




## REVIEW ARTICLE OPEN



# Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare

Davide Cirillo<sup>1,10</sup>, Silvina Catuara-Solarz<sup>2,3,10</sup>, Czuee Morey<sup>3,4</sup>, Emre Guney<sup>5</sup>, Laia Subirats<sup>6,7</sup>, Simona Mellino<sup>3</sup>, Annalisa Gigante<sup>3</sup>, Alfonso Valencia<sup>1,8</sup>, María José Rementería<sup>1</sup>, Antonella Santucci Chadha<sup>3</sup> and Nikolaos Mavridis<sup>3,9</sup>

Precision Medicine implies a deep understanding of inter-individual differences in health and disease that are due to genetic and environmental factors. To acquire such understanding there is a need for the implementation of different types of technologies based on artificial intelligence (AI) that enable the identification of biomedically relevant patterns, facilitating progress towards individually tailored preventative and therapeutic interventions. Despite the significant scientific advances achieved so far, most of the currently used biomedical AI technologies do not account for bias detection. Furthermore, the design of the majority of algorithms ignore the sex and gender dimension and its contribution to health and disease differences among individuals. Failure in accounting for these differences will generate sub-optimal results and produce mistakes as well as discriminatory outcomes. In this review we examine the current sex and gender gaps in a subset of biomedical technologies used in relation to Precision Medicine. In addition, we provide recommendations to optimize their utilization to improve the global health and disease landscape and decrease inequalities.

*npj Digital Medicine* (2020)3:81; <https://doi.org/10.1038/s41746-020-0288-5>

## INTRODUCTION

Precision Medicine, as opposed to the preponderant one-size-fits-all approach, attempts to find personalized preventative and therapeutic strategies by taking into account differences in genes, environment and lifestyle, throughout the lifespan. The value and impact of this approach makes Precision Medicine one of the most promising health initiatives in our society<sup>1</sup>.

Both biological (sex) and socio-cultural (gender) aspects (see Supplementary Note 1 “Sex and gender”) constitute relevant sources of variation in a number of clinical and subclinical conditions, affecting risk factors, prevalence, age of onset, symptomatology manifestation, prognosis, biomarkers and treatment effectiveness<sup>2</sup>. Evidence of sex and gender differences has been reported in chronic diseases such as diabetes, cardiovascular disorders, neurological diseases<sup>3</sup>, mental health disorders<sup>4</sup>, cancer<sup>5</sup>, autoimmunity<sup>6</sup>, as well as physiological processes such as brain aging<sup>7</sup> and sensitivity to pain<sup>8</sup>. Moreover, differences in lifestyle factors that are associated with sex and gender, such as diet, physical activity, tobacco use and alcohol consumption, also correlate with the epidemiology of diseases<sup>9–11</sup>. Nonetheless, there are still open questions regarding health differences across the gender spectrum, reflected by the scarcity of studies dedicated to intersex, transgender and nonbinary individuals<sup>12,13</sup>. Initiatives, such as the Global Trans Research Evidence Map<sup>14</sup>, foster research access in this area to improve our understanding of the effects of medical interventions on health and life quality across the gender spectrum. Additionally, such clinical differences are accompanied by sex and gender gaps in the use and access of medical services and tools as well as affordability to medical costs<sup>15</sup>.

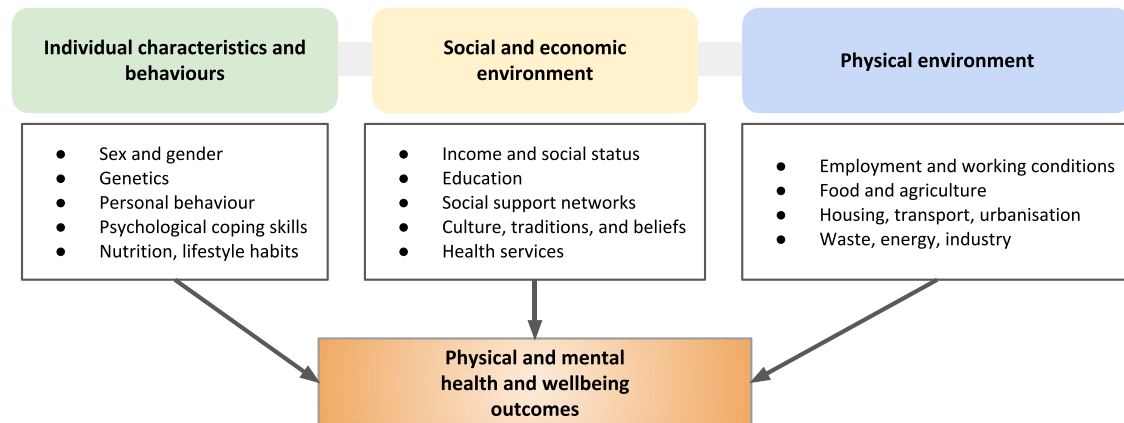
The study of sex and gender differences represents an increasingly significant line of research<sup>16</sup>, involving all levels of biomedical and health sciences, from basic research to population studies<sup>17</sup>, and also fueling debate regarding its sociological implications<sup>18,19</sup>. Observed sex and gender differences in health and wellbeing are influenced by complex links between both biological and social-economic factors (see Fig. 1), which are often surrounded by confounding variables such as stigma, stereotypes, and the misrepresentation of data. Consequently, health research and practices can be entangled with sex and gender inequalities and biases<sup>20</sup>.

In recent years, the social awareness of such biases has increased and they have become even more evident with the introduction of widespread advance in biomedical artificial intelligence (AI). In this regard, one could argue that AI technologies act as a double-edged sword. On one hand, algorithms can magnify and perpetuate existing sex and gender inequalities if they are developed without removing biases and confounding factors. On the other hand, they have the potential to mitigate inequalities by effectively integrating sex and gender differences in healthcare if designed properly. The development of precise AI systems in healthcare will enable the differentiation of vulnerabilities for disease and response to treatments among individuals, while avoiding discriminatory biases.

The purpose of this review is to highlight the main available biomedical data types and the role of several AI technologies to understand sex and gender differences in health and disease. We address their existing and potential biases and their contribution to create personalized therapeutic interventions. We examine the sex and gender issues involved with the generation and collection

<sup>1</sup>Barcelona Supercomputing Center (BSC), C/ Jordi Girona, 29, 08034 Barcelona, Spain. <sup>2</sup>Telefonica Innovation Alpha Health, Torre Telefonica, Plaça d'Ernest Lluch i Martin, 5, 08019 Barcelona, Spain. <sup>3</sup>The Women's Brain Project (WBP), Guntershausen, Switzerland. <sup>4</sup>Wega Informatik AG, Aeschengraben 20, CH-4051 Basel, Switzerland. <sup>5</sup>Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Research Institute and Pompeu Fabra University, Dr. Aiguader, 88, 08003 Barcelona, Spain. <sup>6</sup>Eurecat - Centre Tecnològic de Catalunya, C/ Bilbao, 72, Edifici A, 08005 Barcelona, Spain. <sup>7</sup>eHealth Center, Universitat Oberta de Catalunya, Rambla del Poblenou, 156, 08018 Barcelona, Spain. <sup>8</sup>ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain. <sup>9</sup>Interactive Robots and Media Laboratory (IRML), Abu Dhabi, United Arab Emirates. <sup>10</sup>These authors contributed equally: Davide Cirillo, Silvina Catuara-Solarz. ✉email: [davide.cirillo@bsc.es](mailto:davide.cirillo@bsc.es)

## Determinants of physical and mental health and wellbeing



**Fig. 1 The key determinants of health.** Health and wellbeing of individuals and communities are influenced by several factors, which include the person's individual characteristics and behaviours and the socio-economic, and physical environment, according to the World Health Organization (WHO) ([www.who.int/hia/evidence/doh/en/](http://www.who.int/hia/evidence/doh/en/)). Sex and gender differences interact with the whole spectrum of health determinants.

of experimental, clinical and digital data. Furthermore, we review a number of technologies to analyze and deploy this data, namely Big Data Analytics, Natural Language Processing and Robotics. Those technologies are becoming increasingly relevant for Precision Medicine while being exposed to potential sex and gender biases. In addition, we surveyed Explainable AI and algorithmic Fairness, which ensure the trustworthy delivery of AI solutions that can account for sex and gender differences in the patient's wellbeing. Finally, we provide a summary to incorporate the sex and gender dimension into biomedical research and AI technologies to accelerate the developments that will enable the creation of effective strategies to augment populations' health and wellbeing.

### Desirable vs. undesirable biases

Despite the fact that the term "bias" has gained a negative connotation due to its association to unfair prejudice, the differential consideration and treatment to specific biomedical aspects is a necessary course of action in the context of Precision Medicine. Therefore, here we defined two main categories of sex and gender biases: desirable and undesirable (see Fig. 2). The difference between them is found in the impact that these biases have on the patients' wellbeing and healthcare access.

A desirable bias implies taking into account sex and gender differences to make a precise diagnosis and recommend a tailored and more effective treatment for each individual. This represents a much more accurate approach than collapsing all sex and gender categories into a single one, such as data generated from mixed sex or gender cohorts<sup>16</sup>. Table 1 reports illustrative examples of clinical conditions and biomedical techniques in which desirable biases would be beneficial for both basic and clinical research as well as diagnosis and treatment.

