatory frameworks in the EU will be studied. Through critical thinking, this section will explore the complex dynamics and interconnections of the Artificial Intelligence Act (EU AI Act) and the Medical Device Regulation (MDR). By showcasing the legal complexities in the world of healthcare AI, the section offers a legal exercise of double compliance through a hypothetical case. The analysis reveals a significant regulatory gap in the EU, where the absence of precise guidelines on navigating AI-related legislation and the rapid evolution of AI creates ongoing governance challenges. This underscores the need for the EU to refine its digital and AI regulatory frameworks to maintain global competitiveness and effectively harness AI's potential in healthcare.

Navigating the Ethical Landscape of Healthcare AI

We have said so far that the integration of AI into healthcare systems offers transformative possibilities, ranging from improved diagnostic accuracy to personalised treatment protocols. However, we also explored that alongside these advancements, AI integration introduces a spectrum of risks and ethical considerations.

Both benefits and risks emerge in various layers and affect multiple stakeholders. As a result, the integration of AI technologies into healthcare systems should be analysed from various levels to provide an optimal and harmonious environment for such integration.

In this approach, there are three Levels of Abstraction (LoAs) through which these benefits and risks can be explored and compared (Floridi, L., 2008).

i) Issues related to the individuals participating in the healthcare system. We call this the individual level.

ii) Issues related to the healthcare providers and the healthcare system itself. We call this the systemic level.

iii) Finally, issues related to the healthcare system's relation to its broader environment, be that on the national, the supernational, or the international level. These dilemmas occur on the inter-systemic level.

Below, we explore the possible ethical dilemmas under each LoA through hypothetical scenarios.

A main concern at the individual level regards data privacy. Personal medical data are considered sensitive, and the need to protect patient privacy is widely recognised. Individuals have a vested interest in ensuring their medical history and related data remain confidential, as sharing such data could put them in vulnerable positions (Abouelmehdi, K. et al., 2018). When handling anonymised or pseudonymised medical data to build computational models, we enter more complex scenarios that shift the focus from data privacy at the individual level.

Consider a hypothetical AI system designed to recognise cancer. From the individual's perspective, one issue is whom to trust more: a human doctor or an AI system (LaRosa, E. & Danks, D., 2018). Several questions arise in this regard: Is it possible to recognise human bias, competence, and work ethics? Can we detect if the medical professionals are overworked, overconfident, or superficial? Do we trust the educational system to train competent doctors? Ultimately, if healthcare professionals disagree with the prediction of an AI model trained on a much larger dataset, which opinion is to be trusted more?

Other issues may emerge in relation to trust. First, for patients to trust AI systems' predictions, these systems should be explainable and provide clear reasoning for their diagnoses. Further, the systems should be robust and aim for high accuracy[4], and they should be trained on fair and inclusive datasets in order to have a good performance for every individual.

If we only considered the individual level, it would not be so difficult, at least in theory, to define such rules for this hypothetical AI model, but in practice this can be a complicated task. Most of the high-level AI 'white papers' broadly articulate these wishes as demands to have a safe and beneficial AI, without providing clear guidelines.

From the perspective of healthcare providers, additional concerns arise, particularly at the systemic level where decisions are both statistical and ethical. Measuring the success rate of physicians to establish a benchmark for AI in diagnosis or prognosis is challenging but not impossible. For instance, reliable data on false-negative and false-positive diagnoses in cancer diagnostics is available in many regions, though it can vary by type of cancer, demographic, and other factors. Cancer registries and health informatics systems provide significant data that can be used to track performance metrics, but gaps still exist.

Often, we must rely on estimates and assumptions due to incomplete data. Without precise metrics on human performance, setting an appropriate benchmark for AI systems is difficult (Babushkina, D., 2023). This issue is significant because the impact of diagnoses extends beyond

[4] It is worth mentioning that an error rate of nearly zero might indicate overfitting, leading to poorer performance on new data.

individual accuracy rates to broader community implications, such as public health outcomes and healthcare equity.

