QUEER-RESPONSIVE REGULATION FOR ARTIFICIAL INTELLIGENCE IN HEALTHCARE: A COMPARATIVE STUDY

SERGIO SULMICELLI*

This article employs a queer theoretical framework to analyse algorithmic discrimination in healthcare-related artificial intelligence, with particular attention to the biases affecting sexual and gender minorities. It offers a comparative analysis of three regulatory models: principle-based; technical-oriented; and sociotechnical-oriented, by examining the European Union Artificial Intelligence ('AI') Act, the Council of Europe Framework Convention on AI, and the Brazilian AI Bill. This article argues for a queer-responsive regulatory approach capable of addressing biases embedded in AI systems and promoting more inclusive and equitable healthcare technologies.

I INTRODUCTION

The issue of algorithmic discrimination has received extensive attention in legal and policy discussions at national, international, and comparative levels. However, the debate has predominantly focused on formalistic approaches, highlighting the limitations of traditional equality and anti-discrimination law, which rely mainly on technical or legalistic solutions. Algorithmic discrimination, however, is a sociotechnical phenomenon that demands solutions which integrate both social and technological considerations to counter algorithmic bias.

This article employs a queer approach[1] to examine how artificial intelligence ('AI') systems, particularly in healthcare, produce and perpetuate discrimination

* PhD in Comparative and European Legal Studies, University of Trento. Academic Fellow in Comparative Public Law, Bocconi University. The author is grateful to the anonymous reviewers for their insightful comments and constructive suggestions which have greatly improved the final version of this article, and to the Editor of the Issue, Bisesh Belbase, for editorial support.

1 On the queer approach and queer theories more broadly, see the foundational work of Teresa de Lauretis: Teresa de Lauretis, *Queer Theory: Lesbian and Gay Sexualities* (1991) 3(2) *Differences* iii <https://doi.org/10.1215/10407391-3-2-iii>. See also Eve Kosofsky Sedgwick, *Epistemology of the Closet* (University of California Press, 1990). For queer legal theory, see Francisco Valdes, 'Afterword and Prologue: Queer Legal Theory' (1995) 83(1) *California Law Review* 344 <https://doi.org/10.2307/3480882>; Brenda Cossman, 'Gender Performance, Sexual Subjects and International Law' (2002) 15(2) *Canadian Journal of Law and Jurisprudence* 281 <https://doi.org/10.1017/S0841820900003623>; Damir Banović, 'Queer Legal Theory' in Dragica Vujadinović, Antonio Álvarez del Cuvillo and Susanne Strand (eds), *Feminist Approaches to Law: Theoretical and Historical Insights* (Springer, 2023) 73 <https://doi.org/10.1007/978-3-031-14781-4>.

against sexual and gender minorities. A queer approach challenges fixed categories of sex, gender, and sexuality, emphasising their socially constructed nature. Queer theory questions the neutrality of law, exposing its role in maintaining structures of discrimination and inequality against those who deviate from dominant norms. Additionally, it challenges the neutrality of AI, recognising how AI systems replicate and intensify existing biases due to their reliance on binary classifications. As a matter of fact, AI's predictive and autonomous decision-making capabilities, combined with technical complexities such as 'black box' algorithms and biased datasets, risk exacerbating the vulnerabilities of already marginalised groups, including women, transgender, intersex, and non-binary individuals.

To explore this, the article first defines key concepts, including algorithmic bias, and outlines a framework for analysing discrimination through both technical and sociotechnical lenses. The discussion then progresses by reviewing specific examples within healthcare, such as biases in diagnostic algorithms and treatment recommendations. Subsequently, this article conducts a comparative analysis of three regulatory models: principle-based; technical-oriented; and sociotechnical-oriented, to assess their responsiveness to queer-specific challenges posed by AI biases.

In conclusion, this article advocates for a comprehensive, queer-responsive approach that integrates substantive equality into AI regulation, emphasising diversity, participation, and a queer understanding of sex, gender, and sexuality data.

This article is structured as follows: Part II delves into the concepts of algorithmic bias and discrimination in AI systems, critically examining how these biases emerge through technical, foundational, and sociotechnical lenses, particularly emphasising the distinct sources of bias relevant to sex, gender, and sexuality. Part III focuses specifically on healthcare, providing an in-depth analysis of sex and gender bias within AI applications, supported by concrete case studies that illustrate the real-world impacts of such biases on healthcare outcomes for gender and sexual minorities.

Part IV undertakes a comparative analysis of regulatory approaches designed to address algorithmic bias and discrimination. It assesses three regulatory frameworks: a principle-based model exemplified by the Council of Europe *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* ('*The Convention*');[2] a technical-oriented model illustrated by the *European Union Artificial Intelligence Act* ('*EU AI Act*');[3] and a sociotechnical-oriented model represented by Bill Nº 2338 of 2023 (Providing for the Use of Artificial Intelligence) ('Brazilian AI Bill').[4] Each model is analysed for its strengths and limitations to address queer-specific biases.

---

2 *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law,* opened for signature 5 September 2024, CETS No 225 (not yet in force) ('*The Convention*').

3 *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)* [2024] OJ L 2024/1689 ('*EU AI Act*').

4 Projeto de Lei Nº 2338, de 2023 (Dispõe sobre o uso da Inteligência Artificial) [Bill Nº 2338 of 2023 (Providing for the Use of Artificial Intelligence)] ('Brazilian AI Bill') [tr author].

Finally, the conclusion (Part V) synthesises the findings, advocating for a comprehensive, queer-responsive approach that integrates substantive equality into AI regulation. It argues for policy frameworks that emphasise diversity, meaningful participation of affected communities, and a queer-informed understanding of sex, gender, and sexuality data in AI governance.

II THE CONCEPT(S) OF BIAS AS A SOURCE OF AI-BASED DISCRIMINATION

Generally speaking, the term 'bias' 'simply refers to a deviation from the standard'.[5] In the context of AI, bias is often necessary to identify and weight statistical patterns within data, allowing the system to classify and distinguish between different instances effectively.[6] Bias enables AI models to make predictions by emphasising certain features or trends in the data.[7]

By relying on the work of Batya Friedman and Helen Nissenbaum, I will use the term 'biased AI systems' to indicate AI tools that:

> systematically and unfairly discriminate against certain individuals or groups of individuals in favor of others. A system discriminates unfairly if it denies an opportunity or a good or if it assigns an undesirable outcome to an individual or group of individuals on grounds that are unreasonable or inappropriate.[8]

Notably, scholarship in the field of AI discrimination, from sociological, ethical, legal, or technical aspects, has long proposed several typologies of 'entry points' of bias in AI systems that can lead to a discriminatory outcome.[9]

In their work,[10] David Danks and Alex John London have distinguished three main causes for biases: i) bias in modelling, referring to the process of deliberately introducing bias to mitigate and compensate for bias in the data; ii) bias in training, referring to the fact that algorithms learn to make decisions or predictions based

---

5 David Danks and Alex John London, 'Algorithmic Bias in Autonomous Systems' (Conference Paper, International Joint Conference on Artificial Intelligence, 19–25 August 2017) 4692 <https://doi.org/10.24963/ijcai.2017/654>, highlighting also that '[c]rucially, the very same thing can be biased according to one standard, but not according to another'.

6 Xavier Ferrer et al, 'Bias and Discrimination in AI: A Cross-Disciplinary Perspective' (2021) 40(2) *Institute of Electrical and Electronics Engineers Technology and Society Magazine* 72.

7 Ibid 72–3.

8 Batya Friedman and Helen Nissenbaum, 'Bias in Computer Systems' (1996) 14(3) *Association for Computing Machinery Transactions on Information Systems* 330 (emphasis omitted) (citations omitted).

9 The literature on the topic is vast. In this article I will refer to some of the taxonomies recently proposed as they will be more relevant for the analysis. However, it seems to be necessary to at least account for the other scholarship that have worked on the topic from various perspectives, as all have contributed to my understanding of the topic: see, eg, Tal Z Zarsky, 'An Analytic Challenge: Discrimination Theory in the Age of Predictive Analytics' (2017) 14(1) *Information Systems* 11; Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan, 'Semantics Derived Automatically from Language Corpora Contain Human-Like Biases' (2017) 356(6334) *Science* 183 <https://doi.org/10.1126/science.aal4230>; Ignacio N Cofone, 'Algorithmic Discrimination Is an Informational Problem' (2019) 70(6) *Hastings Law Journal* 1389; Emilio Ferrara, 'Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies' (2024) 6(1) *Sci* 3:1–15 <https://doi.org/10.3390/sci6010003>.

10 Danks and London (n 5).

on datasets that can contain and reflect existing prejudices. Moreover, bias can emerge from datasets that do not correctly represent the characteristics of different populations (a problem known as 'unequal ground truth'); iii) eventually, bias can result from the usage of algorithmic systems in situations for which they were not intended. In this case, the bias can be classified as a problem of 'transfer context bias', meaning the issue arising from using an algorithm trained within one population and applied to a different one, or a problem of 'interpretation bias', meaning the issue emerging from the misinterpretation of the output.[11]

Solon Barocas and Andrew Selbst notably proposed a five-tier taxonomy[12] including: i) target variables, what the AI system aims to predict or optimise. If the chosen target inherently reflects societal biases or historical inequalities, the AI model may perpetuate or even exacerbate these biases; ii) training data, meaning the data used to train AI models that can carry biases present in the real world. This includes imbalances in representation, historical prejudices, or errors in data collection; iii) relevant features, meaning input variables used by the AI model to make predictions. Selecting which features are considered relevant can introduce bias, especially if important attributes are excluded or irrelevant ones are included; iv) proxy variables are attributes that are not explicitly sensitive but correlate strongly with sensitive attributes (eg, race, gender). These proxies can unintentionally introduce bias by allowing the model to infer and discriminate based on protected characteristics; and v) intentional discrimination ('masking'), which occurs when bias is deliberately embedded into the AI system, either through the design choices or by masking discriminatory practices to appear fair.[13]

From a sociotechnical point of view, Friedman and Nissenbaum have distinguished three categories of bias: i) pre-existing bias, related to historical discrimination against disadvantaged groups and minorities; ii) technical bias, related to technical design such as design choice and technical constraints; and iii) emergent bias, resulting from the context of use, as a result of changing societal knowledge, population, or cultural values.[14] These categories do not operate in isolation but frequently interact, reinforcing and amplifying one another.

For instance, pre-existing bias, such as the historical under-representation of trans or intersex individuals in medical datasets, can shape the technical architecture of an AI system when those data gaps lead to exclusionary modelling choices or proxy features that misclassify people. This exclusion then becomes embedded in the algorithm's decision logic, generating technical bias that appears neutral but replicates the initial inequality. Later, when the AI system is deployed in new or changing environments, its rigid classifications may fail to adapt to evolving understandings of sex, gender, or healthcare needs, giving rise to emergent bias. Similarly, an example can also be found in AI-based gender recognition systems. These systems often rely on facial analysis models trained predominantly on

11 Ibid.

12 Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact' (2016) 104 *California Law Review* 671, 677–94 <https://doi.org/10.2139/ssrn.2477899>.

13 Ibid 677 ff.

14 Friedman and Nissenbaum (n 8) 333–6.

datasets composed of cisgender individuals, with an over-representation of white male subjects. This reflects a pre-existing bias, the historical exclusion of trans, non-binary, and racialised individuals from technological datasets. The models then encode this imbalance through design choices, such as binary gender classification (male/female) and reliance on facial features stereotypically associated with those categories, generating technical bias. When deployed in contexts such as airport security or access to services, these systems may repeatedly misgender or fail to recognise individuals whose appearance does not conform to binary norms, resulting in emergent bias tied to the system's inability to adapt to social and cultural understandings of gender diversity.