Conversely, an undesirable bias is that which exhibits unintended or unnecessary sex and gender discrimination. This occurs when claims are made in relation with sex or gender and medical conditions despite the lack of exhaustive evidence to support them or based on skewed evidence.

For instance epidemiological studies indicate that there is a higher prevalence of depression among women, however, this may result from a skewed diagnosis due to clinical scales of depression measuring symptoms that occur more frequently among women<sup>21</sup>. Another source of undesirable bias is the misrepresentation of the target population, leaving minorities out.

An example of this is the case of the insufficient representation of pregnant women in psychiatric research<sup>22</sup>.

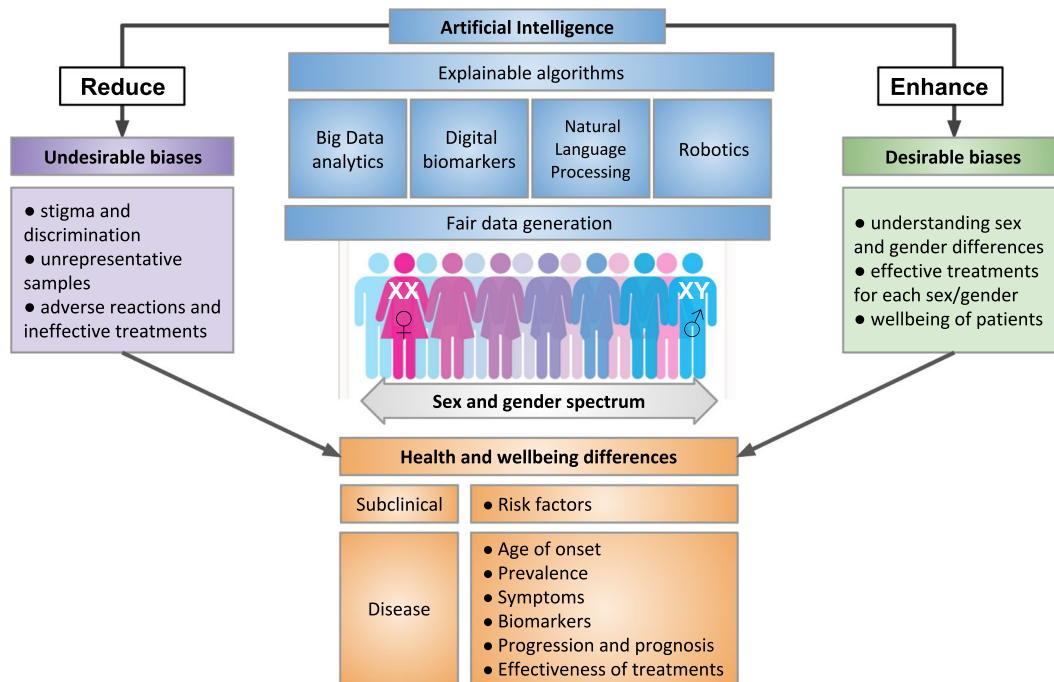
There are multiple sources of undesirable biases that could accidentally be introduced in AI algorithms<sup>23</sup> (see Table 2). The most common one is the lack of a representative sample of the population in the training dataset. In some cases, a bias may exist in the overall population as a consequence of underlying social, historical or institutional reasons. In other cases, an algorithm itself, and not the training dataset, can introduce bias by obscuring an inherent discrimination or inducing an unreasoned or irrelevant selectivity.

### SOURCES AND TYPES OF HEALTH DATA

#### Experimental and clinical data

In the early days of biomedical research and drug discovery, sex-specific biological differences were neglected and both experimental and clinical studies were fundamentally focused on male experimental models or male subjects<sup>24</sup>. Even nowadays, male mouse models are overall more represented than female models in basic, preclinical, and surgical biomedical research<sup>25</sup>. A recent analysis of data on 234 phenotypic traits from almost 55,000 mice showed that existing findings were influenced by sex<sup>26</sup>. The lack of representation of female models and patients is partly due to technical and bioethical considerations, such as the attempt to reduce the impact of estrous cycle in experimental studies and protective policies for women of childbearing age in clinical research. Consequently, some of the treatments that currently exist for several diseases are not adequately evaluated in women<sup>27,28</sup> who are likely to be underrepresented in clinical trials<sup>29,30</sup>, especially in Phases I and II<sup>31,32</sup>.

Differences in the physiology of sexes<sup>33</sup> might translate into clinically relevant differences in pharmacokinetics and pharmacodynamics of drugs. These differences, taken together with the underrepresentation of women in clinical trials, can explain why women typically report more adverse event reactions compared with men<sup>34</sup>. An illustrative example of the discrepancy between sexes in clinical trials is zolpidem, a sleep medication<sup>35</sup>, which shows slower drug metabolism and high secondary effects in women, increasing their health risks compared with men<sup>34,36,37</sup>. In 2013, the FDA recommended a weight-based dosing zolpidem for women due to potential sex-specific impairments<sup>38</sup>, proving how a stratified consideration of sexes enables a better understanding of differential drug toxicity. The design of preclinical and clinical



**Fig. 2 Desirable and undesirable biases in artificial intelligence for health.** Fair data generation and explainable algorithms are fundamental requirements for the design and application of artificial intelligence to optimize for health and wellbeing across the sex and gender spectrum. This will facilitate the reduction of undesirable biases that propagate inequity and discrimination, and will promote desirable differentiations that help develop Precision Medicine.

studies should have a sex and gender-based approach in order to reduce the time to translate research into clinical practice, as well as to understand and implement precise pharmacological guidelines<sup>39</sup>.

Accounting for sex and gender differences leads to a better understanding of the pharmacodynamic and pharmacokinetic action of a drug. It also carries substantial economic implications<sup>40</sup> as conducting studies on large population-based trials is generally more expensive<sup>41</sup>, and often requires post-trial analyses to identify and categorise the factors that explain the varying drug response across individuals.

In summary, although there is a significant gap between two sexes on the availability of clinical data and the knowledge on the effects of drugs, recent clinical guidelines and initiatives hint to a fairer landscape that accounts for sex differences in biomedical research and clinical practice.

#### Digital biomarkers

Digital biomarkers are physiological, psychological and behavioral indicators based on data including human-computer interaction (e.g. swipes, taps, and typing), physical activity (e.g. gait, dexterity) and voice variations, collected by portable, wearable, implantable or even ingestible devices<sup>42</sup>. They can facilitate the diagnosis of a condition, the assessment of the effects of a treatment and the predicted prognosis for a particular patient. In addition, some digital biomarkers can inform on patient adherence to treatment.

There are many digital biomarkers that are currently being developed or already approved or cleared by the U.S. Food and Drug Administration (FDA) for use cases such as risk detection, diagnosis, and monitoring of symptoms and endpoints in clinical trials<sup>42</sup> (see Table 3).

A particular therapeutic area where digital biomarkers are becoming beneficial is that of neurological and mental health disorders. Since digital devices can acquire health related data in real-time, they can enable a continuous monitoring of an individual's health parameters in a cost-effective way that is more

granular, ecological and objective than the currently clinically used self-reports, questionnaires or psychometric tests. Digital biomarkers are becoming especially relevant for those clinical conditions where small fluctuations in daily symptoms or performance are clinically meaningful. This is the case, for example, of early detection of neurodegenerative disorders such as Alzheimer's disease (AD), in which key indices of preclinical stages are cognitive, motor and sensory changes that occur 10 or 15 years prior to its effective diagnosis<sup>43,44</sup>.

Despite the progress that has been made on digital biomarkers in the last years, sex and gender differences in these indices of health and disease have not been examined yet. Considering that several studies have shown that there are significant sex differences on neurodegenerative, physiological and cognitive aspects during the preclinical stages of AD<sup>45</sup>, it is reasonable to expect that further sex differences will be found in the digital biomarkers for this and other clinical conditions.

In some cases the analysis of sex differences on digital biomarkers is prevented by undesired biases in the datasets used by the models that provide the health indicators. For instance, current studies that test digital biomarkers are often performed with small sample sizes in the range of tens to hundreds of subjects and tend to show insufficient demographic information on sex and gender<sup>46</sup>. For example, in a study assessing digital biomarkers for Parkinson's disease (PD), only 18.6% were women<sup>47</sup>. As a consequence, if an algorithm is trained with a dataset overrepresented by male patients, it may lead to a more accurate detection of those symptoms that are more frequently manifested by male PD patients (rigidity and rapid-eye movement) in comparison to those symptoms that are more frequently manifested by female PD patients (dyskinesias and depression)<sup>48</sup>.

In other cases, the undesired biases arise from the digital device itself, such as in the case of a pulse oximetry which showed errors in the predicted arterial oxyhemoglobin saturation associated with sex and skin colour of the subjects<sup>49</sup>.

An additional source of undesired biases in digital biomarkers is the unbalanced access, and use of digital devices among people

**Table 1.** Illustrative examples of clinical conditions and studies in which desirable biases would be beneficial for both basic and clinical research as well as diagnosis and treatment.