Establishing benchmarks for AI requires robust, scientifically accurate data to ensure these systems can effectively complement or enhance human performance. Comprehensive studies comparing human and AI diagnostic accuracy do exist (Oakden-Rayner, L., et al, 2022 ; Fonseca, Â., et al, 2024), but the variability in data quality and availability poses challenges. Thus, while substantial progress has been made, there is still a need for ongoing improvement in data collection and reporting to fully leverage AI in healthcare.

At the systemic level, ethical dilemmas are abundant, and justifying one set of preferences over another is challenging. Consider a scenario where a hypothetical AI model demonstrates higher accuracy for white individuals than for black individuals due to underrepresentation in the training dataset.[5] The state, as the primary healthcare provider (as in the EU), encounters a dilemma: Should the implementation of this AI model be delayed until it can provide equally accurate results for both white and black individuals? The answer to this question depends on whether the state involved prioritises utilitarian principles over principles of individual rights.

Using this AI tool is problematic because it is racially biased, being less accurate for black individuals. Nevertheless, the technology is affordable and increases access to diagnostics. If integrated into universal healthcare, it could reduce waiting times and costs. The state can potentially deploy the AI system to leverage its economic and healthcare benefits, ensuring transparency about the system's limitations and potential biases. By openly communicating these issues and actively pursuing mitigation strategies, such as prioritising screenings for historically underserved groups and continuously improving the AI model's accuracy across all demographics, the government can maximise the benefits of the technology while addressing ethical and fairness concerns.

[5] Real-life examples of AI systems showing racial, or gender bias exist and usually mirror the systemic inequities of society. See, for instance (Parikh, R. B. et al., 2019; Buolamwini, J., & Gebru, T. 2018; Cirillo, D., et al., 2020).

Statistically, economically disadvantaged black individuals are more affected and less likely to receive preventive screenings through traditional methods (Ward, E., et al., 2004 ; Sambamoorthi, U., & McAlpine, D. D., 2003).

If a hypothetical government decides to deploy the AI model despite its bias, this results in improved statistics and timely diagnosis for black individuals. The AI serves as a pre-selection tool, enabling a broader pool of patients to be screened, with those marked 'positive' proceeding with traditional diagnostic methods.

While many lives can be saved, the prevalence of false negatives among black individuals creates a subgroup with near-zero chances of receiving a diagnosis. These individuals are incorrectly labelled as healthy and could even get at the bottom of the waiting list due to incorrect labelling, a particularly troubling issue leading to further systemic discrimination.

From a utilitarian perspective, deploying the AI model offers significant economic benefits despite its ethical issues. The primary economic advantage lies in the model's ability to reduce healthcare costs by streamlining diagnostics, thereby reducing the need for expensive human labour in preliminary screenings. This cost-saving can be redirected to other critical areas within the healthcare system, potentially improving overall healthcare quality and accessibility. The trade-off involves balancing these economic benefits against the need for accuracy, fairness, explainability, and transparency.

The AI's higher accuracy for white individuals over black individuals poses a fairness issue, as it may perpetuate existing healthcare disparities. Additionally, the lack of explainability and transparency in the AI's decision-making process, often referred to as the *black box* problem (Castelvecchi, D., 2016), complicates trust and accountability.

Examining this hypothetical example in its broader context, namely the inter-systemic level, adds further complexity. Up to this point, our discussion has revolved around the assumption of an anonymised dataset serving as the foundational basis for training AI algorithms. While we have explored potential challenges posed by a hypothetical diagnostic AI tool, we have yet to delve into the entity responsible for its development and the management of the underlying dataset.

Let us consider a scenario where a private company, designated as Company A, has collected, anonymized, and stored a dataset containing sensitive medical information pertaining to a specific population. This dataset holds the potential to train models for identifying a particular type of cancer. Consequently, questions regarding data ownership and governance come to the forefront.

