GENDER, AI AND HEALTHCARE: A SUBSTANTIVE EQUALITY APPROACH

Queer-Responsive Regulation for AI in Healthcare

Presenting findings from a comparative study of AI regulation frameworks through a queer theoretical lens:
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5292012

Sergio Sulmicelli

Bilbao, June 6, 2025

"MEDICINE+AI, Law and Ethics for an Augmented and Human-Centered Medicine" (2022YB89EH) – CUP E53D23007020006.

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

UNIVERSITÀ DI TRENTO

I. Introduction

- **Research Question:**
  The central research question is: How can a queer-responsive regulatory approach effectively address algorithmic bias in healthcare-related AI, particularly biases affecting sexual and gender minorities?

- **Methodology:**
  queer theoretical framework that:
  - Challenges fixed categories of sex, gender, and sexuality as socially constructed rather than biologically essential.
  - Reveals how both law and AI systems replicate and intensify biases against those who deviate from dominant norms.
  - Integrates comparative legal analysis of three regulatory models, each exemplifying a different orientation—principles-based, technical-oriented, and sociotechnical-oriented.

- **Overview of Sections:**
  II. introduces and categorizes sources of algorithmic bias.
  III. focuses on AI in healthcare, examining how sex and gender biases manifest in diagnostic tools, treatment recommendations, and specific case studies
  IV. undertakes a comparative analysis of three regulatory frameworks, evaluating them against a queer-responsive criterion: Do they explicitly or implicitly address the complexity of sex and gender, data, and do they promote participation from affected communities?
  V. draws conclusions on which model most effectively addresses queer-specific biases

II. Defining Algorithmic Bias and Discrimination

- "Bias" → a deviation from some neutral standard. In AI, however, bias is a necessary feature: it allows algorithms to identify and weight statistical patterns in complex datasets.

- Friedman and Nissenbaum define a "biased AI system" as one that "systematically and unfairly discriminates against certain individuals or groups [by] denying opportunities or assigning undesirable outcomes on grounds that are unreasonable or inappropriate."

- **1. Technical Sources of Bias**

Solon Barocas and Andrew Selbst:
- Target Variables: If the objective the AI optimizes already reflects historical inequities.
- Training Data: Imbalances or historical prejudices in data collection can skew predictions.
- Relevant Features: Input variables chosen for the model.
- Proxy Variables: Attributes correlated with sensitive characteristics →proxies enable the model to discriminate indirectly.
- Intentional Discrimination ("Masking")

II. Defining Algorithmic Bias and Discrimination (Part 2)

- **2. Sociotechnical Sources of Bias**
  Nissenbaum and Friedman

- Preexisting Bias: Structural, historical inequalities, such as the longstanding exclusion of trans and intersex people from medical research datasets.
- Technical Bias: Arising from design choices and technical constraints, like relying on binary sex categories in algorithm architecture.
- Emergent Bias: Produced by the deployment context, when an algorithm operates in a setting that differs from its original design assumptions, failing to account for shifting cultural norms or evolving social knowledge.

- **3. Foundational Sources of Bias**
  two foundational sources of bias that underlie and precede the others:

- Lack of diversity in the tech field: AI researchers, developers, and engineers predominantly come from similar backgrounds
- Misinterpretation and bias in sex, gender, and sexuality data: Big data are never neutral; the cultural contexts in which data are gathered and interpreted influence every aspect of the dataset, from what counts as “male” or “female” to how “gender” is recorded.

II. Defining Algorithmic Bias and Discrimination (Part 2)

Therefore, we can synthesize three broad categories of bias:

- **Foundational Bias**: Lack of diversity among AI professionals; Misinterpretation of sex, gender, and sexuality as data points.

- **Technical Bias**: Architectural choices (e.g., feature selection, binary classifiers); Data quality issues (e.g., biased training sets, missing datasets on trans, intersex or non-binary individuals).

- **Implementation & Interpretation Bias**: Contextual misuse (applying an algorithm outside its intended population); Misreading outputs due to rigid categories (e.g., misgendering in facial recognition).