Clinical conditions and studies	Current status without the desirable bias	Utility of the desirable bias
Autistic spectrum disorder	There is a current lack of consideration of the demonstrated age-dependent sex differences in the symptomatology related with impairments in social communication and interaction, expressive behaviour, reciprocal conversation, non-verbal gestures for diagnostic purposes <sup>123</sup> .	Differential diagnostic criteria for males and females could facilitate the identification of the clinical diagnosis leading to appropriate treatment.
Cardiovascular disorders	Although it has been documented that men and women respond differently to many cardiovascular medications such as statins, angiotensin-converting enzyme inhibitors and $\beta$ -Blockers among others, adopted treatments do not consider sex differences <sup>124</sup> . Despite the fact that Coronary heart disease (CHD) is the leading cause of death among women <sup>125</sup> , the majority (67%) of patients enrolled in clinical trials for cardiovascular devices are male <sup>126</sup> .	Making prescriptions according to the sex of the patient could lead to improved health benefits. The application of a desirable bias towards women would lead to a more accurate representation of sex differences in clinical research.
Genome-wide association studies (GWAS)	Most of genome-wide association studies (GWAS) focus on white male subjects <sup>127</sup> and those that explore sex differences in complex traits are scarce <sup>128</sup> .	The introduction of desirable biases to deliberately include female subjects and other ethnicities in GWAS could lead to better account for potential sex differences in disease that are currently unknown because of being overlooked.
Human immunodeficiency virus (HIV)	The observed lower female representation in HIV clinical trials depends, among other factors, from the disadvantaged awareness about treatment and enrolment options compared with men <sup>129–131</sup> .	Promoting empowerment initiatives in those patients with disadvantages will increase their exposure to treatment options and clinical trial enrolment.

with different sexes and genders as well as education and income levels and age<sup>50</sup>. In fact, in low and middle income countries, women are 10% less likely to own smartphones (see Fig. 3) and 26% less likely to use the internet compared with men, and 1.2 billion women do not even have access to mobile internet<sup>51</sup>. This creates uneven datasets that promote misrepresentation of digital biomarkers. Awareness and efforts into the identification of sex and gender differences in digital biomarkers will lead to more accurate indicators for prevention and diagnosis of disease, as well as more effective treatment monitoring.

## TECHNOLOGIES FOR THE ANALYSIS AND DEPLOYMENT OF HEALTH DATA

### Big Data analytics

Big Data analytics is a body of techniques and tools to collect, organize and examine large amounts of data. Common Big Data analytics processes and approaches include the creation of data management infrastructures and the application of data-driven algorithms and AI solutions<sup>52</sup>. Biomedical and clinical Big Data have the potential of providing deeper insights into health and disease at an unprecedented scale. Moreover, the availability of longitudinal health Big Data enables the characterization of the transitions between health and disease states as well as their similarities and differences among sexes and genders. Large international research infrastructures, such as ELIXIR<sup>53</sup> and NIH Big Data to Knowledge (BD2K)<sup>54</sup>, provide robust, long-term sustainable biomedical resources that will enable identifying differential patterns for health and disease transitions including the sex and gender dimension.

For instance, data from GWAS targeting smoking behaviour have shown sex-associated genetic differences that influence smoking initiation and maintenance<sup>55</sup>. Interestingly, these differences complement the differential effectiveness of tobacco control initiatives based on the sex of the individuals that receive the preventative messages<sup>56</sup>. Similarly, genomic studies in large human cohorts revealed chromosomal factors related to sex differences in excess body fat accumulation<sup>57</sup>, interlinking recent

insights on obesity from different Big Data types such as social media, retail sales, commercial data, geolocalization, transport and digital devices<sup>58</sup>.

Big Data analytics focused on health under the sex and gender lens are carried out worldwide by several initiatives such as Data2x ([www.data2x.org](http://www.data2x.org)). This collaborative platform explores female wellbeing through statistical analysis of data covering demography, education, health, geolocation, in order to map indices disaggregated by gender.

For instance, significant sex differences in behavioral and social patterns related with communication such as the number and duration of phone calls and the degree of social networking callers have been observed<sup>59</sup>. Furthermore, quantitative analysis into sex and cultural differences uncovered associations with mental health and social networks<sup>60</sup>, showing men express higher negativity and lower desire of social support on social media compared with women.

Awareness of sex and gender differences through biomedical Big Data could lead to a better risk stratification. For example, a query of sex and gender differences in heart diseases revealed that in women enhanced parasympathetic and decreased sympathetic tones appear to be greater and also defensive during cardiac stress<sup>61</sup>, while key reproductive factors associated with coronary heart disease only modestly improve risk prediction<sup>62</sup>.

The caveat of these resources is that the exploitation of their biomedical Big Data can magnify existing undesirable biases, for instance by introducing inferential errors due to sampling, measurement, multiple comparison, aggregation, and systematic exclusion of information<sup>63</sup>. For example, biases may be introduced in clinical decision support algorithms that rely on data obtained from the large reservoirs of electronic health records (EHRs), which may display missing data, unbalanced representation, and implicit selectivity in patient factors such as sex and gender<sup>24</sup>.

In agreement with the Findability, Accessibility, Interoperability, and Reusability (FAIR) recommendations for responsible research and gender equality<sup>64</sup>, biomedical Big Data requires innovative procedures for bias corrections<sup>65</sup>, including sex and gender bias, as well as algorithm interpretability<sup>66</sup> (see Valuable outputs of

**Table 2.** Source of undesirable bias in Artificial Intelligence with examples in health research and practice.

Source of bias in artificial intelligence	Description
Historical bias	Arises even if the data is perfectly measured and sampled, when the world as it is leads a model to produce outcomes that are not desired. e.g. incorrectly assuming that HIV is inherently linked to homosexual and bisexual men as its prevalence is higher in this population <sup>132</sup> .
Representation bias	Occurs when certain parts of the input space are underrepresented. e.g. European male populations are the primary focus in genomics research and its derived clinical findings, neglecting other ethnicities and populations <sup>133</sup> .
Measurement bias	Occurs when measured data are often proxies for some ideal features and labels. e.g. the use of clinical, social, and cognitive variables to detect the prodromal phase in schizophrenia and other psychotic disorders despite of observed sex differences in the expression of those symptoms and their associated risk for psychosis <sup>134</sup> .
Aggregation bias	Arises when a one-size-fits-all model is used for groups with different conditional distributions. e.g., for the diagnosis and monitoring of diabetes, haemoglobin A1c (HbA1c) levels are routinely used, despite of differences associated with ethnicities <sup>135</sup> and gender <sup>136</sup> .
Evaluation bias	Occurs when the evaluation and/or benchmark data for an algorithm does not represent the target population. e.g. underperformance of commercial facial recognition algorithm in dark-skinned female faces as most benchmark face image datasets come from white men <sup>137</sup> .
Algorithmic Bias	Occurs when bias is introduced in the algorithm consciously or unconsciously in ad-hoc solutions. e.g. by using health care cost as a proxy feature for health status without correcting for existing inequalities in health access, a commercial algorithm to predict health care needs was found to exhibit significant racial discrimination <sup>138</sup> .

health data), facing mounting pressure in data processing and privacy with the pursuit of “equal opportunity by design”<sup>67</sup>. Fair big data analytics will facilitate the identification of sex and gender differences in health as well as accurate indicators for prevention and diagnosis, and effective treatment.

#### Natural Language Processing

Natural Language Processing (NLP) consists of computational systems aimed at understanding and manipulating written and spoken human language for purposes like machine translation, speech recognition and conversational interfaces<sup>68</sup>.

In relation to biomedical research, NLP techniques allow processing of voice recordings and transcripts as well as large volumes of scientific knowledge accumulated in the textual forms, such as biomedical literature, electronic medical records, clinical trials and pathology reports. This automatic processing enables, for instance, the creation of major knowledge bases such as NDEx (<https://home.ndexbio.org/>), OncoKB (<https://oncokb.org/>), and Literome<sup>69</sup>.

As for Precision Medicine, these technologies allow to make predictions that can contribute to clinical decisions, such as diagnosis, prognosis, risk of relapse, and symptomatology fluctuations in response to treatments. Examples of applications of NLP to Precision Medicine comprise the identification of personalised drug combinations<sup>70</sup>, the knowledge-based curation of clinical significance of variants<sup>71</sup>, and patient trajectory modelling from clinical notes<sup>72</sup>. Activities to overcome some of the main challenges in NLP, such as complex semantics extraction and reasoning, entail automated curation efforts, such as Microsoft Project Hanover (<https://www.microsoft.com/en-us/research/project/project-hanover/>), and evaluation campaigns, such as BioCreative<sup>73</sup>.

The sex and gender dimension is crucial for the development of effective NLP solutions for health since multiple sex and gender differences have been documented in written and spoken language<sup>74</sup>. In fact, major differences are observed in dialogue structure<sup>75</sup>, word reading<sup>76</sup>, and even in children’s linguistic tasks<sup>77</sup>. Although the reasons for the differential use of language between men and women needs further investigations<sup>78</sup>, the existence of such differences can either facilitate or complicate the development of NLP technologies. For instance, while it is possible to accurately categorize texts based on the author’s gender<sup>79</sup>,

performances of sentiment analysis of male- and female-authored texts are extremely variable<sup>80</sup> and potentially biased<sup>81</sup>. Thus, knowing the sex and gender of the author enables a better targeted prediction of symptoms conveyed through natural language (text or speech). An example of this is the case of personalised healthcare for transgender and gender nonconforming patients based on EHRs analysis<sup>82</sup>.

