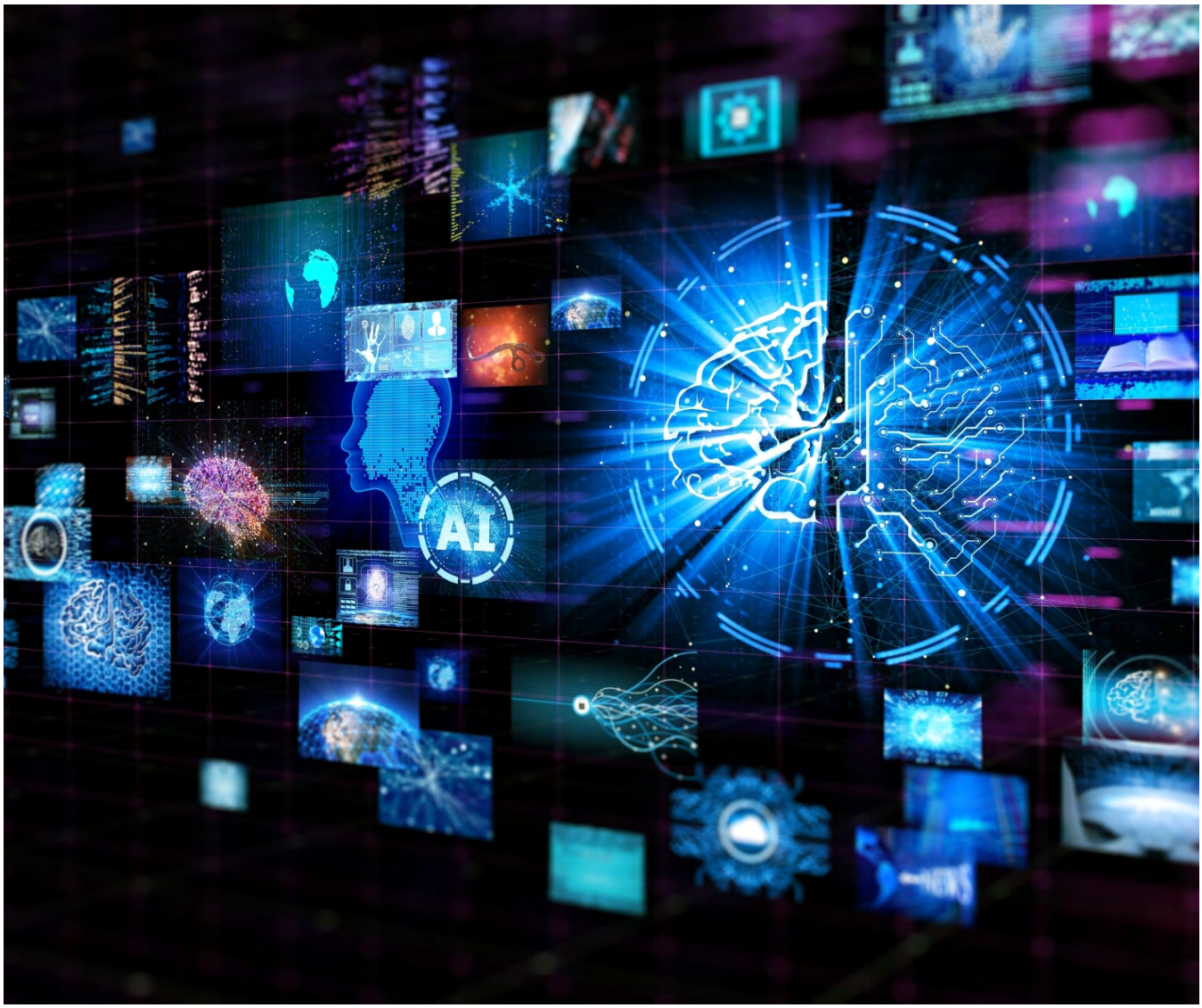
SMART CITIES, ARTIFICIAL INTELLIGENCE AND DIGITAL TRANSFORMATION LAW

A Handbook for Students and Professionals



Edited by E. E. Akin, S. Klimbacher, G. Ziccardi

ILS
INFORMATION, LAW & SOCIETY

Milano University Press

SMART CITIES, ARTIFICIAL INTELLIGENCE AND DIGITAL TRANSFORMATION LAW

A Handbook for Students and Professionals

Edited By
Eylül Erva Akin, Simona Klimbacher and Giovanni Ziccardi

Milano University Press

Smart Cities, Artificial Intelligence and Digital Transformation Law: A Handbook for Students and Professionals / Edited by Eylül Erva Akin, Simona Klimbacher and Giovanni Ziccardi. Milano: Milano University Press, 2024. (Information, Law & Society; 1)