On one hand, there are compelling benefits to sharing such datasets. Should Company A opt to sell its dataset to Company B, which possesses its own dataset, the integration of these datasets could lead to improved AI-models. This amalgamation has the potential to enhance diagnostic accuracy and unveil significant correlations. As companies merge and datasets accumulate, the field of diagnostics and treatment may witness substantial advancements, potentially ushering in a revolution in healthcare (Raghupathi & Raghupathi, 2014). The prevention of such accumulation could impede innovation and postpone the discovery of cures for complex diseases such as dementia and Parkinson's.

On the other hand, centralising medical data poses risks that could render the population vulnerable (Jha, A. 2023). The potential misuse of such data to create bioweapons underscores the complexity of predicting the scope and nature of potentially harmful medical data. (Egan & Rosenbach, 2023). Furthermore, data localization issues introduce additional complications, as medical datasets are frequently delocalized, complicating regulatory oversight. Legal restrictions on data accumulation may prove insufficient in preventing malicious use, particularly given the lower standards for developing harmful AI applications compared to beneficial ones.

Moreover, data centralisation fosters dependency. For instance, a large international conglomerate that accumulates vast medical datasets could wield significant influence in healthcare advancement, potentially shifting power dynamics away from nation-states towards a select few dominant companies. This prompts the question of whether it is feasible to regulate data centralisation without inadvertently consolidating power. While the state, by virtue of its privileged position in healthcare data collection, could theoretically manage datasets more effectively, political realities and concerns regarding privacy and trust may impede such initiatives.

Additionally, the challenge of global cooperation arises. Addressing both the benefits and risks of medical data centralisation may necessitate international alliances or deeper cooperation within existing frameworks, such as the European Union.

The deployment of AI technologies in healthcare brings forth significant ethical and societal challenges at individual, systemic, and inter-systemic LoAs. Regardless of whether a regulatory framework favours corporate-led AI innovation or a public institution-controlled approach, it is imperative to avoid dependence without adequate oversight. Implementing checks and balances, distinguishing between data concentration and power concentration, and proactively addressing the risks associated with malevolent AI use are essential steps in navigating this intricate landscape.

The next section will demonstrate the transition of ethical principles to legal norms. It will explore through critical thinking, the complex dynamics, and interconnections of the EU AI Act MDR. By showcasing the legal complexities in the world of healthcare AI, the section offers a legal exercise of double compliance through a hypothetical case.

Navigating the Regulatory Challenges of Healthcare AI in the EU

In the European Union, ethical guidelines emphasise the development of AI systems free from bias, ensuring fair treatment for all individuals regardless of race, gender, or socioeconomic status. The European Commission's ethical guidelines for trustworthy AI highlight the importance of mitigating biases in AI training datasets and algorithms to prevent discriminatory outcomes (European Commission, 2019).

These guidelines also address the protection of patient data privacy and confidentiality, which are paramount in healthcare. The General Data Protection Regulation (GDPR) in the European Union sets a benchmark for personal data privacy, including strict rules on data consent and anonymization before AI processing. These regulations ensure that AI systems respect and uphold individuals' privacy rights (European Parliament and Council of the European Union, 2016).

Transparency and accountability are crucial aspects of ethical AI in healthcare. AI systems should be transparent in their operations and decision-making processes, with mechanisms in place to hold developers and users accountable. The IEEE's Ethically Aligned Design principles stress the importance of transparency to facilitate scrutiny and ethical assessment of AI technologies (IEEE, 2019).

Floridi and Cowls have developed an ethical framework for AI based on the core bioethical principles of beneficence, non-maleficence, autonomy, and justice, as outlined by Beauchamp and Childress (2012). Floridi (2013) argues that bioethics closely parallels digital ethics in addressing the ecological impacts of new types of agents, patients, and environments.

To this foundation, Floridi and Cowls add the principle of explicability, which includes both intelligibility and accountability. Intelligibility ensures that AI systems are understandable to non-experts, such as patients and business customers, as well as experts like product designers and engineers. Accountability ensures that those who design, deploy, and manage AI systems can be held responsible for their actions and decisions (Floridi & Cowls, 2019).

