Accounting for diversity in AI for medicine

Fosch Villaronga, E.; Drukarch, H.; Khanna, P.; Verhoef, T.; Custers, B.H.M.

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/CLSR

Computer Law & Security Review

Accounting for diversity in AI for medicine

Eduard Fosch-Villaronga[a,*], *Hadassah Drukarch*[a], *Pranav Khanna*[a], *Tessa Verhoef*[b], *Bart Custers*[a]

[a] *eLaw Center for Law and Digital Technologies, Leiden University, the Netherlands*
[b] *Creative Intelligence Lab (CIL) & Leiden Institute of Advanced Computer Science, Leiden University, the Netherlands*

ARTICLE INFO

Keywords:
Artificial intelligence
Medicine
Gender
Bias
Diversity
Inclusion
Discrimination
AI governance

ABSTRACT

In healthcare, gender and sex considerations are crucial because they affect individuals' health and disease differences. Yet, most algorithms deployed in the healthcare context do not consider these aspects and do not account for bias detection. Missing these dimensions in algorithms used in medicine is a huge point of concern, as neglecting these aspects will inevitably produce far from optimal results and generate errors that may lead to misdiagnosis and potential discrimination. This paper explores how current algorithmic-based systems may reinforce gender biases and affect marginalized communities in healthcare-related applications. To do so, we bring together notions and reflections from computer science, queer media studies, and legal insights to better understand the magnitude of failing to consider gender and sex difference in the use of algorithms for medical purposes. Our goal is to illustrate the potential impact that algorithmic bias may have on inadvertent discriminatory, safety, and privacy-related concerns for patients in increasingly automated medicine. This is necessary because by rushing the deployment of AI technologies that do not account for diversity, we risk having an even more unsafe and inadequate healthcare delivery. By promoting the account for privacy, safety, diversity, and inclusion in algorithmic developments with health-related outcomes, we ultimately aim to inform the Artificial Intelligence (AI) global governance landscape and practice on the importance of integrating gender and sex considerations in the development of algorithms to avoid exacerbating existing or new prejudices.

1. Introduction

Officially coined in 1956 (Haenlein and Kaplan, 2019). Artificial Intelligence (AI) knows many definitions which changed as the field experienced many ups and downs. For instance, the European Commission defined AI in 2018 as "*systems that display intelligent behavior by analyzing their environment and taking actions - with some degree of autonomy - to achieve specific goals*" (European Commission, 2018).

AI, which involves - among other things - machine learning and natural language processing, serves exceptionally well in revolutionizing knowledge-intensive sectors such as healthcare (Garbuio and Lin, 2019; Lee and Yoon, 2021). AI has gained in popularity within the healthcare domain, where it has shown to have clear potential for stimulating the develop-

* Corresponding author.
E-mail address: e.fosch.villaronga@law.leidenuniv.nl (E. Fosch-Villaronga).

https://doi.org/10.1016/j.clsr.2022.105735

ment of new medical treatments for a wide variety of diseases and disorders, for improving both the standard and accessibility of care, for enhancing patient health outcomes, and for filling quantitative care gaps, supporting caregivers, and aiding healthcare workers (Fosch-Villaronga and Drukarch, 2022). Different medical domains previously reserved for human experts are increasingly augmented or transformed completely thanks to the integration of AI in clinical practice. This includes disease diagnosis, automated surgery, patient monitoring, translational medical research encompassing advances in drug discovery, drug repurposing, genetic variant annotation, and the automation of specific biomedical research tasks such as data collection, gene function annotation, or literature mining (Yu et al., 2018; Ahuja, 2019). Moreover, AI is well-suited to handle repetitive work processes, managing large amounts of data, and can provide another layer of decision support to mitigate errors, allowing for improvement of patient outcomes while reducing treatment costs (Frost and Sullivan, 2016; Accenture, 2017). More specifically, AI promises to find and use complex underlying relationships between the way humans work and how to care for them to improve care, discover new treatments, and advance scientific hypotheses even if we as humans do not understand those underlying relationships (Price and Nicholson, 2019).

Although these advances may entail incredible progress for medicine and healthcare delivery soon, more research is needed for these systems to perform well in the wild (Gruber, 2019). Room for improvement is in the area of diversity and inclusion. In healthcare, such considerations are crucial because they affect individuals' health and disease differently, as well as their response to treatment (Nielsen et al., 2021). Yet, most algorithms deployed in the healthcare context do not consider these aspects and do not account for bias detection (Cirillo et al., 2020). Missing these dimensions in algorithms used in medicine is a huge point of concern, as neglecting these aspects will inevitably produce far from optimal results and generate errors that may lead to misdiagnosis and potential discrimination (Cirillo et al., 2020).

Questions around the consequences of missing the gender and sex dimensions in algorithms that support decision-making processes are nevertheless particularly poorly understood and often underestimated (Buolamwini and Gebru, 2018; Keyes, 2018), also in the field of medicine (Saddler et al., 2021). While, here, AI is used to predict, address or support a health-related decision, errors may compromise safety and allow for misdiagnosis, a massive problem that, paradoxically, AI is trying to solve. The technical literature focuses on how algorithms can infer user gender from user traits for several purposes (Nieuwenhuis & Wilkens, 2018; Garibo-Orts, 2018; Pasti & Castro, 2016; Fink et al., 2012), often lacking an in-depth reflection of the implications those inferences have on society (Hamidi et al., 2018; Keyes, 2018). For instance, automated gender recognition systems try to identify the gender of a person objectively. However, this clashes with the idea that gender is subjective and internal, often leading to misgendering outcomes that may have ulterior adverse effects for large parts of the population, including the transgender, intersex, and non-binary community (Fosch-Villaronga et al., 2021).

Although different communities focus on AI diversity and inclusion (Stathoulopoulos and Mateos-Garcia, 2019), these investigation efforts are still very much scattered and rarely compared to other research strains that focus on safety or data protection (The EUGenMed et al., 2015; Malgieri and Niklas, 2020). There is also little insight into how all this research applies to specific contexts, such as in the medical field. There are obvious differences between individuals specially in terms of gender, sex, race, and socio-economical background. In medicine, such considerations are crucial because they affect individuals' health and disease and their response to treatment and drug response differently, resulting often in detrimental health outcomes and increased health costs (Franconi et al., 2007; LeBreton, 2013; Weiner et al., 2020). For instance, in the development of COVID-19 trials, sex disparities in genetics, immunological responses, and hormonal mechanisms are relevant, as they underly the substantially higher fatality rates reported in male COVID-19 patients (Schiffer et al., 2020). Not accounting for diversity aspects in today's medicine raises questions about patient representation (Carnevale et al., 2021), discrimination (Rotenstein and Jena, 2018), autonomy, and, most significantly, safety, as it can result in harmful outcomes, including death (Muñoz et al., 2020), to many, but especially to women and marginalized communities such as the transgender community, which has been historically disregarded and discriminated against at best (Bird et al., 2012; Sizemore-Barber, 2020; Barbee et al., 2022). Such consequences stem from the traditional understanding of concepts such as sex and gender, which are usually reduced to a binary opposite outcome - masculine vs. feminine (Nielsen et al., 2021) and often confused and put together even by the main medical community (The EUGenMed et al., 2015).

With an emphasis on the potential for AI/ML bias in medicine (Willson, 2017; Noble, 2018; Ito, 2019), this paper explores the legal and regulatory implications of missing diversity and inclusion considerations in the context of AI for medicine. To do so, we bring together notions and reflections from computer science, queer media studies, and legal insights to understand the magnitude of failing to consider gender and sex differences in the use of algorithms for medical purposes. Our paper touches upon a mutual flaw in legal, computational, and clinical settings: the accounting for gender and sex considerations in medicine. To do so, we give a first, ambitious look at existing technical challenges posed by algorithms in medicine in the context of intersectional justice, i.e., when two or multiple personal characteristics operate simultaneously and interact inextricably, producing distinct and specific forms of discrimination (Council of Europe, 2022). The purpose of our paper is to forward this discussion to clinical data-driven healthcare in light of current theorizations about Responsible Research and Innovation (RRI) and a rapidly growing environment of AI-related policy-making. While we acknowledge that our effort may appear to many as incomplete, we think it is timely and necessary to promote and raise attention to the account for privacy, safety, diversity, and inclusion in algorithmic developments with health-related outcomes to ensure these systems are safe to use. Since AI developments are currently in full swing, there is still time to incorporate these considerations into their design (rather than patch them later, which is usually more costly and less effective). As a result, we ultimately aim to inform the AI global

governance landscape on the importance of integrating gender and sex considerations in the development of algorithms to avoid discrimination and exacerbating existing biases.

This paper is structured as follows. After this introduction, we provide an overview of recent developments in the use of AI for medicine in Section 2. Here, we also highlight some of the applications of AI systems in medicine, we explain how these systems work, and the role that data plays within this context, and we address algorithmic accounts of gender and sex considerations. Section 3 clarifies some of the concepts used in this paper, including sex, gender, and sexuality, and covers some of the sex and gender implications of AI in medicine, thereby specifically focusing on inadvertent discriminatory, safety, and privacy-related concerns for patients in increasingly automated medicine. In Section 4, we propose to address these sex and gender implications of AI in medicine by offering technological solutions for redesigning AI in medicine, stressing the importance of responsible research and innovation, and highlighting the need for specific and sufficiently adequate legal and regulatory frameworks. The paper concludes with some final remarks in Section 5.