ISBN 979-12-5510-164-2 (print)
ISBN 979-12-5510-166-6 (PDF)
ISBN 979-12-5510-167-3 (EPUB)
DOI 10.54103/infolawsoc.186

Pubblicato da:
Milano University Press
Via Festa del Perdono 7 – 20122 Milano
Sito web: https://milanoup.unimi.it
e-mail: redazione.milanoup@unimi.it

L'edizione cartacea del volume può essere ordinata in tutte le librerie fisiche e online ed è distribuita da Ledizioni (www.ledizioni.it)

Table of contents

PART II
ARTIFICIAL INTELLIGENCE

PART III
SMART CITIES AND DIGITAL SOCIETY

Chapter XI
Artificial Intelligence and Healthcare

by Malwina Anna Wójcik-Suffia*

Index: 1. The opportunities and risks of "black-box medicine". – 2. Regulating AI in healthcare – the EU perspective. – 3. AI in healthcare and the four principles of bioethics. – 4. Conclusion.

1. The opportunities and risks of "black-box medicine"

The integration of Artificial Intelligence (AI) into the medical field marks a transformative era in healthcare, public health, and biomedical research. Firstly, machine learning (ML) algorithms are increasingly used in patient care for diagnosis, prognosis, and treatment recommendations. Secondly, AI is facilitating effective planning and management of healthcare systems and public health surveillance. Thirdly, advancements in AI technologies offer benefits for health research and drug development. For instance, AI has enabled the prediction of protein folding, a key element of drug discovery (WHO, 2021).

The latest developments in healthcare AI involve large foundation models, which are trained on broad data at scale and subsequently adapted to a variety of clinical and administrative tasks, including analysis of different modalities of data, information retrieval, medical chatbots, and education tools (WHO, 2024).

Unfortunately, the operation of certain ML systems in healthcare raises transparency and safety concerns, as they fall into the category of "black box medicine" (W. Nicholson Price II 2017). As opposed to traditional, knowledge-based systems which operate on the rules of logic, ML algorithms work by drawing complex correlations between huge amounts of data. These correlations are not always clearly identifiable and explainable, making algorithmic decision-making inherently opaque. For instance, researchers discovered that ML models can detect self-reported race from medical images with a high degree of accuracy. However, research is not yet able to explain how the system reaches these conclusions (Banerjee *et al.*, 2022). Moreover, since it is very common for ML algorithms to find correlations between numerous data points, it might not be feasible to confirm them in clinical trials (Ferretti *et al.*, 2018). Finally, the

* PhD Researcher at the University of Bologna and the University of Luxembourg. Research Fellow at the Information Society Law Center (ISLC).

developers of AI might refuse to disclose how the model works to protect their intellectual property rights.

This chapter offers a review of the legal and ethical challenges associated with the use of black-box medicine. It starts with a brief introduction to the key legal instruments regulating healthcare AI in the EU. Following the legal overview, the chapter adopts a bioethical perspective to scrutinise the intricacies involved in deploying AI in healthcare settings. It does so by analysing the impact of AI on the four principles of bioethics: beneficence, non-maleficence, autonomy and justice (Beauchamp and Childress, 2019).

2. Regulating AI in healthcare – the EU perspective

In the EU, AI applications in healthcare are governed by a complex regulatory framework. Its core elements include:

- the General Data Protection Regulation (GDPR);
- the Medical Devices Regulation (MDR);
- the AI Act;
- the European Health Data Space (EHDS).

2.1 The General Data Protection Regulation

The GDPR plays a key role in regulating health data which constitute the lifeblood of medical AI. According to Art. 9 of the GDPR, genetic data, biometric data, and data concerning health fall into the category of sensitive data. Thus, their processing is permissible only based on specific grounds. These include, among others, patient consent, public interest in the safety of medicinal products or medical devices, and scientific research. Moreover, further processing of data collected for the purpose of providing healthcare is possible in the case of scientific research. However, there are two important caveats. Firstly, the extent to which the commercial development of AI technologies in healthcare can be considered scientific research under the GDPR is debatable and is likely to depend on the presence of substantial public interest (Meszaros and Ho, 2021). Secondly, it is crucial to note that GDPR allows Member States to introduce further restrictions concerning the processing of genetic data, biometric data or data concerning health, potentially obstructing the re-use and cross-border flow of health data.

