frontiers | Frontiers in Genetics

Ethical layering in AI-driven polygenic risk scores—New complexities, new challenges

Marie-Christine Fritzsche[1,2]*, Kaya Akyüz[3,4], Mónica Cano Abadía[3], Stuart McLennan[1,2], Pekka Marttinen[5], Michaela Th. Mayrhofer[3] and Alena M. Buyx[1,2]

[1]Institute of History and Ethics in Medicine, TUM School of Medicine, Technical University of Munich, Munich, Germany, [2]Department of Science, Technology and Society (STS), School of Social Sciences and Technology, Technical University of Munich, Munich, Germany, [3]Biobanking and Biomolecular Resources Research Infrastructure Consortium - European Research Infrastructure Consortium (BBMRI-ERIC), Graz, Austria, [4]Department of Science and Technology Studies, University of Vienna, Vienna, Austria, [5]Helsinki Institute for Information Technology HIIT, Aalto University, Helsinki, Finland

Researchers aim to develop polygenic risk scores as a tool to prevent and more effectively treat serious diseases, disorders and conditions such as breast cancer, type 2 diabetes mellitus and coronary heart disease. Recently, machine learning techniques, in particular deep neural networks, have been increasingly developed to create polygenic risk scores using electronic health records as well as genomic and other health data. While the use of artificial intelligence for polygenic risk scores may enable greater accuracy, performance and prediction, it also presents a range of increasingly complex ethical challenges. The ethical and social issues of many polygenic risk score applications in medicine have been widely discussed. However, in the literature and in practice, the ethical implications of their confluence with the use of artificial intelligence have not yet been sufficiently considered. Based on a comprehensive review of the existing literature, we argue that this stands in need of urgent consideration for research and subsequent translation into the clinical setting. Considering the many ethical layers involved, we will first give a brief overview of the development of artificial intelligence-driven polygenic risk scores, associated ethical and social implications, challenges in artificial intelligence ethics, and finally, explore potential complexities of polygenic risk scores driven by artificial intelligence. We point out emerging complexity regarding fairness, challenges in building trust, explaining and understanding artificial intelligence and polygenic risk scores as well as regulatory uncertainties and further challenges. We strongly advocate taking a proactive approach to embedding ethics in research and implementation processes for polygenic risk scores driven by artificial intelligence.

1 Introduction

Machine learning (ML) techniques, in particular deep neural networks (DNNs), are increasingly being developed to generate polygenic risk scores (PRSs) using electronic health records (EHRs) as well as genomic and other health data (Ho et al., 2019; Badré et al., 2021; Elgart et al., 2022). While this may allow greater accuracy, performance and prediction ability of PRSs, it also presents a range of increasingly complex ethical challenges.

PRSs are defined as "a weighted sum of the number of risk alleles an individual carries" (Lewis and Vassos, 2020). In medicine, PRSs estimate an individual's risk of a specific condition or disease based on their genetic makeup. Even though the genomes of individuals are to a large extent similar, there are genetic differences, which are called genetic variants (Broad Institute, 2021). If a genetic variant is more common in individuals who have a specific disease, it may be associated with an increased risk of that disease (Broad Institute, 2021). A PRS takes into account all these risk variants, however minimal their effect, to estimate an individual's risk of developing a disease (Broad Institute, 2021). Recently, PRSs have been developed to offer targeted risk prediction for a rapidly increasing number of conditions, including complex common diseases and conditions, such as breast cancer (Mavaddat et al., 2019), type 2 diabetes mellitus (Läll et al., 2017), coronary heart disease (Khera et al., 2016; Inouye et al., 2018), obesity (Khera et al., 2019), depression (Mitchell et al., 2021) and schizophrenia (Trubetskoy et al., 2022). Researchers aim to develop PRSs as a tool to prevent and more effectively treat serious diseases, disorders and conditions by identifying those at high risk who would benefit from targeted therapies.

