Vulnerability in the age of artificial intelligence: addressing gender bias in healthcare

Laura Piva*

ABSTRACT: Gender bias represent a source of vulnerability in current clinical practice as they can harm patients, especially those identified as gender minorities. This risk is increased in the age of AI-powered healthcare, which calls for redefining vulnerability and strategies to prevent algorithms from exacerbating already existing inequalities. To this end, we analyse the solutions proposed by the European Union's recent AI Act. Moreover, we consider whether AI itself can be a solution to these issues.

KEYWORDS: Vulnerability; artificial intelligence; gender bias; healthcare; technosolutionism.

1. Introduction

Vulnerability is a multifaceted concept which is of central importance to numerous areas, including medicine. Identifying vulnerable groups within the healthcare context represents the first step towards guaranteeing that everybody enjoys high-quality care and fair access to medical services. Ultimately, this means safeguarding and promoting the fundamental rights to health and equality, which are strongly linked[1].

In this article, we argue that the disruptive advent of artificial intelligence (AI) calls for reshaping the concept of vulnerability and rethinking the ways to deal with it.

On one hand, the digitalisation of our world exposes individuals and groups to new or increased physical and psychological risks and forms of inequality[2]. With AI, this happens partly due to the algorithmic discrimination phenomenon but also due to the uneven distribution of these technologies and the "digital divide".

---

* PhD Student, University of Trento. Mail: laura.piva-1@unitn.it. The article was subject to a double-blind peer review process.

[1] L. BUSATTA, *La salute sostenibile. La complessa determinazione del diritto ad accedere alle prestazioni sanitarie*, Torino, 2018, 1-10; M. TOMASI, *Sperimentazioni cliniche e medicina di genere: la ricerca dell'euguaglianza attraverso la valorizzazione delle differenze*, in B. PEZZINI, A. LORENZETTI (eds.), *70 Anni dopo tra uguaglianza e differenza: una riflessione sull'impatto del genere nella Costituzione e nel costituzionalismo*, Torino, 2019, 215-230.

[2] S.S. YILMAZ, Z.Ö. KOTIL, *Final exit before the bridge in AI: Strengthening the right to human oversight of vulnerable subjects with two agents,* in *European Journal of Privacy Law & Technology - Observatory,* 2023, available at: https://universitypress.unisob.na.it/ojs/index.php/ejplt/article/view/1857 (last accessed: 21/11/2024).

On the other hand, there is the hope that AI will shed light on human bias and mitigate or even overcome them.

The present article assesses these assumptions by taking gender bias in healthcare as a case study. After investigating the origin and role of gender bias in healthcare, we tried to understand how they can be embedded in algorithms and how AI can reproduce or expand them, with a negative impact on individuals' right to health and equality.

This operation implied delving into human and algorithmic bias, for which consulting interdisciplinary literature relating to medicine, bioethics and technology was mandatory.

After having defined this problem, we will review some possible solutions. First, we will draw some reflections on whether AI itself could be used to overcome gender bias in healthcare. Then we will focus on the way the European Union's law deals with this issue with the recently approved Regulation on Artificial Intelligence (AI Act). In the conclusions, we consider how AI can play a role in raising awareness towards existing (gender) bias, thus encouraging developers, policymakers and regulators to seriously address healthcare disparities abandoning the myth of AI solutionism.

2. Vulnerable patients

Although its meaning is not univocal, vulnerability can generally be understood as the propensity to be easily physically or mentally hurt, influenced or attacked[3].

If we adopt an ontological perspective, this is a condition universally experienced by human beings[4] due to their «shared biological fragility»[5] and their relational nature, which makes them dependent upon others and, thus, in need of care[6].

Besides being universal, vulnerability can be understood as context-specific as well, meaning that individuals or groups with certain characteristics are exposed to an increased risk of harm due to social, political, economic, or environmental factors[7].

Both these perspectives are reflected in the healthcare context. On one hand, experiences such as illness, pain and hospitalisation make all patients vulnerable. Moreover, patients depend upon

---

[3] "Vulnerability", Cambridge Dictionary: https://dictionary.cambridge.org/dictionary/english/vulnerability (last accessed: 01/07/2024).