The High-Level Expert Group on Artificial Intelligence (HLEG), established by the European Commission, has outlined seven key principles for ethical AI: Human Agency and Oversight, Technical Robustness and Safety, Privacy and Data Governance, Transparency, Diversity, Non-discrimination, and Fairness, Societal and Environmental Well-being, and Accountability. These principles aim to guide the development and deployment of AI systems in a manner that is trustworthy and beneficial for society (High-Level Expert Group on Artificial Intelligence, 2019).

Building on these principles, the EU sets its regulatory foundation for governing AI systems, ensuring that they are developed and deployed ethically and responsibly.

First and foremost, there is the landmark regulation on laying down harmonised rules on Artificial Intelligence (EU AI Act). Focusing on high-risk applications, including those in healthcare, the AI Act classifies AI systems according to the risk they pose to safety and fundamental rights, imposing stricter requirements on high-risk applications to ensure they are transparent, traceable, and guarantee a high level of data protection.

As of now, together with the AI Act, the regulatory landscape for medical AI tools in the EU is governed by the 2017/745 Medical Devices Regulation (MDR) and the 2017/746 In Vitro Diagnostic Medical Devices Regulation (IVDR). These regulations mandate rigorous pre-market controls, enhanced clinical investigation requirements, strengthened surveillance throughout the device's lifecycle, and increased transparency through the creation of a European database for medical devices. However, they fall short in addressing specific challenges related to AI, such as the continuous learning capabilities of AI models and the detection of algorithmic biases. The adaptive nature of AI, which evolves as it processes more data, necessitates innovative approaches to continuously monitor and manage the associated risks.

The risks associated with AI can be analysed and classified based on the potential severity and frequency of the harm they might induce. In the healthcare sector, the spectrum of AI risks varies considerably. Some risks are infrequent and manageable, causing only limited harm to patients and healthcare systems, such as an AI tool that inaccurately delineates the boundaries of the heart in a cardiac image, requiring manual correction by a cardiologist. However, other risks are more severe, potentially leading to irreversible damage, such as an AI tool failing to diagnose a life-threatening condition, which could have dire consequences on patient health and clinical outcomes.

To effectively mitigate these risks while maximising the benefits of AI in healthcare, it is crucial to identify, analyse, understand, and monitor potential risks on a case-by-case basis for each new AI algorithm and application, from the design to deployment phases. A systematic risk assessment procedure should be established, classifying identified risks into categories that reflect different levels and types of risk. For example, in the design of the AI algorithm, the provenance, quality, representativeness and reliability of training data should be assessed to identify any possible issue. Additionally, the potential for algorithmic bias, data security vulnerabilities, and ethical concerns related to patient privacy should be evaluated.

For each risk category, appropriate tests and regulatory measures should be specified. Higher risk classes, such as those involving direct patient diagnosis or treatment recommendations, would require more stringent testing and regulation, including extensive clinical trials and continuous post-deployment surveillance. Lower risk categories, such as administrative support functions, might necessitate less intensive mitigation measures, like routine audits and compliance checks. Implementing a suitable risk classification system will enable manufacturers, healthcare providers, and regulators to intervene appropriately to safeguard patient safety and rights without unnecessarily stifling innovation. For instance, a high-risk AI system used in surgical decision-making would undergo rigorous validation and monitoring, whereas an AI tool designed for scheduling patient appointments might only require standard quality assurance procedures. This balanced approach ensures that patient care remains safe and effective while fostering the development and implementation of innovative AI technologies in healthcare.

The AI Act does not specifically tailor its provisions to AI in healthcare but indicates that AI-driven medical devices are likely to be classified as high-risk due to significant safety and privacy concerns. This classification means that future medical AI tools will need to comply not only with the existing requirements of the MDR, but also with additional stipulations outlined in Chapter II of the AI regulation. These include the use of high-quality and representative data, comprehensive technical documentation and traceability, transparency requirements, human oversight, a quality management system, and thorough conformity assessments.