II. Defining Algorithmic Bias and Discrimination (Part 2)

To illustrate how these interact:
A *foundational* bias may be the underrepresentation of transgender bodies in medical imaging data.
→ When that dataset informs an algorithm's architecture (e.g., a binary classifier that recognizes only "male" or "female"), we see a *technical* bias, the rigid design reproduces the original data's exclusions → Once deployed the system misclassifies transgender patients, generating an *implementation/interpretation* bias that can lead to misdiagnosis or denial of care.

III. Sex and Gender Bias: AI in Healthcare as Case Study (Part 1)

- **Biological and Social Data in Healthcare AI**

Medical practice undoubtedly relies on biological data. However, if AI systems reduce everyone (and every biological significance) to "male" or "female," they may:

- <u>Over-Inclusive Errors: </u>Assuming that anyone categorized as "female" shares identical anatomical or physiological traits, thereby ignoring variations (e.g., trans women on hormone therapy, intersex variations).
- <u>Under-Inclusive Errors: </u>Excluding or misclassifying intersex, transgender, and non-binary individuals whose bodies or self-identities do not fit neatly into binary categories.

Sex Data: When medical records record "sex" as a single category, we do not know whether it refers to sex assigned at birth, legal sex marker, current anatomy, or hormonal status. This ambiguity creates foundational bias, because one data point cannot capture all clinically relevant information about a patient's biological and social reality.

Gender Data: Traditionally, "gender" has been viewed as the social expression of "biological" sex. A queer perspective underscores that gender is performed daily, through norms, roles, expressions and that those who do not adhere to binary norms face stigma and discrimination. This means that simply recording "man" or "woman" in medical records fails to capture the lived experiences of trans, intersex, and gender-nonconforming individuals.

A queer-responsive approach to healthcare AI thus calls for:
- Precise, diverse biological markers (hormones, anatomy, genetics).
- Self-identified gender categories that go beyond binary options.
- Contextual information about social determinants of health (stigma, discrimination, access barriers).

III. Sex and Gender Bias: AI in Healthcare as Case Study (Part 2)

Key Concepts from Albert & Delano (2021)

Sex/Gender Slippage
Occurs when “sex” (biological characteristics such as chromosomes, gonads, hormones) is used interchangeably with “gender” (social identity, roles, and expressions). In medical records, terms like “male” or “female” (biological sex) frequently get conflated with “man” or “woman” (gender identity).

→ This conflation erases trans and intersex experiences, assuming sex and gender are concordant. A “male” record might hide a transgender woman’s status, leading to inappropriate clinical decisions (e.g., ignoring hormone replacement therapy effects).

Sex Confusion
Highlights the ambiguity of a single sex marker in Electronic Health Records (EHRs). Does it mean sex assigned at birth, legal sex, or current physiological status? There is no universal standard, so an AI system cannot reliably interpret a solitary data point.

→ A transgender man undergoing a PAP test might be recorded as “female,” even if he no longer has a cervix. Similarly, intersex individuals’ anatomical variations get entirely ignored.

Sex Obsession
Denotes the overemphasis on sex assigned at birth as the primary or sole relevant variable. Clinicians and algorithms may fixate on “birth sex” even when current physiology or identity is more clinically relevant.

→ Transgender and non-binary individuals face misdiagnoses because their care needs differ from cisgender norms —for instance, a transgender woman on estrogen therapy may have different cardiovascular risk factors than cisgender women or men.

III. Sex and Gender Bias: AI in Healthcare as Case Study (Part 2)

Key Concepts from Albert & Delano (2021)

Albert and Delano illustrate these biases vividly in HIV/PrEP risk prediction models. PrEP (pre-exposure prophylaxis) requires accurate estimation of individual risk for HIV infection. However:

- Many ML models rely on datasets that exclude or misclassify transgender women, non-binary people, and sexual minority groups.

- Transgender women and men who have sex with men (MSM) often have unique risk factors—such as specific sexual practices or higher levels of social stigma—yet are omitted from training sets.

As a result, HIV risk prediction tools fail to flag at-risk queer individuals, barring them from timely PrEP access.