In the context of NLP for voice recognition, the relevance of sex differences is evident in applications such as the prediction of suicidal behaviour<sup>83</sup>, especially considering the reported inconsistent and incomplete responses by popular conversational agents (Apple, Samsung, Microsoft) to suicidal ideation<sup>84</sup>.

A case of undesirable biases in NLP is the use of text corpora containing imprints of documented human stereotypes that can propagate into AI systems<sup>85</sup>. For instance, dense vector representations called word embeddings<sup>86</sup> are able to capture semantic relationships between words, such as sex, gender and ethnic relationships<sup>87</sup>, thus absorbing biases existing in the training corpus<sup>88</sup>. Methods for bias mitigation in NLP have been recently reviewed, including learning gender-neutral embeddings and tagging the data points to preserve the gender of the source<sup>89</sup>.

A flourishing area of NLP is that of medical chatbots, aiming to improve users’ wellbeing through real-time symptom assessment and recommendation interfaces. A dialogue of a chatbot can be modelled with available metadata to adjust to features of the replier in terms of gender, age, and mood<sup>90</sup>. In the context of mental health, medical chatbots include Woebot, which proved to relieve feelings of anxiety and depression<sup>91</sup>, and Moodkit, which recommends chatting and journaling activities through text and voice notes<sup>92</sup>. Although both proved to be effective in clinical trials, the lack of data on their long-term effects is raising certain concerns. These include the risk of oversimplifying mental conditions and therapeutic approaches, without considering potentially important factors such as sex and gender differences in non-verbal communication.

Of note, affective computing (i.e. passively estimating human emotional states in real-time) has started to be integrated in automated systems for educational and marketing purposes<sup>93</sup>, as well as voice-activated assistants for mental health support like Mindscape ([www.cultmindscape.com](http://www.cultmindscape.com)). In this regard, potential undesirable biases may undermine the automatic detection of sex-associated speech fluctuation in cognitive impairment<sup>94</sup>.

**Table 3.** Categories of digital biomarkers.

Category	Definition	Corresponding digital biomarker examples
Susceptibility and risk biomarker	A biomarker that indicates the potential for developing a disease or medical condition in an individual who does not currently have clinically apparent disease or medical condition.	<sup>a</sup> Detect cognitive changes in healthy subjects at risk of developing Alzheimer's disease using a video game platform <sup>139</sup> .
Diagnostic biomarker	A biomarker used to detect or confirm the presence of a disease or condition of interest or to identify individuals with a subtype of the disease.	<sup>a</sup> Diagnose ADHD in children using eye vergence metrics <sup>140</sup> . <sup>a</sup> Detect depression and Parkinson's disease using vocal biomarkers <sup>141</sup> . <sup>a</sup> Diagnose asthma and respiratory infections using smartphone-recorded cough sounds <sup>142</sup> .
Monitoring biomarker	A biomarker measured serially for assessing the status of a disease or medical condition or for evidence of exposure to (or effect of) a medical product or an environmental agent.	<sup>a</sup> Quantify Parkinson's disease severity using smartphones and machine learning <sup>143</sup> . <sup>b</sup> Track time and location of short-acting beta-agonist inhaler use through an attached wireless sensor <sup>144</sup> . <sup>a</sup> Detection of nocturnal scratching movements in patients with atopic dermatitis using accelerometers and recurrent neural networks <sup>145</sup> . <sup>b</sup> Measurements of sympathetic nervous impulses at the skin and inference of parasympathetic activity from heart rate variation to detect tonic-clonic epileptic seizures and immediately alert care providers <sup>146</sup> . <sup>b</sup> Portable electrocardiogram sensor associated to a smartphone app to monitor atrial fibrillation, bradycardia, tachycardia or normal heart rhythm and inform the clinician <sup>147</sup> . <sup>a</sup> Measure adherence in treatment of schizophrenia and bipolar disorder with an ingestible digital pill <sup>148</sup> .
Endpoint digital biomarkers in clinical trials	Endpoints generated by the use of mobile technologies in clinical setting.	<sup>a</sup> Accelerometer-derived motor abnormalities for use in Parkinson's disease <sup>47</sup> . <sup>b</sup> Monitoring of multiple sclerosis patients with digital technologies by using active and passive tests (ClinicalTrials.gov Identifiers: NCT03523858; NCT02952911) <sup>b</sup> Virtual Reality Functional Capacity Assessment Tool as co-primary and secondary endpoint in schizophrenia and major depressive disorder <sup>139</sup> .

<sup>a</sup>Digital biomarker under development (in feasibility/exploratory stages).  
<sup>b</sup>Digital biomarker in use in a clinical trial or an FDA cleared/approved digital health product, or a digital health app in use not requiring approval.

In the development and application of biomedical NLP systems, awareness of sex and gender differences is a crucial step in our understanding of women's and men's relative use of language, which could lead to a better patient management and more effective risk stratification.

### Robotics

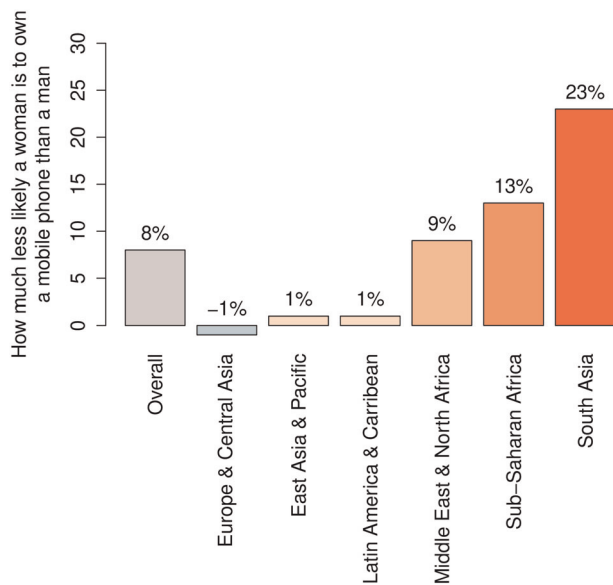
Robots can serve a diverse range of roles in improving a human's tasks, health and quality of life. In the context of Precision Medicine robots are expected to provide personalised assistance to patients according to their specific needs and preferences, at the right time and in the right way. Robotics for health are becoming increasingly impactful, in particular in neurology<sup>95</sup>, rehabilitation<sup>96</sup>, and assistive approaches for improving the quality of life of patients and caregivers<sup>97</sup>.

In a personalised robot-patient interaction both the gender of the patient and the "gender" of the robot have to be taken into account. While there is not a lot of research on how to personalise the behaviour of a robot (e.g. speech style) to an individual's gender, several studies explored how the gendered appearance of a robot differentially affects human-robot interactions. For instance, a recent study revealed sex differences in how children interact with robots<sup>98</sup> with implications for their use in paediatric hospitalization<sup>99</sup>.

The application of robots in human society makes the discussion on humanoids' gender extremely relevant and significantly variable across cultures<sup>100,101</sup>. While some robots are genderless, such as Pepper (Softbank), ASIMO (Honda), and Ripley (MIT), others are designed to display explicit gendered features, such as the females Sophia (Hanson Robotics), Sarah the FaceBot<sup>102</sup>, and male Ibn Sina Robot, a culture-specific historical humanoid<sup>103</sup>. This opened a strong debate regarding the commonalities among humans and robots on physical, sociological and psychological gender<sup>104</sup>.

It has been demonstrated that the outcome of a humanoid robot's task can be affected by its gender, as in the case of female charity robots receiving more donations from men in comparison to women<sup>100</sup>. Indeed, the fact that the traits of a gendered robot are developed in accordance with the perceived gender role of both the developer and the final user, could emphasize social constructs and stereotypes. Gender representation in robots should evade social stereotypes and serve functionally human-robot interactions<sup>102</sup>. An illustrative effort towards gender neutrality in robotics is the creation of a genderless digital voice (<https://www.genderlessvoice.com/>), designed using a gender-neutral frequency range (145–175 Hz).

Awareness of sex and gender differences in patients and in robots could lead to a better healthcare assistance and effective human-machine interactions for biomedical applications as well as a better translation of ethical decision-making into machines<sup>105</sup>.



**Fig. 3 The digital divide in access to mobile technology around the globe.** The bar plot reports how less likely a woman is to own a mobile phone than a man, according to a survey analysis on mobile ownership conducted by the Global System for Mobile Communications Association (GSMA) in low- and middle-income countries (LMIC) in 2019, by geographical area (source: GSMA “The Mobile Gender Gap Report 2020”<sup>51</sup>). For instance, in South Asia women are 23% less likely than men to be the owner of a mobile phone, while in Europe and Central Asia women are 1% more likely to be the owner of a mobile phone. Across LMICs (“Overall”), women are 8% less likely than men to own a mobile phone.

## VALUABLE OUTPUTS OF HEALTH TECHNOLOGIES

### Towards explainable artificial intelligence

In the context of Precision Medicine, the expected outputs of AI models consist of predictions of risk and diagnosis of medical conditions or recommendations of treatments, with profound influence in people’s lives and health.

Despite the progress of AI models in recent years, the complexity of their internal structures has led to a major technological issue termed the ‘Black box’ problem. It refers to the lack of explicit declarative knowledge representations in machine learning models<sup>106</sup>, meaning their inability to provide a layman-understandable explanation and/or interpretation to respond to “how” or “why” questions regarding their output.