2. AI in medicine

In the context of healthcare, AI is poised to play an increasingly prominent role in medicine and healthcare because human biology is tremendously complex, and our tools for understanding it are limited (Price and Nicholson, 2019). Due to advances in computing power, learning algorithms, and the availability of large datasets (big data) sourced from medical records and wearable health monitors, digitized medicine - which has proved to be useful in overcoming this significant shortcoming - has become more readily available in several healthcare areas (Ahuja, 2019; Custers, 2006), and AI in medicine has started to proliferate (Bakkar et al., 2018; Kaul et al., 2020). Thanks to the processing of vast amounts of health data from electronic health records, AI could support predictive models that can be ulteriorly used to diagnose diseases as accurately as experienced healthcare providers. AI could assist pediatricians (Liang et al., 2019), predict therapeutic response, and potentially preventative medicine in the future (Amisha et al., 2019), predict women at high risk of postpartum depression (Zhang et al., 2020), or give triage advice safer than that of human specialists (Razzaki et al., 2018). AI improves diagnostic accuracy, efficiency in provider workflow and clinical operations, facilitates better disease and therapeutic monitoring, and enhances procedure accuracy and overall patient outcomes (Kaul et al., 2020).

Owing to recent advances in medicine, AI has impacted medical approaches towards chronic disease management and clinical decision-making (Bresnick, 2016), and is now increasingly used for risk stratification, genomics, imaging and diagnosis, precision medicine, and drug discovery (Fosch-Villaronga and Drukarch, 2022). Clinical domains in which AI is currently being put to use include radiology (Bakkar et al., 2018; Wang et al., 2017), oncology (Houssami et al., 2017; Patel et al., 2018), pathology (Cruz-Roa et al., 2017; Yu et al., 2016; Wong and Yip, 2018; Capper et al., 2018), dermatology (Haenssle et al., 2018), ophthalmology (Gulshan et al., 2016; Roach, 2017), cardiology (Zhang et al., 2018; Petrone, 2018), gastroenterology (Wang et al., 2018), surgery (Hashimoto et al., 2018), and mental health (Topol, 2019a; 2019b). Bearing this in mind, advances in AI technologies may entail incredible and unprecedented progress for medicine and healthcare delivery, both in terms of quantity and quality, that could eventually help repair diagnostic errors and their very high consequences for society soon (Singh et al., 2014).

Generally speaking, AI in medicine can be divided into two subtypes: virtual and physical (Amisha et al., 2019). The virtual part ranges from electronic health record systems to neural network-based guidance in treatment decisions. In contrast, the physical part deals with robots assisting in performing surgeries, intelligent prostheses for people with physical disabilities, and elderly care. As such, AI-enabled computer applications will help primary care physicians to better identify patients who require extra attention and provide personalized protocols for each individual.

Examples of such technologies are smartwatches that are capable of detecting atrial fibrillation (Buhr, 2017), and smartphone exams with AI are being pursued for a variety of medical diagnostic purposes, including skin lesions and rashes, ear infections, migraine headaches, and retinal diseases, such as diabetic retinopathy and age-related macular degeneration (e.g., AiCure)[1] (Levine and Brown, 2018). Another example in the context of robotic surgery is IBM's Watson which created an intelligent surgical assistant that uses unlimited medical information and natural language processing to clarify surgeons' doubts about surgery performance (Fosch-Villaronga and Drukarch, 2022; Fosch-Villaronga et al. 2021a). Moreover, in terms of social and physical assistance, the application of AI in (robotic) healthcare delivery has been the driver of significant progress. For instance, according to Fosch-Villaronga & Drukarch (2022), "the increased capabilities with respect to advanced data acquisition, processing, and control techniques based on AI enable the construction of robust control strategies that outperform classic approaches in biomechatronic systems, including Physically Assistive Robots (PARs)." In the same vein, the value of AI in the domain of social and medical assistance is also increasingly being acknowledged, with virtual and robotic AI agents not merely being deployed for low-level mental health support (e.g., comfort or social interaction). They are also for high-level therapeutic interventions with sensitive patient groups (e.g., people with dementia or children who have ASD, Autism Spectrum Disorder) that previously required interventions by highly trained, skilled health professionals (Inkster et al., 2018). Finally, beyond anticipating major outcomes (Topol, 2019a), AI applications are deployed in medical administration to automate non-patient care activities and undertake repetitive routine tasks, such as patient data entry and automated review of laboratory data and imaging results, writing chart notes, prescribing medications, ordering tests, and assist hospitals in predicting the duration of patient stays at the pre-admission stage, thereby lessening the burden on clinicians (Snyder et al., 2011), allowing healthcare providers to cut documentation time, improve reporting quality and free time for clinicians to

[1] *See* https://aicure.com/.

provide direct care (Ahuja, 2019), and enabling hospitals to use their stretched resources more efficiently and appropriately (Topol, 2019a; Fosch-Villaronga and Drukarch, 2022).

Despite all the promises of AI technology, it has shown formidable obstacles and pitfalls in its adoption and implementation in the healthcare setting, especially when it pertains to validation and readiness for implementation in patient care (Topol, 2019b). A recent example of this is IBM Watson Health's cancer AI algorithm. When fed with very limited input (actual data) from clinicians, the potential for significant harm to patients and medical malpractice by a flawed algorithm arises. At the same time, large amounts of uncurated data may be embedded with certain hegemonic patterns around gender, race, ethnicity, and disability status which can have adverse effects on communities at the margins (Bender et al., 2021). This highlights already existing concerns about the dangers resulting from so-called 'black-box algorithms' and stresses the need for systematic debugging, audit, extensive simulation, and validation, along with prospective scrutiny before the relevant AI algorithm is unleashed in clinical practice (Topol, 2019b). This opaqueness has led to an increased demand for transparency and explainability in AI environments (Felzmann et al., 2020) (e.g., see the explicit requirements for transparency laid down in the European Union's General Data Protection Regulation, GDPR) before an algorithm can be used for patient care in practice. Nevertheless, caution has been made in relation to calls for ML systems to become more explainable and transparent, as the results of such explainability and transparency may have perverse effects (Smith, 2019). This is because, as noted by Smith (2019), efforts towards developing "self-explaining" or "interpretable" neural networks, may unintentionally decrease their performance, and drive them toward unwarranted reliance on binary or discrete categories and towards the implicit or explicit reliance on formal ontology and its inadequacies when applied in practice.

3. Failing to account for diversity in AI for medicine

3.1. *A definitional framing of sex and gender*

There are multiplicities of understanding, accepting, and legalizing the intricate relations between sex, gender, and sexuality (Hooper, 2001; Haas and Hwang, 2007; Randall and Waylen, 2012; Klein, 2013). This is also true for how gender is understood and utilized in and through algorithms (Fosch-Villaronga et al., 2021) and by the law, which has progressively evolved in integrating such dimensions in laws against discrimination or data protection. Designers of AI cannot easily ignore existing gender norms because they are so deeply embedded in how we navigate the world that they even translate into the design of algorithms and robots with different embodiments (O'Neil, 2016; Nomura, 2017). At the technical level, algorithms usually work in binary terms (e.g., 'yes/no,' 'black/white,' 'moves/does not move,' 'man/woman'), as if the world were a simple classification problem to be solved. However, the world is not black and white, and there are many intricacies between the various concepts we refer to in order to better understand our surroundings.

Definitions play a crucial role in creating more clarity and avoiding misunderstandings when discussing a particular subject. Nevertheless, not all concepts are easy to describe, particularly in those fields intersecting law and new technologies (Fosch-Villaronga and Drukarch, 2022). Here, the use and meaning of words differ entirely in different contexts and according to the communities by which they are used. An example of this can be found in the use of the term *transparency* within the legal and computer science domains. While the legal domain defines this term as "easy to perceive or detect," within the context of computing, it is defined as 'of a process or interface functioning without the user being aware of its presence.' This indicates that while both domains make extensive reference to this term, they understand and apply it differently (Felzmann et al., 2019). The same applies to using terminologies such as "sex, gender, and sexuality" and "male, female, intersex." There is considerable disagreement in defining these terminologies between members of different research traditions. This inevitably leads to confusion when terms used by one community seem to be juxtaposed to the understanding that another community has attributed to the very same concept. Although it goes beyond the scope of this paper to provide an in-depth definitional framework for the concepts of gender and sex, it bears important to clarify some of these concepts to understand better how they influence and impact the development of new technologies and how the law should frame them and account for them when regulating the highly complex landscape involving new technologies (Deaux, 1985; Pryzgoda and Chrisler, 2000; Shotwell and Sangrey, 2009; Dembroff, 2019; Fiane and Serpe, 2020; Lips, 2020):

- *Sex* tends to be associated with the assigned gender at birth based on medical factors such as genitalia, chromosomes, and hormones. In short, it is *anatomical sex*. Sex can be male, female, or intersex, and it is changeable via medical gender transition. It is common to see the initials AAB or 'assigned at birth' accompanying this term.
- Also called *gender identity*, *gender* is a person's subjective experience of their gender and links to social, cultural, and legal factors. The current understanding of *gender* includes cisgender (for those whose anatomical sex and gender align, thus male, female, or intersex), transgender (for those whose gender does not align to their sex assigned at birth), gender neutral, non-binary (if it does not identify exclusively as male or female), agender, pan/omnigender, genderqueer/third gender, gender-fluid (if it varies over time), two-spirit, gender non-conforming/expansive (for those free to not fit into a specific societal norm), gender-void (for those not feeling their gender).
- *Sexuality* means the 'physical, romantic, and emotional attraction to another person.' In the law, this is often called *sexual orientation* and it is considered a special category of data within the EU General Data Protection Regulation (GDPR).

While the scientific community broadly supports the narrative that integrating gender and sex factors in research

makes better science (Schiebinger, 2014; Tannenbaum et al., 2019), many disciplines struggle to account for diversity. For instance, algorithmic systems usually take *sex* as a primary reference point and usually focus on *male* and *female* categories, mainly disregarding *intersex* people. The belief that gender is rooted in physiological terms harms transgender people by essentializing the body as the source of gender and harms the non-binary community, who cannot be accurately classified (Keyes, 2018; Fergus, 2020). In the medical context, Barbee et al. (2022) warn that this "could exacerbate existing health disparities, facilitate risky health behaviors, and lead to preventable deaths."