2.2 The Medical Device Regulation

AI tools that are either independent software or an accessory to a medical device (*e.g.,* software for a wearable device) fall within the scope of the Medical Device Regulation if they are intended, among others, for "diagnosis, prevention,

monitoring, prediction, prognosis, treatment or alleviation of disease (Art. 2 MDR)." Medical devices need to undergo a conformity assessment procedure to confirm their safety before being placed on the market, and they are subject to rigorous post-authorisation oversight. The level of scrutiny depends on the classification of medical devices into four categories: I (low risk), IIa (moderate risk), IIb (medium risk), and III (high risk). The level of potential risk is based on the device's intended purpose. While class I requires only a self-assessment by the manufacturers, classes IIa, IIb, and III require an intervention by a notified body.

Software that is driving a medical device falls within the same class as the device itself. Standalone software is classified as low risk unless it is used for medical diagnosis, therapy, or to monitor physiological processes. In these cases, it falls under class IIa (moderate risk), IIb (medium risk), or III (high risk), depending on its function and the level of possible harm to the patient. Software will be classified as high risk if it can cause death or an irreversible deterioration of a person's state of health.

2.3 The AI Act

Many applications of AI in healthcare will be classified as high-risk systems under the AI Act, triggering requirements about risk management, quality of data requirements, technical documentation, transparency and provision of information to users, quality management systems, human oversight, robustness, accuracy, and cybersecurity.

There are two ways in which an AI system can be classified as high risk under the AI Act.

Firstly, a system will fall within the high-risk category if it is a product (or a safety component of a product) subject to third-party *ex ante* conformity assessment under EU harmonisation legislation listed in Annex I, which includes the MDR. Put simpler, software that is qualified as a medical device of class IIa, IIb or III (moderate to high risk according to the MDR) will be qualified as a high-risk system under the AI Act. Thus, these systems will need to comply with both the MDR and the AI Act.

Secondly, a system will fall within the high-risk category if it is a standalone system listed in Annex III. In the healthcare domain, these systems include systems intended:

- to be used by public authorities or on behalf of public authorities to evaluate individual eligibility for healthcare services and benefits,
- to dispatch emergency medical aid and emergency triage systems; and
- for risk assessment and pricing determination for life and health insurance.

These systems will always be considered high risk if they perform profiling of natural persons. Otherwise, they can be deemed non-high-risk if they are intended to perform certain simple administrative or supervisory tasks

2.4 The European Health Data Space

The EHDS deals specifically with data concerning health, governing their primary and secondary use. The EHDS is relevant for regulating medical AI for two main reasons.

Firstly, it introduces requirements aimed at increasing the availability of health data. It does so by creating trustworthy mechanisms for the re-use of health data to facilitate, among others, research and innovation in the field of medical AI.

Secondly, the EHDS introduces quality, safety, and interoperability requirements concerning electronic health records (EHRs), a major source of health data. The regulation states that EHRs will need to comply with common specifications elaborated by the EU Commission.

Moreover, medical devices and high-risk AI systems under the AI Act that claim compatibility with EHRs will also need to comply with selected requirements under the EHDS.

3. AI in healthcare and the four principles of bioethics

In highly connected environments in which black-box medicine is deployed, it is useful to refer to the four well-established principles of bioethics (beneficence, non-maleficence, autonomy, and justice) to guide the ethical and legal assessment of challenges posed by AI.

a) Beneficence

Beneficence is the first core principle of bioethics. It requires healthcare professionals to "do good", acting in the best interest of the patient. The list of potential benefits of medical AI for patient care and the organisation of healthcare systems is long. AI solutions can contribute to a quicker diagnosis and effective risk assessment, allowing healthcare systems to implement preventive care to reduce the burden of disease and healthcare spending. Advancements in AI-enabled personal mobile devices and apps monitoring the health of patients increase their safety and comfort of living. The analysis of vast amounts of medical data leads to the development of increasingly personalised medicine. Last but not least, the automatisation of repetitive administrative tasks and improving clinical workflow allows doctors to focus on the human side of medicine – presence, empathy, trust, and caring (Eric Topol, 2019).

However, the beneficial effects of AI are not equally distributed among society. The digital divide in healthcare technologies is present both within states and globally.