Getting an explicable justification of how and why these AI models reach their conclusions is now becoming more and more crucial since there is an increasing need to understand the specific parameters used to draw clinical conclusions with relevant impact on patients’ lives. Indeed, the EU directive 2016/680 General Data Protection Regulation (GDPR) states the “right to an explanation” about the output of an algorithm<sup>107</sup>.

In regards to the scope of this review, explainability in AI would help justify algorithms’ clinical predictions and recommendations when they are differential for patients with different sex and genders. On one hand, an explanation of the decisional process would enable to find potential mistaken conclusions derived by training an algorithm with misrepresented data. This will facilitate the identification of undesirable biases generally found in clinical data with unbalanced sex and gender representation. On the other hand, an explanation of the decisional processes will help the discovery of sex and gender differences in clinical data that is representative, therefore promoting the desired biases for personalised preventative and therapeutic interventions.

Different features such as interpretability and completeness (see Supplementary Note 2 “Explainable Artificial Intelligence”) in

AI have been established as explainability requirements to contribute to relevant aspects of general medicine such as confidence, safety, security, privacy, ethics, fairness and trust.

The term explainable artificial intelligence (XAI) is used to refer to algorithms that are able to meet those requirements. XAI is a relatively young field of research and their applications so far have not been particularly involved with sex and gender differences.

An example of XAI is a recent study where a machine learning algorithm made referral recommendations on dozens of retinal diseases, highlighting the specific structures in optical tomography scans that could lead to ambiguous interpretation<sup>108</sup>. Another example is a deep learning model for predicting cardiovascular risk factors based on images of the retina, indicating which anatomical features, such as the optic disc or blood vessels, were used to generate the predictions<sup>109</sup>. XAI is also useful in basic research, for instance, efforts in creating “visible” deep neural networks that provide automatic explanations of the impact of a genotypic change on cellular phenotypic states<sup>110</sup>.

XAI represents a promising technology to assist in the identification of sex and gender differences in health and disease, and to dissociate the underlying sources from biased datasets or social inequalities.

### Bias detection frameworks for fairness

One of the main challenges to develop trustworthy AI is to define the meaning of fairness in the practice of machine learning<sup>111</sup>. Indeed, many approaches have been proposed to achieve fair algorithmic decision-making, some of which not always meet the expected outcome.

For instance, a widely used approach to ensure fairness in data processing is to remove some sensitive information, such as sex or gender, and all other possible correlated features<sup>112</sup>. However, if inherent differences exist in the underlying population, such as sex differences in disease prevalence, this procedure is undesirable as the outcome would be less fair towards specific minorities. Indeed, the learned patterns that apply to the majority group might be invalid for a minority one.

On the contrary, the explicit use of sex and gender information enables to reach an outcome that is fairer towards minorities, which is a desirable procedure when inherent differences exist. A theoretical implementation of such approach, also called fair affirmative action, has been proposed as an optimisation problem to obtain, at the same time, both group fairness (a.k.a statistical parity) and individual fairness<sup>113</sup>.

Although affirmative action represents a remedy for unfair algorithmic discrimination, ensuring the quality of the data used for algorithm training is also crucial. For instance, a study found that only 17% of cardiologists correctly identified women as having greater risk for heart disease than men<sup>114</sup>. Indeed, physicians are typically trained to recognise patterns of angina and myocardial infarction that occur more frequently in men, resulting in women being typically under-diagnosed for coronary artery disease<sup>115</sup>. Consequently, training an algorithm on available data on diagnosed cases could be influenced by an implicit sex and gender bias.

Fairness is highly context-specific and requires an understanding of the classification task and possible minorities. Awareness and deep knowledge of sex and gender differences as well as the related socio-economical aspects and possible confounding factors are of paramount importance to establish fairness in algorithmic development.

The development and application of fair approaches will be critical for the implementation of unbiased and interpretable models for Precision Medicine<sup>106,116</sup>. In this regard, the use of visualizations, logical statements, and dimensionality reduction techniques can be implemented in computational tools to achieve interpretability<sup>23</sup>.

Mitigating undesirable bias to achieve fairness might require an explicit instruction to the artificial learning engine including rules of appropriate conduct, as proposed in the domain of cognitive robotics<sup>117</sup>. In addition, caution should be used particularly with the unsupervised learning components of AI given the wide availability of biased data sets and self-learning algorithms. Recent developments in bias detection and mitigation also include methods such as adopting re-sampling<sup>118</sup>, adversarial learning<sup>119</sup>, and open-source toolkits such as IBM AI Fairness 360 (AIF360) (aif360.mybluemix.net) and Aequitas (dsapp.uchicago.edu/projects/aequitas).

## DISCUSSION

Technological advances in machine learning and AI are transforming our health systems, societies, and daily lives<sup>120</sup>. In the context of biomedicine, such systems can sometimes either neglect desired differentiations, such as sex and gender, or amplify undesired ones, such as reinforcing existing socio-cultural discriminations that promote inequalities.

The ambitious goals set by Precision Medicine will be achieved using the latest advances in AI to properly identify the role of inter-individual differences. This will include the impact of sex and gender in health and disease, as well as eradicating existing undesirable sex and gender biases from data sets, algorithms and experimental design. The proper use of innovative technologies will pave the way towards tailored and personalised disease prevention and treatment, accounting for sex and gender differences and extending towards generalized wellbeing. Actions that foster the effective utilization of AI systems will not only enable the acceleration towards Precision Medicine, but most importantly, will significantly contribute to the improvement of the quality of life of patients of all sexes and genders.

Ethical standards will have to continue to be considered by governments and regulatory organisations to guarantee the preservation of personal data privacy and security as well as to determine the way new technological tools should be employed, data should be collected, and models improved<sup>121,122</sup>. Governments and regulatory organisations are establishing the guidelines for actions in this direction, such as the case of AI-WATCH (<https://ec.europa.eu/knowledge4policy/ai-watch>), an initiative of the European Commission to monitor the socio-economic, legal and ethical impact of AI and robotics.

Based on the information surveyed in this work, we provide the following recommendations to ensure that sex and gender differences in health and disease are accounted for in AI implementations that inform Precision Medicine:

1. Distinguish between desirable and undesirable biases and guarantee the representation of desirable biases in AI development (see Introduction: Desirable vs. Undesirable biases).
2. Increase awareness of unintended biases in the scientific community, technology industry, among policy makers, and the general public (see Sources and types of Health data and Technologies for the analysis and deployment of Health data).
3. Implement explainable algorithms, which not only provide understandable explanations for the layperson, but which could also be equipped with integrated bias detection systems and mitigation strategies, and validated with appropriate benchmarking (see Valuable outputs of Health technologies).
4. Incorporate key ethical considerations during every stage of technological development, ensuring that the systems maximize wellbeing and health of the population (see Discussion).

## DATA AVAILABILITY

No datasets were generated or analyzed during the current study.

Received: 18 July 2019; Accepted: 28 April 2020;