However, the classification of all medical AI tools as high-risk may not be entirely appropriate. For instance, numerous AI applications in radiology aim to expedite tasks like the contouring of organs and lesions on medical images—a process crucial for subsequent quantification and diagnosis, such as outlining the cardiac ventricles or lung tumours. These AI-powered tools, while integral to clinical practice and enhancing efficiency, might not inherently demand the same level of transparency as other AI applications. Clinicians are typically able to visually verify and correct the outputs of such tools, thereby mitigating the risks associated with their use. Consequently, there is a valid argument for a more nuanced classification system within the regulatory framework that distinguishes between low- and high-risk AI applications in healthcare. This would support ongoing innovation and investment in the sector without unnecessarily stifling technological advancement with stringent regulations.

As manufacturers develop AI-based medical devices, they encounter an important question: What regulatory obligations must their products meet to leverage AI effectively? And how do these regulations interact with one another?

This query becomes increasingly complex when considering the intersecting regulatory frameworks, notably Regulation (EU) 2017/745 on medical devices (MDR) and Regulation (EU)

2017/746 on in vitro diagnostic medical devices (IVDR), alongside the European Regulation introducing harmonised standards for artificial intelligence (AI Act). It has been pointed out that the AI Act and the MDR may be overlapping, and in some cases inconsistent, leading to necessity for double compliance.

In order to clarify how these regulations may interact, let us consider a hypothetical example of a medical device. A well-known use of AI in cancer diagnosis are image recognition technologies. Let us consider an AI-enabled Diagnostic Imaging System, with the purpose to analyse and interpret medical images of patients. This device incorporates AI algorithms that assist radiologists in identifying anomalies, lesions, or abnormalities in the images.

Article 3 of the AI Act defines 'AI system' as "a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments;" In our case, the system falls under said definition.

Under the AI Act Article 6.1, the AI-enabled Diagnostic Imaging System shall be considered to be high-risk where both of the following conditions are fulfilled:

"(a) (…) the AI system is itself a product, covered by the Union harmonisation legislation listed in Annex I;

(b) (…) the AI system itself as a product, is required to undergo a third-party conformity assessment, with a view to the placing on the market or the putting into service of that product pursuant to the Union harmonisation legislation listed in Annex I."

Annex I offers a list of the Union harmonisation legislations, amongst which the MDR.

The AI-enabled Diagnostic Imaging System can be defined as a medical device under Article 2 of the MDR, as it is intended to be used for diagnosing or monitoring. This means that if also the second condition regarding a conformity assessment outlined in Article 6.1 (b) of the AI Act is fulfilled, the system is considered as high risk under the AI Act.

For understanding whether or not the system needs to go through a conformity assessment, we look at the MDR again. Article 52 of the MDR sets the conformity assessment procedures, for which the risk classification of the system has to first be determined. Article 51 classifies systems into I, IIa, IIb and III classes, taking into account the intended purpose of the devices and their inherent risks, in accordance with Annex VIII. According to Rule 10 of Annex VIII, the AI-enabled

Diagnostic Imaging System is a class IIa, as a device "intended for direct diagnosis or monitoring of vital physiological processes", and if it is "used for diagnosis in clinical situations where the patient is in immediate danger", it is classified as Class IIb. Let us consider the first case as a hypothetically valid classification.

Under article 52.4 of the MDR, "Manufacturers of class IIa devices (…), shall be subject to a conformity assessment as specified in Chapters I and III of Annex IX, and including an assessment of the technical documentation as specified in Section 4 of that Annex (…)."

This legal exercise may seem overcomplicated, and intentionally so. The EU regulatory attempts to govern AI include lots of pinpointing instances, just as witnessed in our hypothetical case. There are many conditions to evaluate, addressed in not a very straightforward or clear way, leading to legal syllogism and uncertainty.

In a potential attempt to overcome these issues, going back to the AI Act, according to Article 8.2, providers must make sure their product follows the rules in both MDR and AI Act:

"In ensuring the compliance of high-risk AI (…) with the requirements set out in this Section, and in order to ensure consistency, avoid duplication and minimise additional burdens, providers shall have a choice of integrating, as appropriate, the necessary testing and reporting processes, information and documentation they provide with regard to their product into documentation and procedures that already exist and are required under the Union harmonisation legislation listed in Section A of Annex I."