The ethical and social implications of many PRS applications in medicine have already been widely discussed (e.g., Adeyemo et al., 2021; Knoppers et al., 2021; Slunecka et al., 2021). However, their confluence with ML has not yet been sufficiently considered in either literature or practice. We argue that the interaction between different and novel layers of ethical and social concerns pertaining to artificial intelligence (AI) and big data, as well as PRSs in research and translation into the clinical setting, stand in need of urgent consideration. This includes ethical aspects of AI as well as ethical and social implications of precision medicine and PRSs. We highlight potentially increasing complexities and the need to explore which new ethical and social issues arise from increased use of AI techniques for different PRS applications. We do so in the hope that those who aim to embed PRSs in healthcare systems take a proactive approach to embedding ethics during the research and implementation process. After giving a brief overview of the background to AI-driven PRSs, we consider the many ethical layers involved, beginning with the ethical and social implications of PRSs, then moving on to the challenges in AI ethics, and finally, exploring potential complexities of AI-driven PRSs.

2 Background to PRSs and AI-driven PRSs

Early studies on PRSs (Purcell et al., 2009; Dudbridge, 2013) applied the so-called *classic PRS method* (Choi et al., 2020), where the risk is calculated as a weighted sum (i.e., a linear regression) of a set of genetic risk alleles for given single nucleotide polymorphisms (SNPs) (see also Figure 1). The relevant subset of SNPs is selected using a genome-wide association study (GWAS), usually conducted in a cohort different from the target cohort, such that SNPs exceeding a certain p-value threshold are included in the calculation of the risk in the target population. Instead of using a subset corresponding to the significant SNPs, it is possible to include a much larger number of SNPs in the weighted sum to calculate the risk. When so many SNPs are included, it is necessary to prevent overfitting by applying shrinkage on the linear regression weights using either classic

FIGURE 1
Classic PRSs and ML-driven PRSs the polygenic risk score for a target individual and phenotype of interest (y) is based on the individual's genetic data (x_g) but can also include other data types (x_e). The score is calculated using a linear regression (with weights β) or a machine learning model f_θ (e.g. a neural network with parameters θ). The parameters (β, θ) are learned using a separate training cohort. Note, however, that while the linear regession cofficients β are often publicly available or can be derived from published summary statistics, to train the neural network f_θ it is necessary to have access to individual level data in the training cohort.

techniques such as the LASSO or ridge regression (Mak et al., 2017) or Bayesian methods (Ge et al., 2019), the latter having given rise to some of the most popular implementations today (Vilhjalmsson et al., 2015). The SNP weights in a PRS can be derived from effects sizes published for the GWAS cohort, where the effect of each SNP on the risk has been estimated one-SNP-at-a-time, by accounting for linkage disequilibrium (LD) between the SNPs (Choi et al., 2020). Therefore, to apply classic PRS, individual-level data are only needed for the target individuals, but not from the GWAS cohort.

Recent years have witnessed attempts to replace the linear regression in PRS calculations with more sophisticated ML methods, which promise increased accuracy due to less restrictive modeling assumptions (Ho et al., 2019; Elgart et al., 2022). For example, DNNs which belong to the broader class of deep learning (DL), have been tested in PRSs for breast cancer, leading to improved scores compared to other statistical and ML estimation methods (Badré et al., 2021). A DNN processes input SNP data by passing them successively through multiple layers, where each layer takes the features from the previous layer as input, updates them, and passes the updated features forward to the next layer. In this way, features in higher layers can represent arbitrary, non-linear combinations of SNPs instead of the simple linear summation in conventional PRSs, which may better reflect the underlying biology.

Besides applying DL to modeling the genetic component, DL can alternatively be used to extract additional predictive features from EHRs (Miotto et al., 2016), which can be combined with the genetic data as input in PRS calculation (Dias and Torkamani, 2019). For example, using non-genetic risk factors together with genetic data improves the accuracy of breast cancer (Lee et al., 2019) and coronary artery disease (Inouye et al., 2018) risk modeling with the potential to enhance risk-based screening. However, current models typically build on combining genetics and EHR features additively (i.e., a simple summation), leaving room for the development of more complete approaches, for example a DNN that takes as input the different risk factors jointly to learn about the complex interplay between them.