Published online: 01 June 2020

## REFERENCES

1. Ginsburg, G. S. & Phillips, K. A. Precision Medicine: from science to value. *Health Aff.* **37**, 694–701 (2018).
2. Regitz-Zagrosek, V. Sex and gender differences in health. Science & society series on sex and science. *EMBO Rep.* **13**, 596–603 (2012).
3. Ferretti, M. T. et al. Sex differences in Alzheimer disease - the gateway to precision medicine. *Nat. Rev. Neurol.* **14**, 457–469 (2018).
4. Kuehner, C. Why is depression more common among women than among men? *Lancet Psychiatry* **4**, 146–158 (2017).
5. Kim, H.-I., Lim, H. & Moon, A. Sex differences in cancer: epidemiology, genetics and therapy. *Biomol. Ther.* **26**, 335–342 (2018).
6. Natri, H., Garcia, A. R., Buetow, K. H., Trumble, B. C. & Wilson, M. A. The pregnancy pickle: evolved immune compensation due to pregnancy underlies sex differences in human diseases. *Trends Genet.* **35**, 478–488 (2019).
7. Guggenmos, M. et al. Quantitative neurobiological evidence for accelerated brain aging in alcohol dependence. *Transl. Psychiatry* **7**, 1279 (2017).
8. Dance, A. Why the sexes don't feel pain the same way. *Nature* **567**, 448 (2019).
9. Linn, L., Oliel, S. & Baldwin, A. Women and men face different chronic disease risks. PAHO/WHO. [https://www.paho.org/hq/index.php?option=com\\_content&view=article&id=5080:2011-women-men-face-different-chronic-disease-risks&Itemid=135&lang=en](https://www.paho.org/hq/index.php?option=com_content&view=article&id=5080:2011-women-men-face-different-chronic-disease-risks&Itemid=135&lang=en) (2011).
10. Vari, R. et al. Gender-related differences in lifestyle may affect health status. *Ann. Dell'Istituto Super. Sanità.* [https://doi.org/10.4415/ANN\\_16\\_02\\_06](https://doi.org/10.4415/ANN_16_02_06) (2016).
11. Torres-Rojas, C. & Jones, B. C. Sex differences in neurotoxicogenetics. *Front. Genet.* **9**, 196 (2018).
12. Jones, T. Intersex studies: a systematic review of international health literature. *SAGE Open* **8**, 215824401774557 (2018).
13. Scandurra, C. et al. Health of non-binary and genderqueer people: a systematic review. *Front. Psychol.* **10**, 1453 (2019).
14. Marshall, Z. et al. Documenting research with transgender, nonbinary, and other gender diverse (Trans) individuals and communities: introducing the global trans research evidence map. *Transgender Health* **4**, 68–80 (2019).
15. Ensuring the Health Care Needs of Women: A Checklist for Health Exchanges. *The Henry J. Kaiser Family Foundation.* <https://www.kff.org/womens-health-policy/issue-brief/ensuring-the-health-care-needs-of-women-a-checklist-for-health-exchanges/> (2013).
16. Shansky, R. M. Are hormones a “female problem” for animal research? *Science* **364**, 825–826 (2019).
17. Rich-Edwards, J. W., Kaiser, U. B., Chen, G. L., Manson, J. E. & Goldstein, J. M. Sex and gender differences research design for basic, clinical, and population studies: essentials for investigators. *Endocr. Rev.* **39**, 424–439 (2018).
18. Eliot, L. Neurosexism: the myth that men and women have different brains. *Nature* **566**, 453–454 (2019).
19. Ferretti, M. T., Santuccione-Chadha, A. & Hampel, H. Account for sex in brain research for precision medicine. *Nature* **569**, 40–40 (2019).
20. Hay, K. et al. Disrupting gender norms in health systems: making the case for change. *Lancet* S0140673619306488. [https://doi.org/10.1016/S0140-6736\(19\)30648-8](https://doi.org/10.1016/S0140-6736(19)30648-8) (2019).
21. Martin, L. A., Neighbors, H. W. & Griffith, D. M. The experience of symptoms of depression in men vs women: analysis of the National Comorbidity Survey Replication. *JAMA Psychiatry* **70**, 1100–1106 (2013).
22. Mental health aspects of women's reproductive health: a global review of the literature. (World Health Organization, 2009).
23. Tjoa, E. & Guan, C. A survey on explainable artificial intelligence (XAI): towards medical XAI. Preprint at <https://arxiv.org/abs/1907.07374> (2019).
24. McGregor, A. J. et al. How to study the impact of sex and gender in medical research: a review of resources. *Biol. Sex. Differ.* **7**, 46 (2016).
25. Yoon, D. Y. et al. Sex bias exists in basic science and translational surgical research. *Surgery* **156**, 508–516 (2014).
26. Karp, N. A. et al. Prevalence of sexual dimorphism in mammalian phenotypic traits. *Nat. Commun.* **8**, 15475 (2017).
27. Holdcroft, A. Gender bias in research: how does it affect evidence based medicine? *J. R. Soc. Med.* **100**, 2–3 (2007).
28. Clayton, J. A. Studying both sexes: a guiding principle for biomedicine. *FASEB J.* **30**, 519–524 (2016).



29. Melloni, C. et al. Representation of women in randomized clinical trials of cardiovascular disease prevention. *Circ. Cardiovasc. Qual. Outcomes* **3**, 135–142 (2010).
30. Geller, S. E. et al. The More Things Change, the More They Stay the Same: A Study to Evaluate Compliance With Inclusion and Assessment of Women and Minorities in Randomized Controlled Trials. *Acad. Med. J. Assoc. Am. Med. Coll.* **93**, 630–635 (2018).
31. Raz, L. & Miller, V. M. Considerations of sex and gender differences in preclinical and clinical trials. *Handb. Exp. Pharmacol.* 127–147. [https://doi.org/10.1007/978-3-642-30726-3\\_7](https://doi.org/10.1007/978-3-642-30726-3_7) (2012).
32. McGregor, A. J. Sex bias in drug research: a call for change. *Pharmaceutical J.* <https://www.pharmaceutical-journal.com/opinion/comment/sex-bias-in-drug-research-a-call-for-change/20200727.article> (2016).
33. Tower, J. Sex-specific gene expression and life span regulation. *Trends Endocrinol. Metab.* **28**, 735–747 (2017).
34. Tharpe, N. Adverse drug reactions in women's health care. *J. Midwifery Women's Health* **56**, 205–213 (2011).
35. Simon, V. Wanted: women in clinical trials. *Science* **308**, 1517–1517 (2005).
36. Light, K. P., Lovell, A. T., Butt, H., Fauvel, N. J. & Holdcroft, A. Adverse effects of neuromuscular blocking agents based on yellow card reporting in the U.K.: are there differences between males and females? *Pharmacoepidemiol. Drug Saf.* **15**, 151–160 (2006).
37. Oertelt-Prigione, S. The influence of sex and gender on the immune response. *Autoimmun. Rev.* **11**, A479–485 (2012).
38. Norman, J. L., Fixen, D. R., Saseen, J. J., Saba, L. M. & Linnebur, S. A. Zolpidem prescribing practices before and after Food and Drug Administration required product labeling changes. *SAGE Open Med.* **5**, 205031211770768 (2017).
39. Franconi, F. & Campesi, I. Pharmacogenomics, pharmacokinetics and pharmacodynamics: interaction with biological differences between men and women: pharmacological differences between sexes. *Br. J. Pharm.* **171**, 580–594 (2014).
40. Miller, V. M., Rocca, W. A. & Faubion, S. S. Sex differences research, precision medicine, and the future of women's health. *J. Women's Health* **2002** **24**, 969–971 (2015).
41. Schork, N. J. Personalized medicine: time for one-person trials. *Nature* **520**, 609–611 (2015).
42. Coravos, A., Khozin, S. & Mandl, K. D. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *Npj Digit. Med.* **2**, 14 (2019).
43. Sperling, R., Mormino, E. & Johnson, K. The evolution of preclinical Alzheimer's disease: implications for prevention trials. *Neuron* **84**, 608–622 (2014).
44. Kourtis, L. C., Regele, O. B., Wright, J. M. & Jones, G. B. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. *Npj Digit. Med.* **2**, 9 (2019).
45. Koran, M. E. I., Wagener, M., Hohman, T. J. & Alzheimer's Neuroimaging Initiative. Sex differences in the association between AD biomarkers and cognitive decline. *Brain Imaging Behav.* **11**, 205–213 (2017).
46. Snyder, C. W., Dorsey, E. R. & Atreja, A. The best digital biomarkers papers of 2017. *Digit. Biomark.* **2**, 64–73 (2018).
47. Lipsmeier, F. et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial: remote PD testing with smartphones. *Mov. Disord.* **33**, 1287–1297 (2018).
48. Miller, I. N. & Cronin-Golomb, A. Gender differences in Parkinson's disease: clinical characteristics and cognition: gender differences in Parkinson's disease. *Mov. Disord.* **25**, 2695–2703 (2010).
49. Feiner, J. R., Severinghaus, J. W. & Bickler, P. E. Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender. *Anesth. Analg.* **105**, S18–23 (2007).
50. Reid, A. J. *The Smartphone Paradox: Our Ruinous Dependency in the Device Age* (Springer International Publishing, 2018).
51. Rowntree, O. et al. GSMA The Mobile Gender Gap Report. <https://www.gsma.com/r/gender-gap/> (2020).
52. Fan, W. & Bifet, A. Mining big data: current status, and forecast to the future. *SIGKDD Explor. Newsl.* **14**, 1–5 (2013).
53. Durinx, C. et al. Identifying ELIXIR core data resources. *F1000Research* **5**, ELIXIR–2422 (2016).
54. Bourne, P. E. et al. The NIH big data to knowledge (BD2K) initiative. *J. Am. Med. Assoc.* **22**, 1114–1114 (2015).
55. Matoba, N. et al. GWAS of smoking behaviour in 165,436 Japanese people reveals seven new loci and shared genetic architecture. *Nat. Hum. Behav.* **3**, 471–477 (2019).
56. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990–2015. A systematic analysis from the Global Burden of Disease Study 2015. *Lancet Lond. Engl.* **389**, 1885–1906 (2017).
57. Zore, T., Palafox, M. & Reue, K. Sex differences in obesity, lipid metabolism, and inflammation—A role for the sex chromosomes? *Mol. Metab.* **15**, 35–44 (2018).
58. Timmins, K. A., Green, M. A., Radley, D., Morris, M. A. & Pearce, J. How has big data contributed to obesity research? A review of the literature. *Int. J. Obes.* **42**, 1951–1962 (2018).
59. Frias-Martinez, V., Frias-Martinez, E. & Oliver, N. A gender-centric analysis of calling behavior in a developing economy using call detail records. AAAI Spring Symposium Series, North America. <https://www.aaai.org/ocs/index.php/SSS/SSS10/paper/view/1094/1347> (2010).
60. De Choudhury, M., Sharma, S. S., Logar, T., Eekhout, W. & Nielsen, R. C. Gender and cross-cultural differences in social media disclosures of mental illness. In *Proc. 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. (eds Poltrock S. & Lee C. P.) 353–369 (Association for Computing Machinery, Portland, OR, USA, 2017).
61. Calvo, M. et al. In *Sex-Specific Analysis of Cardiovascular Function*. Vol. 1065 (eds Kerkhof, P. L. M. & Miller, V. M.) 181–190 (Springer International Publishing, 2018).
62. Parikh, N. I. et al. Reproductive Risk Factors and Coronary Heart Disease in the Women's Health Initiative Observational Study. *Circulation* **133**, 2149–2158 (2016).
63. Wang, W. & Krishnan, E. Big data and clinicians: a review on the state of the science. *JMIR Med. Inform.* **2**, e1 (2014).
64. European Commission. Turning FAIR data into reality. (Publications Office of the European Union, 2018).
65. Harford, T. Big data: a big mistake? *Significance* **11**, 14–19 (2014).
66. Price, W. N. Big data and black-box medical algorithms. *Sci. Transl. Med.* **10**, eaao5333 (2018).
67. Podesta, J., Pritzker, P., Moniz, E. J., Holdren, J. & Zients, J. *Big Data: Seizing Opportunities, Preserving Values*. (White House, Washington DC, 2014).
68. Liddy. Natural Language Processing. In *Encyclopedia of Library and Information Science* (Marcel Decker, Inc., NY, 2001).
69. Poon, H., Quirk, C., DeZiel, C. & Heckerman, D. Literome: PubMed-scale genomic knowledge base in the cloud. *Bioinformatics* **30**, 2840–2842 (2014).
70. Sutherland, J. J. et al. Co-prescription trends in a large cohort of subjects predict substantial drug-drug interactions. *PLOS ONE* **10**, e0118991 (2015).
71. Lee, K. et al. Scaling up data curation using deep learning: an application to literature triage in genomic variation resources. *PLOS Comput. Biol.* **14**, e1006390 (2018).
72. Miotto, R., Li, L., Kidd, B. A. & Dudley, J. T. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci. Rep.* **6**, 26094 (2016).
73. Hirschman, L., Yeh, A., Blaschke, C. & Valencia, A. Overview of BioCreAtivE: critical assessment of information extraction for biology. *BMC Bioinforma.* **6**, S1 (2005).
74. Larson, B. Gender as a Variable in Natural-Language Processing: Ethical Considerations. In *Proc. First ACL Workshop on Ethics in Natural Language Processing*. (eds Hovy, D., Spruit, S., Mitchell, M., Bender, E. M., Strube, M., Wallach, H.) 1–11 (Association for Computational Linguistics, Valencia, Spain, 2017).
75. Garimella, A., Banea, C., Hovy, D. & Mihalcea, R. Women's Syntactic Resilience and Men's Grammatical Luck: Gender-Bias in Part-of-Speech Tagging and Dependency Parsing. In *Proc. 57th Annual Meeting of the Association for Computational Linguistics*. (eds Hovy, D., Spruit, S., Mitchell, M., Bender, E. M., Strube, M., Wallach, H.) 3493–3498 (Association for Computational Linguistics, Valencia, Spain, 2019).
76. Wirth, M. et al. Sex differences in semantic processing: event-related brain potentials distinguish between lower and higher order semantic analysis during word reading. *Cereb. Cortex* **17**, 1987–1997 (2007).
77. Burman, D. D., Bitan, T. & Booth, J. R. Sex differences in neural processing of language among children. *Neuropsychologia* **46**, 1349–1362 (2008).
78. Newman, M. L., Groom, C. J., Handelman, L. D. & Pennebaker, J. W. Gender differences in language use: an analysis of 14,000 text samples. *Discourse Process.* **45**, 211–236 (2008).
79. Koppel, M. Automatically categorizing written texts by author gender. *Lit. Linguist. Comput* **17**, 401–412 (2002).
80. Thelwall, M. Gender bias in sentiment analysis. *Online Inf. Rev.* **42**, 45–57 (2018).
81. Kiritchenko, S. & Mohammad, S. M. Examining gender and race bias in two hundred sentiment analysis systems. In *Proc. Seventh Joint Conference on Lexical and Computational Semantics*. (eds Nissim, M., Berant, J., Lenci, A.) 518–2005 (Association for Computational Linguistics, New Orleans, Louisiana, USA, 2018).
82. Burgess, C., Kauth, M. R., Klemm, S., Shanawani, H. & Shipherd, J. C. Evolving sex and gender in electronic health records. *Fed. Pract. Health Care Prof. VA DoD. PHS* **36**, 271–277 (2019).
83. Oquendo, M. A. et al. Sex differences in clinical predictors of suicidal acts after major depression: a prospective study. *Am. J. Psychiatry* **164**, 134–141 (2007).
84. Miner, A. S. et al. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Intern. Med.* **176**, 619–625 (2016).