[4] M.A FINEMAN, *The vulnerable subject: Anchoring equality in the human condition,* in *Yale Journal of Law and Feminism,* 20, 1, 2008, 1-23.

[5] This expression is taken from W. ROGERS, *Vulnerability in Bioethics*, in C. MACKENZIE, W. ROGERS, S. DODDS (eds.), *Vulnerability: New Essays in Ethics and Feminist Philosophy*, Oxford, 2013, 71.

[6] C. BOTTI, *Vulnerabilità, relazioni e cura. Ripensare la bioetica*, in *Etica & Politica / Ethics & Politics*, 18, 3, 2016, 33-57. This is in line with the Oxford Reference's definition of "vulnerability", namely «the position of relative disadvantage, which requires a person to trust and depend upon others», see: https://www.oxfordreference.com/display/10.1093/oi/authority.20110803120303277 (last accessed: 01/07/2024).

[7] B. WISNER, P. BLAIKIE, T. CANNON, I. DAVIS, *At Risk. Natural Hazards, People's Vulnerability and Disasters*, second ed., London, 2004 describe vulnerability as the set of characteristics of an individual or group that, combined with the situation or context in which they live, influence their ability to adapt, resist and anticipate the impact of adverse events. Similarly, ten Have refers to a limited capacity or resilience to absorb, adapt to or recover from that harm: H. TEN HAVE, *Vulnerability: Challenging bioethics*, London, 2016. See also: C. MACKENZIE, W. ROGERS, S. DODDS (eds.), *Vulnerability: New Essays in Ethics and Feminist Philosophy*, Oxford, 2013.

Special issue

physicians due to the information asymmetry existing between them[8], even though the shift from medical paternalism to more participatory models of care has partly reduced this gap[9].

At the same time, vulnerability is indisputably shaped by power dynamics[10] and, therefore, some categories of patients historically disadvantaged have been exposed to higher risks or more detrimental consequences than others based on their ethnicity, religion, age, income, sex and gender.

While acknowledging that all these variables – and their intersection[11] – are important to identify vulnerable subjects in healthcare and address the disparities they face and their specific needs, this paper focuses on sex and gender only.

First, we will try to understand which gender biases exist in healthcare and why they can be a source of vulnerability (i.e., of increased risk of harm or discrimination) in this field.

2.1. The role of sex and gender in medicine

Medical literature has proved sex and gender[12] to be relevant in medicine, as they can influence the likelihood of developing a certain condition, its risk factors and development, symptoms, therapeutic

---

[8] L.A. COYLE, S. ATKINSON, *Vulnerability as practice in diagnosing multiple conditions*, in *Medical Humanities*, 45, 3, 2019, 278-286.

[9] This process, which has led to greater recognition for patients' autonomy, has however not affected everybody equally, as noted by L. BUSATTA, C. CASONATO, S. PENASA, M. TOMASI, *Le "maschere" della vulnerabilità nella cura della persona*, in AA. VV. (eds.), *Liber Amicorum per Paolo Zatti*, vol. 1, Napoli, 2023, pp. 651-663 in relation to minors, elderly, disabled and prisoned patients. See also W. ROGERS, *op. cit.*: «the greater the knowledge and skills imbalance between the practitioner and patient, the more vulnerable the patient is to harm from their medical attendant, and the more important it is that practitioners are bound to protect patient interests». For an overview of different models of care: E.J. EMANUEL, L.L. EMANUEL, *Four Models of the Physician-Patient Relationship,* in *JAMA*, 267, 16, 1992, 2221-2226.

[10] K. TIERNEY, *Disasters: A Sociological Approach*, Cambridge, 2019, 127: «people are not born vulnerable, they are made vulnerable […] different axes of inequality combine and interact to form systems of oppression – systems that relate directly to differential levels of social vulnerability».