Current research aims to pool and assess genomic data from biobanks, cohorts or registries on an unprecedented scale by combining it with environmental, other -omics data and health data such as EHRs. Considering the increasing heterogeneity of data that is used in the development of PRSs, more complex uses of AI have also been employed, such as making use of deep phenotypic information in medical images and EHRs to support downstream genetics analyses (Dias and Torkamani, 2019). Currently, PRSs typically only involve the genetic component, which is easier to interpret. The challenges in interpretation mainly occur when other data types are included, such as EHRs or gene expression data, the latter being different from SNPs that are currently used as data for PRSs (see also Figure 1). Other such data types are likely to increase in use, so the major challenges regarding the black box nature of the DNN models will probably be more relevant in the (near) future. Although researchers aim to reveal more and more causal relations, to date, analyses with AI for PRSs are mainly limited to correlations and improving predictions, which can result in inconclusive evidence (see Section 4). Barriers to the explainability of AI for PRSs also exist due to the statistical-probabilistic properties and the difficulty of the model to uncover the more complex biological, chemical and physical mechanisms that influenced it. In addition, there is a risk of potentially superfluous or inflated correlations due to the

limitations of the method through phenomena such as the recently observed "cross-trait assortative mating" (Border et al., 2022). The risks of misinterpretation of (AI-driven) PRSs by clinicians, patients and other stakeholders involved should not be underestimated, especially as there may be a risk of drawing conclusions about causal relationships too quickly and where knowledge of statistics and causality/correlation claims is too low in many groups involved. Although the difference between causation and correlation is well understood by scientists, authors point to the need for education of the public about such differentiations for PRSs (Slunecka et al., 2021).

3 Ethical and social implications of PRSs

The potential benefits of the clinical use of PRSs may be manifold, both for individuals and/or society: identifying individuals at risk, improving the precision and range of differential diagnoses and treatments, as well as promoting the development of intervention thresholds. Incorporating polygenic risk profiles into population screening is expected to increase efficiency in contrast to screening stratified by age (Chowdhury et al., 2013; Torkamani et al., 2018; Kopp et al., 2020), while use of combined PRSs for various conditions in healthcare systems may contribute to early identification of potential non-genetic interventions and increased life expectancy (Meisner et al., 2020). Thus, PRSs may benefit individuals and represent a dramatic improvement of public health with potential socio-economic impacts. This has led to demands by PRS advocates within the medical community for a radical rethinking of PRSs as clinical instruments that could inform clinical decisions, such as in the prioritisation of psychosocial or pharmaceutical interventions "rather than treat/not treat decisions" (Lewis and Vassos, 2020).

While they come with important benefits, discussions in the literature on the multiple ethical and social implications for the medical use of PRSs range from social and distributive justice questions to debates on scientific validity and clinical utility (Babb de Villiers et al., 2020; Lewis and Vassos, 2020; Knoppers et al., 2021; Lewis and Green, 2021; Slunecka et al., 2021; Widen et al., 2022). In the context of PRS development and clinical implementation, ethical debates reflect those on monogenic genetic findings (Lewis and Green, 2021). Common concerns relate, for example, to genetic determinism as well as the concepts of ancestry/ethnicity, where tools such as AI for risk stratification may not be representative of human diversity and whose development and use may distract attention from the social determinants of health (Lewis and Vassos, 2020; Knoppers et al., 2021; Lewis and Green, 2021). Particular concerns about the risk of genetic discrimination and eugenics are raised with regard to the application of PRSs for embryo screening (Treff et al., 2019; Tellier et al., 2021; Turley et al., 2021); most recently for pre-implantation genetic testing (PGT) (Kozlov, 2022) and premature direct-to-consumer genetic testing/genetic counselling (Docherty et al., 2021), which are also intertwined with marketability and commercialisation. Furthermore, due to underrepresentation of already underserved communities in the research process, some authors note that health disparities could increase through the use of PRSs in the clinical setting (Martin et al., 2019a).

There has been extensive discussion of the clinical and/or personal utility of PRSs (Torkamani et al., 2018; Lambert et al., 2019; Wald and Old, 2019; Lewis and Vassos, 2020; Moorthie et al., 2021; Sud et al.,

2021). Scientific and clinical validity are challenges on multiple levels (Janssens, 2019; Lewis and Vassos, 2020; Knoppers et al., 2021), which touch ethical as well as epistemic concerns. PRSs, for example, do not cover the full risk for certain diseases because of the multiple factors involved. This includes e.g. environmental factors (Slunecka et al., 2021) and complex interactions between environments and PRSs (Domingue et al., 2020). Due to this complexity, interpretation of PRSs poses serious challenges, especially in relation to minors (Palk et al., 2019). From an ethical point of view, the necessity of communicating the limitations of risk prediction with PRSs therefore has to be considered in clinical applications. To this end, "effective and clear risk communication by trained professionals" should "minimize potential psychosocial effects" (Adeyemo et al., 2021). However, in this context, there is a lack of standardised PRS disclosure for individuals (Brockman et al., 2021; Lewis et al., 2022) as well as for kin, such as cascade screening for family members (Reid et al., 2021). Tools for standardisation of PRS disclosure have been developed for certain diseases, such as coronary artery disease (Widen et al., 2022), but the need for additional research on a broader range of populations and better standardisation has been emphasised (Brockman et al., 2021).