The accuracy of algorithmic decision-making lies in the early stages of development and training. The importance of accounting for diversity from the outset should not be underestimated. Specifically, deep learning systems are trained to recognize patterns during the development stage by being subjected to many data sets or training data such as pictures, object characteristics, or situations that humans may have already labeled or classified. The systems learn from these pre-classified data sets to recognize and classify objects and examples that the system may have never encountered before (Pew Research Center, 2019). However, algorithms fail to integrate gender considerations, primarily because algorithms perform poorly in recognizing objectively internal and subjective aspects tied to social and cultural factors (Fosch-Villaronga et al., 2021). Instead, *gender-sensitive* research only accounts for differences between men and women (Decataldo and Ruspini, 2016) and algorithms can misclassify users, which may lead to several consequences depending on the context of the application. For social media, misgendering users can cause feelings of rejection, which can ulteriorly impact one's self-esteem, confidence, and authenticity and increase social stigmatization (Hamidi et al., 2018; Keyes, 2018). If failing to account for sex and gender considerations in algorithmic systems is a point of concern in social media practices, failing to do so in remarkably sensitive domains of application like healthcare where these considerations are essential in determining patient safety and healthcare outcomes is appalling. Despite clear evidence to the contrary, science holds onto the promise that these systems will help deliver safer care (Yu et al., 2018; Ahuja, 2019).

The binary understanding of sex has traditionally been considered the point of departure for many legal provisions. Take as an example *gender stereotyping*. Gender stereotyping "refers to the practice of ascribing to an individual 'woman' or 'man' specific attributes, characteristics, or roles by reason only of their membership in the social group of 'women or men'" (UN, 2022). However, gender is not limited to the simple binary classification of being solely a "man" or a "woman." It is a social construction that encompasses many typologies and experienced inner understandings of what is a person's gender identity. In this sense, gender stereotyping is a complex process grounded in solid beliefs of what gender should be and is often used and comprehended too simplistically (Kachel et al., 2016). For instance, lesbian women are frequently categorized as 'butchers' or 'truck drivers' and put together with traditional men stereotypes. Gay men can also be hyper-sexualized (the masculine promiscuity stereotype) or feminized if they are perceived as feminine and fall into traditional female stereotypes. While the contemporary understanding of sex and gender reveals an increasing sensitivity towards the topic from different streams of knowledge, information that defines the true self of a person is not recognized as a sensitive data under the GDPR, even if scholarship continues to highlight its sensitivity (Wachter and Mittelstadt, 2019).

3.2. *Sex and gender considerations in precision medicine*

Precision medicine implies a deep understanding of inter-individual differences in health and disease inherent to genetic and environmental factors, there is a growing need to implement different types of technologies based on AI (Cirillo et al., 2020). In such a context, generating fair and unbiased classifiers becomes of paramount importance (Larrazabal et al., 2020), mainly because puzzling variables such as stigma, stereotypes, and data misrepresentation, health research, practices, and robot and algorithmic design are inevitably tangled with sex and gender inequalities and biases (Søraa, 2017). Despite the significant scientific advances achieved so far, most of the currently used AI technologies in medicine do not account for sex and gender considerations, meaning they do not consider health and disease differences among different individuals (Cirillo et al., 2020).

This understanding has especially gained ground in the context of rising inequities and bias in healthcare today, which does not provide adequate care for all, explicitly excluding minority groups in society like the transgender and the intersex communities (Barbee et al., 2022) . Intertwined with this concern of exacerbating pre-existing inequities, including gender inequalities, is embedded bias present in many algorithms due to the lack of inclusion of minorities in datasets (Topol, 2019b). For example, AI used in dermatology to diagnose melanoma lacks the inclusion of skin color (Esteva et al., 2017), and the use of the corpus of genomic data, which so far has seriously underrepresented minorities (Wapner, 2018). Furthermore, there is a multitude of sex-based differences in the prevelance of certain skin diseases and autoimmune conditions that AI applications need to take into account. For example, in females, melanomas are more likely to occur on the hip and lower extremities compared to males (Olsen et al., 2020). These sex-based differences have varying impacts ranging from different symptoms in males and females to varying 'Time To Diagnosis' and are sometimes crucial to the outcome of the treatment provided (Sun et. al, 2020). These findings indicate that much work is still needed in the area of diversity in AI for medicine to eradicate embedded prejudice in AI and strive for medical research that provides a true representative cross-section of the population (Topol, 2019b).

Just like in dermatological settings, significant differences between women and men exist in several other human diseases. These include diabetes, cardiovascular disorders, neurological conditions, mental health disorders, cancer, autoimmunity, as well as physiological processes such as brain aging and sensitivity to pain (Wagner et al., 2019; Cirillo et al., 2020). Also, research highlights the robust sex and gender influences that exist across leading causes of death and morbidity globally (Mauvais-Jarvis et al., 2020). These disparities are noted in epidemiology, pathophysiology, clinical manifestations, disease progression, and response to treatment.

For example, studies have revealed that female athletes are more susceptible to injuries of the anterior cruciate ligaments than men, owing to the difference in pelvis position (Ireland ML, 2002). If this naturally existent difference is not considered and recognized by an AI system, it can adversely affect underrepresented population groups such as intersex and transgender (Tomasev et al., 2021). For example, according to the Irish Heart Foundation, heart attacks and strokes, which happen to be a significant health risk for women, are often missed as the primary symptoms experienced by women happen to be different from men (Rosamond et al., 2008). The most prominent symptoms in women are nausea and back pain, while for men, it could be crushing pain in the chest that extends down the arm (Shannon, 2018). Such crucial differences can be vital when it comes to critical conditions and directly impact patient safety.

Although AI shows great promise in imaging and diagnostics that could support ulterior decisions in cardiovascular medicine, a distinct challenge in this context is the significant heterogeneity in diagnostic studies (Tat et al., 2020). The investigations often include limited sample sizes not validated by others outside of the research or contrasted with other populations including women, men, or intersex. Some of these techniques also still require human interpretation and do not follow a standardized protocol, which varies by the institution and machine vendor (Tat et al., 2020). Furthermore, adequate sex or gender consideration evaluating drug safety disparities and efficacy is mainly absent from clinical trials, although it should be present in AI for drug discovery (Mauvais-Jarvis et al., 2020).

The sociocultural dimension of gender also plays a significant role in influencing the awareness of a particular disease, the attitudes towards it, the manifestation of disease symptoms, or the interpretation of signs and symptoms of the disease (Regitz-Zagrosek, 2017). This sociocultural dimension also affects other essential aspects such as access to healthcare, the doctors' attitudes toward patients, or even pain communication. Differences in lifestyle factors that are associated with sex and gender (e.g., diet, perceived stress, smoking, and physical activity, and affect health and disease susceptibility) influence the behavior of communities, clinicians, and patients. For instance, gender roles represent the behavioral norms applied to men and women in society, thereby influencing individuals' everyday actions, expectations, and experiences (Mauvais-Jarvis et al., 2020; Cirillo et al., 2020). In this sense, Obermeyer et al. (2019) found that black patients with the same risk level determined by an algorithm were sicker than white patients. The algorithm used health costs as a proxy for health needs, resulting in an appalling discriminatory result: "less money is spent on black patients who have the same level of need, and the algorithm thus falsely concludes that black patients are healthier than equally sick white patients" (Obermeyer et al., 2019).

Although these differences are evident to the inner biological workings of each gender, a complex intertwinement between biological and social-economic factors affects and determines sex and gender differences in health and well-being (Cirillo et al., 2020). In this sense, part of the community would prefer to separate the effects of biology, sex, gender, and the disease's sociocultural mechanisms. However, this is not always possible in medicine since most environmental stresses leave traces in epigenetic modifications. In other words, the environment, including nutrition, stress, and behavior, dramatically impacts our bodies' biology. Consequently, establishing a clear distinction between the effects of sex and gender is nearly impossible, forcing medicine to cover all the different dimensions of gender because the distribution of gender-related attributes within populations of men and women can affect health differently from biological sex (Mauvais-Jarvis et al., 2020). Programming all of these aspects in algorithmic systems could help overcome the complexity in the processing of the vast amount of information these may generate; although the possibility for errors and false positives may also increase dramatically (Bhavnani and Harzand, 2018). This is particularly salient in the context of AI for medicine, since studies reveal that marginalization and social exclusion are major factors causing avoidance and underutilization of healthcare by minorities, as a consequence of which they end up being less healthy than the general population (Vermeir et al., 2018). In a way, the tendency of AI systems to learn from biased models, which reproduce social stereotypes and underperform in minority groups, may be especially dangerous in the context of healthcare (Larrazabal et al., 2020).

3.3. *Inaccuracies and biases in the training data*

As the accuracy of the training data determines the quality of algorithmic decision-making, the training data must be representative of the real world. A data bias may result in a skewed decision from the system, resulting in a decision difficult to anticipate or understand (O'Neil, 2016). For algorithms that map individual patients' multiple characteristics and medical conditions to make diagnosis and treatment recommendations, such inaccuracy may result in adverse events that may harm patients at the very least.

Two illustrative examples in this context are language technologies and imaging technologies. Natural Language Processing (NLP) is a rapidly emerging AI application that will likely play an increasing role in medicine to get an online consultation or a pre-screen. Developers use massive corpora of human-produced texts to train NLP models to help AI algorithms understand human language. To make sure these algorithms do not only learn about the structure of language but also about the meaning of sentences, networks of related words are created based on co-occurrence statistics, such as GloVe (Pennington et al., 2014) and Word2Vec (Mikolov et al., 2013). The word embeddings encoded in such machine representations reliably reproduce meaningful associations that make sense to most humans, for instance, that flowers are more pleasant than insects (Caliskan et al., 2017).