[11] C.H.A. KURAN ET AL.*, Vulnerability and vulnerable groups from an intersectionality perspective,* in *International Journal of Disaster Risk Reduction*, 50, 101826, 2020, 1-8, https://doi.org/10.1016/j.ijdrr.2020.101826; F. LUNA, *Identifying and evaluating layers of vulnerability – a way forward,* in *Developing World Bioethics*, 19, 2, 2019, 86-95.

[12] Defining "sex" and "gender", and how such terms will be used in the present paper is of paramount importance. "Sex" refers to the individual's biological characteristics (sexual chromosomes XX and YX, genitalia, sexual hormones) at birth, whereas "gender" indicates the behaviours, attitudes, feelings and role perception that a person has of herself (gender identity) or that a culture attributes to an individual (D. CIRILLO ET AL., *Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare*, in *npj Digital Medicine*, 3, 81, 1 June 2020, 1-11, https://doi.org/10.1038/s41746-020-0288-5). Thus, sex is a biological construct while gender a psychological and social one. As for what concerns this article, the examples that will be reported – which were derived from relevant medical literature – mainly address differences between males/men and females/women, however, this does not imply that sex or gender are binary constructs (see: T.E. MADSEN ET AL., *Article Commentary: Sex- and Gender-Based Medicine: The Need for Precise Terminology*, in *Gender and the Genome*, 1, 3, 2017, 122-128; S.S. RICHARDSON, *Sex Contextualism*, in *Philosophy, Theory, and Practice in Biology*, 14, 2, 2022, 1-17, https://doi.org/10.3998/ptpbio.2096). "Gender bias" will be used in relation to prejudices and assumptions concerning both sex and gender.

needs, drug effectiveness and mortality[13]. Moreover, gender is a social determinant of health which affects aspects like lifestyle, access to healthcare and help-seeking behaviours[14]. Being aware of such differences is pivotal to realise accurate prevention, diagnosis and treatment, as well as healthcare policies directed at achieving equity.

Notwithstanding, medicine has long been an androcentric discipline, ignoring the specificities and needs of all of those deviating from the standard cisgender male model[15].

Even if in the last decades such issues have started being addressed by gender-specific medicine[16], problems persist as clinical trials' samples are still under-representative of women and sex and gender minorities and research towards their specific health conditions and needs gains scarce funding and limited attention[17].

In any case, disparities based on gender can not only be explained by the lack of representative medical data. For instance, healthcare practitioners' attitudes and responses can change depending on whether they perceive the patient as a man or a woman, undermining fair healthcare access and de-livery[18].

2.2. Gender bias as a source of vulnerability

What has been described so far highlights two opposite and equally damaging attitudes in medicine: not accounting for sex and gender when such variables have an actual impact on the illness experience and, on the contrary, hyper-focusing on them when they do not[19].

---

[13] S. GARATTINI, R. BANZI, *Una medicina che penalizza le donne*, Cinisello Balsamo, 2022; M.J. LEGATO, P.A. JOHNSON, J.E. MANSON, *Consideration of sex differences in medicine to improve health care and patient outcomes*, in *JAMA*, 316, 18, 2016, 1865-1886; S. GREGO ET AL., *"Sex and gender medicine": il principio della medicina di genere*, in *Giornale Italiano di Cardiologia*, 21, 8, 2020, 602-606.

[14] The WHO individuated gender as one of the core social determinants of health (together with income, education, occupation, social class and ethnicity) in O. SOLAR, A. IRWIN, *A conceptual framework for action on the social determinants of health. Social determinants of health discussion paper 2 (policy and practice),* Geneva, 2010. On the same topic: F. MAUVAIS-JARVIS ET AL., *Sex and gender: modifiers of health, disease, and medicine*, in *Lancet*, 396, 10250, 2020, 565-582; N. BUSLÓN, A. CORTÉS, S. RACIONERO-PLAZA, *Sex and gender inequality in precision medicine: Socioeconomic determinants of health*, in D. CIRILLO, S. CATUARA SOLARZ, E. GUNEY (eds.), *Sex and Gender Bias in Technology and Artificial Intelligence: Biomedicine and Healthcare Applications*, Amsterdam, 2022, 35-54.