In sum, AI's "double-edged sword" potential means: on one side, it can advance personalized medicine by capturing biological and social differences; on the other, if foundational and technical biases remain unaddressed, healthcare AI can exacerbate existing disparities for queer populations.

IV. Comparative Analysis of Queer-Responsiveness in AI Regulation (Part 1)

We define “**queer-responsiveness”** as the extent to which a law or policy:
- Recognizes the social construction of sex, gender, and sexuality data.
- Mandates participation and representation of queer communities in AI governance.
- Addresses both technical (data quality, algorithm design) and sociotechnical (cultural meanings, cognitive biases) sources of bias.

We will analyze:
- A Principles-Based Model: the Council of Europe Framework Convention on AI.
- A Technical-Oriented Model: the EU Artificial Intelligence Act.
- A Sociotechnical-Oriented Model: the Brazilian AI Bill.

For each, we will consider:
- Scope & Source Type
- Equality & Non-Discrimination Provisions
- Technical Measures: Data governance, bias mitigation by design, impact assessments.
- Sociotechnical Measures: Participatory requirements, diversity mandates, recognition of foundational biases.

IV. Comparative Analysis of Queer-Responsiveness in AI Regulation (Part 2)

- **A. Principles-Based Model: Council of Europe Framework Convention on AI**

  1. Nature and Scope
  First legally binding international treaty on AI, called the Framework Convention on AI, Human Rights, Democracy, and the Rule of Law ("the Convention").
  Technology-neutral and principle-driven—does not prescribe technical specifications.
  Applies equally to public authorities and private actors when performing public functions or operating under delegation. States must decide how to apply principles to private sector activities, potentially through national laws or voluntary measures.

  2. Equality and Non-Discrimination (Article 10)
  Article 10 mandates adoption or maintenance of measures ensuring all AI lifecycle activities comply with equality, including gender equality, and prohibition of discrimination under international/domestic law.
  Goes beyond *ex post* redress; requires proactive obligations aiming for "fair, just, and equitable outcomes" by tackling structural inequalities.

  3. Specific Measures Mentioned
  Preamble explicitly signals concern about "risks of discrimination in digital contexts," including harms faced by women and "persons in vulnerable situations."
  Article 16: Participatory dimension, parties must "take into account, where appropriate, the perspectives of relevant stakeholders, in particular persons whose rights may be affected."
  Digital Literacy: Encouraging digital skills across all populations, targeting those responsible for identifying and mitigating AI risks.
  Risk Monitoring and Documentation: Parties must monitor and document risks, test AI systems pre-deployment and upon significant modifications.

  4. Queer-Responsiveness Assessment

  - Strong high-level commitment to equality, emphasizing structural and social inequities.
  - Participatory principle (Article 16) invites inclusion of affected individuals.
  - Addresses "social bias" explicitly as failure to incorporate historical inequalities.

IV. Comparative Analysis of Queer-Responsiveness in AI Regulation (Part 3)

- **B. Technical-Oriented Model: EU Artificial Intelligence Act**

  1. Nature and Scope
  A binding EU Regulation—the first comprehensive legislative framework for AI across all sectors in the EU.
  Employs a risk-based classification.
  Directly applies to providers, deployers, and importers of AI systems in the EU.

  2. Discrimination (Recitals 7 & 48; Annex III)
  Recital 48 emphasizes the right to non-discrimination and gender equality as key public interests.

  3. Equality by Design: Data Governance (Article 10)
  Article 10 requires:
  - Datasets used for training, validation, and testing to be "relevant, representative, accurate, and complete".
  - Data governance practices that detect, prevent, and mitigate biases likely to lead to unlawful discrimination.
  - Consideration of geographical, contextual, behavioral, and functional settings where the AI is used—addressing transfer context bias.

  4. Fundamental Rights Impact Assessment (FRIA) (Article 27)
  Deployers of high-risk AI (public bodies & private entities delivering public services) must:
  - Describe intended use and deployment processes; Identify categories of affected individuals or groups; Assess specific risks of harm (including discrimination); Document human oversight measures; Detail mitigation plans if risks materialize.