Moreover, Caliskan et al. (2017) found a well-known bias in machine word associations: European American names are perceived as more pleasant significantly more often than African American names. Besides, female words were more often associated with family terms, while male words became more often linked to career terms. In terms of imaging technologies, while the research community of medical image computing is making significant efforts in developing more accurate algorithms to assist medical doctors in the difficult task of disease diagnosis, little attention is paid to the types

of collected data, the way they are collected in databases, and how this may influence the performance of AI systems (Larrazabal et al., 2020). For instance, empirical studies on the performance of AI trained on skewed data sets have confirmed that when female populations are under-represented in the training stage of the AI, gender gaps are created in the way the algorithm performs, making them a minority in the field of healthcare (Larrazabal et al., 2020). Specifically, the lack of a representative sample in the data set used to train the algorithm may result in unreasoned or irrelevant selectivity. Moreover, human bias, such as gender and racial bias, may also be inherited and amplified by AI systems in multiple contexts (Caliskan et al., 2017). If text- and image-based medical AI applications are fueled with machine semantics, these problems risk perpetuating existing cultural stereotypes and may exacerbate existing biases if not appropriately addressed.

3.4. *Privacy and data protection considerations in the data cycle*

Apart from discriminatory outcomes, the use of large amounts of personal data usually triggers questions regarding the (data) privacy of the people to whom the data relate. Although for most people it is intuitively obvious that processing large amounts of health data may raise privacy and confidentiality issues, it may sometimes be hard to articulate precisely what the privacy concerns are. There are issues regarding informed consent, professional secrecy, and data security that are well-known and well-documented (though not always solved). However, on top of these well-known privacy issues, in the context of AI some of these issues may be exacerbated or put in a new perspective. Here we list four typical privacy issues, but we stress that this is not a complete list, as researchers are still exploring the full privacy exacerbations of the use of AI (Manheim and Kaplan, 2019; Price and Cohen, 2019; Fosch-Villaronga et al., 2020).

The first issue is that of predictions and inferred data. The aggregated, combined and analyzed data provides novel insights and added value (Custers and Bachlechner, 2018). Typical examples here are epidemiological data and DNA research, which are often analyzed at the level of aggregated data. Also research on very rare diseases, with very limited prevalence and incidence, suddenly becomes possible when using big data and AI. This allows for building epidemiological profiles, risk profiles for people attracting certain diseases, and assessments of which therapies and treatments are effective for particular diseases. After the large amounts of data are collected, they are analyzed, usually in automated ways, using tools like data mining and machine learning (Kamiran et al., 2013). Typical issues with these tools are that profiling based on datasets from various sources that contain large amounts of inferred data may propagate any existing biased patterns, leading to disparate impact (Barocas and Selbst, 2016). Moreover, reusing inferred data as input for data analytics, particularly profiling processes, may turn profiling processes into amplifiers with positive (i.e., self-reinforcing) feedback loops (Custers, 2018).

Effects of minor disturbances (like incorrect or incomplete data or flaws in the data analysis) may increase the magnitude of perturbations. Such disturbances may occur in various forms, among which certain almost undetectable perturbations as a result of which severe artifacts in the reconstruction may materialize; small structural changes which may not be captured; and the paradoxical phenomenon whereby an increased amount of samples may yield poorer performance (Antun et al., 2020). This may lead to the identification of false positives and negatives which may have serious consequences for the subjects if algorithmic systems entail health-related outcomes, for instance in the case of breast cancer detection (Pisano, 2020). Algorithmic systems deployed for medical purposes can predict highly sensitive factors, such as ethnicity, sexual orientation, use of illegal substances, and risks to attract specific diseases (Kosinski et al., 2012). The predictions may even concern information that persons did not know about themselves, such as life expectancy or risks to attract certain forms of cancer). Some people may not even want to know specific information about themselves (such as personalized cancer survival rates or genetic diseases that may impact close relatives). Bearing this in mind, the performance of AI algorithms is highly dependent on the population used in the training sets, and it is therefore essential that a representative sample of the general population be used in the development of such technology to ensure that the results are broadly applicable (Pisano, 2020).

A second issue is that informed consent to privacy interferences needed for processing personal data in medicine is much more complicated when AI is used. This makes it harder for people to manage their privacy and control their data. The essence of AI technologies is that they are autonomous and self-learning, which makes it much harder (and sometimes impossible) to explain the technology's inner workings to people when asking for (informed) consent. Also, informing people about the consequences can be complicated, as it may not always be predicted what outcomes of processing large amounts of data by AI may have. For instance, profiling for relatively innocuous diseases (to which people may easily consent) may suddenly reveal risks for serious diseases (to which people may not so easily consent). Thus, while the added value of many AI-related applications is that they yield novel, unexpected results, which can be very beneficial, they can also entail a lack of transparency (Felzmann et al., 2019) which prevents practitioners from seeking patients' consent easily for the use of AI applications in medical environments.

A third issue is that anonymity, one of the traditional ways of privacy protection, is rapidly becoming ineffective in the context of big data and AI; in particular, in precision medicine. As mentioned above, it is increasingly easy to predict missing attributes. This can also be applied to identifying characteristics. In other words, in the case of anonymized data, it is not very hard to indicate to whom the data is related, effectively de-anonymizing it (Ohm, 2009; Brasher, 2018). For instance, on the basis of trivial data such as postal codes, it may be possible to predict all kinds of sensitive health characteristics. As a result, anonymization can provide a false sense of privacy protection (Jensen, 2013). Even worse, anonymization may diminish any existing legal privacy protections: the EU General Data Protection Regulation (GDPR) only applies to personal data. Anonymized data are explicitly out of scope, which means this solid legal instrument does not protect them for privacy.

A fourth issue is that it is hard for individuals to address any privacy violations after they took place. If personal data is

being processed that a person did not consent to or no longer consents to (someone may change his or her mind and withdraw consent), legal provisions allow for addressing this. For instance, a person can lodge a complaint at the data controller (for instance, the hospital) or supervisory authorities. In case of severe violations of privacy and data protection law, supervisory authorities can impose significant fines upon data controllers under the EU GDPR, up to 10 or 20 million euros or 2% or 4% of the data controller's annual worldwide turnover (whichever is higher). However, in practice, awareness among data subjects on which data is being processed and its consequences is limited. Furthermore, many people are not aware of their rights under EU data protection law (Eurobarometer 2011; Custers et al. 2014; Soumelidou and Tsohou, 2021).

4. Addressing sex and gender implications of AI in medicine

In medicine, AI is what Cirillo et al. (2020) described as a confounding "double-edged sword" because it may exacerbate and perpetuate existing biases for sex and gender, but could play a significant role in mitigating these inequalities. However, this may have catastrophic consequences for patient safety, privacy, and discrimination. So, on the one hand, discrimination in AI for medicine is desired, i.e., we may want AI to account for sex and gender differences between individuals because it may lead to improved performance and precision medicine. On the other hand, avoiding unwanted discrimination and preserving privacy should be an essential proactive part of these advancements. Because AI applications are the outcomes of political, scientific, and technical interactions, there is a pressing need to address these to mitigate risks in health-related outcomes best. Below, we discuss several ways in which sex and gender implications of AI in medicine could be addressed.

4.1. *Legal frameworks should account for diversity*

While the technical, scientific, and medical communities are often criticized for the tardiness in accounting for diversity (Calleja et al., 2022a), the global landscape of AI ethics guidelines and legal frameworks do not seem to provide adequate guidance either in addressing the potential implications of missing gender and inclusivity considerations in AI development in medicine (Dillon and Collett, 2019). For instance, ISO 13482:2014, the leading standard for robots used in personal care, does not consider any special safety requirements for users with different sex, gender, sizes, shapes, and medical conditions, although stating such a need in the introduction (Fosch-Villaronga, 2016; Calleja et al., 2022b). In this respect, it could be that a personal care robot is certified under the standard with disregard for its safety or accessibility. Understanding the impact of such a miss is easier thanks to the physical embodiment of such technologies, the challenge will be to understand 'the impact of AI tools on gender issues (…) an area in which global guidance is currently lacking' (Schwalbe and Wahl, 2020) and that it comes with its own problems such as opacity and apparent neutrality (Selbst, and Barocas, 2018).

In a recent policy review, Jobin et al. (2019) identified the main ethical principles in AI guidelines globally: transparency, justice, and fairness, non-maleficence, responsibility, privacy, beneficence and autonomy, trust, sustainability, freedom, dignity, and solidarity (see also La Fors et al. 2019). They acknowledge the importance of diversity as a relevant factor in realizing justice, fairness, and equity. However, current legal frameworks and healthcare policies usually overfocus on physical safety, neglecting other essential aspects like security, privacy, psychological aspects, and diversity, which play a crucial role in robot safety (Martinetti et al., 2021). As a result, developers struggle to implement them in their algorithms and fail to provide an adequate level of safety, especially in healthcare applications (Gruber, 2019).

Moreover, sex and gender have not traditionally been considered sensitive personal characteristics in related frameworks, such as the GDPR (Fosch-Villaronga et al., 2021), where no specific mention of such personal attributes is given these aspects are not given that much importance as a safety parameter. Bearing this in mind, part of the community claims that sex and gender considerations should be incorporated in international guidelines. According to Mauvais-Jarvis et al. (2020), one of such corpora that could revisit the absence of sex and gender considerations in evaluating drug safety disparities and efficacy is the International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use.[2] In recent years, efforts have been made towards better accounting for the risks posed by AI systems. In the following, we highlight two of the most recent efforts.

In April 2021, the European institutions released a proposal for a regulation laying down harmonized rules on artificial intelligence (AI Act, 2021). Before, there was an absence of specific AI or robot regulation in which clear procedures, boundaries, and requirements for AI developers are explained, challenging how they can integrate these considerations into their design to make them safe (Holder et al., 2016; Fosch-Villaronga, 2019). The AI Act (2021) establishes as 'high-risk' those 'AI systems that pose significant risks to the health and safety of fundamental rights of persons' (p. 3). As such, one would think that algorithmic systems that generate health-related outcomes will generally be considered high-risk. However, while the AI Act (2021) in Annex III lists high-risk applications, they do not include any application considering healthcare or medicine.[3] If that was the case, not being categorized as 'high-risk' means that the requirements that would typically apply to high-risk systems do not apply to such applications. These requirements refer to the high quality data, documentation and traceability, transparency, human oversight, accuracy and robustness, which are strictly necessary

[2] *See* https://www.ich.org/.