In this way, historical inequality (pre-existing bias) influences technological design (technical bias), which in turn creates new forms of exclusion or harm in applied settings (emergent bias).

Engaging with the work of Nissenbaum and Friedman and building on these frameworks, this article will expand this taxonomy to include two other sources of bias that are systemic and structural and so precede and influence the others: the i) lack of diversity in the technology field as a discipline and practice and the ii) misinterpretation and bias in the meaning of sex, gender, and sexuality as data by those who operate both with algorithmic input and output.

One of the primary sources of bias can be traced to the insufficient presence of minority groups in technical domains like 'AI research, development, and engineering'.[15] Feminist scholars have long pointed out how this under-representation, particularly of women, shapes the field. When those designing AI systems come from a narrow set of backgrounds, the systems themselves risk reflecting limited worldviews,[16] ultimately constraining the inclusivity and responsiveness of AI technologies.[17]

The second source of bias builds on the idea proposed by critical data scholars observing that '[b]ig data … is never a neutral tool' as it 'always shapes and is shaped by a contested cultural landscape in both creation and interpretation'.[18] In this sense, it has been argued that 'big data's ambition for total knowledge often overlooks the effects of homophobia, biphobia and transphobia on historical and contemporary data practices and is ill-suited to qualitative methods, which tend to predominate studies of LGBTQ people'.[19] As a matter of fact, bias and assumptions that inform the meaning of sex, gender, and sexuality as data influence all the other

15 Meenakshi Punia, 'Challenges for Women in Artificial Intelligence: Promoting Gender Equality and Inclusivity' (2023) 6(3) *International Journal of Law, Management and Humanities* 3252, 3253 <https://doij.org/10.10000/IJLMH.115248>.

16 Mirjam Gruber and Roland Benedikter, 'The Role of Women in Contemporary Technology and the Feminization of Artificial Intelligence and Its Devices' in Tugrul Keskin and Ryan David Kiggins (eds), *Towards an International Political Economy of Artificial Intelligence* (Palgrave Macmillan, 2021) 17.

17 Punia (n 15) 3253.

18 Craig Dalton and Jim Thatcher, 'What Does a Critical Data Studies Look Like, and Why Do We Care?', *Society and Space* (Essay, 12 May 2014) <https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care>.

19 Kevin Guyan, *Queer Data: Using Gender, Sex and Sexuality Data for Action* (Bloomsbury Publishing, 2022) 115 <https://doi.org/10.5040/9781350230767>.

factors assumed to be sources of bias in AI (from the choice of the data input to the choice of outcome and its interpretation).[20] This cannot be solved by technical solutions alone, nor by expanding the amount of queer data,[21] but it requires a substantive approach to equality in its recognition dimension. For this reason, data sources should be complemented with 'rigorous qualitative research'.[22] As Kate Crawford posits, '[s]ocial science methodologies may make the challenge of understanding big data more complex, but they also bring context-awareness to our research to address serious signal problems'.[23]

In this sense, we can identify three categories of bias: i) foundational bias related to the lack of diversity in the AI field and to the misinterpretation and bias in the meaning of data; ii) technical bias related to both the architecture of the algorithm (selection of relevant features and the modelling of the algorithm) and bias related to the quality of the data (in training activities, in the composition of the data); and finally, iii) bias in implementation and interpretation of the outcome.

As a matter of fact, while human discrimination may stem from explicit prejudice or unconscious bias, algorithmic discrimination emerges from structural and technical sources, such as skewed datasets, inadequate representation, proxy variables, or design assumptions that reflect dominant norms. AI systems do not invent bias but inherit it from the world and the people who build them. Moreover, as noted by legal scholars: 'AI does not discriminate in an equivalent way to humans, which disrupts established methods for detecting, investigating, and preventing discrimination.'[24] In healthcare, where classification systems often rely on binary categories, foundational bias in the meaning of sex and gender, technical bias related to the selection of features, quality of data, and the architecture of algorithms, can translate into discriminatory clinical outcomes, even without intent.

A queer-responsive regulatory framework is thus essential not only to reveal but to redress the embedded logic of exclusion in algorithmic healthcare tools.

---

20 On the topic see, eg, Hannah Devinney, Jenny Björklund and Henrik Björklund, 'Theories of "Gender" in NLP Bias Research' (Conference Paper, Association for Computing Machinery, 20 June 2022) 2083 <https://doi.org/10.1145/3531146.3534627>.

21 Guyan (n 19) 115.

22 Kate Crawford, 'The Hidden Biases in Big Data', *Harvard Business Review* (Web Page, 2 April 2013) <https://hbr.org/2013/04/the-hidden-biases-in-big-data>.

23 Ibid.

24 Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Why Fairness Cannot Be Automated: Bridging the Gap between EU Non-Discrimination Law and AI' (2021) 41 *Computer Law and Security Review* 105567:1–31 <https://doi.org/10.1016/j.clsr.2021.105567>.

III SEX AND GENDER BIAS: AI IN HEALTHCARE AS A CASE STUDY

Crawford and Alexander Campolo have observed how '[w]hen applied in critical social domains', (deep learning) technology 'may deepen existing power imbalances between those who create the technologies, and those on whom they act'.[25]

AI, in its multifaceted forms, is 'a form of power without knowledge'.[26]

As a matter of fact, as it has been stated from an ethics perspective, the queer reconceptualisation of AI serves to 'overcome top-down categorization and be more accessible to diverse subjectivities and layered perspectives'.[27]

Queer research is situated within 'conceptual frameworks that highlight the instability of taken-for-granted meanings and resulting power relations'.[28] Queer data pertain to information about gender, sex, and sexuality that oppose biological determinism, emphasising that these categories are neither static nor predetermined.[29] While biological characteristics related to sex undoubtedly play a fundamental role in healthcare, given their influence on disease risk, clinical presentation, and treatment responses[30] (for instance, prostate cancer screenings are crucial for individuals with prostates irrespective of gender identity, and cervical cancer screenings for those with a cervix), the traditional binary categorisation of sex as exclusively male/female risks producing both over-inclusive and under-inclusive outcomes when embedded in AI systems. Over-inclusivity occurs when AI models wrongly assume all individuals categorised as female or male uniformly share specific anatomical or physiological traits, potentially ignoring significant variations within these groups. Conversely, under-inclusivity arises when rigid binary assumptions exclude or misrepresent intersex, transgender, and other individuals whose anatomical, hormonal, or chromosomal profiles do not align neatly with this binary division.

Queer theory specifically seeks to deconstruct such biologically deterministic assumptions, highlighting how binary sex categories are socially constructed

25 Alexander Campolo and Kate Crawford, 'Enchanted Determinism: Power without Responsibility in Artificial Intelligence' (2020) 6 *Engaging Science, Technology, and Society* 1, 5 <https://doi.org/10.17351/ests2020.277>. See also Kate Crawford and Ryan Calo, 'There Is a Blind Spot in AI Research' (2016) 538(7625) *Nature* 311 <https://doi.org/10.1038/538311a>.

26 Campolo and Crawford (n 25) 5.

27 Eduard Fosch-Villaronga and Gianclaudio Malgieri, 'Queering the Ethics of AI' in David J Gunkel (ed), *Handbook on the Ethics of Artificial Intelligence* (Edward Elgar Publishing, 2024) 301 <https://doi.org/10.4337/9781803926728.00026>.

28 Kath Browne and Catherine J Nash, 'Queer Methods and Methodologies: An Introduction' in Kath Browne and Catherine J Nash (eds), *Queer Methods and Methodologies: Intersecting Queer Theories and Social Science Research* (Routledge, 2016) 4 <https://doi.org/10.4324/9781315603223>.

29 Guyan (n 19) 8–9.

30 See Belinda Bennett et al, 'Gender Inequalities in Health Research: An Australian Perspective' in Michael Freeman (ed), *Law and Bioethics* (Oxford University Press, 2008) 409, 416 <https://doi.org/10.1093/oso:acprof/9780199545520.003.0022>: '[t]here are recognized sex differences in the causes, incidence, response to treatment, and prognosis of diseases, such as HIV/AIDS, coronary heart disease, depression, tropical infectious diseases, and tuberculosis.'

simplifications rather than universally accurate biological facts.[31] From this perspective, healthcare should rely on more specific and diverse biological data points, such as hormonal levels, anatomical variations, genetic markers, and physiological characteristics, instead of simplistically equating the broad and reductive labels of male or female with fixed sets of biological attributes.

Thus, a queer-responsive approach to AI and healthcare advocates recognising the medical relevance of detailed biological data, rather than relying on binary sex classifications that inadequately capture the full spectrum of human biological diversity.

Traditionally, sex is understood as relating to a person's identity based on primary and secondary sex characteristics. However, as Judith Butler explains, even the category of sex does not exist outside the social meanings we assign to it.[32] Kevin Guyan further elucidates that a queer approach does not aim to erase or rewrite sex as a category but seeks to highlight how sex, connected to biological characteristics, is significant in specific contexts, such as medical practice.[33] At the same time, the queer approach reminds us that sex cannot be considered in isolation from other factors.[34] Therefore, the definition of sex should not be confined to the binary of man/woman but must include space within and beyond these two poles for people with sex variations and non-binary individuals.

Gender has long been regarded, even legally, as the social expression of biological data, constructed around the man/woman binary directly linked to sex binarism.[35] A queer approach to gender emphasises that gender is performed daily through adherence to norms, roles, and relationships.[36] Those who do not fit within the expected gender norms often face stigma, prejudice, and discrimination. Therefore, gender data must also consider the experiences of individuals who do not identify with the expected gender corresponding to the sex assigned at birth, as well as those who do not consider themselves as belonging to any gender or who express gender differently.[37]

Similarly, sexuality data must be recognised for its fluid and dynamic nature, detached from biological determinism that envisages only heterosexual or homosexual orientations. It should encompass sexual or affective attraction and actions directed toward people of the same sex or gender, different sexes or genders, multiple sexes or genders, or no sex or gender.[38]

---

31 Anne Fausto-Sterling, 'The Five Sexes, Revisited' (2000) 40(4) *Sciences* 18 <https://doi.org/10.1002/j.2326-1951.2000.tb03504.x>.

32 Judith Butler, *Gender Trouble: Feminism and the Subversion of Identity* (Routledge, 1990) ('*Gender Trouble*').

33 Guyan (n 19) 9.

34 Judith Butler, *Bodies That Matter: On the Discursive Limits of 'Sex'* (Routledge, 1993).

35 See, eg, Mary Anne C Case, 'Disaggregating Gender from Sex and Sexual Orientation: The Effeminate Man in the Law and Feminist Jurisprudence' (1995) 105(1) *Yale Law Journal* 1, 2–3 <https://doi.org/10.2307/797140>.