Given that PRSs are still an emerging field, there is remarkable heterogeneity around their application and reporting, thus constraining the implementation of PRSs in clinical settings (Slunecka et al., 2021). Publicly accessible catalogues and reporting standards for PRSs have been developed that are responsive to the current research landscape to allow reporting on the design and validation of PRSs within the literature (Lambert et al., 2021; Slunecka et al., 2021; Wand et al., 2021), such as the NHGRI-EBI, an extensive database of summary statistics of GWAS (Buniello et al., 2018). One aim of these efforts is to generate comparable PRSs metrics of performance (Lambert et al., 2021). This should increase the reproducibility and transparency of the PRS development process as well as support studies evaluating the clinical utility of the respective PRSs (Lambert et al., 2021). External and systematic PRS studies with benchmarking should also contribute to these aims (Wand et al., 2021). Another practical ethical issue is that the application of PRSs for medical purposes is presently uncertain under the majority of legal frameworks (Lewis and Vassos, 2020; Adeyemo et al., 2021).

Moreover, some authors also point out the importance of seeing PRSs in the respective context (Chatterjee et al., 2016; Torkamani et al., 2018; Slunecka et al., 2021), considering that the scope and diversity of available data (for instance, ancestry) and the techniques used to produce and use the scores are continuously changing (Trubetskoy et al., 2022). This therefore necessitates consideration, e.g., of the particular PRSs and the disease for which the PRSs were designed and the sophistication of the PRS itself. Consequently, the ethical and social implications need to be explored, taking into account the respective context. For example, specific ethical concerns in PRSs have been increasingly described for psychiatric conditions from informational risks in the use of the PRS in clinical setting, to the research showing links between the condition and social factors such as socioeconomic status or potential use in prenatal testing among others (Agerbo et al., 2015; Loh et al., 2015; Martin et al., 2019b; Palk et al., 2019; Docherty et al., 2021; Murray et al., 2021). This may differ for other conditions, for instance, in terms of actionability or potential for stigmatisation.

4 Challenges in AI ethics

There is much debate on ethical aspects around AI in healthcare (Morley et al., 2020), the role that AI should play (Rigby, 2019), the role and ethical implications of "explainability for AI in healthcare" (Amann et al., 2020), and ethical challenges of ML (Vayena et al., 2018) and of DL in healthcare (Char et al., 2018; Miotto et al., 2018). In particular, the following ethical and social challenges are often discussed in AI ethics (Mittelstadt et al., 2016; Floridi et al., 2021; Tsamados et al., 2021): How to ensure fairness and justice, overcome biases, ensure explainability, transparency, traceability, accountability, privacy, confidentiality, data protection and patient safety–how to design AI for the common good.

In AI ethics, not only are there normative concerns about algorithms such as "unfair outcomes" and "transformative effects", but also epistemic concerns such as "inconclusive evidence", "inscrutable evidence" and "misguided evidence" (Mittelstadt et al., 2016; Tsamados et al., 2021), and often epistemic and normative concerns come together as in the case of traceability. Issues such as the black box problem, accountability and transparency can be subsumed under inscrutable evidence (Mittelstadt et al., 2016). The black box problem in ML hinges on the lack of explainability as to how results are generated. The importance of this is also reflected in European law like the EU General Data Protection Regulation (GDPR) (European Parliament and Council of the European Union, 2016), which entails a general "right to explanation" (Goodman and Flaxman, 2017) for users and a future where explainability could become a legal requirement for ML specifically. The proposed Artificial Intelligence Act of July 2021 explicitly includes the requirement that AI systems be explainable for high-risk sectors (European Commission, 2021). The literature in recent years has repeatedly underlined the need for explainable AI (xAI) in medicine (Ribeiro et al., 2016; Hudec et al., 2018; Holzinger et al., 2019; Azodi et al., 2020), which is seen as (part of) a possible solution to many of the above-mentioned challenges in AI applications in healthcare.

