COMBATING ALGORITHMIC BIAS IN HEALTHCARE: TOWARDS A EUROPEAN REGULATORY FRAMEWORK BASED ON THE RIGHT TO HEALTH AND THE RIGHT TO SCIENCE

Malwina Anna Wójcik

INTRODUCTION

In their opinion piece for Quartz magazine, Andy Coravos and colleagues argued that "algorithms are like drugs" – they can have a profound impact on human life, their performance varies according to different demographics and, perhaps most crucially, they can cause serious side effects.[1] However, unlike prescription drugs, algorithms do not feature a warning label, they rarely come with instruction for use, and are not yet subject to a comprehensive regulatory framework. This contrast becomes even more apparent when we consider the application of artificial intelligence (AI) in healthcare itself.

The growing use of AI technologies, such as machine learning (ML) and deep learning, in medical prognostics, diagnostics, benefit allocation and research offers a promise of faster, cheaper and more accurate decision making in healthcare. For example, AI systems are successfully utilised to detect diabetic retinopathy,[2] predict a patient's response to different anti-depressant drugs,[3] or recommend a combination of chemotherapy for cancer patients.[4] An emerging field of research aims to augment motor, communicational and cognitive functions through brain-computer interface.[5]

However, just like in cases of prescription drugs, innovation comes with a price. Deployment of AI in healthcare raises many important ethical and legal concerns, including,

---

[1] Andy Coravos, Irene Chen, Ankit Gordhandas, and Ariel Dora Stern, *We should treat algorithms like prescription drugs*, QUARTZ (Feb. 19, 2019), https://qz.com/1540594/treating-algorithms-like-prescription-drugs-could-reduce-ai-bias/ (last visited Jan. 7, 2022).

[2] Ulrich M. Gassner and Ulrich Juknat, *Regulatory Approaches to AI in Medical Practice,* 3 EUR. PHARM. L. J. 176, 177-78 (2019).

[3] Kevin P. Nguyen et al., *Patterns of Pre-Treatment Reward Task Brain Activation Predict Individual Antidepressant Response: Key Results from the EMBARC Randomized Clinical Trial*, 91 BIOLOGICAL PSYCHIATRY 550, 558 (2021).

[4] Sahar Takshi, *Unexpected Inequality: Disparate-Impact from Artificial Intelligence in Healthcare Decisions*, 34 J.L. & HEALTH 215, 219 (2021).

[5] *See generally* Marcello Ienca and Karolina Ignatiadis, *Artificial Intelligence in Clinical Neuroscience: Methodological and Ethical Challenges*, 11(2) AM. J. OF BIOETHICS NEUROSCIENCE 77.

inter alia, fairness, patient autonomy, privacy and liability for harm. This essay will focus on one "side-effect" of medical algorithms – algorithmic bias. Adopting the European legal perspective, it will explain why specific focus on healthcare is needed in the ongoing effort to design a comprehensive regulation for AI in the European Union (EU). Then, it will propose how a framework based on the right to health and the right to science could improve the resilience of the European Artificial Intelligence Act (AIA)[6] against algorithmic discrimination in healthcare.

I. HOW ALGORITHMS DISCRIMINATE AGAINST VULNERABLE PATIENTS

The fact that AI technology can reflect and exacerbate existing discrimination patterns is already well documented. A notable example includes the Gender Shades project, which uncovered racial and gender bias in leading facial recognition technologies.[7] As highlighted by the recent World Health Organization Guidance, healthcare AI is not immune from discriminatory outcomes.[8] Scholars distinguish two main sources of algorithmic bias – training data[9] and "unequal ground truth."[10]

There are various ways in which training data can lead to unfair and discriminatory outcomes. For example, the so-called "garbage-in-garbage-out" principle is exemplified by a ML system deployed to predict the risk of hospital patients developing pneumonia, which wrongly suggested to send patients with asthma home.[11] The algorithm classified those

---

[6] *Proposal for the Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM (2021) 206 final (Apr. 21, 2021).