[15] For instance, medical research and clinical trials have historically been exclusively or predominantly conducted on men, and the results thereby obtained have been generalised to the whole population. Ironically, this was partly due to the willingness to protect "vulnerable women" from the risks of human experimentation but ended up drawing incorrect and often harmful conclusions towards them and other patients' groups. See: M. FASAN, C.M. REALE, *Genere e sperimentazioni cliniche: il Regolamento (UE) n. 536/2014, un'occasione mancata?* in *Bio-Law Journal – Rivista di BioDiritto*, 4, 2022, 251-276; T.E. MADSEN ET AL., *op. cit.*, 122-128.

[16] G. BAGGIO ET AL., *Gender medicine: a task for the third millennium*, in *Clinical Chemistry and Laboratory Medicine*, 51, 4, 2013, 713-727; S. GREGO ET AL., *op. cit.*, 602-606.

[17] M. TOMASI, *op. cit.*, 215-230.

[18] For example, K. HAMBERG, *Gender bias in medicine*, in *Women's Health*, 4, 3, 2008, 237-243 highlighted that physicians tend to interpret men's symptoms as organic and women's as psychosocial, leading to inaccurate or delayed diagnosis for the latter.

[19] M. SUNDAI ET AL., *Law, policy, biology, and sex: Critical issues for researchers*, in *Science*, 376, 6595, 2022, 802-804; I. STRAW, *The automation of bias in medical Artificial Intelligence (AI): Decoding the past to create a better*

Hence, we can identify two types of gender bias negatively affecting healthcare: those leading to "gender blindness" (i.e., ignoring true biological and social differences)[20] and those creating false or discriminatory assumptions.

When medical decision-making is tainted by such prejudices, there can be serious consequences such as missed or inaccurate diagnoses, ineffective or less effective treatments, adverse effects, and wrong prioritisation during triage or emergency admissions (e.g., due to the downplay of symptoms or underestimation of predictive factors). All of this can put individuals' right to health in danger. This is particularly critical when those most at risk are the same patients who already experience poorer health and/or face barriers to accessing healthcare services due to historical or pre-existing inequities. Thus, gender bias can be qualified as a source of vulnerability for patients.

These issues reach paramount and renovated relevance in the age of AI-powered healthcare, as algorithms are capable of replicating and amplifying such bias and their discriminatory effects.

3. Vulnerable algorithms

Although AI carries the promise of yielding medical decision-making more accurately and fairly by correcting human biases[21] such technology is not neutral[22].

Conversely, AI can incorporate values and prejudices and, thus, be vulnerable to bias «that may disproportionately affect model performance in a certain subgroup»[23].

This means that gender bias illustrated in the previous section can be embedded in algorithms, according to the well-known "garbage in, garbage out" principle[24]. For instance, machine learning (ML) algorithms trained on datasets under-representative of the female population showed sex-related performance disparities in predicting heart failure[25] or the likelihood of developing liver diseases[26]. In particular, the systems produced a significantly higher rate of false negative results when applied to women,

---

[20] D. CIRILLO ET AL., *op. cit.*, 1-11.

[21] C.R. SUNSTEIN, *Algorithms, correcting bias*, in *Social Research: An International Quarterly*, 86, 2, 2019, 499-511.

[22] P. TRAVERSO, *Breve introduzione tecnica all'Intelligenza Artificiale*, in *DPCE online*, 1, 2022, 155-167; E. STRADELLA, *Stereotipi e discriminazioni: dall'intelligenza umana all'intelligenza artificiale*, in AA. VV. (eds.) *Liber Amicorum per Pasquale Costanzo Costanzo – Diritto Costituzionale in trasformazione Vol. I – Costituzionalismo, Reti e Intelligenza artificiale*, Genova, 2020, 391-400.