On a domestic level, the low digital literacy of vulnerable patient groups, such as the elderly or immigrant communities, often prevents them from seizing the benefits of AI solutions. Moreover, the spatial accessibility of state-of-the-art

medical AI, especially in the public healthcare systems, is limited to well-served, urban areas. This often results in excluding patients coming from lower socio-economic backgrounds.

On a global level, resource constraints and the lack of digital medical infrastructure make lower and middle-income countries less likely to fully benefit from AI technologies. At the same time, AI's potential to revolutionise healthcare is the greatest in these areas.

In Europe, some measures aimed at increasing universal access to AI benefits can be highlighted. For example, in Germany, a system of reimbursement for "apps on prescription" has been introduced to enable more equitable access to technology.

The problems of inequitable distribution of benefits of AI are closely related to the principle of justice, explained in the paragraphs that follow.

b) Non-maleficence

The principle of non-maleficence entails the duty not to harm the patient, as expressed in the Hippocratic Oath itself – "first, do no harm".

Despite their potential benefits, medical AI systems can still exhibit risks of errors and inaccuracies, leading to failed or misguided interventions. Firstly, noise in the input data, such as poor quality of medical images or inconsistent labelling, can cause AI tools to draw incorrect conclusions. Secondly, AI is particularly prone to memorizing spurious correlations. For example, deep learning systems detecting Covid-19 from chest X-rays often yield incorrect diagnoses because they associate the presence of the disease with clinically irrelevant factors (De Grave *et al.*, 2021). Thirdly, AI systems do not generalize well. In other words, highly accurate ML algorithms trained on a dataset coming from hospitals X and Y can yield inadequate results when applied to a new set of data in hospital Z (Zech *et al.*, 2018).

Misguided algorithmic recommendations can lead clinicians to commit medical errors, causing harm to patients. At the same time, the legal regimes of medical negligence and product liability are not well suited to address the complex, multi-actor problem of AI-enabled decision-making in healthcare. In particular, the allocation of liability between doctors, healthcare providers and developers of AI tools remains an open question. Although the proposed Product Liability Directive (PLD) and AI Liability Directive (AILD) constitute a crucial step toward developing a unified European approach to liability issues, they fail to protect patients when harm occurs due to the non-interpretable character of the AI system (Duffourc and Gerke, 2023).

As discussed, the lack of explainability is an inherent feature of some AI models. Thus, it does not constitute a defect under the PLD's strict liability rules. Similarly, it cannot be considered a fault of the manufacturer or healthcare provider under the AILD. The lack of a coherent liability framework for

AI-perpetuated harm in medicine can lead to stifling innovation in healthcare, as healthcare professionals avoid implementing AI solutions for fear of liability.

Another harm is associated with the potential misuse of AI systems in healthcare. Firstly, the majority of commercially available m-Health apps are not qualified as medical devices, and thus they do not undergo rigorous testing. Yet, many patients rely on them for diagnosis and medical advice despite their questionable accuracy (Lekadir, 2022). Secondly, the lack of sufficient education and training results in low technical literacy of healthcare professionals. This makes doctors and healthcare providers prone to automation and confirmation biases, showing a preference for algorithmic decisions without properly reviewing them. For instance, one of the main healthcare insurers in the US has been recently sued for prematurely terminating coverage based on the recommendation of an algorithm and contrary to the advice of healthcare professionals (Barrows *et al. v.* Humana, Inc. (3:23-cv-00654) Kentucky Western District Court; Estate of Gene B. Lokken *et al. v.* UnitedHealth Group, Inc. et al. (0:23-cv-03514) Minnesota District Court).

Finally, the use of AI in healthcare is connected with considerable privacy and cybersecurity risks. Potential leaks of data relating to health can expose patients to various forms of cyberattacks. The misuse of healthcare data could lead to serious consequences for data subjects, such as denial of insurance or job opportunities based on one's state of health. While data concerning health are rightfully protected as sensitive data under the GDPR, access to them is also necessary for the training and validation of AI systems. Thus, Art. 10 of the AI Act allows developers of AI to process sensitive data subject to appropriate safeguards and to the extent it is strictly necessary to detect and correct bias. Moreover, the EHDS Proposal enables the sharing of healthcare data for secondary purposes, including the training of medical AI. However, privacy requires that patients understand and consent to the re-purposing of their health data. For instance, in 2016, a scandal was caused by the transfer of data of 1,6 million UK patients to Google's DeepMind to develop an AI kidney disease detection system. The transfer took place without the data subjects' consent, violating patient privacy (Lekadir, 2022).