85. Stubbs, M. *Text and Corpus Analysis: Computer-assisted Studies of Language and Culture*. (Blackwell Publishers, 1996).
86. Mikolov, T., Yih, W. & Zweig, G. Linguistic regularities in continuous space word representations. In *Proc. 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (eds Vanderwende, L., Daumé, H. III., Kirchhoff, K.) 746–751 (Association for Computational Linguistics, Atlanta, Georgia, USA, 2013).
87. Garg, N., Schiebinger, L., Jurafsky, D. & Zou, J. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proc. Natl Acad. Sci. USA* **115**, E3635–E3644 (2018).
88. Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V. & Kalai, A. Man is to computer programmer as woman is to homemaker? Debiasing Word Embeddings. In *Proc. 30th International Conference on Neural Information Processing Systems*. (eds Lee, D. D., Luxburg, U. V., Garnett, R., Sugiyama, M., Guyon, I. M.) 4356–4364 (NIPS, Barcelona, Spain, 2016).
89. Sun, T. et al. Mitigating gender bias in natural language processing: literature review. In *Proc. 57th Annual Meeting of the Association for Computational Linguistics* (eds Nakov, P., Palmer, A.) 1630–1640 (Association for Computational Linguistics, Florence, Italy, 2019).
90. Jiwei Li, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, Bill Dolan. A Persona-Based Neural Conversation Model. In *Proc. 54th Annual Meeting of the Association for Computational Linguistics*. Vol 1: Long Papers (eds Erk, K., Smith, N. A.) 994–1003 (Association for Computational Linguistics, Berlin, Germany, 2016).
91. Fitzpatrick, K. K., Darcy, A. & Vierhile, M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Ment. Health* **4**, e19 (2017).
92. Bakker, D., Kazantzis, N., Rickwood, D. & Rickard, N. A randomized controlled trial of three smartphone apps for enhancing public mental health. *Behav. Res. Ther.* **109**, 75–83 (2018).
93. Calvo, R. A. & D’Mello, S. Frontiers of affect-aware learning technologies. *IEEE Intell. Syst.* **27**, 86–89 (2012).
94. Mirheidari, B., Blackburn, D., Walker, T., Reuber, M. & Christensen, H. Dementia detection using automatic analysis of conversations. *Comput. Speech Lang.* **53**, 65–79 (2019).
95. Kim, G. H. et al. Structural brain changes after traditional and robot-assisted multi-domain cognitive training in community-dwelling healthy elderly. *PLoS ONE* **10**, e0123251 (2015).
96. Volpe, B. T. et al. Intensive sensorimotor arm training mediated by therapist or robot improves hemiparesis in patients with chronic stroke. *Neurorehabil. Neural Repair* **22**, 305–310 (2008).
97. Khan, A. & Anwar, Y. In *Advances in Computer Vision*. Vol. 944 (eds Arai, K. & Kapoor, S.) 280–292 (Springer International Publishing, 2020).
98. Kory-Westlund, J. M. & Breazeal, C. A Persona-Based Neural Conversation Model. In *Proc. 18th ACM Interaction Design and Children Conference (IDC)*. (ed. Fails, J. A.) 38–50, (ACM Press, Boise, Idaho, US, 2019).
99. Logan, D. E. et al. Social robots for hospitalized children. *Pediatrics* **144**, e20181511 (2019).
100. Robertson, J. Gendering humanoid robots: robo-sexism in Japan. *Body Soc.* **16**, 1–36 (2010).
101. Mavridis, N. et al. Opinions and attitudes toward humanoid robots in the Middle East. *AI Soc.* **27**, 517–534 (2012).
102. Mavridis, N. et al. FaceBots: Robots utilizing and publishing social information in Facebook. In *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* 273–274 (2009).
103. Riek, L. D. & Ahmed, Z. Ibn Sina Steps Out: Exploring Arabic Attitudes Toward Humanoid Robots. *Proc. 2nd Int. Symp. New Front. Human-robot Interact. AISB Leic.* Vol. 1, (2010).
104. Soraa, R. A. Mechanical genders: how do humans gender robots? *Gend. Technol. Dev.* **21**, 99–115 (2017).
105. Deng, B. Machine ethics: the robot’s dilemma. *Nature* **523**, 24–26 (2015).
106. Holzinger, A., Biemann, C., Pattichis, C. S. & Kell, D. B. What do we need to build explainable AI systems for the medical domain? Preprint at: <https://arxiv.org/abs/1712.09923> (2017).
107. Towards trustable machine learning. *Nat. Biomed. Eng.* **2**, 709–710. <https://www.nature.com/articles/s41551-018-0315-x> (2018).
108. De Fauw, J. et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **24**, 1342–1350 (2018).
109. Poplin, R. et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2**, 158–164 (2018).
110. Ma, J. et al. Using deep learning to model the hierarchical structure and function of a cell. *Nat. Methods* **15**, 290–298 (2018).
111. Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G. & Chin, M. H. Ensuring fairness in machine learning to advance health equity. *Ann. Intern. Med.* **169**, 866–872 (2018).
112. Zemel, R., Wu, Y., Swersky, K., Pitassi, T. & Dwork, C. Learning fair representations. In *Proc. 30th International Conference on International Conference on Machine Learning*. Vol 28 III–325–III–333 (eds Dasgupta, S. & McAllester, D.) (JMLR.org, Atlanta, Georgia, USA, 2013).
113. Dwork, C., Hardt, M., Pitassi, T., Reingold, O. & Zemel, R. Fairness through awareness. In *Proc. 3rd Innovations in Theoretical Computer Science Conference on - ITCS '12* 214–226 (eds Dasgupta, S. & McAllester, D.) (ACM Press, Atlanta, Georgia, USA, 2012).
114. Mosca, L. et al. National Study of Physician Awareness and Adherence to Cardiovascular Disease Prevention Guidelines. *Circulation* **111**, 499–510 (2005).
115. Daugherty, S. L. et al. Implicit gender bias and the use of cardiovascular tests among cardiologists. *J. Am. Heart Assoc.* **6**, e006872 (2017).
116. Hamburg, M. A. & Collins, F. S. The path to personalized medicine. *N. Engl. J. Med.* **363**, 301–304 (2010).
117. Hanheide, M. et al. Robot task planning and explanation in open and uncertain worlds. *Artif. Intell.* **247**, 119–150 (2017).
118. Amini, A., Soleimany, A., Schwarting, W., Bhatia, S. & Rus, D. Uncovering and mitigating algorithmic bias through learned latent structure. In *Proc. 2019 AAAI/ACM Conference on AI, Ethics, and Society*. (eds Conitzer, V., Hadfield, G. & Vallor, S.) 289–295 (Association for Computing Machinery, Honolulu, HI, USA, 2019).
119. Zhang, B. H., Lemoine, B. & Mitchell, M. Mitigating Unwanted Biases with Adversarial Learning. In *Proc. 2018 AAAI/ACM Conference on AI, Ethics, and Society*. (eds Furman, J., Marchant, G., Price, H. & Rossi, F.) 335–340 (Association for Computing Machinery, New Orleans, LA, USA, 2018).
120. Iacobacci, N. *Exponential Ethics*. (ATROPOS PRESS, 2018).
121. Can, A. I. Help reduce disparities in general medical and mental health care? *AMA J. Ethics* **21**, E167–E179 (2019).
122. Suresh, H. & Gutttag, J. V. A framework for understanding unintended consequences of machine learning. Preprint at <https://arxiv.org/abs/1901.10002> (2019).
123. Werling, D. M. & Geschwind, D. H. Sex differences in autism spectrum disorders. *Curr. Opin. Neurol.* **26**, 146–153 (2013).
124. Stock, E. O. & Redberg, R. Cardiovascular disease in women. *Curr. Probl. Cardiol.* **37**, 450–526 (2012).
125. Xhyheri, B. & Bugiardini, R. Diagnosis and treatment of heart disease: are women different from men? *Prog. Cardiovasc. Dis.* **53**, 227–236 (2010).
126. Dhruva, S. S., Bero, L. A. & Redberg, R. F. Gender bias in studies for food and drug administration premarket approval of cardiovascular devices. *Circ. Cardiovasc. Qual. Outcomes* **4**, 165–171 (2011).
127. Whose genomics? *Nat. Hum. Behav.* **3**, 409. <https://www.nature.com/articles/s41562-019-0619-1> (2019).
128. Khramtsova, E. A., Davis, L. K. & Stranger, B. E. The role of sex in the genomics of human complex traits. *Nat. Rev. Genet.* **20**, 173–190 (2019).
129. Coakley, M. et al. Dialogues on diversifying clinical trials: successful strategies for engaging women and minorities in clinical trials. *J. Women’s Health* **21**, 713–716 (2012).
130. Squires, K. et al. Insights on GRACE (Gender, Race, And Clinical Experience) from the Patient’s Perspective: GRACE Participant Survey. *AIDS Patient Care STDs* **27**, 352–362 (2013).
131. Schott, A. F., Welch, J. J., Verschraegen, C. F. & Kurzrock, R. The national clinical trials network: conducting successful clinical trials of new therapies for rare cancers. *Semin. Oncol.* **42**, 731–739 (2015).
132. Centers for Disease Control and Prevention. HIV Surveillance Report, 2017. **29**, (2018).
133. Bentley, A. R., Callier, S. & Rotimi, C. N. Diversity and inclusion in genomic research: why the uneven progress? *J. Community Genet.* **8**, 255–266 (2017).
134. Barajas, A., Ochoa, S., Obiols, J. E. & Lalucat-Jo, L. Gender differences in individuals at high-risk of psychosis: a comprehensive literature review. *Sci. World J.* **2015**, 1–13 (2015).
135. Cavagnoli, G., Pimentel, A. L., Freitas, P. A. C., Gross, J. L. & Camargo, J. L. Effect of ethnicity on HbA1c levels in individuals without diabetes: systematic review and meta-analysis. *PLoS ONE* **12**, e0171315 (2017).
136. Bae, J. C. et al. Hemoglobin A1c values are affected by hemoglobin level and gender in non-anemic Koreans. *J. Diabetes Investig.* **5**, 60–65 (2014).
137. Buolamwini, J. & Geburu, T. Gender shades: intersectional accuracy disparities in commercial gender classification. *Conf. Fairness, Accountability Transparency* **81**, 77–91 (2018).
138. Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453 (2019).