However, it should be noted that on the one hand the conformity assessment under the MDR in itself is complicated, and on the other hand, the fundamental rights-oriented focus of the AI Act is not reflected in the MDR, as the latter is a more technical regulation.

For instance, Annex IV of the AI Act regarding technical documentation referred to in Article 11(1), requires providers to include: detailed information about the foreseeable unintended outcomes and sources of risks to health and safety, fundamental rights and discrimination.

In the meantime, the MDR merely sets requirements for the technical functions of the devices for its conformity assessment, without for instance considering discrimination in a fundamental rights perspective.

This takes us to think back at the levels discussed in the previous section. Namely the individual, systemic and inter-systemic levels.

Going back to our example of an AI-enabled image diagnosis system, let us imagine a scenario where the system shows suboptimal performances in its diagnosis. For example, consider that the system discriminates against a certain group of people sharing certain characteristics, due to the uneven representation of society in data. For example, the system may have a lower accuracy for transgender women due to the lack of representation in medical data.

Discrimination is problematic at the individual level, as it threatens a fundamental right, and individuals' trust in the healthcare system. In this case, which regulation prevails in protecting the individual? And what redress mechanisms does the regulation offer?

Of course, if the system's provider has done a conformity assessment under the MDR, and their system is technically robust and accurate, there is no further investigation needed, even though the system is discriminatory, they comply with the AI Act automatically. This means that compliance with the harmonisation legislation will trigger presumption of compliance with the AI Act. However, if the system proved to be discriminatory, there would be no compliance anymore. An interesting question to be asked here is what tools- if any- do individuals have to actually contest compliance?

Or imagine that the system works perfectly accurately 98% of the time, but for a certain individual, it gives the wrong diagnosis, for instance predicting that the individual does not have cancer, while they do, without any discrimination, just because it is a false negative. From a technical perspective, the system is functioning very well under the MDR, but from an ethics-oriented and AI Act approach, the conformity assessment is not complete (even though the AI Act deems it to be).

At the systemic level too, the co-existence of the AI Act and MDR may not be sustainable. Of course, we acknowledge that each of these regulations are thought to govern different things, and in some cases, the same things from a different level of abstraction. However, there is a clear lack of guidance on how to approach compliance in certain cases such as the example mentioned. The excessive burdens of compliance with the regulations can be discouraging. There is a risk that the EU is being left behind in the pace of innovation, as together with existing privacy requirements, manufacturers of AI systems must comply with a broad regulatory environment.

Finally, on the inter-systemic level, and the healthcare system's relation to its broader environment, be that on the national, supernational or the international level, the dilemma is trades and relationships. If EU-based manufacturers are discouraged due to over-regulation and legal uncertainty, AI systems will not be manufactured within the EU. In that case, who will be the providers of AI systems and what would compliance look like then?

Concluding Remarks and Future Research

In these sections, we learned about the vast potential of AI to revolutionise healthcare, offering significant improvements in diagnostics, personalised medicine, and operational efficiency. We also pointed out that the integration of AI into healthcare systems presents complex challenges that must be addressed to harness its benefits fully. Ethical concerns, such as data privacy, algorithmic bias, and the need for explainability, transparency and accountability, are critical. From the readings of different existing legal frameworks, we realised that one of the primary challenges in governing AI used in healthcare in the EU revolves around the absence of precise legislation. The dynamic nature and swift evolution of AI present ongoing challenges in terms of regulatory frameworks. In other words, regulations usually lag behind technological innovation due to the pacing problem of law. Current regulations often prove insufficient in adequately addressing the complexities inherent in AI applications within healthcare. This lack of clear and detailed legislation fosters uncertainty, potentially impeding innovation and deterring investments due to concerns regarding compliance and liabilities. To address these challenges, the EU is actively working to regulate AI, including its application in healthcare. However, this approach has been criticised for hampering innovation and not progressing at the global level.