Inconclusive evidence (Mittelstadt et al., 2016) involves ethical issues of causality and correlation, probabilities and predictions. Inconclusive evidence and incorrect causal associations and correlations are a problem for any statistical model, which can be the result, e.g. of biased sampling or hidden contamination. Authors generally point to the need to understand causality of the representations in ML systems (Pearl, 2009; Gershman et al., 2015; Peters et al., 2017; Holzinger et al., 2019). Furthermore, as substructures from genomic and population data are correlated, this can potentially result in false causal associations (Sohail et al., 2019) and misleading information based on bias embedded in genomic data (see Section 5.1). Increasing the robustness of the detected effects across different populations would go some way towards separating true causal effects from spurious associations. In genetics, replicating the findings in multiple cohorts is usually a stipulation, but more work is required to ensure inclusion of more diverse populations (see Section 5.1).

The topic of "misguided evidence leading to bias" (Mittelstadt et al., 2016) and "unfair outcomes leading to discrimination" (Mittelstadt et al., 2016) are key issues in AI ethics. In medical AI, biases (Obermeyer et al., 2019) abound, and the replication of biases and the amplification of real-world injustices by algorithms poses a serious risk.

There are many different proposals for frameworks on how the challenges of applying AI in medicine should be addressed ethically, which principles and values are of particular importance and which guidelines should be followed. Ethical challenges exist in terms of principles, not only regarding which principles should be considered crucial, but also in terms of differences in what the principles mean, e.g. what justice encompasses, as there are many different forms of justice derived from different philosophical theories and different underlying values (Whittlestone et al., 2019). Furthermore, there is the question of what "for the good of society" means—What would AI that is focused on the common good look like? This would need to be discussed and defined in each context (Whittlestone et al., 2019).

Another challenge usually arises when principles conflict with each other, as is often the case with AI in healthcare. Explainability is often not technically possible, and the benefits of AI can vary in significance, so the trade-off would have to be weighed up for each AI system and context. Another major ethical challenge around AI is putting principles into practice. Authors point out that attention needs to be paid to the tensions and conflicts that arise in this process and that these need to be addressed (Whittlestone et al., 2019) so that risks can be avoided and the benefits of AI can be reaped.

5 Bringing ethical and social aspects of PRSs and AI ethics together—New complexities for AI-driven PRSs?

In bringing ethical and social implications of PRSs and of AI ethics together, we would like to point out potential new complexities for AI-driven PRSs. Particularly around the following topic clusters which will be discussed in detail in what follows.

1) More complexity regarding fairness and justice
2) Challenges in building trust, communication and education
3) Privacy and autonomy challenges
4) Regulatory uncertainties and further challenges

5.1 More complexity regarding fairness and justice

Although many researchers point out the opportunities of xAI and interpretable ML (iML), two ethically relevant issues with respect to explanatory methods remain generally difficult to solve: different biases within datasets leading to biased DNN and suspicion of bias in results leading to unfairness (Ras et al., 2018). This could apply also to ML application for PRSs on multiple levels: many biases in PRS development can be linked to biases in the combination of EHRs with genomic and further health data as well as in the substructures of this data.

Firstly, the majority of genetic studies lack diversity (Sirugo et al., 2019). PRSs have mainly been developed with datasets from European populations and predictions of genetic risk are susceptible to unequal outputs (performance levels) across different populations as they are underrepresented in training data, which hinders generalisability (Martin et al., 2019a). Authors observe that research infrastructures like biobanks may suffer from "recruitment bias" as a risk which "infringes on the principle of justice, influences representativity of biobank collections and has implications for the generalizability of

research results and ability to reach full statistical power" (Akyüz et al., 2021).

Secondly, further data biases can be linked to many other factors. There is a considerable gap in medical studies on the representation of women (Daitch et al., 2022) as the case of cardiovascular disease also shows (Burgess, 2022). More broadly, gender bias can be found in written documents used for certain ML techniques (Bolukbasi et al., 2016). Gender bias may also occur when heteronormative paradigms are not met, e.g., when data on gender and sex do not match and are therefore automatically excluded for analysis, which is currently a common practice in genomics (American Medical Association, 2018; Ganna et al., 2019). EHRs can contain multiple biases resulting e.g., from physician bias or certain delivery of care (Ching et al., 2018; Gianfrancesco et al., 2018), and even laboratory measurements (which are considered less biased) can show bias resulting from the patient health state and healthcare process (Pivovarov et al., 2014)—although they may be representative regarding population (Kerminen et al., 2019; Adeyemo et al., 2021). Overall, there are substructures in genomics and other health data that can be linked to actual differential causal relationships between health outcomes and putative risk factors. Other substructures can be traced to external factors such as cultural practices, socioeconomic status and other non-causal factors that relate to healthcare provision, access to medicine and clinical trials (Gianfrancesco et al., 2018; Dias and Torkamani, 2019; Sirugo et al., 2019).