The AI Act acknowledges and addresses the potential harms of AI. For instance, Art. 14 states that providers and deployers of high-risk systems should ensure that natural persons who oversee their functioning are aware of automation and confirmation bias. Moreover, Art. 15, introduces requirements regarding accuracy, robustness, and cybersecurity.

c) Autonomy

The principle of autonomy entails the right to freely decide about one's medical treatment and participation in research. A central component of autonomy is informed consent to treatment, often described as the cornerstone of medical

ethics. The right to informed consent is protected by domestic and supranational law. The latter includes the Council of Europe's Oviedo Convention on Human Rights and Biomedicine (Art. 5) and the EU's Charter of Fundamental Rights (Art. 3). However, the application of the right in the context of AI poses many doubts explored by practitioners and scholars.

The first key debate concerns the doctor's duty to disclose the fact that a diagnosis was made with the help of AI. Interpretations of the disclosure question might vary under domestic law. Arguably, the more prevalent AI solutions become in healthcare, the less controversial the fact of non-disclosure becomes. After all, informed consent does not require doctors to voluntarily disclose all the sources of their decision, such as practical experience or the books they have consulted. Therefore, it could be argued that it is unlikely that the disclosure is mandatory unless a patient explicitly enquires about the involvement of AI (Glen I. Cohen, 2019). However, we might imagine specific fact patterns that point in the direction of disclosure.

Firstly, disclosure can be considered mandatory when AI plays a substantial role in decision-making. On the EU level, Art. 22 of the GDPR enshrines the right of data subjects, including patients, not to be subject to a decision "based solely on automated processing" and a corresponding right to be informed when such a decision is made. However, it's important to underline that decisions based solely on algorithmic output are very rare in the medical domain. Thus, a degree of human involvement in decision-making could prevent patients from relying on the GDPR to support a duty to disclose the use of medical AI.

Nevertheless, it could be argued that such a duty would exist under domestic informed consent law in cases in which AI plays a substantial role in the decision process. This is especially true in jurisdictions in which the appropriate legal test of informed consent is patient-centric, as opposed to doctor-centric. Studies of patient attitudes toward AI decision-making suggest that they tend to find automated medical diagnostics "dehumanising" (Formosa *et al.*, 2022). Thus, it is likely that a reasonable patient would like to know whether a key recommendation concerning her or his treatment was formulated by a machine.

Secondly, a duty to disclose might exist when the use of AI would be analogous to an experimental treatment (Glen I. Cohen, 2019). That could be the case if healthcare professional lacks certainty about the accuracy of the tool. For instance, an AI system that detects cancers accurately for a subset of the population on which it was tested, might not be effective when run on different demographics. The doctor should therefore inform the patient when the potential risk of misdiagnosis is high. Thus, it could be argued that the duty to disclose the use of AI to a patient would be conditioned on the doctor possessing enough evidence about the performance of the system in question. To this end, different jurisdictions have proposed transparency solutions enabling healthcare professionals to access basic information about the system's functioning.

For instance, Art. 13 of the AI Act puts certain transparency obligations on providers of high-risk systems, enabling users to receive basic information about the system's functioning. Similarly, in the US, a new executive rule holds that healthcare IT should be accompanied by information that enables clinical users to determine its utility in a given context.

Thirdly, a duty to disclose the use of AI might exist when the system is designed to optimise its decisions for a specific variable, such as cost-effectiveness. (Glenn I. Cohen, 2019). This would potentially limit a patient's right to choose between alternative treatments.

The debate on informed consent gets complex once a patient starts asking more specific questions about the diagnosis or recommendation. Informed consent requires the medical professional to ensure the patient can interpret the information to make a fully informed choice about the treatment. However, physicians who use AI for decision support might face difficulties in explaining the basis of the algorithmic decision to patients due to algorithmic opacity.

Both computer science and law aim to address the problem of black-box medicine, dealing with the challenge of adjusting the explanation mechanisms to the needs of different stakeholders. From the technical point of view, researchers working in the field of explainable AI (XAI) elaborate different methods to interpret the outputs of algorithmic systems (Chaddad *et al.*, 2023). From the legal point of view, scholars and legislators discuss the right to explanation of decisions taken with the use of AI. Although Art. 15 of the GDPR contains the right to request "meaningful information about the logic involved" in fully autonomous decisions, scholars argue that the right does not extend to rationales for specific decisions. Rather, it focuses on general system functionality (Wachter *et al.*, 2017). In addition to the GDPR, the AI Act introduces a right to explanation of decisions taken with the use of high-risk AI in certain healthcare contexts. However, the practical implications of this provision remain to be explored.