[3] The high-risk categories, according to annex III AI Act (2021) are: (1) Biometric identification and categorisation of natural persons, (2) Management and operation of critical infrastructure, (3) Education and vocational training, (4) Employment, workers management and access to self-employment (5) Access to and enjoyment of essential private services and public services and benefits, (6) Law enforcement, (7) Migration, asylum and border control management, (8) Administration of justice and democratic processes. *See* https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=75789.

to mitigate the risks to fundamental rights and safety posed by AI (AI Act, 2021). Moreover, it also means that non high-risk AI systems will not have to comply with a set of horizontal mandatory requirements for trustworthy AI that the High-level Expert Group on AI established in 2019. In these guidelines, they proposed a 'Trustworthy AI assessment list' aimed at operationalizing the critical requirements of (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination, and fairness, (6) environmental and societal well-being, and (7) accountability (HLEG AI, 2019). Especially in the context of AI and diversity, the guidelines refer to unfair bias avoidance, accessibility, universal design, and stakeholder participation. The guidelines also mention that factors linked to one's identity, such as sex and gender considerations, determine whether a person is vulnerable or is part of a vulnerable group. They also remind us of the Art. 21 of the Charter of Fundamental Rights of the EU, on non-discrimination on the basis of sex, race, color, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age, and sexual orientation. This should apply to any AI development, including for those algorithmic systems used in medicine. Although these guidelines do not provide concrete answers on how to address developers' questions, they stimulate reflection on how they can put into practice these trustworthy AI requirements.

Said this, a closer look at the AI Act (in particular Article 6 AI Act[4] and Annex II 11 AI Act) reveals that all AI medical devices that must undergo a conformity assessment by a Notified Body are considered high-risk within the AI Act. However, this classification overlooks that AI-powered medical devices are usually software and are primarily subject to classification IIa or higher (Annex VIII, Chapter III, Rule 11 of the MDR), making nearly all medical software "high-risk AI systems" within the purpose of the AI Act (Tietjen and Woedtke, 2021).

This disconnect reveals that there is an underlying question about the communication, compatibility, and overlaps between all these frameworks and the compatibility with existing medicine frameworks that future work should address (Amann et al., 2020; Vollmer et al., 2020). This issue is pressing especially in providing legal certainty about the compatibilities between the new proposed AI Act and the new Medical Device Regulation, which put forward different safety, performance, and quality requirements; different definitions of user severe incidence that are not fit for purpose AI definitions; different risk classification and risk management requirements; different and limited number of notified bodies; and similar incident reporting (Beckers et al., 2021; MedTech Europe, 2022; Niemiec, 2022).

4.2. *Supporting more diverse research via responsible innovation tools*

The proposed EU regulation on AI ('the AI Act') is designedly technology-neutral,[5] laying down essential requirements to be complied with, without designating any specific technical solution to comply with those provisions. In this sense, although moving a step forward in framing the development of AI technologies, developers struggle to translate these provisions into concrete, practical, and widely adopted actions for informing their creations. However, given the ulterior consequences of not integrating diversity considerations in medicine, developers should make serious efforts towards understanding how to mitigate these throughout the entire research process, including the data-cycle. The goal is to anticipate any undesired outcome of the subsequent research.

One of the tools that the EU has established for some time now is the Responsible Research and Innovation (RRI). According to Schomberg (2013), this is "a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view on the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)." RRI focuses on identifying ways in which innovations' societal impact can be proactively addressed and constructively shaped by the collaboration of different stakeholders in the innovation process to prevent avoidable harms and create benefits. The underlying assumption for RRI is that for new technologies to become accepted by society and integrated into societal practices, they have to be aligned with societal needs and values. The focus in RRI is on achieving stakeholder input, especially by bringing researchers, technology developers, organizations, and societal representatives together. Moreover, incentivising RRI could help businesses realize opportunities while also leaving positive economic, societal and environmental impacts, thereby substantially benefiting both businesses as well as the society (Gurzawska et al., 2017).

In this respect, the RRI approach provides a suitable framework that includes the principles of inclusion, anticipation, reflection, and responsiveness to guide all the social actors involved in research and innovation (R&I) processes (Stilgoe et al., 2013; Stahl, and Coeckelbergh, 2016; Aymerich-Franch and Fosch-Villaronga, 2020):

- **Inclusion** refers to conducting research involving a wide range of stakeholders from the early stages of the R&I process. In practice, this could translate into more diverse R&I teams, instead of only 'white males'. This would include having more diverse data and test cases, instead of testing only adult males, also, include data on children, preg-

[4] *Art 6 AI Act: Irrespective of whether an AI system is placed on the market or put into service independently from the products referred to in points (a) and (b), that AI system shall be considered high-risk where both of the following conditions are fulfilled: (a) the AI system is intended to be used as a safety component of a product, or is itself a product, covered by the Union harmonisation legislation listed in Annex II; (b)the product whose safety component is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment with a view to the placing on the market or putting into service of that product pursuant to the Union harmonisation legislation listed in Annex II. 2. In addition to the high-risk AI systems referred to in paragraph 1, AI systems referred to in Annex III shall also be considered high-risk.*

[5] This is a term used within the context of legislative technique to stress that "legislation should abstract away from concrete technologies to the extent that it is sufficiently sustainable and at the same provides sufficient legal certainty" (Koops, 2006).

nant women, etc., in a secure and privacy-preserving way, of course.

- **Anticipation** encourages R&I social actors to ask "what if" questions to help them anticipate any adverse consequence and devise contingency plans accordingly. In this particular question, researchers should focus on foreseeable scenarios in which these algorithms could work. Working in simulators or testing zones that simulate real-world scenarios would help.
- **Reflection** encourages researchers to think about their work mindfully to identify prevailing assumptions to identify biases and frame issues and problems constructively. Multidisciplinary R&I teams could create room for more inclusive intersectional reflections (Søraa and Fosch-Villaronga, 2020).
- **Responsiveness** refers to the possibility to reshape R&I processes in response to events that no longer align with the continually evolving needs of society. This would mean to react upon the current understanding of certain concepts, including gender.
- **Transparency** encourages open-access dissemination of the results and conclusions to enable public scrutiny and dialogue.

These principles within the RRI framework should not be understood as a simple static framework but rather as a living exercise that demands revision and adaptation to particular cases and throughout the whole life-cycle of the system.

In this respect, a number of initiatives are being developed and implemented around the globe. The Government of Catalonia (2022), in Spain, has recently published a tool geared towards incorporating the perspective of sex and gender in basic science, clinical, health services, and public health studies research content. The tool considers the different phases of the research process in which researchers should incorporate sex or gender considerations, including during the problem identification, study design, analysis, and results and translation of knowledge. In it, the researchers can find questions addressing sex and gender integration in research and some examples to help researchers understand how to integrate the aspects into their research. The Canadian Institute of Health Research have developed and implemented an online tool that provides training to researchers for integrating sex and gender analysis into biomedical research. Another example is the global collaborative project *Gendered Innovations* (2009) between Stanford University, the European Commission, and the US National Science Foundation that has inspired recent EU practices towards gender and innovation (European Commission, 2020). Under this initiative, practical methods and tools have been developed for researchers to understand the implications and impact of sex and gender considerations in scientific discovery and innovation. Such tools could help realize the goals of a more responsible, inclusive, and diverse research and support the goals for intersectional justice.

4.3. *Technical account for diversity in AI for medicine*

According to Carr (2020) developers "are usually so intent on solving a particular problem or untangling some thorny scientific or engineering dilemma that they don't see the broader implications of their work." He continues by explaining that in addition to this, "the users of the technology are also usually oblivious to its ethics". Rather, he holds, they "(…), too, are concerned with the practical benefits they gain from employing the tool". As a result, it is not uncommon to see research and innovation failing to reflect on the consequences and missing essential considerations in their research, such as sex and gender may entail for society. In this respect, approximating diversity and inclusion in algorithmic systems can generate recommendations that are informed and more attuned to the social context in which they occur (Mitchell et al., 2020).

Accounting for intersectionality in data labeling is extremely difficult and it can only be achieved through both technical, organizational, and legal manners. As a starting point, discrimination-aware algorithms which do not yield discriminating patterns, such as gender-based patterns or profiles (Kamiran et al., 2013) should form the point of departure in developing algorithms for medical purposes. For instance, the algorithms can be designed so that they look at certain features from an intersectional point of view, like gender as a non-binary characteristic and by representing individual differences through more fine-grained dimensions. Also, sensitive information relating to, for instance, gender, sex, or race should only be used in specific and regulated applications, where it is proven they matter (Fosch-Villaronga et al., 2021). As far as possible, gender-neutral biomarkers could also be used by AI for decision-making. This would be more in line with other principles that aim at minimizing the amount of collected data. Alternatively, algorithms can be designed so that they are discrimination-aware (Kamiran et al., 2013) or privacy-preserving (Lindell and Pinkas, 2002), also in the context of medicine (Cirillo et al., 2020). In this way, biases can be eliminated from the data used to train the AI by ensuring there is an equal representation of examples, and diversity can be better accounted for. This requires the development of datasets that are carefully curated and documented (Bender et al., 2021), which will prevent the encoding of existing biases and also benefit our understanding of what the algorithm is being trained on and how machine recommendations should be interpreted.