139. Gold, M. et al. Digital technologies as biomarkers, clinical outcomes assessment, and recruitment tools in Alzheimer's disease clinical trials. *Alzheimers Dement. Transl. Res. Clin. Inter.* **4**, 234–242 (2018).
140. Varela Casal, P. et al. Clinical validation of eye vergence as an objective marker for diagnosis of ADHD in children. *J. Atten. Disord.* **23**, 599–614 (2019).
141. Ghosh, S. S., Ciccarelli, G., Quatieri, T. F. & Klein, A. Speaking one's mind: vocal biomarkers of depression and Parkinson disease. *J. Acoust. Soc. Am.* **139**, 2193–2193 (2016).
142. Diagnosing respiratory disease in children using cough sounds 2 - ClinicalTrials.gov. <https://clinicaltrials.gov/ct2/show/NCT03392363> (2018).
143. Zhan, A. et al. Using smartphones and machine learning to quantify Parkinson disease severity: the mobile Parkinson disease score. *JAMA Neurol.* **75**, 876–880 (2018).
144. Barrett, M. A. et al. Effect of a mobile health, sensor-driven asthma management platform on asthma control. *Ann. Allergy Asthma Immunol.* **119**, 415–421.e1 (2017).
145. Moreau, A. et al. Detection of nocturnal scratching movements in patients with atopic dermatitis using accelerometers and recurrent neural networks. *IEEE J. Biomed. Health Inf.* **22**, 1011–1018 (2018).
146. Picard, R. W. Improvement of a convulsive seizure detector relying on accelerometer and electrodermal activity collected continuously by a wristband. *MIT Media Lab.* <https://www.media.mit.edu/publications/improvement-of-a-convulsive-seizure-detector-relying-on-accelerometer-and-electrodermal-activity-collected-continuously-by-a-wristband/> (2016).
147. Halcox, J. P. J. et al. Assessment of Remote Heart Rhythm Sampling Using the AliveCor Heart Monitor to Screen for Atrial Fibrillation: The REHEARSE-AF Study. *Circulation* **136**, 1784–1794 (2017).
148. Commissioner, O. of the. FDA approves pill with sensor that digitally tracks if patients have ingested their medication. *FDA.* <http://www.fda.gov/news-events/press-announcements/fda-approves-pill-sensor-digitally-tracks-if-patients-have-ingested-their-medication> (2018).

## ACKNOWLEDGEMENTS

This work is written on behalf of the Women's Brain Project (WBP) ([www.womensbrainproject.com/](http://www.womensbrainproject.com/)), an international organization advocating for women's brain and mental health through scientific research, debate and public engagement. The authors would like to gratefully acknowledge Maria Teresa Ferretti and Nicoletta Iacobacci (WBP) for the scientific advice and insightful discussions; Roberto Confalonieri (Alpha Health) for reviewing the manuscript; the Bioinfo4Women programme of Barcelona Supercomputing Center (BSC) for the support. This work has been supported by the Spanish Government (SEV 2015–0493) and grant PT17/0009/0001, of the Acción Estratégica en Salud 2013–2016 of the Programa Estatal de Investigación Orientada a los Retos de la Sociedad, funded by the Instituto de Salud Carlos III (ISCIII) and European Regional Development Fund (ERDF). EG has received funding from the Innovative Medicines Initiative 2 (IMI2) Joint Undertaking under grant agreement No 116030 (TransQST), which is supported by the European Union's

Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations (EFPIA).

## AUTHOR CONTRIBUTIONS

A.S.C., N.M., S.C.S. and D.C. conceived the study. A.V., M.J.R., A.G. supervised the project. All the authors contributed to the writing of the article, assisting with specific sections based on their expertise (E.G., Experimental and clinical data; S.C.S., C.M., S.M., Digital biomarkers; L.S., Big Data; D.C., Natural Language Processing; N.M., Robotics; S.C.S., Explainable A.I.; D.C., Fairness).

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41746-020-0288-5>.

**Correspondence** and requests for materials should be addressed to D.C.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020