Apart from the bias in data, *machine bias* has to be mentioned in the context of AI use for PRSs. This encompasses the biases that are learned by the models (Ching et al., 2018; Dias and Torkamani, 2019). In this context, one criterion for iML for genetic risk prediction could be whether a certain model is adequately interpretable for bias to be detected (Ching et al., 2018; Dias and Torkamani, 2019). Authors call for standards of fairness in order to diminish disparities caused by bias of ML in genetic risk prediction (Dias and Torkamani, 2019; McInnes et al., 2021). Moreover, they point to the necessity for careful application of AI and differentiation between the various forms of bias arising when AI is applied to genetic risk prediction (Dias and Torkamani, 2019). Tools are already being developed to help eliminate machine bias. This is not only intended to eliminate bias of ML, but also to create diagnostic systems that are much freer from human bias than classical diagnostics by physicians allow (Shen et al., 2019). These and further innovative sorts of techniques should also be consistently considered for ML use for PRSs.

In addition to injustice due to biases, *injustice and unfairness regarding data access and sharing data and algorithms* is also an issue for AI-driven PRSs. In this regard, biased processes and results are co-produced, potentially sustaining existing inequalities and unfairnesses. Further, apart from comprehensibility, accessibility can be considered as the second main component of transparency in generating information about how algorithms function (Mittelstadt et al., 2016). While many advances have been made thanks to international initiatives and large interdisciplinary research consortia, authors still highlight the ongoing need to collect, harmonise and share data in genomics and healthcare (Diao et al., 2018; Lambert et al., 2021). The Polygenic Risk Score Task Force of the International Common Disease Alliance has called for the "GWAS research community, global biobank collaborations, and private direct-to-consumer companies" (Adeyemo et al., 2021) to create requirements for public sharing of summary statistics using standardised formats, with the aim of avoiding the exacerbation of

worldwide health inequalities (Adeyemo et al., 2021). However, sharing DL models with the biomedical data and health records of individuals not only faces legal and technical barriers but also poses a major "cultural challenge" (Ching et al., 2018). A culture that rewards discovery rather than the production of data will have a difficult time motivating researchers to share their hard-earned datasets (Ching et al., 2018). However, as is pointed out in recent articles, it is this data that would drive DL (Ching et al., 2018).

Apart from well-known privacy regulations and standards in medical and biological research (Ching et al., 2018), factors such as the costs related to regulations for medical devices may also play an important role in access to PRSs, creating inequalities among populations, subgroups and countries (Adeyemo et al., 2021). Not only does global cooperation contribute to more equity in medical research and healthcare, it also serves an important role for the improvement of clinical validity and utility of PRSs (Adeyemo et al., 2021; Knoppers et al., 2021). Moreover, an open exchange of AI models for genetic risk prediction with the medical and scientific communities is called for to enhance transparency, where the model sharing should include details such as model weights, source codes and meta diagrams (Dias and Torkamani, 2019). Synthetic genetic and phenotypic data (Abadi et al., 2016) is suggested for genomic projects (Moorthie et al., 2021) and is already being tested in PRS development to provide greater diversity in genetic data, avoid biases and privacy issues. Furthermore, protecting data and privacy are very relevant for public-private partnerships (Murdoch, 2021), which play an increasingly important role for the implementation and dissemination of PRSs.

5.2 Challenges in building trust, communication and education

One of the greatest challenges in translating PRSs to the clinical setting is the communication of PRSs. This includes communication to and dialogues with the public(s) and patients as well as educating all other stakeholders involved. The challenge of communicating PRSs in the clinical setting, particularly for doctors (Fiske et al., 2019), is magnified when explaining AI-driven PRSs.