[7] *See, e.g.,* Joy Boulamwini and Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. OF MACH. LEARNING RSCH. 1 (2018).

[8] *Ethics and Governance of Artificial Intelligence for Health*, WORLD HEALTH ORGANIZATION [WHO] 70-73 (2021), https://www.who.int/publications/i/item/9789240029200 (last visited Mar. 1, 2022).

[9] Philipp Hacker, *A legal framework for AI training data—from first principles to the Artificial Intelligence Act,* 13 L., INNOVATION AND TECH. 257, 261 (2021).

[10] Daniel Schönberger, *Artificial Intelligence in Healthcare: A Critical Analysis of The Legal and Ethical Implications*, 27 INT'L J.L. & INFO. TECH. 171, 178 (2019).

[11] *See generally* Rich Caruana et al., *Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-day Readmission*, KDD '15:

patients as low risk because their files were missing from the training dataset, as they were often admitted directly to the intensive care unit. Similarly, when an algorithm is trained on data that fail to represent the target patient population, selection bias occurs. Thus, a skin cancer detection model trained primarily on white patients underperformed on people of colour.[12] Finally, algorithms can entrench existing patterns of discrimination reflected in data; this is called feedback loop bias. An illustrative example is provided by the Impact Pro algorithm, which falsely attributed a lower risk of serious disease to Black patients, replicating biases embedded in historic data.[13]

Worryingly, assuring high quality of data is not enough to avoid algorithmic discrimination. Sometimes, the "ground truth," which can be defined as the closest mathematical expression of reality, is simply unfair. Because finding new correlations in data lies at the very core of AI, algorithms often discriminate in abstract and subtle ways. For example, a medical appointment scheduling algorithm caused overbooking of people of colour because prior no-shows were correlated with lower socio-economic background, unemployment, lack of medical insurance and lack of access to effective transport means.[14] Since Black patients were "overrepresented at the lower socioeconomic status level," prior no-shows became a proxy for race.

Previously unknown patterns of discrimination by algorithms continue to emerge. For instance, the recent work of researchers in Stanford has shown that ML models exhibit a high degree of accuracy in predicting self-reported race from medical images.[15] This algorithmic side effect is highly problematic because a system identifying a patient's race without the health professional knowing can entrench health inequalities. Interestingly, researchers are not sure why the algorithm thrives at predicting race. Moreover, they have not been able to

---

Proc. of the 21th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining, (Aug. 2015).

[12] *See generally* Adewole S. Adamson and Avery Smith, *Machine Learning and Health Care Disparities in Dermatology*, 154 JAMA DERMATOLOGY 1247 (2018).

[13] Ziad Obermeyer et al., *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*, 336 SCIENCE 447 (2019).

[14] Michele Samorani & Linda Goler Blount, *Machine Learning and Medical Appointment Scheduling: Creating and Perpetuating Inequalities in Access to Health Care*, 110 AM. J. PUB. HEALTH 440, 440-41 (2020).

[15] Imon Banerjee et al., *Reading Race: AI Recognizes Patient's Racial Identity in Medical Images*, ARXIV (July 21, 2021) https://arxiv.org/pdf/2107.10356.pdf (last visited Mar. 1, 2022).

eliminate the high accuracy rate of race prediction using de-biasing techniques.

II. REASONS TO REGULATE HEALTHCARE AI THROUGH LAW

As the above examples make clear, the deployment of AI in healthcare contexts carries a significant risk of discriminatory effects for patients coming from vulnerable and marginalised communities, putting their life and well-being at stake. I argue that there are several reasons why European legislators should pay particular attention to the issue of algorithmic discrimination in healthcare.