5. Conclusion

AI can be used to predict, address or support a health-related decision. However, AI in medicine can be a "double-edged sword". On the one hand, it can play a significant role in mitigating existing inequalities by increasing the quantity and quality of healthcare provision (Fosch-Villaronga and Drukarch, 2022), but on the other hand it can also exacerbate and perpetuate existing biases for sex and gender. The more personalized medicine will be, the more personal information will be required, pushing for individual data and collective knowledge to allow for real-time accurate decision-making processes. While the literature has started to highlight the importance of integrating diversity considerations in medicine (Cirillo et al., 2020; Mauvais-Jarvis et al., 2020), medical AI policy ecosystems are oblivious to the vast landscape of gender identity understanding. This ignorance may have potentially

been for the traditional heteronormative configuration of the medical world (usually configured by male doctors and female nurses), which most certainly has played a role in how little the awareness concerning the implications of missing intersectional aspects in medicine in general, had for society. In this respect, there is still much to explore about the implications of missing these considerations in algorithmic developments with health-related outcomes, which we elaborated on in this paper.

Given the risks of AI when it comes to reproducing and exacerbating existing biases, there is a need for developing gender-sensitive AI to deliver good care. This new diverse approach to AI should be effective and responsive to the varying needs of individuals with different genders and offer adequate protection by accounting for sex and gender differences and countering potential undesirable bias. Failure in accounting for these differences will cause several issues. Firstly, it will generate sub-optimal results and produce mistakes and discriminatory outcomes that can have disastrous consequences in healthcare (Cirillo et al., 2020). Secondly, it would adversely impact the quality of healthcare delivered to different groups, including the transgender and the intersex communities (Barbee et al., 2022) . Thirdly, it induces non-transparent and selective decision-making standards in precision medicine which can very much harm patients. In other words, the harms of misgendering may vary from patient-specific physical harms, including death, to broader impacts such as reinforcing gender stereotypes, accentuating gender binarism, undermining autonomy, and leading to toxic cultures and algorithmic bias (Keyes, 2018; Fosch-Villaronga et al., 2021).

Organizations often consider resource efficiency and increased productivity as key parameters in the development of algorithms. However, focusing on these aspects overlooks that automated data processing supports ulterior decision-making processes that can impact individuals and society in many different ways, e.g., invading individual privacy, raising inequality, or harming specific communities (Kasy and Abebe, 2021; Hampton, 2021). In the context of healthcare, these consequences can have alarming implications for the safety of patients that current and proposed frameworks, unfortunately, fail to consider seriously and comprehensively. In this respect, 'modern problems cannot be reduced to mere engineering solutions over the long-term' (Johnston, 2018). That is why we propose a three-level approach combining legal, organizational, and technical measures to help realize the diversity ideals into concrete AI applications in medicine.

Of course, it takes a village to integrate sex and gender considerations more systematically in AI for medicine. In this respect, this contribution starts a conversation within the law and technology community. We raise awareness of the magnitude of this problem and highlight the big unknowns still to be answered, such as how ignoring gender and sex considerations could affect patient safety or how such concerns can be adequately addressed through technical means. While efforts in bridging different disciplines, discussions, and bodies of literature cannot do justice to all the facets of the problem, we nevertheless believe that such an effort needed to start somewhere, somehow, because rushing the deployment of AI technologies that do not account for diversity (Poulsen, Fosch-Villaronga, & Soraa, 2020) risks having an even more unsafe and inadequate healthcare delivery for society.

Declaration of Competing Interest

None.

REFERENCES

Accenture (2017) Artificial Intelligence: Healthcare's New Nervous System. *Accenture Insight Driven Health*. Retrieved from https://www.accenture.com/_acnmedia/PDF-49/Accenture-Health-Artificial-Intelligence.pdf (last accessed 16 February 2021).

Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. PeerJ 2019;7:e7702.

Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020;20(1):1–9.

Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Fam Med Prim Care 2019;8(7):2328–31.

Antun V, Renna F, Poon C, Adcock B, Hansen AC. On instabilities of deep learning in image reconstruction and the potential costs of AI. Proc Natl Acad Sci 2020;117(48):30088–95.

Artificial Intelligence Act (2021) Proposal for a Regulation laying down harmonised rules on artificial intelligence. Retrieved from https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence.

Aymerich-Franch L, Fosch-Villaronga E. A self-guiding tool to conduct research with embodiment technologies responsibly. Front Robot AI 2020;7(22):1–5.

Bakkar N, Kovalik T, Lorenzini I, Spangler S, Lacoste A, Sponaugle K, et al. Artificial intelligence in neurodegenerative disease research: use of IBM Watson to identify additional RNA-binding proteins altered in amyotrophic lateral sclerosis. Acta Neuropathol 2018;135(2):227–47.

Barbee H, Deal C, Gonzales G. Anti-transgender legislation—a public health concern for transgender youth. JAMA Pediatr 2022;176(2):125–6.

Barocas S, Selbst AD. Big data's disparate impact. 104. Calif Law Rev 2016;671:1–62.

Beckers R, Kwade Z, Zanca F. The EU medical device regulation: implications for artificial intelligence-based medical device software in medical physics. Physica Med 2021;83:1–8.

Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: can language models be too big?. Proceedings of the 2021 ACM conference on fairness, accountability, and transparency; 2021. p. 610–23.

Bhavnani SP, Harzand A. From false-positives to technological Darwinism: controversies in digital health. Pers Med 2018;15(04):247–50.

Bird JD, Kuhns L, Garofalo R. The impact of role models on health outcomes for lesbian, gay, bisexual, and transgender youth. J Adolesc Health 2012;50(4):353–7.

Brasher EA. Addressing the failure of anonymization: guidance from the European union's general data protection regulation. Columbia Bus Law Rev 2018;209:1–45.

Bresnick, J. (2016). Big data, artificial intelligence, IoT may change healthcare in 2017. Retrieved https://healthitanalytics.com/news/big-data-artificial-intelligence-iot-may-change-healthcare-in-2017 (last accessed 16 February 2021).

Buhr, S. FDA clears AliveCor's Kardiaband as the first medical device accessory for the Apple Watch. In TechCrunch https://techcrunch.com/2017/11/30/fda-clears-alivecors-kardiaband-as-the-first-medical-device-accessory-for-the-apple-watch/ (2017).

Buolamwini J, Gebru T. Gender shades: intersectional accuracy disparities in commercial gender classification. Proceedings of the 1st conference on fairness, accountability and transparency; 2018. p. 77–91.

Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. Science 2017;356:183–6 6334.

Calleja C, Drukarch H, Fosch-Villaronga E. Diversity observations in an exoskeleton experiment. Proceedings of the inclusive HRI workshop: equity and diversity in design, application, methods, and community, 2022a. https://drive.google.com/file/d/1rBlf_XcqwKb_R1aLE63y5B-CQumk-FpR/view.

Calleja C, Drukarch H, Fosch-Villaronga E. Harnessing robot experimentation to optimize the regulatory framing of emerging robot technologies. Data & policy. Cambridge University Press; 2022b. p. 1–15. https://t.co/arDLOYqNvL.

Capper D, Jones DT, Sill M, Hovestadt V, Schrimpf D, Sturm D, et al. DNA methylation-based classification of central nervous system tumours. Nature 2018;555:469–74 7697.

Carnevale A, Tangari EA, Iannone A, Sartini E. Will big data and personalized medicine do the gender dimension justice? AI Soc 2021;1:1–13. doi:10.1007/s00146-021-01234-9.

Carr N. The shallows: what the internet is doing to our brains. WW Norton & Company; 2020.

Cirillo D, Catuara-Solarz S, Morey C, et al. Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ Digit Med 2020;3:81. doi:10.1038/s41746-020-0288-5.

Council of Europe (2022) Intersectionality - quoting Sandra Fredman May 2016. Retrieved from https://www.coe.int/en/web/north-south-centre/intersectionality, last accessed 16 June 2022.

Cruz-Roa A, et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: a deep learning approach for quantifying tumor extent. Sci Rep 2017;7:46450.

Custers BHM, Tavani Herman. The risks of epidemiological data mining. Ethics, computing and genomics: moral controversies in computational genomics. Boston: Jones and Bartlett Publishers, Inc; 2006.

Custers B, Van der Hof S, Schermer B. Privacy expectations of social media users: the role of informed consent in privacy policies. Policy Internet 2014;6(3):268–95.

Custers BHM, Bayamlioglu E, Baraliuc I, Janssens L, Hildebrandt M. Profiling as inferred data: amplifier effects and positive feedback loops. Being profiled: cogitas ergo sum. Amsterdam: Amsterdam University Press; 2018. p. 112–15.

Custers BHM, Bachlechner D. Advancing the EU data economy; conditions for realizing the full potential of data reuse. Inf Polity 2018;22(4):291–309. doi:10.3233/IP-170419.

Deaux Kay. Sex and gender. Annu Rev Psychol 1985;36:49–81.

Decataldo A, Ruspini E. Gender-sensitive data: the state of the art in Europe. Int Rev Sociol 2016;26(3):407–23.

Dembroff, R. (2019). Beyond binary: genderqueer as critical gender kind. Philosopher's Imprint. Retrieved from http://philsci-archive.pitt.edu/16317/, last accessed 17 June 2022.

Dillon S, Collett C. AI and gender: four proposals for future research. Cambridge: The Leverhulme Centre for the Future of Intelligence; 2019 Retrieved from(last accessed 2 February 2021).

Esteva A, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115–18.

EUGenMed, Cardiovascular Clinical Study Group, Regitz-Zagrosek, V., Oertelt-Prigione, S., Prescott, E., Franconi, F., ... & Stangl, V. (2016). Gender in cardiovascular diseases: impact on clinical manifestations, management, and outcomes. European heart journal, 37(1), 24-34.

Eurobarometer Survey 359 (2011) *Attitudes on Data Protection and Electronic Identity in the European Union*, Brussels, June 2011.

European Commission's High-Level Expert Group on AI (2018) A definition of AI: Main capabilities and scientific disciplines. European Commission. Retrieved from https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf.