[23] J.K. PAULUS, D.K. KENT, *Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities*, in *npj Digital Medicine*, 3, 99, 2020, 4, https://doi.org/10.1038/s41746-020-0304-9. See also: A.J. LARRAZABAL ET AL., *Gender imbalance in medical imaging datasets produces biased classifiers for computer aided diagnosis*, in *Proceedings of the National Academy of Sciences of the United States of America*, 117, 23, 2020, 12592-12594.

[24] R. XENIDIS, L. SENDEN, *EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination*, in U. BERNITZ ET AL. (eds.), *General Principles of EU law and the EU Digital Order*, Alphen aan den Rijn, 2020, 151-182.

[25] I. STRAW, G. REES, P. NACHEV, *Sex-based Performance Disparities in Machine Learning Algorithms for Cardiac Disease Prediction: Exploratory Study*, in *Journal of Medical Internet Research*, 26, 26 August 2024, 1-18, https://doi.org/10.2196/46936.

[26] I. STRAW, H. WU, *Investigating for bias in healthcare algorithms: A sex-stratified analysis of supervised machine learning models in liver disease prediction*, in *BMJ Health and Care Informatics*, 29, 1, 2022, 1-8.

resulting in missed diagnoses and a consequent lack of appropriate and timely care for this patient group.

Even when the data sample is diverse, major errors can arise from how such data were interpreted, selected, cleaned, formatted, and labelled before building the training dataset[27].

Besides data, a second entry point for bias is the algorithm's design which encompasses the definition of the system's objectives and target population, the selection of the model's relevant features and their weight, and the choice of training, testing and validation methodologies[28].

We can conclude that algorithms' "technical" vulnerabilities can make patients vulnerable by replicating existing biases which undermine their right to health and equality. In fact, model performance disparities not only put single individuals' safety at stake but can result in discrimination when patients are negatively affected due to belonging to a certain group[29].

At the same time, I argue that AI generates further forms of vulnerability, both from a quantitative and qualitative point of view.

On one hand, ML, DL and generative AI can not only reproduce biases but also multiply them due to "feedback loops"[30]. On the other, algorithms' spurious correlations could generate inaccurate predictions based on different (and even unexpected) characteristics or combinations of them (e.g., gender, ethnicity, income), leading to intersectional discrimination and reinforcing historical inequities[31].

This shows how algorithms can exponentially enhance inequality and normalise it, as the optimism surrounding AI and the opacity that usually characterises it make it very likely that such phenomena will go undetected[32].

Thus, medical AI calls for looking for ways to target algorithmic discrimination as a source of vulnerability.

---

[27] For instance, medical data (e.g., parameters and biochemical thresholds) could have been aggregated so that sex and gender indicators do not emerge. Likewise, they could only account for binary definitions of such categories. See: A. Gerybaite, S. Palmieri, F. Vigna, *Equality in Healthcare AI: Did Anyone Mention Data Quality?*, in *BioLaw Journal – Rivista di BioDiritto*, 4, 2022, 385-409.

[28] For instance, an AI model could be built around and item, such as an x-ray scan, which is more informative for the male population than for the female one in detecting a certain medical condition: M. Ganz, S.H. Holm, A. Feragen, *Assessing Bias in Medical AI*, in *Workshop on Interpretable ML in Healthcare at International Connference on Machine Learning (ICML)*, 2021, available at: https://www.cse.cuhk.edu.hk/~qdou/pub-lic/IMLH2021_files/64_CameraReady_ICML_2021_Interpretable_Machine_Learning_in_Healthcare_work-shop.pdf (last accessed: 21/11/2024).

[29] P.L. Lau, *AI Gender Biases in Women's Healthcare: Perspectives from the United Kingdom and the European Legal Space*, in E. Gill-Pedro, A. Moberg (eds.), *YSEC Yearbook of Socio-Economic Constitutions 2023: Law and the Governance of Artificial Intelligence*, Cham, 2023, 247-274; E. Stradella, *op. cit.*, 391-400.

[30] "Feedback loops" occur when previous AI's outputs influence the future ones.