36 Butler, *Gender Trouble* (n 32).

37 Guyan (n 19) 9.

38 Ibid.

Therefore, the original challenge of this contribution is to reconfigure how those who design, use, and regulate AI systems conceptualise gender, sex, and sexuality so that, concerning queer life, AI becomes 'power with knowledge' in order to guarantee equality in recognition, participation, and representation for sex, gender, and sexual minorities. Failure to do so will lead to the perpetuation of discrimination and oppression towards underrepresented groups.[39]

Queer theories aim not only to address the experiences of marginalised groups but also to challenge and deconstruct the frameworks governing all genders and sexualities, including those considered normative or dominant. By dismantling the norms and stereotypes that keep sex, gender, and sexual minorities in a subordinate position to cisgender men, a queer perspective challenges the traditional binary frameworks underpinning AI systems.[40]

In the following subsections, I will present a case study and research that demonstrates how the lack of a queer approach to AI due to foundational, technical, and implementation bias leads to discriminatory practices against sexual, gender, and sex minorities in a sector that poses a significant risk to fundamental rights such as healthcare.

A Sex and Gender AI Bias in Healthcare: A Case Study

Artificial intelligence in healthcare holds great potential[41] as it can improve diagnostic accuracy, personalise treatments, and optimise the efficiency of

39 Fosch-Villaronga and Malgieri (n 27) 301–2. See also, Adam Poulsen, Eduard Fosch-Villaronga and Roger Andre Søraa, 'Queering Machines' (2020) 2(2) *Nature Machine Intelligence* 152, 152 <https://doi.org/10.1038/s42256-020-0157-6https>: '[i]f the perspectives of queer users are not considered in robot and AI development, implicit biases will persist. Moreover, queer people will remain mostly invisible, silent, powerless and unable to understand how these technologies may affect them.'

40 Foad Hamidi, Morgan Klaus Scheuerman and Stacy M Branham, 'Gender Recognition or Gender Reductionism: The Social Implications of Automatic Gender Recognition Systems' (Conference Paper, Association for Computing Machinery Conference on Human Factors in Computing Systems, 19 April 2018) <https://doi.org/10.1145/3173574.3173582>; Janine Aldous Arantes and Mark Vicars, 'Missing in Action: Queer(y)ing the Educational Implications of Data Justice in an Age of Automation' (2023) 48(2) *Learning, Media and Technology* 213 <https://doi.org/10.1080/17439884.2023.2207141>; Dawn McAra-Hunter, 'How AI Hype Impacts the LGBTQ+ Community' (2024) 4(3) *Artificial Intelligence and Ethics* 771 <https://doi.org/10.1007/s43681-024-00423-8>.

41 Eric Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again* (Basic Books, 2019). From a legal perspective, the use of AI in healthcare raises issues across multiple domains, including data protection, regulatory compliance, and, as analysed in this article, discrimination. Another crucial dimension concerns the responsibility of medical professionals: see Bernadette Richards, Susannah Sage Jacobson and Yves Saint James Aquino, 'Regulation of AI in Health Care: A Cautionary Tale Considering Horses and Zebras' (2021) 28(3) *Journal of Law and Medicine* 645. Ethically, AI should be understood within the existing framework of the professional duty of care and does not displace clinical judgement: see Enrico Coiera, Maureen Baker and Farah Magrabi, 'First Compute No Harm', *The BMJ Opinion* (Blog Post, 19 July 2017) <https://blogs.bmj.com/bmj/2017/07/19/enrico-coiera-et-al-first-compute-no-harm/>. However, the legal question of medical responsibility, particularly in terms of civil liability, falls outside the scope of this analysis. This article does not aim to address the legal obligations of clinicians in malpractice contexts but rather focuses on the regulatory mechanisms through which algorithmic bias can be identified and corrected in AI systems used in healthcare.

healthcare services.[42] In fact, advances in machine learning and big data analytics have already made it possible to apply AI in various areas of healthcare today, including diagnosis, research, and patient management.[43]

However, research has shown how data-related biases can affect AI, leading to discriminatory practices towards minorities in healthcare practice.[44] For instance, evidence of ethnic bias has been identified in a widely used algorithmic system that targets patients for 'high-risk care management' programs.[45]

When it comes to sex and gender, the issue of AI biases in healthcare is multifaceted. These biases involve not only data-related and modelling issues, but also foundational and implementation biases related both to a lack of diversity in the field and to the medical understanding and interpretation of sex and gender as data points.[46]

Sex and gender data are fundamentally crucial for ensuring personalised access to care and clinical practices.[47] In the context of precision medicine,[48] which aims to tailor healthcare to the unique characteristics of each patient, accurate sex and gender data are essential.

AI, with its advanced capacity for predictive analysis, has the potential to significantly enhance this personalised approach to medicine,[49] leading to more precise diagnoses, effective treatments, and improved patient outcomes. However, despite these promising advantages, certain biases, both technical and sociotechnical, can hinder this development.[50]

Technical biases arise from unrepresentative or incomplete datasets used to train AI models. If the data does not adequately represent all segments of the

---

42 Nuffield Council on Bioethics, 'Artificial Intelligence (AI) in Healthcare and Research' (Policy Briefing Note, 15 May 2018) 1, 3–4.

43 Eduard Fosch-Villaronga et al, 'Accounting for Diversity in AI for Medicine' (2022) 47 *Computer Law and Security Review* 105735:1–15 <https://doi.org/10.1016/j.clsr.2022.105735>.

44 Sara Gerke, Timo Minssen and Glenn Cohen, 'Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare' in Adam Bohr and Kaveh Memarzadeh (eds), *Artificial Intelligence in Healthcare* (Elsevier, 2020) 295 <https://doi.org/10.1016/B978-0-12-818438-7.00012-5>.

45 Ziad Obermeyer et al, 'Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations' (2019) 366(6464) *Science* 447, 447 <https://doi.org/10.1126/science.aax2342>. See also Anjali Jain et al, 'Awareness of Racial and Ethnic Bias and Potential Solutions to Address Bias with Use of Health Care Algorithms' (2023) 4(6) *Journal of the American Medical Association Health Forum* e231197:1–13 <https://doi.org/10.1001/jamahealthforum.2023.1197>.

46 Nataly Buslón et al, 'Raising Awareness of Sex and Gender Bias in Artificial Intelligence and Health' (2023) 4 *Frontiers in Global Women's Health* 4:1–8 <https://doi.org/10.3389/fgwh.2023.970312>. Buslón et al identify four major strategic areas of action in AI: i) lack of data; ii) social impact and awareness; iii) AI biases; and iv) regulatory aspects: at 3. They also propose specific recommendations which focus on scientific, educational, and political strategies including enhancing female representation in science, technology, engineering, and mathematics fields: at 2.

47 Davide Cirillo et al, 'Sex and Gender Differences and Biases in Artificial Intelligence for Biomedicine and Healthcare' (2020) 3 *Nature Partner Journals Digital Medicine* 81:1–11 <https://doi.org/10.1038/s41746-020-0288-5>.

48 Ibid 1, defining 'Precision Medicine' as the approach 'to find personalized preventative and therapeutic strategies by taking into account differences in genes, environment, and lifestyle'.

49 Kevin B Johnson et al, 'Precision Medicine, AI, and the Future of Personalized Health Care' (2021) 14(1) *Clinical and Translational Science* 86 <https://doi.org/10.1111/cts.12884>.

50 Cirillo et al (n 47) 1.

population, the AI may perform poorly for under-represented groups. For instance, if an AI system is trained predominantly on data from men, it may not provide accurate predictions for women. As Natalia Norori et al argued, for example, heart attacks are 'overwhelmingly misdiagnosed in women'[51] due to biases in the data used to train diagnostic algorithms. This issue arises because women often exhibit different symptoms of heart attacks than men, and these differences are not adequately represented in the data that inform clinical guidelines and AI models.[52]

Another study has assessed the gender performance of AI, particularly deep learning algorithms, in assisting the diagnosis of diseases from medical images such as X-rays.[53] The research demonstrated that a considerable number of medical imaging datasets are not gender-balanced, exhibiting a higher proportion of images from male patients than female patients.[54] Models trained on gender-imbalanced datasets exhibited notable performance disparities between male and female patients. When the training data were skewed towards males, the model's performance declined significantly.[55]

At the same time, sociotechnical biases stem from the way sex and gender are conceptualised and recorded in medical data. Traditional medical practices often rely on binary classifications of sex and gender,[56] which do not capture the full spectrum of human diversity.[57] This can lead to the exclusion or misrepresentation of people with sex variations, transgender people, and non-binary individuals in healthcare data.[58]

As a matter of fact, assumptions that sex and gender are binary, static, and concordant are embedded in medical systems and thus influence AI in healthcare.

Scholars have explored the challenges associated with the use of sex and gender data in machine learning applications within healthcare, particularly when using electronic health records ('EHRs').[59]

---

51 Natalia Norori et al, 'Addressing Bias in Big Data and AI for Health Care: A Call for Open Science' (2021) 2(10) *Patterns* 100347:1–9, 2 <https://doi.org/10.1016/j.patter.2021.100347>. See also Giorgia Guerra, 'Evolving Artificial Intelligence and Robotics in Medicine, Evolving European Law: Comparative Remarks Based on the Surgery Litigation' (2021) 28(6) *Maastricht Journal of European and Comparative Law* 805 <https://doi.org/10.1177/1023263X211042470>.

52 Norori et al (n 51) 2.

53 Agostina J Larrazabal et al, 'Gender Imbalance in Medical Imaging Datasets Produces Biased Classifiers for Computer-Aided Diagnosis' (2020) 117(23) *Proceedings of the National Academy of Sciences* 12592 <https://doi.org/10.1073/pnas.1919012117>.

54 Ibid 12593–4.

55 Ibid 12592–3.

56 Fosch-Villaronga et al (n 43). See also Sivaniya Subramaniapillai et al, 'Sex and Gender in Health Research: Intersectionality Matters' (2024) 72 *Frontiers in Neuroendocrinology* 101104:1–6 <https://doi.org/10.1016/j.yfrne.2023.101104>.

57 Fosch-Villaronga et al (n 43) 5.

58 Ibid. See also Clair A Kronk et al, 'Transgender Data Collection in the Electronic Health Record: Current Concepts and Issues' (2022) 29(2) *Journal of the American Medical Informatics Association* 271 <https://doi.org/10.1093/jamia/ocab136>.

59 Melissa McCradden et al, 'What's Fair is… Fair: Presenting JustEFAB, an Ethical Framework for Operationalizing Medical Ethics and Social Justice in the Integration of Clinical Machine Learning' (Conference Paper, Association for Computing Machinery Conference on Fairness, Accountability, and Transparency, 12 June 2023) <https://doi.org/10.1145/3593013.3594096>.

In their work, Kendra Albert and Maggie Delano introduce three key concepts that are used to describe the problematic ways in which sex and gender are handled in machine learning healthcare research.[60]

The expression 'sex/gender slippage' is used to describe the frequent substitution of sex-related terms for gender-related terms, the first one being related to biological characteristics such as gonads, anatomy, chromosomes, hormone levels, and the second one to a person's perceived and experienced identity, influenced both by internal and external factors.[61] In medical practice, this reflects an underlying assumption that sex and gender are the same and interchangeable, based on the false idea of concordance of sex as a biological factor with gender as a social factor.[62] For example, medical records may use the terms male and female (which are related to biological sex) interchangeably with man and woman (which are related to gender). This practice fails to recognise the distinction between biological attributes and social identities and it fails to account for people who have chromosomal, gonadal, anatomical, and secondary characteristics that do not align with the defaults for male or female sex, and for people whose gender does not fit into the man/woman binary (non-binary), or people whose gender does not align with the sex assigned at birth (trans).[63]

'Sex confusion' highlights the ambiguity surrounding what a sex variable represents in medical records. A single sex marker in EHRs might refer to sex assigned at birth, legal sex, or current physiological characteristics, and may not correspond to any specific anatomy or hormonal status.[64] This confusion arises because sex is context-dependent and cannot be accurately captured by a singular, static data point. The authors emphasise that sex has always been a context-dependent variable, and relying on a single marker ignores the complexities of individual experiences.[65]

Finally, 'sex obsession' denotes the overemphasis on sex assigned at birth as the primary or sole relevant variable in medical inquiries.[66] This fixation can lead to inappropriate assumptions and decisions in healthcare, particularly affecting transgender and non-binary individuals. The medical system's focus on sex assigned at birth can result in practices that overlook or misunderstand the needs of patients whose experiences do not align with traditional binary notions.