First, the use of AI can exacerbate existing patterns of discrimination in healthcare. Significant health inequalities persist across Europe, and patients are not adequately protected against discriminatory practices. Studies by the Fundamental Rights Agency (FRA)[16] and EQUINET[17] underline that the nature of discrimination in healthcare is often intersectional, which means that vulnerable individuals face disadvantages based on a unique combination of protected grounds. For example, Roma women often encounter specific obstacles in access to healthcare because they are "unequal both as a woman within Roma society, and as a Roma woman among other women."[18] Since complex patterns of disadvantage are inevitably woven into the data fabric, algorithmic decision making can amplify intersectional discrimination.[19] Moreover, systems which rely on algorithmic profiling, correlating specific individuals with a protected group, can foster intersectional disadvantage through interactions between labels.[20] Thus, intersectional minorities are one of the most likely targets of automated discrimination. In addition, discrimination in healthcare remains largely underreported, because vulnerable

[16] European Union Agency for Fundamental Rights, *Inequalities and multiple discrimination in access to and quality of healthcare*, at 15 (2013), https://fra.europa.eu/sites/default/files/inequalities-discrimination-healthcare_en.pdf (last visited Mar. 1, 2022).
[17] Niall Crowley, *Equality, diversity and non-discrimination in healthcare: Learning from the work of equality bodies*, EUROPEAN NETWORK OF EQUAL. BODIES (2021), https://equineteurope.org/wp-content/uploads/2021/06/Health-Perspective.pdf (last visited Mar. 1, 2022).
[18] European Union Agency for Fundamental Rights, *supra* note 16, at 58.
[19] Raphaele Xenidis, *Tuning EU equality law to algorithmic discrimination: Three pathways to resilience*, 27 MAASTRICHT J. OF EUR. AND COMPAR. L. 736, 740 (2021).
[20] *Id.*

patients fear retaliation, lack knowledge about the available complaint procedures, are deterred by their complexity or doubt their effectiveness.[21] Deployment of algorithmic decision making in healthcare is likely to exacerbate this phenomenon. Due to the complex nature of algorithmic discrimination, patients might never know they have been treated unfairly, possibly presuming that a decision taken by machines is fairer than a decision taken by health professionals. Those who decide to make a discrimination complaint are likely to face unsurmountable evidentiary challenges due to the 'black box' nature of some algorithms, which makes it impossible to explain the rationale behind an automated decision.[22]

Second, the patchwork nature of the EU anti-discrimination legal regime does not offer adequate protection to patients experiencing algorithmic discrimination. There are crucial problems relating to the scope and application of anti-discrimination law. The anti-discrimination directives applicable in the context of healthcare offer only three grounds for protection – race, ethnic origin[23] and gender.[24] Both the aforementioned directives and Article 21 of the EU Charter of Fundamental Rights, which is an open-ended anti-discrimination provision, apply only when a matter falls within the scope of EU law.[25] Moreover, the EU anti-discrimination law, which is based on specific protected grounds, does not accommodate patterns of discrimination prevalent in algorithmic decision making. For example, the Court of Justice of the European Union's (CJEU) stance on discrimination by proxy remains unclear and incoherent.[26] Additionally, in *Parris*, the CJEU rejected the

[21] European Union Agency for Fundamental Rights, *supra* note 16, at 8.
[22] *See* Philipp Hacker, *Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law*, 55 COMMON MARKET L. REV. 1143, 1186 (2018).
[23] Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin, O.J. (L 185) [hereinafter *Race Equality Directive*].
[24] Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services, O.J. (L 373) [hereinafter *Goods and Services Directive*].
[25] *Id.* at art. 3; *Race Equality Directive*, *supra* note 23, art. 3; EU Charter of Fundamental Rights, Dec. 1, 2009 O.J. (C 326) art. 21.
[26] Xenidis, *supra* note 19, at 746.

notion of intersectionality, reinforcing a single-axis model of anti-discrimination law.[27]

Third, without explicit legal requirements in place, manufacturers and users of healthcare AI systems might lack sufficient incentives to minimise bias.[28] First, an attempt to eradicate certain correlations might be detrimental to accuracy. Thus, there exists a trade-off between utility and minimisation of bias. Second, bias detection and correction techniques are expensive. Therefore, it might not be economically viable to apply them, especially if an algorithm performs well in a given context. Lastly, by continuously denying opportunities to protected groups and exacerbating existing patterns of discrimination, algorithms can contribute to self-fulfilling prophecies.