While a lot of scholarly attention in the context of medical AI is devoted to the right to know, it should be underlined that autonomy also comprises the right not to know (Andorno, 2004). Such right is often invoked in genetic cases when patients express the will not to be informed about potential hereditary and incurable diseases. The pervasiveness of AI-enabled medical screening technologies could violate the right not to know, potentially leading to overdiagnosis, increased stress, and anxiety.

Lastly, it is important to mention that AI-powered developments in neurotechnology, including brain-computer interfaces enabled through external or implantable medical devices, pose serious risks to human autonomy. As pointed out by researchers, the processing of "mental data" creates a danger of third-party intrusion and manipulation of the mental sphere (Ienca and Malgieri, 2022).

d) Justice

The principle of justice requires a fair distribution of burdens and benefits in the healthcare system. Put simply, patients should not face discrimination in access to and provision of healthcare. They should also be able to freely participate in medical research and benefit from its results.

Unfortunately, using AI to guide decisions in both diagnostic and resource allocation scenarios disrupts the core tenets of justice. AI has the propensity to encode and perpetuate data gaps and biases concerning vulnerable patient groups, leading to unlawful discrimination.

Firstly, the datasets on which AI systems are trained and validated often lack reliable data on groups historically excluded from research. A leading example is women, whose bodies have been largely understudied despite evidence of numerous sex differences in organ anatomy and disease patterns (Caroline Criado Perez, 2020). The data gaps contribute to the misdiagnosis of many common conditions, such as stroke. Consequently, AI systems deployed in cardiology are likely to work less accurately for female patients (Fosch-Villaronga *et al.*, 2022). Similarly, skin cancer detection systems that do not take into account sex and gender-related characteristics risk reinforcing bias (Lee *et al.*, 2022).

Secondly, even if the data used to train and validate the algorithm includes information on vulnerable groups, it might still be tainted by harmful stereotypes. A good example is the use of race-adjusted clinical algorithms. This practice, spanning across various medical disciplines, entails differential diagnosis of patients based on their race (Vyas *et al.*, 2020). For instance, in nephrology, the eGFR formula used to measure kidney function based on the level of creatinine is designed to return higher scores for Black patients. A high score is associated with better kidney function, potentially resulting in de-prioritisation of Black patients in the prevention and treatment of nephrological diseases. The rationale for race correction in eGFR is based on the observation that the level of creatine is typically higher in Black patients. This in turn is often explained by Black people being allegedly more muscular.

Even if adopted with equity and accuracy in mind, many of the race-correction algorithms are based on the flawed premise of biological differences between races. Race, however, has been proven to be a socio-cultural concept, distinct from genetic ancestry (Malinowska and Żuradzki, 2022). Thus, many of the presumed differences between White and non-White patients stem from the legacy of racism and eugenics which sought to portray Black bodies as inferior. The race-based medicine led to the development of stereotypes such as the thickness of Black skin, higher pain tolerance, or even the difference in brain size. Researchers show that some of these harmful premises are replicated by state-of-the-art LLMs in healthcare (Omiye *et al.*, 2023).

Thirdly, it is important to note that training algorithms on data that adequately represent reality can still lead to unfair decisions when this reality

reflects inequitable access to healthcare. Using historic data often means that instead of moving towards more fair and equitable decisions, AI systems simply embrace the status quo, fuelling the vicious circle of exclusion. A real-life example is provided by the Impact Pro algorithm which was used in the US to identify patients with complex health problems, suitable for high-risk care management. The algorithm used healthcare spending as a proxy for illness and falsely attributed a lower risk of serious disease to Black patients (Obermeyer *et al.*, 2019). Note that unlike the clinical algorithms described above, this AI system was race blind. Nevertheless, the data reflected unequal healthcare access experienced by people of colour, leading to the replication of discriminatory outcomes.