In the specific case of HIV and pre-exposure prophylaxis ('PrEP'), Albert and Delano delve into how foundational biases regarding sex and gender manifest in the research and development of machine learning models designed to predict

60 Kendra Albert and Maggie Delano, 'Sex Trouble: Sex/Gender Slippage, Sex Confusion, and Sex Obsession in Machine Learning Using Electronic Health Records' (2022) 3(8) *Patterns* 100534:1–11 <https://doi.org/10.1016/j.patter.2022.100534>.

61 Ibid 2.

62 Ibid 3.

63 Ibid.

64 Ibid 3–4.

65 Ibid 4.

66 Ibid 3–4.

the risk of HIV infection for the purpose of administering PrEP,[67] a preventive medication that significantly reduces the likelihood of contracting HIV.[68]

The risk of HIV transmission is influenced by a complex interplay of biological, behavioural, and social factors closely linked to both sex and gender. Certain sexual practices and societal stigmas disproportionately affect specific communities, such as transgender individuals, men who have sex with men ('MSM'), and other gender and sexual minorities. These groups often face unique challenges that increase their vulnerability to HIV, including discrimination, limited access to healthcare, and higher rates of risk-taking behaviours due to social marginalisation.[69]

Albert and Delano analyse how biases related to sex and gender are embedded in studies utilising machine learning for HIV prevention among at-risk individuals.[70] They highlight that many machine learning models fail to adequately account for the diversity of gender identities, leading to significant shortcomings in predicting HIV risk. For instance, transgender and non-binary people are frequently excluded from datasets or misclassified due to a reliance on binary notions of sex and gender. This exclusion results in models that do not recognise these individuals as being at risk, thereby preventing them from accessing PrEP.[71]

Despite the undeniable progress and effectiveness of AI in predicting, addressing, or supporting health-related decisions, when it comes to data related to sex and gender, AI can act as a 'double-edged sword'.[72] On one edge, AI has the potential to advance personalised medicine by considering individual differences in sex and gender, leading to better health outcomes. On the other edge, if both foundational and technical biases are not adequately addressed, AI can perpetuate or even exacerbate existing disparities in healthcare.

IV COMPARATIVE ANALYSIS OF QUEER-RESPONSIVENESS IN AI REGULATION

The analysis shows that biases in AI systems are not just technical glitches but are deeply rooted in foundational and interpretative biases stemming from societal norms and the under-representation of minorities in the tech industry. These biases influence how data is interpreted, and algorithms are designed, often leading to discriminatory outcomes that perpetuate existing inequalities.

---

67 Ibid. See also Douglas S Krakower et al, 'Development and Validation of an Automated HIV Prediction Algorithm to Identify Candidates for Pre-exposure Prophylaxis: A Modelling Study' (2019) 6(10) *Lancet HIV* 696 <https://doi.org/10.1016/s2352-3018(19)30139-0>.

68 Douglas S Krakower and Kenneth H Mayer, 'Pre-exposure Prophylaxis to Prevent HIV Infection: Current Status, Future Opportunities and Challenges' (2015) 75(3) *Drugs* 243 <https://doi.org/10.1007/s40265-015-0355-4>.

69 See generally AKM Ahsan Ullah and Ahmed Shafiqul Huque, *Asian Immigrants in North America with HIV/AIDS: Stigma, Vulnerabilities and Human Rights* (Springer, 2014) ch 5 <https://doi.org/10.1007/978-981-287-119-0_5>.

70 Albert and Delano (n 60).

71 Ibid 6–7.

72 Cirillo et al (n 47) 1.

Applying a queer theoretical perspective underscores the fluidity and social construction of sex, gender, and sexuality. However, AI systems frequently fail to recognise this fluidity, instead enforcing binary and exclusionary norms that marginalise queer identities.[73] This misalignment between the lived realities of queer individuals and the rigid frameworks used by AI technologies exacerbates discrimination and hinders the realisation of substantive equality in its multidimensional form, entailing also recognition, representation, and participation of minorities.[74]

I involve in my analysis a queer-responsive approach to AI technology and its regulation.[75] With the expression 'queer-responsive', I identify laws and policies that are more attentive and that respond more effectively to the challenges of sex, gender, and sexual minorities. The expression is borrowed from scholarship that has identified the need for 'gender-responsive legislation' as a means to reach gender equality for women.[76]

In this context, I specifically confront the responsiveness of regulation as a means to overcome the issues of technical and sociotechnical bias that affect queer people.

Specifically, by analysing hard law sources,[77] it is possible to identify three models to address AI bias that can be used to frame a queer-responsive regulatory

---

73 Karin Danielsson et al, 'Queer Eye on AI: Binary Systems Versus Fluid Identities' in Simon Lindgren (ed), *Handbook of Critical Studies of Artificial Intelligence* (Edward Elgar Publishing, 2023) 595 <https://doi.org/10.4337/9781803928562.00061>.

74 Sandra Fredman, 'Substantive Equality Revisited' (2016) 14(3) *International Journal of Constitutional Law* 712 <https://doi.org/10.1093/icon/mow043>.

75 José-Miguel Bello y Villarino and Ramona Vijeyarasa, 'International Human Rights, Artificial Intelligence, and the Challenge for the Pondering State: Time to Regulate?' (2022) 40(1) *Nordic Journal of Human Rights* 194 <https://doi.org/10.1080/18918131.2022.2069919>; Prashant Chauhan and Gagandeep Kaur, 'Gender Bias and Artificial Intelligence: A Challenge within the Periphery of Human Rights' (2022) 8(1) *Hasanuddin Law Review* 46 <https://doi.org/10.20956/halrev.v8i1.3569>.

76 See Ramona Vijeyarasa, 'In Pursuit of Gender-Responsive Legislation: Transforming Women's Lives through the Law' in Ramona Vijeyarasa (ed), *International Women's Rights Law and Gender Equality* (Routledge, 2021) 3 <https://doi.org/10.4324/9781003091257>. See also Ramona Vijeyarasa, 'What Is Gender-Responsive Legislation: Using International Law to Establish Benchmarks for Labour, Reproductive Health and Tax Laws That Work for Women' (2020) 29(3) *Griffith Law Review* 334, 335–6 <https://doi.org/10.1080/10383441.2020.1853900>, where the author notes that (citations omitted):

> Gender-responsive legislation entails laws that are more responsive to explicit and implicit gender issues. Such legislation facilitates accountability, in legislative and policy implementation, to the specific needs of different sexes and to different gendered perspectives on pivotal social, economic and political issues.

77 The methodological choice to focus exclusively on sources of hard law responds to a need for systematisation capable of revealing the potential of a queer-responsive regulation. However, it is important to remember that with regard to AI regulation, as already emphasised in Part I, other processes of normative production come into play. These are not confined to regulation through law traditionally understood but are integrated with forms of *co-regulation* or even *self-regulation*, in which technical standards and technological solutions integrate and complete the regulatory framework. Moreover, legal scholarship has highlighted the disruptive challenges of the 'duality of AI' acting in the regulatory environment as both target and tool of regulation. See Simone Penasa, 'Verso un Diritto "Technologically Immersive": la Sperimentazione Normativa in Prospettiva Comparata' [Towards a 'Technologically Immersive' Law: Normative Experimentation in Comparative Perspective] (2023) 57(1) *Diritto Pubblico Comparato ed Europeo Online* 671, where Simone Penasa explains that technology, and its applications across various aspects of social life, does not gain significance solely as an object of legal regulation from

approach for the use of AI in the healthcare sector and beyond with a specific focus on regulation of equality and non-discrimination rules: i) a principle-based responsive model; ii) a technical responsive model; and iii) a sociotechnical responsive model, and to assess their responsiveness to technical or sociotechnical bias as defined in the previous section.

The principle-based model focuses on setting broad principles rather than prescribing detailed technical or sociotechnical requirements. In this approach, policymakers establish high-level values that AI systems must uphold and provide flexibility for regulated entities to interpret and implement these principles in different contexts. In this model, the prototype is identified as *The Convention*.

The technical-oriented model, on the other hand, emphasises embedding bias mitigation directly into the technological framework of the AI system. Here, policymakers take a by-design approach, ensuring that the system's architecture includes specific and legally binding mechanisms to reduce or eliminate bias from the outset. By focusing on technical solutions, this model seeks to proactively address bias at the design stage, making the technology itself a vehicle for promoting equality. The *EU AI Act* is prototypical of this model.

Finally, the sociotechnical-oriented model combines both technical solutions and broader societal considerations. This integrated approach recognises that addressing AI bias requires not only technical adaptations, but also societal engagement and awareness. Prototypes of this approach include the Brazilian AI Bill.

It is important to emphasise that these are models of tendencies determined by the comparative choices made. Therefore, the division follows the idea of the prevalence of one approach over another in a given source: in some cases, which will be accounted for, alongside a predominant approach, there will be minor instances of another.

The choice of sources is determined by the integration of two factors. Firstly, the analysis focuses on legal systems that adhere to the dimensions of formal and substantive equality as defined through constitutional charters and adherence to international or supranational treaties. Secondly, reference will be made to legal systems where the process of normative regulation of AI has already been completed or where normative perspectives have already been advanced. Additionally, the international dimension will also be considered, specifically that of the Council of Europe, as an international treaty instrument that is legally binding.

This comparison meets a need for systematisation based on three fundamental factors regarding the sources analysed: i) type of source; ii) scope of application; and iii) degree of responsiveness to queer-related bias.

These comparative factors are inherently influenced by the authority that issues these instruments and their degree of legal enforceability. For instance, an international treaty, such as *The Convention*, is necessarily principle-based, whereas the *EU AI Act* must provide the specificity required by regulation as a source of EU law directly applicable. Enforcement systems for these instruments

---

a traditional perspective of the regulatory functions performed by legal sources. Instead, from a dynamic and innovative standpoint, it also serves as a tool for regulation (and even for normative creation): at 672 [tr author].

vary significantly from case to case and, therefore, do not form part of the factors upon which my comparative model is built.