III. THE RIGHT TO HEALTH AND THE RIGHT TO SCIENCE AS A FRAMEWORK FOR REGULATING HEALTHCARE AI

The intricate nature of algorithmic bias and the deficiencies of the existing anti-discrimination legal framework call for additional legal safeguards protecting patients from algorithmic discrimination. To establish a coherent legal framework for medical AI, the European legislators need to undertake a holistic review of different areas of law, including anti-discrimination law, data protection law, regulation of medical devices and the legislative proposal concerning the horizontal regulation of AI. I argue that the design of this framework should be guided by the right to the highest attainable standard of health[29] and the right to enjoy the benefits of scientific progress and its applications.[30]

The proceeding sections analyse the content of the respective rights, shedding some light on their utility in the context of regulating healthcare AI. I follow the structure endorsed by the UN Committee on Economic, Social and Cultural Rights (CESCR), which conceptualises the right to health and the right to science as consisting of four elements: availability, accessibility, acceptability, and quality.[31] I explore

[27] Case C-443/15, Parris v. Trinity College Dublin, ECLI:EU:C:2016:493, (June 30, 2016).
[28] Hacker, *supra* note 22.
[29] International Covenant on Economic, Social and Cultural Rights art. 12, Dec. 16, 1966, 993 U.N.T.S. 3.
[30] *Id.* at art. 15.
[31] CESCR, General Comment 14: The Right to the Highest Attainable Standard of Health, U.N. Doc. E/C.12/2000/4 (Aug. 11, 2000) [hereinafter

each of the elements in the context of the European AIA, pointing out the strengths and weaknesses of the proposal. Presented in April 2021, the act foresees different regulatory regimes based on risk-assessment. High-risk systems are subject to strict obligations both before and after being placed on the market. They include AI software, which is integrated into a medical device or is an accessory to the medical device subject to *ex-ante* assessment,[32] as well as AI systems providing "access and enjoyment of essential private and public services," such as allocation of priority medical aid.[33]

a. Availability

In the case of the right to health, availability pertains to the presence of a sufficient number of public healthcare facilities, goods and services, including underling determinants of health, such as trained medical and professional personnel.[34] The right to science requires that States promote scientific progress and make its benefits available, especially to vulnerable and marginalised groups.[35]

The development of science gradually redefines the scope of social determinants of health. For example, it is increasingly accepted that the definition of a "trained" healthcare professional should include technological literacy. When the use of AI in clinical decision making and benefit allocation becomes prevalent, the availability of health data concerning marginalised communities is likely to become a new social determinant of health. The AIA mentions the need to provide technical training and education to the users of high-risk systems,[36] which can also encompass healthcare professionals. However, the proposal does nothing to acknowledge the changing relationship between health, science and fundamental rights. Unlike other sensitive

---

General Comment 14]; CESCR, General Comment 25: On science and economic, social and cultural rights (article 15 (1) (b), (2), (3) and (4) of the International Covenant on Economic, Social and Cultural Rights), U.N. Doc. E/C.12/GC/25 (Apr. 30, 2020) [hereinafter *General Comment 25*].

[32] *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Act*, COM (2021) 206 final, art. 6(1) (Apr. 21, 2021) [hereinafter *AIA*]; Commission Regulation 2017/745 of Apr. 5, 2017, on medical devices, 2017 O.J. (L. 117/1) arts. 2(1), 2(2).

[33] *AIA*, *supra* note 32, art. 6(2).

[34] *General Comment 14*, *supra* note 31, ⁋ 12.