European Commission, Directorate-General for Research and Innovation (2020) *Gendered innovations 2: how inclusive analysis contributes to research and innovation: policy review*, Publications Office, https://data.europa.eu/doi/10.2777/316197.

Felzmann H, Fosch-Villaronga E, Lutz C, Tamò-Larrieux A. Transparency you can trust: transparency requirements for artificial intelligence between legal norms and contextual concerns. Big Data Soc 2019;6(1):1–14 2053951719860542.

Felzmann H, Fosch-Villaronga E, Lutz C, Tamò-Larrieux A. Towards transparency by design for artificial intelligence. Sci Eng Ethics 2020;26:3333–61.

Fergus, J. (2020). Twitter is guessing users' genders to sell ads and often getting it wrong, input, https://www.inputmag.com/tech/twitter-guesses-your-gender-to-serve-you-ads-relevant-tweets-wrong-misgendered, accessed June 7, 2022.

Fiani CN, Serpe CR, Ryan JM. Non-binary identity and the double-edged sword of globalization. Trans lives in a globalizing world. Routledge; 2020. p. 50–65.

Fink, C., Kopecky, J., & Morawski, M. (2012). Inferring gender from the content of tweets: A region specific example. In Proceedings of the International AAAI Conference on Web and Social Media, 6(1), 459-462.

Fosch-Villaronga, E. (2016). ISO 13482:2014 and Its Confusing Categories. Building a Bridge Between Law and Robotics. In Wenger P., Chevallereau C., Pisla D., Bleuler H., Rodić A. (eds) New Trends in Medical and Service Robots, Vol. 39, Series Mechanisms and Machine Science, Springer, 31-44. doi:10.1007/978-3-319-30674-2_3.

Fosch-Villaronga E. Robots, healthcare, and the law: regulating automation in personal care. Routledge; 2019.

Fosch-Villaronga E, Chokoshvili D, Pierce RL, Ienca M, Binz VV, Leenes R, van Brakel R, Gutwirth S, de Hert P. Implementing AI in healthcare: an ethical and legal analysis based on case studies. Computers, privacy, and data protection 2020 – artificial intelligence (2020). Hart Publishing; 2020.

Fosch-Villaronga E, Drukarch H. AI for healthcare robotics. CRC Press; 2022.

Fosch-Villaronga E, Poulsen A, Søraa RA, Custers BHM. A little bird told me your gender: gender inferences in social media. Inf Process Manag 2021;58(3).

Fosch Villaronga E, Drukarch H, Khanna P, Custers BHM. A human in the loop in surgery automation. Nat Mach Intell 2021a;3:368–9. doi:10.1038/s42256-021-00349-4.

Franconi F, Brunelleschi S, Steardo L, Cuomo V. Gender differences in drug responses. Pharmacol Res 2007;55(2):81–95.

Frost & Sullivan (2016) Frost & Sullivan From $600 M to $6 billion, artificial intelligence systems poised for dramatic market expansion in healthcare. Retrieved from https://ww2.frost.com/news/press-releases/600-m-6-billion-artificial-intelligence-systems-poised-dramatic-market-expansion-healthcare/ (last accessed 2 February 2021).

Garbuio M, Lin N. Artificial intelligence as a growth engine for healthcare startups: emerging business models. Calif Manag Rev 2019;61(2):59–83.
Garibo-Orts, O. (2018, September). A big data approach to gender classification in twitter. In Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018). Retrieved from http://ceur-ws.org/Vol-2125/paper_204.pdf.
Government of Catalonia (2022) Sex and gender perspective incorporation tool in research. Retrieved from https://aquas.gencat.cat/ca/ambits/recerca-salut/responsable/genere/eina-incorporacio-perspectiva-genere-recerca/index.html#googtrans(ca%7Cen), last accessed 17 June 2022.
Gruber K. Is the future of medical diagnosis in computer algorithms? Lancet Digit Health 2019;1(1):e15–16.
Gulshan V, Peng L, Voram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016;316:2402–10. doi:10.1001/jama.2016.17216.
Gurzawska A, Mäkinen M, Brey P. Implementation of responsible research and innovation (RRI) practices in industry: providing the right incentives. Sustainability 2017;9(10):1759. doi:10.3390/su9101759.
Haas L, Hwang CP. Gender and organizational culture: correlates of companies' responsiveness to fathers in Sweden. Gend Soc 2007;21(1):52–79.
Hamidi F, Scheuerman MK, Branham SM. Gender recognition or gender reductionism? The social implications of embedded gender recognition systems. Proceedings of the 2018 CHI conference on human factors in computing systems; 2018. p. 1–13.
Hampton, L.M. (2021). Black feminist musings on algorithmic oppression. arXiv preprint arXiv:2101 09869.
Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial Intelligence in surgery: promises and perils. Ann Surg 2018;268(1):70–6. doi:10.1097/SLA.0000000000002693.
Haenssle HA, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol 2018;29:1836–42.
Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. California management review, 61(4), 5-14.
High Level Expert Group on AI (2019) Ethical Guidelines for Trustworthy AI. Retrieved from https://ec.europa.eu/futurium/en/ai-alliance-consultation.
Holder, C., Khurana, V., Harrison, F., & Jacobs, L. (2016). Robotics and law: Key legal and regulatory implications of the robotics age (Part I of II). Computer law & security review, 32(3), 383-402.
Hooper C. Manly states: masculinities, international relations, and gender politics. Columbia University Press; 2001.
Houssami, Houssami N, Lee CI, Buist D, Tao D, et al. Artificial intelligence for breast cancer screening: opportunity or hype? Breast 2017;2017(36):31–3 2017. doi:10.1016/j.breast.2017.09.003.
Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 2018;6(11):e12106 2018 Nov 23. doi:10.2196/12106.
Ireland ML. The female ACL: why is it more prone to injury? Orthop Clin 2002;33(4):637–51.
Ito, J. (2019). Supposedly 'fair'algorithms can perpetuate discrimination. Wired, April, 2. Retrieved from https://www.wired.com/story/ideas-joi-ito-insurance-algorithms/, last accessed 17 June 2022.
Jensen M. Challenges of privacy protection in big data analytics. Proceedings of the IEEE international congress on big data; 2013. p. 235–8.
Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell 2019;1(9):389–99.
Johnston SF. The technological fix as social cure-all: origins and implications. IEEE Technol Soc Mag 2018;37(1):47–54.
Kamiran F, Calders T, Pechenizkiy M. Techniques for discrimination-free predictive models. Discrimination and privacy in the information society in Custers et al. (eds). Heidelberg: Springer; 2013.
Kachel S, Steffens MC, Niedlich C. Traditional masculinity and femininity: validation of a new scale assessing gender roles. Front Psychol 2016;7:956.
Kasy M, Abebe R. Fairness, equality, and power in algorithmic decision-making. Proceedings of the 2021 ACM conference on fairness, accountability, and transparency; 2021. p. 576–86.
Kaul V, Enslin S, Gross SA. The history of artificial intelligence in medicine. Gastrointest Endosc 2020;92(4):807–12.
Keyes O. The misgendering machines: trans/HCI implications of automatic gender recognition. Proceedings of the ACM on human-computer interaction; 2018. p. 1–22 2(CSCW).
Klein E. Gender politics: from consciousness to mass politics. Harvard University Press; 2013.
Koops BJ. Should ICT regulation be technology-neutral?. TMC Asser Press; 2006. p. 77–108 Bert-Jaap Koops, Miriam Lips, Corien Prins & Maurice Schellekens, eds.
Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behaviour. Proceedings of the national academy of sciences (PNAS), 2012.
La Fors K, Custers BHM, Keymolen E. Reassessing values for emerging big data technologies: integrating design-based and application-based approaches. Ethics Inf Technol 2019;21(3):209–26. doi:10.1007/s10676-019-09503-4.
Larrazabal AJ, Nieto N, Peterson V, Milone DH, Ferrante E. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. Proc Natl Acad Sci 2020;117(23):12592–4.
LeBreton M. The erasure of sex and gender minorities in the healthcare system. BioéthiqueOnline 2013;2. retrieved from http://hdl.handle.net/1866/9811.
Lee D, Yoon SN. Application of artificial intelligence-based technologies in the healthcare industry: opportunities and challenges. Int J Environ Res Public Health 2021;18(1):1–18 271.
Levine, B. & Brown, A. Onduo delivers diabetes clinic and coaching to your smartphone. In Diatribe https://diatribe.org/onduo-delivers-diabetes-clinic-and-coaching-your-smartphone (2018).
Liang H, Tsui BY, Ni H, Valentim CC, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med 2019;25(3):433–8.
Lindell, Y., & Pinkas, B. (2000, August). Privacy preserving data mining. In Annual International Cryptology Conference. Springer, Berlin, Heidelberg, 36-54.
Lips, H. M. (2020). Sex and gender: An introduction. Waveland Press.
Malgieri G, Niklas J. Vulnerable data subjects. Comput Law Secur Rev 2020;37.
Manheim KM, Kaplan L. Artificial intelligence: risks to privacy and democracy. Yale J Law & Technol 2019;21:107–88.
Martinetti, A., Chemweno, P. K., Nizamis, K., & Fosch-Villaronga, E. (2021). Redefining safety in light of human-robot interaction: A critical review of current standards and regulations. Frontiers in chemical engineering, 32, 1-12.
Mauvais-Jarvis, F., Merz, N.B., Barnes, P.J., Brinton, R.D., Carrero, J.J., DeMeo, D.L., ... & Suzuki, A. (2020). Sex and gender: modifiers of health, disease, and medicine. The Lancet, 396(10250), 565-582.