Another significant challenge in AI governance is the reliance of AI applications on extensive sets to optimise their performance. In this context, the EU places a strong emphasis on the protection of personal data through the General Data Protection Regulation (GDPR). However, effectively balancing the utilisation of AI's potential with ensuring robust data protection continues to be a persistent challenge.

As AI technology evolves, so must regulatory frameworks. These mechanisms need to be flexible and adaptive to address new ethical challenges that arise with technological advancements. Moving forward, it is crucial to refine these regulatory frameworks to avoid unnecessary burdens on AI developers and healthcare providers. Ensuring that the regulations are not overly restrictive will help maintain the EU's competitiveness in the global AI landscape. By striking the right balance between innovation and regulation, AI can be a powerful tool to enhance healthcare delivery, improve patient outcomes, and address some of the most pressing challenges facing healthcare systems worldwide.

Finally, it is worth mentioning that the European Digital Strategy, which aims to make the EU a global leader in digital innovation, has not yet been fully realised. The slow progress in harmonising AI regulations and addressing the specific needs of healthcare AI indicates that there is still much work to be done. The strategy's goals of fostering a robust digital infrastructure and creating a competitive digital economy are hampered by the current regulatory bottlenecks. Such regulatory approaches are far less strict in the USA and China, with a more utilitarian view. Therefore, it is imperative for the EU to accelerate its efforts in refining its digital strategy and AI regulatory frameworks to fulfil its ambition of being at the forefront of digital transformation and innovation.

References

Abouelmehdi, K., Beni-Hessane, A., & Khaloufi, H. (2018). Big healthcare data: preserving security and privacy. *Journal of big data*, *5*(1), 1-18.

Babushkina, D. (2023). Are we justified attributing a mistake in diagnosis to an AI diagnostic system?. *AI and Ethics*, *3*(2), 567-584.

Baeten, R., Spasova, S., Vanhercke, B., & Coster, S. (2018). Inequalities in access to healthcare. *A study of national policies, European Social Policy Network (ESPN). Brussels: European Commission*.

Beauchamp, T. L., & Childress, J. F. (2012). Principles of Biomedical Ethics (7th ed.). Oxford University Press.

Behr, A., & Theune, K. (2017). Health system efficiency: a fragmented picture based on OECD data. *PharmacoEconomics-open*, *1*, 203-221

Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR.

Castelvecchi, D. (2016). Can we open the black box of AI?. *Nature News*, *538*(7623), 20.

Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., ... & Mavridis, N. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *NPJ digital medicine*, *3*(1), 1-11.

Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. New England Journal of Medicine, 372(9), 793-795.

Cristea, M., Noja, G. G., Stefea, P., & Sala, A. L. (2020). The Impact of Population Aging and Public Health Support on EU Labor Markets. *International journal of environmental research and public health*, *17*(4), 1439. https://doi.org/10.3390/ijerph17041439

Egan, J. & E. Rosenbach(2023). *Biosecurity in the Age of AI: What's the Risk? | Belfer Center for Science and International Affairs*. Belfer Center for Science and International Affairs. https://www.belfercenter.org/publication/biosecurity-age-ai-whats-risk#_edn1

European Commission. (2019). Ethics guidelines for trustworthy AI. Retrieved from https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html

European Commission. (2020). AI Act: Regulation (EU) 2020/xxx of the European Parliament and of the Council of [date], laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. Brussels.

European Commission. (2021). (forthcoming) regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts.

European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union, L 119, 1–88.

European Parliament and Council of the European Union. (2017a). Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices. Official Journal of the European Union, L 117, 1-175.

European Parliament and Council of the European Union. (2017b). Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices. Official Journal of the European Union, L 117, 176-332.

Fichman, R. G., Kohli, R., & Krishnan, R. (2011). Editorial overview—The role of information systems in healthcare: Current research and future trends. Information Systems Research, 22(3), 419-428.

Floridi, L. (2008). The method of levels of abstraction. *Minds and machines*, *18*, 303-329.