[31] L. Goetz, N. Seedat, R. Vandersluis, M. van der Schaar, *Generalization – a key challenge for responsible AI in patient-facing clinical applications*, in *npj Ditigal Medicine*, 7, 126, 21 May 2024, 1-4, https://doi.org/10.1038/s41746-024-01127-3.

[32] R. Walker, J. Dillard-Wright, *Algorithmic bias in artificial intelligence is a problem – And the root issue is power*, in *Nursing Outlook*, 71, 102023, 2023, 1-4.

4. Looking for a cure

4.1. Can AI be a solution?

First, we need to consider whether AI itself could be the solution to gender bias and algorithmic discrimination[33]. In fact, one could argue that such issues might be solved by building "better algorithms". Although tempting, the idea of using technology to eliminate bias is utopistic and reductive as it fails to conceive them as a "symptom of power imbalances"[34]. More generally, "technological solutionism"[35] has been criticised as it aims to solve complex social issues – such as healthcare inequities – with technological means only, while they require political and legal action.

This, however, does not imply that algorithms can't play a positive role in overcoming gender bias in medicine. The major aid that AI can give in this respect is to make them evident.

For instance, some authors propose to use algorithms such as linear regression or decision trees to achieve post hoc explanations of opaque AI systems. Turning a black box into a white one or using explainable AI could help identify gender bias, which constitutes the first step to establishing whether they are desirable (i.e., they reflect true biological or social differences) or not[36].

After having established so, a possible strategy might be selective deployment of medical AI tools, meaning that they will be used in relation to the population for which they are able to derive accurate conclusions only[37]. This appears in line with the idea that using AI might not always be desirable, especially when efficiency might be counteracted by discriminatory or dangerous outcomes.

Another possibility might be adopting measures to mitigate bias. Once again, some researchers have proposed to exploit AI to this end, elaborating algorithms suitable not only to detect but even to correct certain biases (e.g., balancing uneven datasets ex-post). However, when problems are not related to datasets' lack of diversity corrections might be difficult to achieve.

Above all, there must be an acknowledgement that technical solutions cannot answer the underlying societal problems that have led to the creation of gender stereotypes or to disregard the importance of conducting medical research adopting a gender perspective.

Moreover, as discussed above, gender bias in healthcare not only originate from social problems but also jeopardise human rights such as the right to health and equality. Therefore, we must now examine how the law could tackle gender bias and vulnerability, focusing on the EU legal landscape.

---

[33] C.R. SUNSTEIN, *op. cit.*, 499-511.

[34] R. WALKER, J. DILLARD-WRIGHT, *op. cit.*, 1-4.

[35] According to E. MOROZOV, *To Save Everything, Click Here: Technology, Solutionism and the Urge to Fix Problems that Don't Exist*, New York, 2013, 5-16, the term refers to the erroneous belief that every political, social, organizational, administrative, and policy problem can be addressed with technological solutions.

[36] R. CONFALONIERI ET AL., *A unified framework for managing sex and gender bias in AI models for healthcare*, in D. CIRILLO, S. CATUARA SOLARZ, E. GUNEY (eds.), *Sex and Gender Bias in Technology and Artificial Intelligence: Biomedicine and Healthcare Applications,* Amsterdam, 2022, 179-204; P. CHANDAK, N.P. TATONETTI, *Using Machine Learning to Identify Adverse Drug Effects Posing Increased Risk to Women*, in *Patterns – Cell Press*, 1, 7, 2020, 1-15.

[37] L. GOETZ ET AL.*, op. cit.*, 1-4.

4.2. The AI Act approach

The European Union has decided to tackle the risks that AI poses to health, safety and fundamental rights with the horizontal Regulation 2024/1689, known as the AI Act[38].

The way the AI Act strives to protect such rights is *ex-ante*, namely by requiring developers and deployers to incorporate certain requirements into their AI systems[39]. As the Regulation follows a risk-based approach, most of the measures indicated therein are mandatory for high-risk systems only. This category includes some systems to be deployed in the healthcare sector, such as medical and in vitro diagnostic medical devices[40], systems «intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for essential public assistance benefits and services, including healthcare services» and «emergency healthcare patient triage systems»[41].