In general, we highlight the need for reflection on epistemological questions around AI use for PRSs and the corresponding normative aspects. It is important to ask what it means to explain, interpret and understand AI-driven PRSs. This should ideally incorporate different perspectives for certain stakeholders and involve further associated questions, e.g., what researchers consider an explanation to be, what kind of explanation users want and need (Slunecka et al., 2021) and what criteria are relevant for explainable PRSs. With the advance of xAI and iML, it is also worth considering how much/what kind of explainability is required for the clinical application of PRSs and how much/what kind of interpretability is clinically meaningful.

With regard to the literature reviewed and the existence of different definitions of explainability, explicability, interpretability and comprehensibility in scientific teams and clinical settings, we argue that awareness of these differences of terms must be raised both in scientific publications and in practice. This would also have the ultimate goal of improving the explainability of the risk scores and the underlying AI mechanisms.

Stakeholders in research and development as well as healthcare areas are constrained to consider the uncertainty of AI-generated PRS predictions and thus need to develop means of dealing with them in a

structured, transparent and responsible way. Even if a more explainable ML for PRSs is developed, the question of how to communicate and generally deal with uncertainty due to lack of explainability of ML for PRSs nevertheless requires discussion and translation into appropriate standards. Embedded ethics approaches (McLennan et al., 2022) in both the research and clinical settings could help resolve the challenge of detecting and reflecting on ethical issues as well as communicating them.

Regarding communication of AI-driven PRSs, there is a clear need for engagement with technical, medical and ethical aspects of PRSs and AI for all the different stakeholders involved. We strongly recommend adopting interactive/participatory engagement practices (Horst et al., 2017), especially between clinicians and patients for AI-driven PRSs. This means limiting or avoiding deficit models of communication, i.e., unchallengeable, non-reflexive (Wynne, 1993) communication, which sees audiences (including any actors other than experts) as deficient both in knowledge and capacity to comprehend (Bell et al., 2008). In light of the developments in e-health, citizen-patients are not considered passive recipients of information, but rather self-informing, active individuals (Felt et al., 2009). Furthermore, the respective educational, socioeconomic and cultural background of individual patients and their families has to be considered when, for example, physicians explain PRSs (Slunecka et al., 2021).

In general, one of the biggest challenges of AI-driven PRSs today is trust in AI/AI-driven PRSs and trust in the medical institutions that will use these technologies on a large scale. However, there is a lack of specificity in the literature on issues of trust in the recently developed AI-driven PRSs. This represents a future issue that will need to be addressed with interdisciplinary teams.

Problems of AI explainability add complexity to matters of trust for AI-driven PRSs. Lack of transparency and lack of human understanding of AI black boxes raises the question of how all kinds of end-users create their relationship with AI. Scholars emphasise the importance of explainable AI (Holzinger et al., 2019) and DL models in medicine by arguing for the trust-building effect they have (Ribeiro et al., 2016). They point to the importance of understanding the rationale underlying the predictions of ML modelling when evaluating trust, which is considered crucial for decisions on the use of new models and actions based on predictions (Ribeiro et al., 2016). Interpretability is reflected in the "fidelity-interpretability trade-off" (Ribeiro et al., 2016) and is key to building trust in AI among healthcare professionals. Practitioners are very unlikely to accept a DL system that they do not understand (Miotto et al., 2018). It is noted that the interpretability of the model in genomics is critical to convincing health professionals of the validity of the actions the prediction system recommends, e.g., to explain which phenotypes drive certain predictions (Miotto et al., 2018).

The High-Level Expert Group on AI of the European Commission proposes trust as one of the defining principles for their AI ethics guidelines (High-Level Expert Group on AI, 2019). However, the technical solutions to the issue of trust, as discussed above, are unlikely to become available in definitive form. We therefore suggest that the social and relational considerations are paramount if we are to create a workable framework for establishing trust. This means the question of how trust is built needs to be addressed by adopting a more reflexive and interdisciplinary perspective. This also includes discussion of the *trustworthiness* of AI use for PRSs. Which is to say, discussions about dependable, trustworthy ML use for the PRSs and what requirements and criteria should be placed on the trustworthiness of AI for PRSs

must perforce address contextual questions, such as what trust means in a particular situation or context. The FUTURE-AI initiative has created dynamic best practices for trustworthy AI in healthcare (Future AI, 2022). Empirical and theoretical studies on the ethical and social issues of AI-driven PRSs and trustworthiness are needed so that this knowledge can then be integrated into the development and application of AI-driven PRSs.