Floridi, L. (2013). The ethics of information. Oxford University Press.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review. https://doi.org/10.1162/99608f92.8cd550d1

Fonseca, Â., Ferreira, A., Ribeiro, L., Moreira, S., & Duque, C. (2024). Embracing the future—is artificial intelligence already better? A comparative study of artificial intelligence performance in diagnostic accuracy and decision-making.*European Journal of Neurology*, e16195.

Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. JAMA Internal Medicine, 178(11), 1544-1547. https://doi.org/10.1001/jamainternmed.2018.3763

High-Level Expert Group on Artificial Intelligence. (2019). Ethics guidelines for trustworthy AI. European Commission. Retrieved from https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

IEEE. (2019). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition. IEEE. Retrieved from https://standards.ieee.org/industry-connections/ec/autonomous-systems.html

Jha, A. (2023). Technological Advances and Evolution of Biowarfare: A Threat to Public Health and Security. *KnE Social Sciences*, 401-416.

Johnson, K. B., Wei, W. Q., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., ... & Snowdon, J. L. (2021). Precision medicine, AI, and the future of personalized health care. *Clinical and translational science*, *14*(1), 86-93.

Ke YY, Peng TT, Yeh TK, et al. Artificial intelligence approach fighting COVID-19 with repurposing drugs. Biomed J 2020; 43: 355–362.

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC medicine*, *17*, 1-9.

Kilkenny, M. F., & Robinson, K. M. (2018). Data quality:“Garbage in–garbage out”. *Health Information Management Journal*, *47*(3), 103-105.

Laitinen, A., & Sahlgren, O. (2021). AI systems and respect for human autonomy. *Frontiers in artificial intelligence*, *4*, 151.

LaRosa, E., & Danks, D. (2018). Impacts on trust of healthcare AI. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 210-215).

Luxton, D. D. (2014). Artificial Intelligence in Behavioral and Mental Health Care. Elsevier.

Mei, X., Lee, H. C., Diao, K. Y., Huang, M., Lin, B., Liu, C., ... & Yang, Y. (2020). Artificial intelligence–enabled rapid diagnosis of patients with COVID-19. *Nature medicine*, *26*(8), 1224-1228.

Mennella, C., Maniscalco, U., De Pietro, G., & Esposito, M. (2024). Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. *Heliyon*.

Oakden-Rayner, L., Gale, W., Bonham, T. A., Lungren, M. P., Carneiro, G., Bradley, A. P., & Palmer, L. J. (2022). Validation and algorithmic audit of a deep learning system for the detection of proximal femoral fractures in patients in the emergency department: a diagnostic accuracy study. *The Lancet Digital Health*, *4*(5), e351-e358.

Obermeyer, Z., Powers, B., Vogeli, C., et al. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453.

Parikh, R. B., Teeple, S., & Navathe, A. S. (2019). Addressing bias in artificial intelligence in health care. *Jama*, *322*(24), 2377-2378.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, *2*(1). https://doi.org/10.1186/2047-2501-2-3

Sambamoorthi, U., & McAlpine, D. D. (2003). Racial, ethnic, socioeconomic, and access disparities in the use of preventive services among women. *Preventive medicine*, *37*(5), 475-484.

Sharara, S., & Radia, S. (2022). Quick response (QR) codes for patient information delivery: a digital innovation during the coronavirus pandemic. *Journal of orthodontics*, *49*(1), 89-97.

Shin, D. (2020). User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explainability. *Journal of Broadcasting & Electronic Media*, *64*(4), 541-565.

Topol, E. J. (2014). Individualized medicine from prewomb to tomb. Cell, 157(1), 241-253.

Ward, E., Jemal, A., Cokkinides, V., Singh, G. K., Cardinez, C., Ghafoor, A., & Thun, M. (2004). Cancer disparities by race/ethnicity and socioeconomic status. *CA: a cancer journal for clinicians*, *54*(2), 78-93.

World Health Organization. (2022). *Health and care workforce in Europe: time to act*. World Health Organization. Regional Office for Europe.