A The Principle-Based Model

The principle-based model is a regulatory approach in which policymakers do not specify detailed rules or a prescribed set of behaviours. Instead, they formulate broad, high-level principles.[78] Under this model, compliance is assessed based on the degree to which the behaviour of regulated entities aligns with these principles, evaluated on a case-by-case basis.[79]

In this framework, emphasis is placed on the desired outcome rather than on the specific processes or procedures to be followed. Regulated parties are expected to interpret the established principles and act in ways that are aligned with the regulatory objectives. This means they must understand the underlying goals of the regulation and ensure their actions consistently support those aims.[80]

The academic literature has recognised both the advantages and limitations of the principle-based model.[81]

Jonas Schuett et al noted that this approach provides the flexibility to establish regulatory principles even in situations where the most appropriate behaviour is not immediately clear. Unlike rule-based models that rely on specific behaviours or actions,[82] a principle-based approach allows regulators to set broad objectives without having to anticipate every possible scenario. This flexibility is particularly useful in rapidly evolving areas where fixed rules can quickly become outdated.[83]

In addition, the principle-based model draws on the expertise of the regulated entities themselves, who often have a deeper understanding of the context and can more effectively identify the actions needed to meet regulatory objectives.[84] This adaptability makes the principle-based model more resilient to future developments,

78 See Jonas Schuett et al, 'From Principles to Rules: A Regulatory Approach for Frontier AI' in Philipp Hacker et al (eds), *The Oxford Handbook of the Foundations and Regulation of Generative AI* (Oxford University Press, forthcoming):1–59, 1, 28 <https://doi.org/10.48550/arXiv.2407.07300>.

79 A principle-based approach has been analysed with reference to others area of law, in particular within the fields of financial market regulation and environmental law: see generally Julia Black, 'Forms and Paradoxes of Principles Based Regulation' (Working Paper No 13/2008, London School of Economics and Political Science, 23 September 2008) <https://doi.org/10.2139/ssrn.1267722>; Neil Gunningham and Darren Sinclair, 'Integrative Regulation: A Principle-Based Approach to Environmental Policy' (1999) 24(4) *Law and Social Inquiry* 853 <https://doi.org/10.1111/j.1747-4469.1999.tb00407.x>.

80 Schuett et al (n 78) 22–3.

81 Ibid.

82 Ibid. See also Ruth B Carter and Gary E Marchant, 'Principles-Based Regulation and Emerging Technology', in Gary E Marchant, Braden R Allenby and Joseph R Herkert (eds), *The Growing Gap Between Emerging Technologies and Legal–Ethical Oversight: The Pacing Problem* (Springer, 2011) 157 <https://doi.org/10.1007/978-94-007-1356-7_10>.

83 Schuett et al (n 78). This framework aligns with the principle of responsive regulation set forth in Ian Ayres and John Braithwaite, *Responsive Regulation: Transcending the Deregulation Debate* (Oxford University Press, 1992) <https://doi.org/10.1093/oso/9780195070705.001.0001>.

84 Schuett et al (n 78) 22–3.

allowing for modifications and new applications in response to technological and social progress.[85]

In addition, focusing on high-level principles helps avoid the pitfalls of under- or over-inclusiveness.[86] By not limiting compliance to a narrow set of behaviours, the model can better target undesirable actions without unnecessarily restricting legitimate practices.

However, this flexibility comes at the cost of reduced legal certainty.[87] Since the principle-based model does not provide explicit rules, regulated entities may find it challenging to discern the precise requirements for compliance.[88]

This approach enables what scholars termed as the integration of legislative law with other normative sources,[89] both internal and external to the legal order, as an essential element in regulating technical and scientific phenomena.[90]

In a principle-based model, external integration beyond the legislative sphere becomes particularly significant. This model brings into play the judicial component, which is tasked with assessing whether specific behaviours adhere to the established principles on a case-by-case basis. Additionally, it involves forms of technical and scientific expertise that further inform and shape the principles' content, allowing the regulation to adapt to the complexities of evolving technical knowledge.[91]

1 The Convention

Within this model, *The Convention* can be incorporated.[92]

The Convention represents the first-ever legally binding international treaty in this field,[93] aiming to 'ensure that activities within the lifecycle of artificial

85 Sofia Ranchordás and Mattis van't Schip, 'Future-Proofing Legislation for the Digital Age' in Sofia Ranchordás and Yaniv Roznai (eds), *Time, Law, and Change: An Interdisciplinary Study* (Hart Publishing, 2020) 347 <https://doi.org/10.5040/9781509930968.ch-016>.

86 See also Frederick Schauer, *Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life* (Clarendon Press, 1991) 4–5.

87 Timothy AO Endicott, 'Linguistic Indeterminacy' (1996) 16(4) *Oxford Journal of Legal Studies* 667 <https://doi.org/10.1093/ojls/16.4.667>.

88 Rosie Cooney and Andrew TF Lang, 'Taking Uncertainty Seriously: Adaptive Governance and International Trade' (2007) 18(3) *European Journal of International Law* 523 <https://doi.org/10.1093/ejil/chm030>.

89 Cristie Ford, 'Principles-Based Securities Regulation in the Wake of the Global Financial Crisis' (2010) 55(2) *McGill Law Journal* 257, 261 <https://doi.org/10.7202/045086ar>,arguing that '[p]rinciples-based regulation is premised on concepts of "co-regulation", or "enforced self-regulation"'.

90 Simone Penasa, 'Alla Ricerca di un Lessico Comune: Inte(g)razioni tra Diritto e Scienze della Vita in Prospettiva Comparata' [In Search of a Common Language: Interactions between Law and Life Sciences in Comparative Perspective] (2020) 44(3) *Diritto Pubblico Comparato ed Europeo Online* 3307 <https://doi.org/10.57660/dpceonline.2020.1079>. Penasa argues that technical–scientific expertise shapes law both inside legislation and, where statutes are lacking, through exo-legislative sources – above all case law – which, while addressing complex bio-law issues, raises concerns about separation of powers, legal certainty, and predictability at 3309 [tr author].

91 Ibid 3309–10.

92 *The Convention* (n 2).

93 Ibid. This was drafted by the 46 member states of the Council of Europe, with the participation of observer states (Canada, Japan, Mexico, the Vatican City and the United States), the European Union, and non-member states (Australia, Argentina, Costa Rica, Israel, Peru and Uruguay). So far, *The Convention*

intelligence systems are fully consistent with human rights, democracy, and the rule of law, while being conducive to technological progress and innovation'.[94]

The Convention's scope covers both the use of AI systems by public and private actors. While *The Convention* does not regulate specific technologies and remains fundamentally technology-neutral to ensure adaptability, it applies broadly to both public authorities, defined as any public law entity at any level, including supranational, state, regional, provincial, municipal, and independent public entities, as well as any private entity exercising official authority and private actors.[95]

In this context, the primary factor that situates this instrument within the proposed model perspective is its adherence to a regulation that, rather than directly shaping the technological construction of AI systems by prescribing specific technical requirements or configurations, mandates adherence to overarching principles[96] for the use of AI systems by public authorities and private actors.

As for public actors, *The Convention* obliges Parties to ensure that AI system activities comply with its provisions when undertaken by public authorities or by private actors acting on their behalf,[97] therefore including situations where public authorities delegate responsibilities to private entities or direct them to act, such as through contracts with private actors for public services as well as public procurement and contracting.[98]

Additionally, *The Convention* requires all Parties to address risks and impacts on human rights, democracy, and the rule of law in the private sector,[99] extending obligations to private actors even where these do not act as public actors. As a matter of fact, to meet this requirement, each Party must submit a declaration to the Secretary General of the Council of Europe upon signing, ratification, acceptance, approval, or accession, specifying how it intends to fulfill these obligations towards private actors.[100] This duty to provide can be achieved either by applying the

---

has been signed by 10 states (Andorra, Georgia, Iceland, Israel, Montenegro, Norway, Moldova, San Marino, the United Kingdom, the United States) and by the European Union. See also Elżbieta Hanna Morawska, 'Council of Europe Standards and Activities Related to AI: Towards a Framework Convention on AI and Human Rights?' in Michał Balcerzak and Julia Kapelańska-Pręgowska (eds), *Artificial Intelligence and International Human Rights Law: Developing Standards for a Changing World* (Edward Elgar Publishing, 2024) 25 <https://doi.org/10.4337/9781035337934.00009>. For a reconstruction of the history behind the approval of the treaty, see Costanza Nardocci, 'Artificial Intelligence at the Crossroads between the European Union and the Council of Europe: Who Safeguards What and How?' (2024) 16(1) *Italian Journal of Public Law* 165.

94 *The Convention* (n 2) art 1.

95 Ibid art 3.

96 To this end, the principles recognised in *The Convention* are fundamental principles: i) human dignity and individual autonomy; ii) equality and non-discrimination; iii) respect for privacy and personal data protection; iv) transparency and oversight; v) accountability and responsibility; vi) reliability; vii) safe innovation.

97 *The Convention* (n 2) art 31(a).

98 Council of Europe, *Explanatory Report to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* (Report, 5 September 2024) 7 [28] ('*Convention Explanatory Report*').

99 *The Convention* (n 2) art 31(b).

100 Ibid.

principles and obligations outlined in *The Convention* to private actors' activities as they come or by implementing other appropriate measures, such as specific legislation, or administrative and voluntary measures.[101]

Having defined the nature and scope of the source, I now turn to examine the reach of the principle of equality within the regulatory instrument analysed. From this perspective, the principle is articulated in a way that encompasses both technical and sociotechnical considerations.

As a matter of fact, in *The Convention*, algorithmic discrimination emerges as a central concern.[102] The Preamble explicitly highlights the issue, noting serious concern about

> the risks of discrimination in digital contexts, in particular those involving artificial intelligence systems, and their potential effect of creating or aggravating inequalities, including those experienced by women and persons in vulnerable situations, in the enjoyment of their human rights and their full, equal and effective participation in economic, social, cultural and political affairs.[103]

The Convention explicitly addresses equality and non-discrimination in article 10. In the article, Parties are required to adopt or maintain measures to ensure that all activities within the AI lifecycle uphold equality, including gender equality, and comply with the prohibition of discrimination as established under relevant international and domestic laws.[104] Additionally, each Party commits to

---

101 *Convention Explanatory Report* (n 98) 7 [29].

102 See Nardocci (n 93).

103 *The Convention* (n 2) Preamble para 7.

104 Ibid art 10. At the global level, each Party shall comply to various relevant frameworks, including several key provisions from international human rights instruments. These encompass article 2, article 24, and article 26 of the *International Covenant on Civil and Political Rights*, opened for signature 16 December 1966, 999 UNTS 171 (entered into force 23 March 1976); articles 2, 3, and 7 of the *International Covenant on Economic, Social and Cultural Rights*, opened for signature 16 December 1966, 993 UNTS 3 (entered into force 3 January 1976); and specialised legal instruments such as the *International Convention on the Elimination of All Forms of Racial Discrimination*, opened for signature 21 December 1965, 660 UNTS 195 (entered into force 4 January 1969); the *Convention on the Elimination of All Forms of Discrimination Against Women*, opened for signature 18 December 1979, 1249 UNTS 13 (entered into force 3 September 1981); the *Convention on the Rights of the Child*, opened for signature 20 November 1989, 1577 UNTS 3 (entered into force 2 September 1990); and the *Convention on the Rights of Persons with Disabilities*, opened for signature 13 December 2006, 2515 UNTS 3 (entered into force 3 May 2008). At the regional level, each Party may also look to specific frameworks within their jurisdiction. In Europe, for instance, relevant provisions include article 14 of the *Convention for the Protection of Human Rights and Fundamental Freedoms*, opened for signature 4 November 1950, 213 UNTS 221 (entered into force 3 September 1953), as amended by *Protocol No 12 to the Convention for the Protection of Human Rights and Fundamental Freedoms*, opened for signature 4 November 2000, ETS No 3 (entered into force 1 April 2005); certain sections of the *European Social Charter*, opened for signature 18 October 1961, ETS No 35 (entered into force 26 February 1965), including paragraphs 20 and 27 of part I, article 20 of part II, and article E of part V; and specialised Council of Europe legal instruments, such as article 4 of the *Framework Convention for the Protection of National Minorities*, opened for signature 1 February 1995, ETS No 157 (entered into force 1 February 1998) and article 4 of the *Convention on Preventing and Combating Violence Against Women and Domestic Violence*, opened for signature 11 May 2011, CETS No 210 (entered into force 1 August 2013). The European Union provides additional guidance through title III of the *Charter of Fundamental Rights of the European Union*, opened for signature 7 December 2000, OJ C 326/01 (entered into force 1 December 2009); specific EU treaties (notably article 2 of the *Treaty on the European Union*, opened for signature 26 October 2012, OJ C 326/13, and article 10 of the *Consolidated Version of the Treaty on the Functioning of the European Union*, opened for signature

implementing measures designed to reduce inequalities, with the goal of achieving fair, just, and equitable outcomes.