Regrettably, EU antidiscrimination law is often ill-suited to address algorithmic discrimination in healthcare (Wójcik, 2022). Firstly, although discrimination can occur on many different grounds, the material scope of EU antidiscrimination law is limited. The anti-discrimination directives applicable in the context of healthcare protect patients on just three grounds: race, ethnic origin and sex (Di Federico, 2017). Article 21 of the Charter of Fundamental Rights, which contains an open list of non-discrimination grounds only applies to Member States when they implement EU law. Since organisation and delivery of healthcare is a primary competence of Member States, the applicability of the Charter will likely be limited to cross-border healthcare scenarios (Di Federico, 2017). Secondly, the EU antidiscrimination law does not protect against discrimination on multiple combined grounds, which is often perpetuated by algorithms (Gerards and Xenidis, 2021). Thirdly, the enforcement mechanisms of antidiscrimination law are highly fragmented (Di Federico, 2017). Fourthly, patients are likely to face difficulties detecting and proving discriminatory treatment, especially in the case of black-box algorithms.

The AI Act contains provisions aimed at reducing the risk of algorithmic bias. Most importantly, Art. 10 addresses bias in training data by introducing quality criteria for training, validation and testing of data sets in data-driven high-risk systems. These data sets must be, *inter alia*, examined for possible biases, relevant, representative, free of errors, complete and contextual, that is trained, validated and tested in a particular geographic, behavioural or functional setting. Moreover, Art. 15 addresses feedback loop bias, providing that high-risk algorithms which continue to learn after being placed on the market, should be accompanied by appropriate bias mitigation measures throughout their life cycle.

Furthermore, the EHDS can contribute to debiasing medical algorithms by increasing the diversity of datasets through data sharing mechanisms.

4. Conclusion

This chapter aimed to systematise the numerous ethical and legal concerns posed by the deployment of AI models in healthcare through the lens of the key principles of bioethics. However, it is important to underline that the complexity of the ethical and regulatory landscape of black-box medicine, can lead to conflicts between the bioethical principles, creating a need to balance competing interests.

For instance, the negotiations concerning the data-sharing mechanisms under the EHDS proposal illustrate the conflict between justice and autonomy. On the one hand, the availability of high-quality data is a prerequisite for creating fairer AI systems. On the other hand, the respect for privacy and informed consent points towards the introduction of an opt-out mechanism for patients who do not want their data to be shared.

Another conflict can arise between justice and non-maleficence in the context of de-biasing algorithms, as researchers acknowledge that there is a trade-off between the fairness and accuracy of AI models (Schönberger, 2019). Thus, imposing fairness constraints on the model can decrease its overall performance.

As a final example, certain uses of AI in mental health can result in a conflict between autonomy and beneficence. For instance, the AI Act carves a therapeutic exception for the general prohibition of algorithmic manipulation and persuasion based on explicit informed consent of patients. While subliminal techniques could offer a benefit to patients in a clinical setting, there is a risk that the therapeutic exception can be abused with a detriment to patient autonomy.

These examples make clear that navigating the ethical and regulatory maze of AI in health requires a collaborative effort among lawyers, policymakers, computer scientists, healthcare professionals, and patients to ensure that AI technologies appropriately balance patient benefit, autonomy, and equity concerns.

Bibliography

Andorno, R. (2004) 'The right not to know: an autonomy based approach', *Journal of medical ethics*, 30(5). Available at: https://doi.org/10.1136/jme.2002.001578.

Banerjee, I. *et al.* (2022) 'Reading Race: AI Recognises Patient's Racial Identity In Medical Images', *The Lancet Digital Health*, 4(6), pp. e406–e414. Available at: https://doi.org/10.1016/S2589-7500(22)00063-2

Beauchamp, T.L. and Childress, J.F. (2019) *Principles of biomedical ethics*. Eighth edition. New York: Oxford University Press.

Chaddad, A. *et al.* (2023) 'Survey of Explainable AI Techniques in Healthcare', *Sensors*, 23(2), p. 634. Available at https://doi.org/10.3390/s23020634.

Cohen, I.G. (2019) 'Informed Consent and Medical Artificial Intelligence: What to Tell the Patient? Symposium: Law and the Nation's Health', *Georgetown Law Journal*, 108(6), pp. 1425–1470.

Criado-Perez, C. (2020) *Invisible women: exposing data bias in a world designed for men*. London: Vintage.

Di Federico, G. (2017) 'Access to Healthcare in the European Union: Are EU Patients (Effectively) Protected Against Discriminatory Practices?', in L.S. Rossi and F. Casolari (eds). *The Principle of Equality in EU Law*. Springer Publishing.