  Limited scope: only certain deployers must conduct FRIA; does not explicitly mandate community participation in assessment procedures.

  5. Additional Technical Requirements

  6. Queer-Responsiveness Assessment
  - Clear, specific requirements for data representativeness and bias detection in high-risk systems, helpful to remedy technical biases (e.g., requiring gender-balanced datasets).
  - FRIA provides a structured mechanism to identify risks to protected groups, including sex and gender minorities.

IV. Comparative Analysis of Queer-Responsiveness in AI Regulation (Part 4)

- **C. Sociotechnical-Oriented Model: Brazilian AI Bill**

  1. Nature and Scope
  A federal bill proposing a regulatory framework for AI across all sectors in Brazil.
  Categorizes AI into "excessive-risk" (prohibited) and "high-risk" (regulated) systems.

  2. Rights and Remedies for Individuals
  Introduces enforceable rights against AI-related discrimination, including:
  - The right to non-discrimination.
  - The right to correct discriminatory biases.
  - The right to clear information prior to system use (especially regarding bias mitigation measures).

  3. Technical and Sociotechnical Provisions (Chapter IV)
  Technical Measures (mirror the EU model):
  - Transparency: Clear human-machine interfaces and disclosures of governance measures.
  - Data Management: Adequate processes to mitigate and prevent discriminatory biases (privacy-by-design/ default).
  - Data Separation & Organization: Appropriate parameters during training, testing, validation.
  - Security Measures: Information security practices from design to deployment.
  - Governance Across Lifecycle: Up-to-date technical documentation for high-risk systems.

  Sociotechnical Measures:
  - Human Cognitive Bias Controls (Article 20(IV)(a)): Require controls during data collection to reduce classification errors, distortions, or underrepresentation of minority groups.
  - Inclusive Team Composition (Article 20(IV)(b)): Mandate diversity in AI design and development teams to broaden worldview and reduce foundational bias.

IV. Comparative Analysis of Queer-Responsiveness in AI Regulation (Part 4)

- **C. Sociotechnical-Oriented Model: Brazilian AI Bill**

  Participatory Requirements:
  - Public bodies at federal, state, and municipal levels must hold prior public consultations and hearings when contracting, developing, or using high-risk AI—providing details on data, operational logic, and test results.
  - Algorithmic Impact Assessment (AIA) extended to include:
    - Publication of processes, results, and mitigation measures, with specific attention to discriminatory impacts.
    - Competent authority can establish additional criteria for AIA, mandating participation from relevant social segments.
    - AIA must be continuous and iterative throughout the AI lifecycle, with mandatory public consultation updates.

  4. Queer-Responsiveness Assessment
  Strengths:
  - Explicit recognition of human cognitive biases, including those in data classification, critical to address sex/gender slippage and sex confusion.
  - Mandating diverse team composition directly tackles foundational bias by injecting queer perspectives into design.
  - Continuous, participatory AIA ensures that queer communities can voice concerns at each stage, data collection, testing, deployment, thus addressing emergent and interpretive biases.
  - Rights-based approach: Individuals can demand correction of discriminatory biases, a powerful redress mechanism for queer healthcare patients harmed by AI.

V. Conclusion

- Overall, the Brazilian model emerges as uniquely well-suited for queer-responsive regulation of healthcare AI because it:

- Integrates Foundational and Sociotechnical Remedies: By requiring diverse teams and actively addressing cognitive biases in data classification, it counters the root causes of queer erasure in AI.

- Mandates Participation and Transparency: Ongoing public consultations and iterative impact assessments ensure that LGBTQ+ voices influence AI governance throughout the lifecycle.

- Secures Individual Rights: The enforceable right to correct discriminatory biases empowers queer patients to seek redress when excluded or misclassified.

REFERENCE: S. Sulmicelli, *Queer-responsive regulation for AI in heathcare: a comparative study*, in UNSW Law Journal, vol. 48 2025, *forthcoming*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5292012

sergio.sulmicelli@unitn.it