The article appears to encompass both *ex post* remedies to discrimination and proactive obligations linked to the achievement of fair, just, and equitable outcomes. The provision makes clear that the required approach should go beyond simply ensuring that individuals are not treated less favourably in certain sectors on the basis of one or more protected characteristics.[105] Rather, Parties commit to adopting new or maintaining existing measures aimed at addressing and overcoming structural and historical inequalities, within the scope of their national and international human rights obligations.

The *Explanatory Report* adds depth to this article by specifying that the issues of equality in the specific AI context include relatively new categories of problems such as 'technical bias', biases that arise from issues in applying machine learning, which can result in additional biases not present in the data used to train the system or make decisions. It also notes 'social bias', or failures to adequately address historical or current inequalities in society during the activities within the AI lifecycle, such as in designing and training models. These inequalities include long-standing and structural barriers to gender equality and equal treatment for individuals belonging to groups that have faced, or continue to face, discrimination or persistent inequality due to their particular characteristics, circumstances or group membership.[106]

As a matter of fact, *The Convention* sets out principles governing measures of assessment and mitigation of risks and adverse effects that have a reach in all our established directions.[107]

With specific reference to equality, article 16 mandates the adoption of measures that 'take into account, where appropriate, the perspectives of relevant stakeholders, in particular persons whose rights may be affected',[108] emphasising a participatory dimension.

Similarly, *The Convention* emphasises that each Party should encourage and promote sufficient digital literacy and skills across all segments of the population.

---

13 December 2007, [2012] OJ C 326/01 (entered into force 1 December 2009) ('*TFEU*'), as well as EU secondary legislation and relevant case-law from the Court of Justice of the European Union. Beyond Europe, relevant regional instruments include article 24 of the *American Convention on Human Rights (the Pact of San José)*, opened for signature 22 November 1969, 1144 UNTS 123 (entered into force 18 July 1978). Other specialised legal frameworks in the Americas address specific forms of discrimination and include Organization of American States, *Inter-American Convention on the Elimination of All Forms of Discrimination Against Persons with Disabilities*, Treaty No A-65, opened for signature 8 June 1999 (entered into force 14 September 2001); Organization of American States, *Inter-American Convention Against Racism, Racial Discrimination, and Related Forms of Intolerance*, Treaty No A-68, opened for signature 5 June 2013 (entered into force 11 November 2017); Organization of American States, *Inter-American Convention on Protecting the Human Rights of Older Persons*, opened for signature 6 June 2015, 55 ILM 989 (entered into force 11 January 2017). See also *Convention Explanatory Report* (n 98) 16 [71]–[73].

105 *Convention Explanatory Report* (n 98) 17 [77].

106 Ibid 16–17 [75]–[76].

107 *The Convention* (n 2) art 16.

108 Ibid art 16(2)(c).

This includes developing specific expertise among individuals responsible for identifying, assessing, preventing, and mitigating the risks associated with artificial intelligence systems. This provision introduces an educational dimension, aiming to enhance understanding and awareness of AI-related risks and impacts throughout society.[109]

Additionally, *The Convention* mandates several key measures to proactively address potential biases and discriminatory impacts in AI systems. These measures include monitoring for risks and adverse impacts; thorough documentation of risks, both actual and potential, alongside the chosen risk management approach; and testing of AI systems before they are deployed for initial use and whenever they undergo significant modifications.[110] While *The Convention* addresses discrimination broadly and includes gender equality as a specific concern, its treatment of sex, gender, and sexuality remains general. From a queer-responsive perspective, its lack of explicit reference to queer individuals and its silence on how AI systems encode binary assumptions limit its capacity to fully address queer biases in healthcare AI. However, its focus on structural inequalities and participatory assessments provides an opening for future interpretation or implementation aligned with queer-informed standards.

B The Technical-Oriented Model

Unlike a principle-based model of regulation, a technical-oriented approach assigns policymakers the responsibility of regulating technology through mechanisms and requirements aimed at reducing or eliminating biases in AI, with a primary focus on technical solutions.[111] This approach is inherently rules-based,[112]

109 Ibid art 20.

110 Ibid arts 16(2)(e)–(g).

111 Riikka Koulu, 'Human Control over Automation: EU Policy and AI Ethics' (2020) 12(1) *European Journal of Legal Studies* 9 <https://doi.org/10.2924/EJLS.2019.019>. See also Matthew U Scherer, 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies' (2015) 29(2) *Harvard Journal of Law and Technology* 353 <https://doi.org/10.2139/ssrn.2609777>; Miriam C Buiten, 'Towards Intelligent Regulation of Artificial Intelligence' (2019) 10(1) *European Journal of Risk Regulation* 41 <https://doi.org/10.1017/err.2019.8>; Marco Almada, 'Regulation by Design and the Governance of Technological Futures' (2023) 14(4) *European Journal of Risk Regulation* 697 <https://doi.org/10.1017/err.2023.37>; Pieter Van Cleynenbreugel, 'EU By-Design Regulation in the Algorithmic Society: A Promising Way Forward or Constitutional Nightmare in the Making?' in Hans-W Micklitz et al (eds), *Constitutional Challenges in the Algorithmic Society* (Cambridge University Press, 2021) 202 <https://doi.org/10.1017/9781108914857.011>; Erica Palmerini et al, 'RoboLaw: Towards a European Framework for Robotics Regulation' (2016) 86 *Robotics and Autonomous Systems* 78 <http://dx.doi.org/10.1016/j.robot.2016.08.026>.

112 Cary Coglianese, 'Rule Design: Defining the Regulator–Regulatee Relationship' in Jean-Christophe Le Coze and Benoît Journé (eds), *The Regulator–Regulatee Relationship in High-Hazard Industry Sectors: New Actors and New Viewpoints in a Conservative Landscape* (Springer Nature, 2024) 89 <https://doi.org/10.1007/978-3-031-49570-0>. See also Brigitte Burgemeestre, Joris Hulstijn, and Yao-Hua Tan, 'Rule-Based Versus Principle-Based Regulatory Compliance' (2009) 205 *Legal Knowledge and Information Systems* 37 <https://doi.org/10.3233/978-1-60750-082-7-37>.

as it relies on specific, prescriptive rules that guide actions to achieve regulatory objectives.[113]

In this sense, a rules-based regulation offers a higher level of specificity regarding the required behaviour.[114] In general, legal doctrine has highlighted both the advantages and disadvantages of this type of regulation in addressing technological challenges.[115]

Among the advantages is that regulated entities have a higher degree of precision regarding their obligations, which enhances the certainty of economic relations.[116] Additionally, this type of regulation allows for stronger enforcement potential and more straightforward verification of compliance by regulated entities.[117] However, downsides include the risk of over- or under-inclusion of certain behaviours.[118] Legal doctrine has also observed that such legislation may foster a 'tick-box mindset' among regulated entities,[119] leading them to 'lose sight of ensuring their actions are consistent with a larger regulatory objective.'[120] Finally, this type of regulation presents challenges related to the expertise and competence of policymakers, who must possess the specialised knowledge required to create effective rules.[121]

In summary, this model incorporates normative solutions with a high level of specificity, prescribing a set of technical requirements for AI development and use to ensure equality by design. In the landscape of AI regulation and proposals, this approach is exemplified in the *EU AI Act*.

1 The EU Artificial Intelligence Act

In the technical-oriented model, the *EU AI Act*[122] serves as a prototypical regulation.

---

113 Christopher Decker, 'Goals-Based and Rules-Based Approaches to Regulation' (Research Paper No 8, Department for Business, Energy and Industrial Strategy, May 2018).

114 Schuett et al, 'From Principles to Rules' (n 78) 28. See also Lyria Bennett Moses, 'How to Think about Law, Regulation and Technology: Problems with "Technology" as a Regulatory Target' (2013) 5(1) *Law, Innovation and Technology* 1 <https://doi.org/10.5235/17579961.5.1.1>.

115 Decker (n 113) 20–7.

116 Ibid 22. See also John Braithwaite, 'Rules and Principles: A Theory of Legal Certainty' (2002) 27 *Australian Journal of Legal Philosophy* 47 <https://doi.org/10.2139/ssrn.329400>.

117 Decker (n 113) 22–3; Louis Kaplow, 'Rules Versus Standards: An Economic Analysis' (1992) 42(3) *Duke Law Journal* 557 <https://doi.org/10.2307/1372840>.

118 Florentin Blanc, 'Tools for Effective Regulation: Is "More" Always "Better"?' (2018) 9(3) *European Journal of Risk Regulation* 465 <https://doi.org/10.1017/err.2018.19>; Lyria Bennett Moses, 'Why Have a Theory of Law and Technological Change?' (2007) 8(2) *Minnesota Journal of Law, Science and Technology* 589.

119 Decker (n 113) 23.

120 Ibid. See also Cass R Sunstein, 'Deciding by Default' (2013) 162(1) *University of Pennsylvania Law Review* 1.

121 Michael Guihot, Anne F Matthew and Nicolas P Suzor, 'Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence' (2017) 20(2) *Vanderbilt Journal of Entertainment and Technology Law* 385 <https://doi.org/10.31228/osf.io/5at2f>.

122 *EU AI Act* (n 3). See Johann Laux, Sandra Wachter, and Brent Mittelstadt, 'Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk' (2024) 18(1) *Regulation and Governance* 3 <https://doi.org/10.1111/rego.12512>; Claudio Novelli et al, 'Taking AI Risks Seriously: A New Assessment Model for the AI Act' (2024) 39(5) *AI and*

This regulation stands as the first comprehensive legislative framework aimed at overseeing AI across all sectors within the EU. Its declared purpose is to enhance the internal market's functioning and foster the adoption of human-centric and trustworthy artificial intelligence, while safeguarding health, safety, and fundamental rights enshrined in the *Charter of Fundamental Rights of the European Union*, including democracy, the rule of law, and environmental protection, from the harmful effects of AI systems.[123] As an EU regulation, the *EU AI Act* has direct effect on providers, deployers, and importers of AI systems within the EU.[124]

The *AI Act* establishes rules and requirements for the development and deployment of AI systems, based on a risk pyramid that categorises AI systems by risk level.[125] Under this structure:

- Unacceptable risk AI systems are prohibited. This includes systems like social scoring and manipulative AI.
- High-risk AI systems, which make up the core focus of the *EU AI Act*, are subject to strict regulatory requirements.
- Limited-risk AI systems are subject to lighter transparency obligations, requiring developers and deployers to ensure that end-users are informed when they are interacting with AI (eg, chatbots and deepfakes).
- Minimal-risk AI systems are not regulated.