Duffourc, M.N. and Gerke, S. (2023) 'The proposed EU Directives for AI liability leave worrying gaps likely to impact medical AI', *npj Digital Medicine*, 6(1), pp. 1–6. Available at: https://doi.org/10.1038/s41746-023-00823-w.

Ferretti, A., Schneider, M. and Blasimme, A. (2018) 'Machine Learning in Medicine: Opening the New Data Protection Black Box', *European Data Protection Law Review (EDPL)*, 4(3), pp. 320–332.

Formosa, P. *et al.* (2022) 'Medical AI and human dignity: Contrasting perceptions of human and artificially intelligent (AI) decision making in diagnostic and medical resource allocation contexts', *Computers in Human Behavior*, 133, p. 107296. Available at: https://doi.org/10.1016/j.chb.2022.107296.

Fosch Villaronga, E. *et al.* (2022) 'Accounting for diversity in AI for medicine', *Computer Law & Security Review*, 47, p. 105735. Available at: https://doi.org/10.1016/j.clsr.2022.105735.

European Commission, Directorate-General for Justice and Consumers, Gerards, J., Xenidis, R. (2021), *Algorithmic discrimination in Europe – Challenges and opportunities for gender equality and non-discrimination law*. Publications Office. Available at: https://data.europa.eu/doi/10.2838/544956.

Ienca, M. and Malgieri, G. (2022) 'Mental data protection and the GDPR', *Journal of Law and the Biosciences*, 9(1), p. lsac006. Available at: https://doi.org/10.1093/jlb/lsac006.

Lee, M.S., Guo, L.N. and Nambudiri, V.E. (2022) 'Towards gender equity in artificial intelligence and machine learning applications in dermatology', *Journal of the American Medical Informatics Association*, 29(2), pp. 400–403. Available at: https://doi.org/10.1093/jamia/ocab113.

Lekadir, K. *et al.* (2022) *Artificial intelligence in healthcare: applications, risks, and ethical and societal impacts*. Brussels: European Parliament.

Malinowska, J.K. and Żuradzki, T. (2022) 'Towards the multileveled and processual conceptualisation of racialised individuals in biomedical research', *Synthese*, 201(1), p. 11. Available at: https://doi.org/10.1007/s11229-022-04004-2.

Meszaros, J. and Ho, C. (2021) 'AI research and data protection: Can the same rules apply for commercial and academic research under the GDPR?', *Computer*

Law & Security Review, 41, p. 105532. Available at : https://doi.org/10.1016/j.clsr.2021.105532.

Obermeyer, Z. *et al.* (2019) 'Dissecting racial bias in an algorithm used to manage the health of populations', *Science (New York, N.Y.)*, 366(6464), pp. 447–453. Available at: https://doi.org/10.1126/science.aax2342.

Omiye, J.A. *et al.* (2023) 'Large language models propagate race-based medicine', *npj Digital Medicine*, 6(1), pp. 1–4. Available at: https://doi.org/10.1038/s41746-023-00939-z.

Price, W.N.I. (2017) 'Regulating Black-Box Medicine', *Michigan Law Review*, 116(3), pp. 421–474.

Schönberger, D. (2019) 'Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications', *International Journal of Law and Information Technology*, 27(2), pp. 171–203. Available at: https://doi.org/10.1093/ijlit/eaz004.

Topol, E.J. (2019) *Deep medicine: how artificial intelligence can make healthcare human again*. First edition. New York: Basic Books.

Vyas, D.A., Eisenstein, L.G. and Jones, D.S. (2020) 'Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms', *New England Journal of Medicine*, 383(9), pp. 874–882. Available at: https://doi.org/10.1056/NEJMms2004740.

Wachter, S., Mittelstadt, B. and Floridi, L. (2017) 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation', *International Data Privacy Law*, 7(2), pp. 76–99. Available at: https://doi.org/10.1093/idpl/ipx005.

Wójcik, M.A. (2022) 'Algorithmic Discrimination in Health Care: An EU Law Perspective', *Health and Human Rights*, 24(1), p. 93.

World Health Organization (2021) *Ethics and governance of artificial intelligence for health: WHO guidance*. Geneva: World Health Organization. Available at https://apps.who.int/iris/handle/10665/341996.

World Health Organization (2024) *Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models*. World Health Organization. Available at https://www.who.int/publications-detail-redirect/9789240084759.

Zech, J.R. *et al.* (2018) 'Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study', *PLoS medicine*, 15(11), p. e1002683. Available at https://doi.org/10.1371/journal.pmed.1002683.