It is among these requirements, in particular the one related to data and data governance (Art. 10) that we find a possible solution to algorithmic discrimination, as well as a reference to vulnerable subjects.

The main idea expressed by Article 10 is that to produce accurate and non-discriminatory results, AI systems must be trained, validated and tested with high-quality datasets. This means, *inter alia*, that data must have been collected and processed correctly, that they must be relevant to the context and, to the best extent, error-free and complete. Also, it is specified that data shall be «sufficiently representative», and have «the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons in relation to whom the high-risk AI system is intended to be used»[42].

The AI Act recalls the need to balance these objectives with the right to data protection. However, it establishes a new exception for processing special categories of data according to the GDPR – which

---

[38] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689&qid=1729690544660#d1e2090-1-1 (last accessed: 21/11/2024). The purpose of the AI Act, which is the world-first horizontal regulation on artificial intelligence, is to «improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence, while ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union and supporting innovation» (Art. 1, Reg. EU 2024/1689).

[39] Art. 3(1) defines AI system as «machine-based system designed to operate with varying levels of autonomy, 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».

[40] As long as they require a third-party conformity assessment, see: Art. 6(1) and Annex I, Reg. EU 2024/1689.

[41] Art. 6(2) and Annex III, 5(a) and (d), Reg. EU 2024/1689. However, some other AI systems used in the medical context might be classified as "limited" or "minimal risk. In any case, those systems will still need to adhere to specific transparency requirements when interacting with natural persons (Art. 50, Reg. EU 2024/1689). Moreover, chatbots providing medical advice might have to implement the requirements indicated for generative AI. For some considerations related to medical AI and (draft) AI Act's provisions: S. PALMIERI, T. GOFFIN, *A Blanket That Leaves the Feet Cold: Exploring the AI Act Safety Framework for Medical AI*, in *European Journal of Health Law*, 30, 4, 2023, 406-427; H. VAN KOLFSCHOOTEN, *EU regulation of Artificial Intelligence: challenges for patients' rights*, in *Common Market Law Review,* 59, 1, 2022, 81-112.

[42] Art. 10(3), Reg. EU 2024/1689.

include data regarding sex and gender – «to the extent that it is strictly necessary for the purpose of ensuring bias detection and correction»[43]. In fact, Article 10 also requires putting in place appropriate measures to detect, prevent and mitigate possible biases, acknowledging that AI can «perpetuate and amplify existing discrimination, in particular for persons belonging to certain vulnerable groups»[44].

A second reference to vulnerability is contained in Recital 165, which states that providers and deployers of AI systems of all risk classes should be «encouraged to apply on a voluntary basis additional requirements related, for example, to the elements of the Union's Ethics Guidelines for Trustworthy AI, […] inclusive and diverse design and development of AI systems, including attention to vulnerable persons […] and diversity of the development teams, including gender balance»[45].

These provisions highlight that EU institutions are aware of the fact that datasets and the design of AI systems constitute entry points for bias and their potential discriminatory effects (see §3).

However, regardless of wishing for more representation inside the developers' teams and for stakeholders' participation in the design of AI systems, the AI Act does not make such requirements mandatory, nor it explains how these results should be achieved.

Therefore, criticism has been made that the solutions proposed by the AI Act target technological issues while overlooking the upstream social problems.

One could thus argue that the EU Regulation on artificial intelligence is, to a certain extent, flawed by technological solutionism as well[46].

At the same time, however, it must be noted that the AI Act is a horizontal product compliance Regulation which shapes the rights of EU citizens as "consumer and (product) safety rights"[47]. Then the question is whether such a legal instrument is the most appropriate one to address the problem of (gender) discrimination in healthcare and, above all, to protect the right to health and other human rights[48].

Certainly, the AI Act can constitute a useful starting point but alone it cannot guarantee that we build fair and trustworthy AI systems that make accurate predictions, diagnoses and treatment recommendations for all patients. Which actors shall integrate this source, at which level and by which means remain open questions.

---

[43] Art. 10(5), Reg. EU 2024/1689.