Most relevant to our analysis are the systems categorised as high-risk. These systems are identified according to a general rule set out in article 6, which specifies criteria for high-risk classification, as well as through a list of specific applications detailed in annex III.[126] As a matter of fact, as stated in recital 7, to ensure consistent

---

Society 2493 <https://doi.org/10.1007/s00146-023-01723-z>; Rostam J Neuwirth, 'Prohibited Artificial Intelligence Practices in the Proposed EU Artificial Intelligence Act (AIA)' (2023) 48 *Computer Law and Security Review* 105798:1–14 <https://doi.org/10.1016/j.clsr.2023.105798>.

123 *EU AI Act* (n 3) recital 1. It is important to notice that the legal basis of the regulation is grounded on article 114 of the *TFEU* (n 104) that is 'the central Treaty provision' for harmonising the laws of EU Member States for achieving the internal market objective such as 'establishing and ensuring the functioning of an area without internal frontiers in which the free movement of goods, persons, services, and capital is ensured'. See Manuel Kellerbauer, 'Article 114 *TFEU*', in Manuel Kellerbauer, Marcus Klamert and Jonathan Tomkin (eds), *The EU Treaties and the Charter of Fundamental Rights: A Commentary* (Oxford University Press, 1st ed, 2019) 1235 <https://doi.org/10.1093/oso/9780198759393.003.212>. See also Sacha Garben, 'Confronting the Competence Conundrum: Democratising the European Union through an Expansion of Its Legislative Powers' (2015) 35(1) *Oxford Journal of Legal Studies* 55, 63 <https://doi.org/10.1093/ojls/gqu021>, arguing that 'the Treaty's functional powers' such as article 114 of the *TFEU* (n 104) (citations omitted):

> can cut horizontally through virtually all policy areas, including those where the EU has no, or only complementary, competence. This means that the EU can, through implied powers, legislate in areas that are considered to fall within national autonomy, leading to what can be called 'harmonisation through the back door'.

124 *EU AI Act* (n 3) art 2(1).

125 On the risk-based approach of the *EU AI Act* (n 3), see Giovanni De Gregorio and Pietro Dunn, 'The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age' (2022) 59(2) *Common Market Law Review* 473 <https://doi.org/10.54648/COLA2022032>. On the risk-based approach to regulation, see generally Julia Black and Robert Baldwin, 'When Risk-Based Regulation Aims Low: Approaches and Challenges' (2012) 6(1) *Regulation and Governance* 2 <http://dx.doi.org/10.1111/j.1748-5991.2011.01124.x>.

126 *EU AI Act* (n 3) art 6(1)–(2), annex III.

and robust protection of public interests regarding health, safety, and fundamental rights, a unified set of rules for high-risk AI systems is necessary.[127]

A critical factor in classifying an AI system as high-risk is the potential adverse impact it may have on fundamental rights protected by the *Charter of Fundamental Rights of the European Union*. Among these rights, recital 48 explicitly highlights the right to non-discrimination and gender equality.[128]

The risk of discriminatory outcomes is indeed acknowledged as a basis for establishing tailored rules and requirements for high-risk systems used in specific contexts, as outlined in annex III. These contexts include, but are not limited to, biometric systems used for categorisation and emotion recognition; educational and vocational training; access to public and private services; law enforcement; employment; migration, asylum, and border management; and the administration of justice.[129]

Among the requirements that high-risk systems must meet, it is possible to identify those specifically aimed at addressing the risk of discrimination.[130] As will be argued, these requirements align closely with an equality by design approach, establishing rules and requirements that pertain to the quality and relevance of datasets used, their management, and the models. These criteria underscore the technical characteristics required of high-risk AI systems, grounding equality protections directly within the design and operational stages of these technologies.[131] The *EU AI Act* recognises in fact that high-quality data is crucial for ensuring that high-risk AI systems perform safely, as intended, and without causing unlawful discrimination.[132] Effective data governance and management practices are therefore required to ensure that training, validation, and testing datasets are relevant, representative, accurate, and complete.[133] Such practices must also account for personal data protection, addressing potential biases in datasets that could reinforce discrimination, particularly for vulnerable groups.

To this end, article 10 establishes a system of data quality and governance through a series of requirements that must be met in the training, validation, and testing data of high-risk AI systems. These data governance measures particularly concern relevant design choices,[134] data collection processes and the origins of

127 Ibid recital 7.

128 Ibid recital 48.

129 Ibid annex III. On the identification of high-risk systems, see Jonas Schuett, 'Risk Management in the *Artificial Intelligence Act*' (2024) 15(2) *European Journal of Risk Regulation* 367 <https://doi.org/10.1017/err.2023.1>.

130 See Luca Deck et al, 'Implications of the *AI Act* for Non-discrimination Law and Algorithmic Fairness' (Research Paper, Central Europe Workshop Proceedings, 29 March 2024) <https://doi.org/10.48550/arXiv.2403.20089>.

131 Alba Soriano Arnanz, 'Creating Non-discriminatory Artificial Intelligence Systems: Balancing the Tensions Between Code Granularity and the General Nature of Legal Rules' (2024) 38 *Revista de Internet, Derecho y Política* 1 <https://doi.org/10.7238/idp.v0i38.403794>.

132 Philipp Hacker, 'A Legal Framework for AI Training Data: From First Principles to the *Artificial Intelligence Act*' (2021) 13(2) *Law, Innovation and Technology* 257 <https://doi.org/10.1080/17579961.2021.1977219>.

133 *EU AI Act* (n 3) recital 67.

134 Ibid art 10(2)(a).

data,[135] and measures to detect, prevent, and mitigate possible biases[136] through examinations aimed at identifying biases likely to lead to discrimination prohibited under EU law.[137]

Moreover, training, validation, and testing datasets must be relevant, sufficiently representative, and, to the greatest extent possible, free of errors and complete concerning the intended purpose.[138] They should possess appropriate statistical properties, including, where applicable, considerations regarding the persons or groups for whom the high-risk AI system is intended to be used.

Clearly, these practices address the problem of technical biases linked to data quality issues I have identified, particularly the lack of representativeness and model-related biases associated with feature selection.[139]

As noted in legal scholarship, the article emphasises specific 'phases of the design process of AI systems to which special attention should be paid'.[140] This focus enhances the chances of identifying and preventing potential discriminatory elements, particularly given that those responsible for designing automated systems 'are often unaware of the risks' these systems pose to equality and non-discrimination rights.[141]

The article also includes two specific measures that extend into what I have identified as sociotechnical issues, related to biases in the meanings attributed to data and implementation biases such as context biases.

Specifically, article 10(2)(d) prescribes that data governance shall also concern the formulation of assumptions, particularly regarding what the data are supposed to measure and represent. Additionally, article 10(4) states that datasets shall take into account, to the extent required by the intended purpose, the characteristics or elements particular to the specific geographical, contextual, behavioural, or functional setting within which the high-risk AI system is intended to be used.

Additionally, the *AI Act* provides another tool known as the Fundamental Rights Impact Assessment ('FRIA'),[142] outlined in article 27. This tool aligns with a by design approach as it aims to ensure the technological quality of AI systems regarding potential violations of fundamental rights, including equality and non-discrimination.[143] Essentially, the FRIA sets a qualitative standard for technological compliance.[144] However, it is important to note that the scope of FRIA's application is limited to deployers of high-risk AI systems that are either public bodies or

135 Ibid art 10(2)(b).
136 Ibid art 10(2)(g).
137 Ibid art 10(2)(f).
138 Ibid art 10(3).
139 Hacker (n 132).
140 Arnanz (n 131) 9.
141 Ibid.
142 On the Fundamental Rights Impact Assessment, see Alessandro Mantelero, 'The Fundamental Rights Impact Assessment (FRIA) in the *AI Act*: Roots, Legal Obligations and Key Elements for a Model Template' (2024) 54 *Computer Law and Security Review* 106020:1–18 <https://doi.org/10.1016/j.clsr.2024.106020>.
143 *EU AI Act* (n 3) art 27.
144 Ibid.

private entities delivering public services.[145] This means the FRIA doesn't extend to all deployers, as was initially outlined in article 29 of the European Parliament's original draft.[146]

The FRIA is structured to ensure technological compliance based on several specific requirements. Deployers must conduct an assessment that includes: i) describing the deployer's processes where the high-risk AI system will be used according to its intended purpose; ii) outlining the time period and frequency for the intended use of each high-risk AI system; iii) identifying categories of individuals or groups likely to be impacted in the specific context; iv) assessing specific risks of harm that may affect the identified individuals or groups, taking into account information provided by the AI system's provider; v) describing human oversight measures implemented, following the system's instructions for use; and vi) detailing measures to be taken if these risks materialise, including arrangements for internal governance and complaint mechanisms.[147]

As noted by Alessandro Mantelero, the *EU AI Act* 'does not give due attention to participation in assessment procedures, contrary to best practices in impact assessment'.[148]

Additionally, the *EU AI Act* specifies other mandatory requirements for high-risk systems aimed at mitigating risks to fundamental rights through a by design approach, that require technical documentation (article 11), record-keeping (article 12), transparency and information provision to users (article 13), human oversight (article 14), and accuracy (article 15).

From a queer-responsive standpoint, the *EU AI Act* provides mechanisms to address technical bias, especially in relation to data quality and representativeness. However, it falls short in addressing sociotechnical biases. For example, while the Act encourages attention to statistical representativeness, it does not mandate the inclusion of diverse sex and gender markers or challenge binary modelling assumptions. Similarly, the FRIA, though potentially powerful, has limited scope and does not guarantee participation from queer communities.

C The Sociotechnical-Oriented Model

The sociotechnical-oriented model combines both technical solutions with broader societal considerations.[149] This integrated approach recognises that addressing AI bias requires not only technical regulatory solutions but also societal engagement and awareness.

---

145 Ibid.

146 See Mantelero (n 142) 8 (citations omitted):

> However, the main difference between the European Parliament's proposal and the adopted text concerns the scope of the FRIA. Under pressure from the other two co-legislators, it was restricted to a limited area, whereas the text proposed by the Parliament referred to all high-risk AI systems as defined in Article 6(2), with the sole exception of systems used for management and operation of critical infrastructure.

147 *EU AI Act* (n 3) art 27.

148 Mantelero (n 142) 8.

149 Mateusz Dolata, Stefan Feuerriegel, and Gerhard Schwabe, 'A Sociotechnical View of Algorithmic Fairness' (2022) 32(4) *Information Systems Journal* 754 <https://doi.org/10.1111/isj.12370>.

A sociotechnical approach to AI regulation starts from the assumption that technology and society should be seen as 'a coherent system' recognising that 'a technology's real-world safety and performance is always a product of technical design and broader social forces'.[150]

In this sense, it has been argued that the problem with a strictly technical-oriented approach is that AI harms do not stem only from the characteristics of the technology but, first and foremost, from the 'broader systems around the technology, ie, from the interaction between the technical and the social'.[151]

Indeed, AI generates power imbalances and inequalities also due to the interaction between humans and machines, as clearly emerges from those sociotechnical biases identified both as foundational, due to the lack of representativeness in the field and the meanings attributed to certain data, and as interpretative biases that depend primarily on the social context of AI use.

Although it remains a rules-based model, as does the technical-oriented approach, the sociotechnical model extends regulatory measures beyond technical specifications to include societal dimensions.[152] These measures are aimed at implementing stakeholder participation, particularly among population groups affected by AI systems, and promoting diversity in the sector. Furthermore, these measures extend to sociotechnical considerations by recognising the need to work on data quality from the perspective of the meaning attributed to the data.