[35] *General Comment 25*, supra note 31, ⁋ 16.

[36] *AIA*, *supra* note 32, art. 9(4).

areas, such as employment or education, healthcare is not directly addressed by the AIA, causing commentators to highlight its "conspicuous absence."[37]

b. Accessibility

For the right to health, accessibility encompasses physical and economic accessibility of health facilities, goods and services, information accessibility and prohibition of discrimination,[38] also by inequitable allocation of resources.[39] Similarly, the right to science presumes non-discrimination in participation in scientific progress, information about risk and benefits of technologies, and access to the applications of science, "particularly when they are instrumental for the enjoyment of other economic, social and cultural rights."[40] The Committee directly recognises a close connection between the right to science and right to health, emphasising the need to prioritise vulnerable groups, such as women, persons with disabilities and those living in poverty, in access to health technologies.[41] In particular, the Committee underlines the need to address multiple discriminations by involving intersectional minorities in decision making procedures concerning science.[42]

Thus, on the one hand, the right to science and the right to health require States to eradicate the digital divide by ensuring that cutting-edge AI technologies, such as personalised medicine, are not disproportionally accessible to the most affluent. On the other hand, States must facilitate involvement of the civil society, including marginalised minorities, in design and deployment of healthcare AI. Regrettably, the AIA does not explicitly address the issue of digital divide, nor does it provide for participatory approaches in design and deployment of new technologies. Researchers underline that the proposal suffers from democratic deficit, as it does not foresee public participation and consultation in conformity assessment and standard setting for high risk AI.[43] This deficiency is

---

[37] Hannah von Kolfschooten, *Conspicuous by its absence: health in the European Commission's Artificial Intelligence Act*, BMJ OPINION BLOG (July 30, 2021), https://blogs.bmj.com/bmj/2021/07/30/conspicuous-by-its-absence-health-in-the-european-commissions-artificial-intelligence-act/ (last visited Dec. 1, 2021).
[38] *General Comment 14*, *supra* note 31, ¶ 12.
[39] *Id.*, ¶ 19.
[40] *General Comment 25*, supra note 31, ¶ 16.
[41] *Id.*, ¶ 70.
[42] *Id.*, ¶ 35.
[43] Nathalie Smuha et al., *The EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial*

problematic, especially in cases of health emergencies, which, according to the proposal, could mandate a temporary derogation from conformity assessment.[44] The example of the COVID-19 pandemic has shown that vulnerable minorities are likely to suffer a disproportionate impact of emergency health policies.

Accessibility of information under the right to health and the right to science requires that patients are shielded from the opacity of "black box" healthcare. The right to science serves as a "mediator between the right to health and the IP rights,"[45] preventing tech companies from invoking trade secrets as a justification for denial of access to the algorithm for the purpose of independent audit or obtaining information about how a decision was reached. Unfortunately, the AIA does not strike the right balance between confidentiality and public interest. The notified bodies, which conduct conformity assessment of high-risk systems, are generally bound by secrecy requirements, unless the disclosure is required by the law.[46] The applicable whistle-blower exception under the Trade Secrets Directive (TSD)[47] mandates disclosure only in case of "misconduct, wrongdoing or illegal activity."[48] This is incompatible with the nature of algorithmic discrimination, which most commonly arises because of mistake or unexpected interactions between the labels. Furthermore, the TSD does not grant access to technical documentation to third parties. While the AIA provides some access rights to competent public authorities,[49] it does not enable external audits by independent researchers. This remains highly problematic, as most of the cases of algorithmic discrimination, especially in healthcare, have been detected by independent auditors in academia and civil society.

c. Acceptability

Acceptability under the right to health means that health facilities, goods and services are respectful of medical ethics, culturally appropriate and designed to improve the health of

---

Intelligence Act, SSRN (Aug. 31, 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3899991 (last visited Mar. 1, 2022), 54.