[44] Recital 67, Reg. EU 2024/1689.

[45] A pivotal role, in this sense, should be played by the AI Office and Member States which are invited to encourage and facilitate the drawing up of codes of conduct (Art. 95, Reg. EU 2024/1689). Specific references to vulnerable persons and to gender balance were not contemplated by the first version of this provision, namely Art. 69 of the AI Act proposal by the EU Commission.

[46] B. PHAM, S.R. DAVIES, *What problems is the AI act solving? Technological solutionism, fundamental rights, and trustworthiness in European AI policy,* in *Critical Policy Studies*, 2 July 2024, 1-19, https://doi.org/10.1080/19460171.2024.2373786.

[47] *Ivi*, 14.

[48] Z. ZŐDI, *The EU AI Act – Can We Protect Human Rights with a Product Compliance Regulation?*, in *IACL-AIDC Blog,* 4 June 2024, available at: https://blog-iacl-aidc.org/2024-posts/2024/6/4/the-eu-ai-act-can-we-protect-human-rights-with-a-product-compliance-regulation (last accessed: 27/06/2024).

5. Conclusions

In this article, we have conceived vulnerability as a greater propensity to suffer damage or prejudices, focusing on a specific context: healthcare. We have also seen that a context-specific approach to vulnerability can serve to identify and respond to inequalities and different needs.

At the same time, we have argued that vulnerability in healthcare has been shaped by power dynamics. As it is evident that AI is a new form of power, its development and deployment in clinical settings cannot be viewed as a miracle solution but must be carefully addressed.

With regard to gender bias and resulting vulnerabilities, AI is a "double-edged sword"[49]. On the one side, algorithms can amplify and act as an "echo chamber" of existing sex and gender inequalities. Thus, they shall be an object of regulation[50], even if the AI Act alone appears inadequate to target the issues relating to discrimination and inequities.

An option could be integrating this source with sectorial laws, policy instruments or professional guidelines for medical researchers and healthcare professionals. In all cases, including a gender perspective could facilitate a fair development, deployment and use of medical AI.

In fact, just as data quality is affected by the biases of researchers, AI design is heavily influenced by those of developers, which is why the need to reach gender balance and diversity within engineering teams is highlighted by several academic articles, soft law and policy documents[51].

On the other side, algorithms have the potential to reduce these inequalities when properly designed, as they can contribute to bias detection and foster social changes[52].

Indeed, all the attention that AI is gaining is shedding light on bias and their impact[53]. AI could therefore provide an opportunity to reflect on measures to address not only algorithmic biases but also our human biases, such as our prejudices about sex and gender. This way, the advent of AI, when accompanied by a serious political reflection, could allow us to move forward from unfair research and healthcare policies and practices and to leave stereotypes rooted in power imbalances behind.

---

[49] D. CIRILLO ET AL., *op. cit.*, 1.

[50] S. PENASA, *Verso un diritto "technologically immersive": la sperimentazione normativa in prospettiva comparata,* in *DPCE online*, 1, 2023, 671-696.

[51] Examples of supranational soft law recommending diversity and inclusion in teams developing medical AI systems are, for instance, the ERPS's report "Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts" (2022), available at: https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2022)729512 (last accessed 21/11/2024) and the WHO's "Guidelines on Ethics and Governance of Artificial Intelligence for Health" (2021), available at: https://www.who.int/publications/i/item/9789240029200 (last accessed 21/11/2024).

[52] H.S. SÆTRA, E. SELINGER, *The Siren Song of Technological Remedies for Social Problems: Defining, Demarcating, and Evaluating Techno-Fixes and Techno-Solutionism*, in *SSRN Electronic Journal*, 19 September 2023, 1-18, http://dx.doi.org/10.2139/ssrn.4576687.

[53] S. LINDGREN, V. DIGNUM, *Beyond AI solutionism*, in S. LINDGREN (ed.), *Handbook of Critical Studies of Artificial Intelligence: sociopolitical process and decisions become computationally streamlined*, Cheltenham, 2023, 167-172.