The sociotechnical model also presents certain challenges. The complexity in implementation due to the integration of multiple approaches can make the development and deployment process more burdensome,[153] as coordination between technical experts and societal stakeholders requires significant effort and resources. There is potential for slower processes due to the participatory elements,[154] as engaging various stakeholders and incorporating their feedback can

---

150 Brian J Chen and Jacob Metcalf, 'Explainer: A Sociotechnical Approach to AI Policy', *Data and Society* (Policy Brief, May 2024) <https://datasociety.net/wp-content/uploads/2024/05/DS_Sociotechnical-Approach_to_AI_Policy.pdf>.

151 Ibid 7. See also Ramona Vijeyarasa, 'Gendered Harms and the Regulation of Artificial Intelligence: A Comparative Assessment of Emerging Legislative Practice' (2023) 5(1) *Notre Dame Journal on Emerging Technologies* 114. See also Eleanor Drage and Kerry Mackereth, 'Does AI Debias Recruitment: Race, Gender, and AI's "Eradication of Difference"' (2022) 35(4) *Philosophy and Technology* 89:1–25, 18 <https://doi.org/10.1007/s13347-022-00543-1>, where the authors argue that 'attempts to "strip" gender and race from AI systems often misunderstand what gender and race are, casting them as isolatable attributes rather than broader systems of power'. See also Robert Holton and Ross Boyd '"Where Are the People? What Are They Doing? Why Are They Doing It?"(Mindell): Situating Artificial Intelligence within a Socio-Technical Framework' (2021) 57(2) *Journal of Sociology* 179 <https://doi.org/10.1177/1440783319873046>.

152 Institute for Trustworthy AI in Law and Society, 'Response to NIST RFI on AI Executive Order 14110: The Importance of a Socio-Technical Approach in AI Development', Submission to the National Institute of Standards and Technology, *Privacy Framework* (2 February 2024).

153 Sheila Jasanoff, *States of Knowledge: The Co-production of Science and Social Order* (Routledge, 2004) <https://doi.org/10.4324/9780203413845>.

154 On participatory design, see Maja van der Velden and Christina Mörtberg, 'Participatory Design and Design for Values' in Jeroen van den Hoven, Pieter E Vermaas and Ibo van de Poel (eds), *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains* (Springer, 2015) 41 <https://doi.org/10.1007/978-94-007-6970-0_33>.

delay the deployment of AI systems.[155] Additionally, measuring compliance across diverse criteria, technical and social, poses challenges for regulators and regulated entities alike. The lack of clear metrics for societal impact can make it difficult to assess whether the intended outcomes are being achieved.[156]

1 The Brazilian AI Bill

The Brazilian AI Bill establishes a broad regulatory framework for the development, implementation, and use of AI systems across all sectors, without focusing on any specific industry.[157] Like the *EU AI Act*, this legislation categorises AI systems based on risk levels, distinguishing between 'excessive-risk' and 'high-risk' AI systems. Excessive-risk AI systems include those AI systems that: (i) employ subliminal techniques to induce behaviour in others that is detrimental or dangerous to their health or safety, or against the principles of Brazil's Proposed AI Regulation; (ii) exploit vulnerabilities of specific groups of persons (e.g., age, or physical or mental disability), to induce behaviour that is detrimental to their health or safety, or against the principles of Brazil's Proposed AI Regulation; or (iii) are implemented by the government for the purposes of social scoring.[158] Such excessive-risk AI systems will be prohibited, while others will be subject to regulation by the competent authority.

High-risk AI systems include AI systems used for certain purposes, such as: security devices in critical infrastructures (such as traffic control, water, and electricity supply networks); credit assessments; certain autonomous vehicles; applications in the healthcare sector; biometric identification systems; and criminal investigation and public security.[159]

The Brazilian AI Bill introduces enforceable rights for individuals impacted by AI systems, including the right to non-discrimination and to the correction of direct, indirect, illegal, or abusive discriminatory biases. To ensure these rights are actionable, providers and operators must establish procedures that enable individuals to exercise them effectively. Individuals affected by AI systems also have the right to receive clear, adequate information prior to the system's use, particularly concerning measures adopted to mitigate the risk of discrimination.

---

155 Other limitations of the participatory design for AI are described by Abeba Birhane et al, 'Power to the People: Opportunities and Challenges for Participatory AI' (Conference Paper, Associations of Computing Machinery Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, 15 September 2022) <https://doi.org/10.1145/3551624.3555290>. See also Archon Fung, 'Varieties of Participation in Complex Governance' (2006) 66(sup 1) *Public Administration Review* 66 <https://doi.org/10.1111/j.1540-6210.2006.00667.x>. Participatory design has been analysed in law reform proposal dealing with decertification of sex and gender from a 'prefigurative' standpoint by Davina Cooper, 'Crafting Prefigurative Law in Turbulent Times: Decertification, DIY Law Reform, and the Dilemmas of Feminist Prototyping' (2023) 31(1) *Feminist Legal Studies* 17 <https://doi.org/10.1007/s10691-022-09515-4>.

156 Ibid.

157 'Brazilian AI Bill' (n 4).

158 Ibid art 14.

159 Ibid art 17.

In addition, chapter IV of the Brazilian AI Bill outlines specific provisions aimed at implementing principles of equal treatment and non-discrimination, addressing these goals from both technical and sociotechnical perspectives.

On the technical side, similar to the technical-oriented approach, the Brazilian AI Bill requires measures to mitigate biases. These include transparency in the use of AI systems in human interactions, which encompasses the use of clear, informative human-machine interfaces, and the transparent disclosure of governance measures adopted by organisations in developing and employing AI systems.

Data management practices also play a crucial role, as providers and operators must implement adequate data handling processes to mitigate and prevent discriminatory biases. Furthermore, the processing of personal data must comply with data protection legislation through privacy-by-design and default measures and techniques that minimise the use of personal data. The Brazilian AI Bill also mandates appropriate parameters for data separation and organisation during the training, testing, and validation stages, ensuring that data handling is robust and does not introduce biases inadvertently. Security measures are similarly emphasised, with requirements for adequate information security practices spanning from the initial design phase to the operational stage of the AI system.

Governance measures apply throughout the entire life cycle of AI systems, from initial design to the end of their activity or discontinuation. High-risk AI systems, in particular, require up-to-date technical documentation to be prepared prior to market entry or service provision, and maintained consistently during use.

For high-risk systems, two measures reflect a purely sociotechnical approach.

Article 20(IV)(a) requires that controls be established to address human cognitive biases during data collection and organisation, aiming to reduce bias risks associated with the significance attributed to specific data categories, to prevent the generation of biases due to classification issues, errors, or lack of information regarding affected groups, lack of coverage, or distortions in representativity according to the intended application.

Furthermore, article 20(IV)(b) calls for inclusive team composition in the system's design and development, prioritising diversity as a guiding principle. These measures demonstrate a comprehensive regulatory approach, recognising the technical and social dimensions essential for mitigating biases in AI systems.

In addition to the governance measures, public bodies and entities at the federal, state, and municipal levels, when contracting, developing, or using high-risk artificial intelligence systems, are requested to conduct prior public consultations and hearings on the planned use of artificial intelligence systems, providing information about the data to be used, the general operational logic, and the results of tests conducted.

Finally, the algorithmic impact assessment tool, already present in the other analysed legislation, is further extended in Brazil's framework to enhance stakeholder participation. First and foremost, the Brazilian AI Bill stipulates that the process and results of tests and assessments, as well as mitigation measures conducted to verify possible rights impacts, with particular attention to potential discriminatory impacts, must be included. Furthermore, the competent authority

may establish additional criteria and elements for conducting impact assessments, incorporating the participation of different affected social segments.

The algorithmic impact assessment will be a continuous, iterative process carried out throughout the entire life cycle of high-risk artificial intelligence systems, requiring periodic updates to remain current. The updating process of the algorithmic impact assessment shall involve public participation through a consultation process with relevant stakeholders. This participatory element reflects an acknowledgement of the importance of community input in the ongoing assessment of AI risks.

From a queer-responsive perspective, the Brazilian AI Bill stands out for its explicit attention to the sociotechnical dimensions of algorithmic harm, particularly through its mandates on team diversity, participatory algorithmic impact assessments, and the recognition of human cognitive biases in data classification. These provisions are especially relevant to queer individuals in healthcare contexts, where foundational biases about sex and gender categories often go unchallenged. By requiring inclusive team composition and continuous stakeholder engagement, including affected communities, the Bill creates regulatory space for queer participation in shaping AI systems. Moreover, its emphasis on recognising human cognitive biases offers a pathway to disrupt binary medical norms that frequently exclude or misclassify trans, intersex, and non-binary people.

V CONCLUSION: ASSESSING QUEER RESPONSIVENESS IN HEALTHCARE

The analysis of AI biases impacting sexual, gender, and sex minorities in healthcare reveals clear distinctions in queer-responsiveness across the three legislative models. The principle-based model, exemplified by *The Convention*, recognises algorithmic discrimination as a central concern, explicitly emphasising the need to ensure gender equality and reduce systemic inequalities through general principles and risk assessments. However, despite its valuable high-level guidance, this approach has critical limitations due to its inherent lack of specificity. The absence of precise implementation guidelines leaves significant uncertainty regarding compliance obligations, weakening practical enforceability and limiting its responsiveness to deeply rooted biases emerging from binary conceptualisations of sex and gender data in healthcare.

In contrast, the technical-oriented model, illustrated by the *EU AI Act*, adopts detailed technical requirements designed explicitly to tackle biases through stringent data governance and technical design measures, such as ensuring data representativeness, accuracy, and proper feature selection. These measures have substantial potential to mitigate purely technical biases identified in healthcare, such as diagnostic inaccuracies and under-representation of diverse groups in medical datasets. Nonetheless, this model falls short of addressing critical sociotechnical biases, particularly foundational biases such as sex/gender slippage, sex confusion, and sex obsession, due to its inadequate consideration of the societal

dimension. Indeed, the *EU AI Act* demonstrates limited attention to educational and participatory mechanisms essential for reshaping the rigid and binary assumptions deeply embedded in healthcare data and practices, as underscored by the minimal role of public participation in the FRIA under article 27.

The sociotechnical-responsive model, represented by the Brazilian AI Bill, emerges as uniquely well-equipped for queer-responsive regulation of AI bias in healthcare, precisely due to its combined emphasis on both technical and social dimensions. Like the EU model, it requires rigorous technical measures aimed at eliminating biases related to data quality and management. Crucially, however, the Brazilian legislation further expands its regulatory toolkit by explicitly mandating participation and education, both of which are essential for confronting foundational and interpretative biases around sex, gender, and sexuality data. Notably, the Brazilian AI Bill includes measures aimed directly at reducing human cognitive biases during data collection (article 20(IV)(a)) and mandates diversity within AI design teams (article 20(IV)(b)). Additionally, the Brazilian AI Bill stipulates iterative, publicly participatory Algorithmic Impact Assessments throughout the AI lifecycle, further integrating societal input into AI deployment decisions. By combining rigorous technical requirements with comprehensive participation and education measures, the sociotechnical model effectively promotes substantive equality in its multidimensional aspects of recognition, representation, and participation, thus demonstrating superior queer-responsiveness in addressing deeply entrenched sociotechnical biases prevalent in healthcare contexts.