[44] *AIA*, *supra* note 32, art. 47(1).

[45] *General Comment 25*, supra note 31, ¶ 69.

[46] *AIA*, *supra* note 32, art. 33(6).

[47] *Id.* art. 70 (1)(a).

[48] Directive 2016/943 of the European Parliament and of the Council of 8 June 2016 on the Protection of Undisclosed Know-how and Business Information (Trade Secrets) Against Their Unlawful Acquisition, Use and Disclosure, 2016 O.J. (L 157/1) art. 5.

[49] *AIA*, *supra* note 32, arts. 33(6), 64(1)-(3), 70(2).

those concerned.[50] The right to science requires the products of science to be "tailored to particularities of populations with special needs," in order to avoid discrimination and preserve cultural diversity and pluralism.[51]

Thus, States must ensure that healthcare AI is free from biases and performs well on patients belonging to marginalised minorities. AI systems should be designed to address health inequalities, reflecting health equity concerns and distributive justice.[52] The AIA addresses the problem of algorithmic discrimination by introducing quality criteria for training, validation and testing of data sets, which 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.[53] However, in spite of its focus on the quality of data, the AIA fails to remedy the deficiencies of anti-discrimination law – it does not make a specific reference to the problem of proxy discrimination and intersectional discrimination. While its preamble mentions that AI can support socially beneficial outcomes in healthcare,[54] the act does not explore the distributive justice concerns in design and deployment of AI.

d. Quality

Finally, health facilities, goods and services must be of high quality, that is, scientifically and medically appropriate.[55] Everybody should enjoy access to the "most advanced, up-to-date and verifiable science."[56] Moreover, the Committee explicitly underlines that regulation and certification might be necessary to foster "the responsible and ethical development and application of science."[57]

Solving the problem of bias in healthcare AI requires regulation which does not only offer adequate redress in case of discriminatory practices, but also prevents them by setting appropriate safeguards. In this regard, AIA suffers from two crucial deficiencies. First, it fails to provide a direct redress

[50] *General Comment 14*, *supra* note 31, ¶ 12.
[51] *General Comment 25*, supra note 31, ¶ 19.
[52] *See generally* Alvin Rajkomar et al., *Ensuring Fairness in Machine Learning to Advance Health Equity*, 169 ANNALS OF INTERNAL MED. 866 (2018).
[53] *AIA*, *supra* note 32, art. 10.
[54] *Id.* recital 3.
[55] *General Comment 14*, *supra* note 31, ¶ 12.
[56] *General Comment 25*, supra note 31, ¶ 18.
[57] *Id.*

mechanism for individuals facing algorithmic discrimination. Second, it focuses heavily on self-assessment, allowing the providers of AI to meet the conformity requirements by following harmonised standards and thus, avoiding independent audit.[58] Over-reliance on self-assessment is likely to lead to discretionary results, letting bias go unnoticed.[59]

CONCLUSION

Algorithmic bias in healthcare AI creates critical dangers for vulnerable minorities, exacerbating existing health inequalities and perpetuating discriminatory practices in Europe. Since the EU anti-discrimination law alone is not well-equipped to deal with the issue of algorithmic discrimination in the context of health, healthcare AI should be subject to a comprehensive regulatory regime. Thus, just like in the case of prescription drugs, the right to health and the right to science can provide a useful framework for regulation, capturing the interdependence between science, health and fundamental rights.[60] Such a framework is suitable to address algorithmic discrimination because it encompasses both access and participation rights, requiring equal distribution of benefits and burdens and a focus on the marginalised.

[58] *AIA*, *supra* note 32, art. 40.
[59] Smuha, *supra* note 43, at 39.
[60] *See generally* Mike Frick & Gisa Dang, *The Right to Science: A Practical Tool for Advancing Global Health Equity and Promoting the Human Rights of People with Tuberculosis, in* THE RIGHT TO SCIENCE THEN AND NOW 246 (2022).