MedTech Europe (2022) The proposed European AI Act and its impact on the medical technology industry. Retrieved from https://library.myebook.com/theparliament/the-parliament-magazine-issue-543-25-october-2021/3691/#page/12, last accessed 21 June 2022.
Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. arXiv preprint *arXiv:13104546*Retrieved from https://arxiv.org/abs/1310.4546 (last accessed 28 March 2021).
Mitchell M, Baker D, Moorosi N, Denton E, Hutchinson B, Hanna A, et al. Diversity and inclusion metrics in subset selection. Proceedings of the AAAI/ACM conference on AI, ethics, and society; 2020. p. 117–23.
Muñoz, D.C., Sant, C., Becedas, R.R., & Fat, D.M. (2020). Dangers of gender bias in CRVS and cause of death data: the path to health inequality, 1–24. Retrieved from https://crvssystems.ca/sites/default/files/assets/files/CRVS_Gender_3.3_COD_e_WEB.pdf, last accessed 16 June 2022.
Nielsen MW, Stefanick ML, Peragine D, Neilands TB, Ioannidis J, Pilote L, et al. Gender-related variables for health research. Biol Sex Differ 2021;12(1):1–16.
Niemiec E. Will the EU Medical Device Regulation help to improve the safety and performance of medical AI devices? Digital Health 2022;8:1–8 20552076221089079.
Nieuwenhuis, M., & Wilkens, J. (2018, September). Twitter text and image gender classification with a logistic regression n-gram model. In Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018). Retrieved from http://ceur-ws.org/Vol-2125/paper_183.pdf.
Noble SU. Algorithms of oppression: how search engines reinforce racism. NYU Press; 2018.
Nomura T, Legato MJ. Chapter 47 - robots and gender. Principles of gender-specific medicine. San Diego: Academic Press; 2017. p. 695–703.
Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366:447–53 6464.
Ohm P. Broken promises of privacy: responding to the surprising failure of anonymization. UCLA Law Review 2009;57:1701.
O'Neil C. Weapons of math destruction; how big data increases inequality and threatens democracy. New York: Crown; 2016.
Olsen CM, Thompson JF, Pandeya N, Whiteman DC. Evaluation of sex-specific incidence of melanoma. JAMA Dermatol 2020;156(5):553–60.
Pasti, R., & Castro, L. N. D. (2016). Gender classification of twitter data based on textual meta-attributes extraction. In New advances in information systems and technologies. Springer, Cham. 1025-1034.
Patel NM, et al. Enhancing next-generation sequencing-guided cancer care through cognitive computing. Oncologist 2018;23:179–85.
Pennington J, Socher R, Manning CD. Glove: global vectors for word representation. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP); 2014. p. 1532–43.
Petrone J. FDA approves stroke-detecting AI software. Nat Biotechnol 2018;36:290.
Pew Research Center (2019) The challenges of using machine learning to identify gender in images. Internet & Technology. Retrieved from https://www.pewresearch.org/internet/2019/09/05/the-challenges-of-using-machine-learning-to-identify-gender-in-images/ (last accessed 2 February 2021).
Pisano, E.D. (2020). AI shows promise for breast cancer screening.
Price WN, Cohen IG. Privacy in the age of medical big data. Nat Med 2019;25(1):37–43.
Price II, Nicholson W. Medical AI and contextual bias. Harv J Law Technol 2019;33(1):1–52.
Pryzgoda J, Chrisler JC. Definitions of gender and sex: the subtleties of meaning. Sex Roles 2000;43(7):553–69.
Randall V, Waylen G. Gender, politics and the state. Routledge; 2012.
Razzaki, S., Baker, A., Perov, Y., Middleton, K., Baxter, J., Mullarkey, D. et al. (2018). A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis. arXiv preprint arXiv:1806.10698.
Regitz-Zagrosek V, Legato MJ. Sex and gender specific aspects—from cells to cardiovascular disease. Principles of gender-specific medicine: gender in the genomic era. Academic Press; 2017. p. 341–62.
Roach L. Artificial intelligence. EyeNet Mag 2017;2017:77–83.
Rosamond W, Flegal K, Furie K, Go A, Greenlund K, Hong Y. Heart disease and stroke statistics—2008 update: a report from the American heart association statistics committee and stroke statistics subcommittee. Circulation 2008;117(4):e25–e146.
Rotenstein LS, Jena AB. Lost Taussigs-the consequences of gender discrimination in medicine. N Engl J Med 2018;378(24):2255–7.
Saddler N, Adams S, Robinson LA, Okafor I. Taking initiative in addressing diversity in medicine. Can J Sci Math Technol Educ 2021;21(2):309–20.
Schiebinger L. Scientific research must take gender into account. Nature 2014;507(7490):9 -9.
Schiffer E, et al. The 'sex gap' in COVID-19 trials: a scoping overview. Lancet eClinicalMed 2020;29. doi:10.1016/j.eclinm.2020.100652.
Schomberg Rvon. A vision of responsible research and innovation. Responsible innovation: managing the responsible emergence of science and innovation in society. John Wiley; 2013. p. 51–74.
Schwalbe N, Wahl B. Artificial intelligence and the future of global health. Lancet N Am Ed 2020;395(10236):1579–86.
Selbst AD, Barocas S. The intuitive appeal of explainable machines. Fordham Law Rev 2018;87:1085.
Shannon, J. (2018). Heart attack – it's different for women. Retrieved 10 April 2021, from https://irishheart.ie/news/heart-attack-its-different-for-women/.
Shotwell A, Sangrey T. Resisting definition: gendering through interaction and relational selfhood. Hypatia 2009;24(3):56–76.
Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf 2014;23(9):727–31.
Sizemore-Barber A. Prismatic performances: queer South Africa and the fragmentation of the rainbow nation. University of Michigan Press; 2020.
Smith BC. The promise of artificial intelligence: reckoning and judgment. Mit Press; 2019.
Snyder CF, Wu AW, Miller RS, Jensen RE, Bantug ET, Wolff AC. The role of informatics in promoting patient-centered care. Cancer J 2011;17(4):211. doi:10.1097/PPO.0b013e318225ff89.
Søraa, R. A. (2017). Mechanical genders: how do humans gender robots?. Gender, Technology and Development, 21(1-2), 99-115.
Søraa RA, Fosch-Villaronga E. Exoskeletons for all: The interplay between exoskeletons, inclusion, gender and intersectionality. Paladyn Journal of Behavioral Robotics 2020;11(1):217–27. doi:10.1515/pjbr-2020-0036.
Soumelidou A, Tsohou A. Towards the creation of a profile of the information privacy aware user through a systematic literature review of information privacy awareness. Telemat Inform 2021;61.
Stathoulopoulos K, Mateos-Garcia JC. Gender diversity in AI research. NESTA; 2019 Retrieved from https://media.nesta.org.uk/documents/Gender_Diversity_in_AI_Research.pdf (last accessed 2 February 2021).
Stahl BC, Coeckelbergh M. Ethics of healthcare robotics: towards responsible research and innovation. Robot Auton Syst 2016;86:152–61.

Stilgoe J, Owen R, Macnaghten P. Developing a framework for responsible innovation. Res Policy 2013;42:1568–80.
Sun, T.Y., Walk IV, O.J., Chen, J.L., Nieva, H.R., & Elhadad, N. (2020). Exploring gender disparities in time to diagnosis. arXiv preprint arXiv:2011.06100.
Tannenbaum C, Ellis RP, Eyssel F, Zou J, Schiebinger L. Sex and gender analysis improves science and engineering. Nature 2019;575(7781):137–46.
Tat E, Bhatt DL, Rabbat MG. Addressing bias: artificial intelligence in cardiovascular medicine. Lancet Digit Health 2020;2(12):e635–6.
Tietjen, D. & Woedtke, N. (2021) Artificial Intelligence Act (AIA) - legal uncertainty for medical device manufacturers. Retrieved from https://www.taylorwessing.com/en/insights-and-events/insights/2021/11/artificial-intelligence-act-rechtliche-unsicherheit-fuer-medizinprodukteherstellerr, last accessed 22 June 2022.
Tomasev, N., McKee, K.R., Kay, J., & Mohamed, S. (2021). Fairness for unobserved characteristics: insights from technological impacts on queer communities. arXiv preprint arXiv:2102.04257.
Topol E. Deep medicine: how artificial intelligence can make healthcare human again. UK: Hachette; 2019a.
Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019b;25:44–56.
United Nations (2022) Gender stereotypes and Stereotyping and women's rights. Retrieved from: https://www.ohchr.org/sites/default/files/Documents/Issues/Women/WRGS/OnePagers/Gender_stereotyping.pdf.
Vermeir E, Jackson LA, Marshall EG. Barriers to primary and emergency healthcare for trans adults. Cult Health Sex 2018;20(2):232–46.
Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ 2020;368:1–12 https://www.bmj.com/content/bmj/368/bmj.l6927.full.pdf.
Wachter S, Mittelstadt B. A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. Columbia Bus Law Rev 2019;2019(2):494–620.
Wagner AD, Oertelt-Prigione S, Adjei A, Buclin T, Cristina V, Csajka C, et al. Gender medicine and oncology: report and consensus of an ESMO workshop. Ann Oncol 2019;30(12):1914–24.
Wang, X. et al. ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. Preprint at https://arxiv.org/abs/1705.02315 (2017).
Wang P, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nat Biomed Eng 2018;2:741–8.
Wapner, J. Cancer scientists have ignored African DNA in the search for cures. In Newsweek https://www.newsweek.com/2018/07/27/cancer-cure-genome-cancer-treatment-africa-genetic-charles-rotimi-dna-human-1024630.html (2018).
Willson, M. (2017). Algorithms (and the) everyday. Information, Communication & Society, 20(1), 137-150.
Wong D, Yip S. Machine learning classifies cancer. Nature 2018;555:446–7.
Yu KH, et al. Predicting non–small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat Commun 2016;7:12474.
Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018;2(10):719–31.
Zhang J, et al. Fully automated echocardiogram interpretation in clinical practice feasibility and diagnostic accuracy. Circulation 2018;138:1623–35.
Zhang Y, Wang S, Hermann A, Joly R, Pathak J. Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women. J Affect Disord 2020;279:1–8.