Overall, we recognise that education and training of AI-driven PRSs would need to cover tech/AI literacy, risk interpretation/ statistical knowledge, genomics/PRS knowledge, communication skills and ethical reflection skills of the stakeholders involved—of course with different granularities depending on the stakeholders: patients, relatives of patients, various public(s), healthcare professionals, medical/nursing students, researchers, technicians, ethics committees, clinical ethics teams, business partners and all the other stakeholders involved in research and development as well as the translation, implementation and application of AI-driven PRSs.

For AI in medicine generally, there is a need to increase education and training for different stakeholders in the healthcare system on applications of technology driven by data (Meskó et al., 2017; Xu et al., 2019). For ML in genomics, authors stress the need to bridge the gaps regarding clinical knowledge and interpreting models (Diao et al., 2018). Others consider the training of clinical staff to be a major challenge for the implementation of PRSs in the clinical setting (Torkamani et al., 2018; Slunecka et al., 2021). The unique nuances of PRSs and GWAS development are mostly unfamiliar to clinicians at this point (Martin et al., 2019a). Concrete suggestions have been made for enhancing education about PRSs for inclusion in the regular curriculum for medical students and in the ongoing education for medical professionals, covering the limitations of PRSs and different forms of risk (Slunecka et al., 2021). In addition, there are different proposals for how experts in genetic risk assessments could be involved in the clinical setting. Furthermore, education of the public(s) is crucial in implementing PRSs for public screening. The website of the National Human Genome Research Institute of the National Institutes of Health (UK), for instance, aims to explain to the public how PRSs work and how to interpret them. Apart from that, sensitivity, reflection and discussion on relationality and power relations of patients, doctors, healthcare and research institutions as well as biotechnology/genomics companies are important issues in the development of AI-driven PRSs. Based on a renewed understanding of how citizens engage with physicians and information technologies in health setting (Felt et al., 2009), empowering citizens and patients is among the key developments for the application of AI-driven PRSs.

5.3 Privacy and autonomy challenges

When large amounts of genomic data and EHRs are used to generate PRSs with AI, privacy is a key issue. A crunch question is whether protection of personal/patient data trumps transparency and right of access to data or *vice versa*. There are also multiple questions revolving around the extent to which anonymisation can be ensured with the large amounts of data used for PRSs, new AI technologies and what informed consent should look like for different uses of PRSs driven by them. For example, the differential privacy method, in which noise is added to data to prevent revealing individual information in case summaries of the data were to be published, does not scale easily to high-dimensional genetic data (Roth and Dwork, 2013). While there are efforts in medicine and PRS development aimed at protecting privacy (Abadi et al., 2016; Simmons

et al., 2016; Ching et al., 2018; Beaulieu-Jones et al., 2019; Zhang et al., 2021), it is unclear how these could be implemented or policed on a large scale for AI-driven PRSs. Despite the discourse of exceptionalism of big data research, privacy is still an issue that is tightly entangled with autonomy (Rothstein, 2015). However, in a data-rich environment, genomic data, which is by definition shared in differing amounts with biological relatives, poses further challenges to our understandings and practices of privacy and autonomy, but also anonymisation or risk of genomic identifiability, raising the necessity for a "post-identifiability" lens (Akyüz et al., 2023). Thus, privacy and autonomy are challenges in their own right due to the peculiarity of genomic data.

5.4 Regulatory uncertainties and further challenges

As for healthcare in general, the need for complementary measures to explainability such as regulation (Markus et al., 2021), enhancing the quality of healthcare data for DL (Miotto et al., 2018) and external validation (Markus et al., 2021) have to be considered for AI-driven PRSs. The need for regulatory measures for PRSs in general is highlighted in the literature reviewed (Adeyemo et al., 2021; Knoppers et al., 2021; Slunecka et al., 2021). Standardisation of regulation frameworks for PRSs as medical devices (Adeyemo et al., 2021) is urgently required. With AI-driven PRSs, it is even more important to establish internationally standardised regulation frameworks which are responsive to the dynamic and fast-evolving technical and scientific findings around PRSs. Flexible, on-demand "*ad hoc*" guidance to positively enhance ongoing algorithm improvement (Vayena et al., 2018; Dias and Torkamani, 2019) would support the ethically sound development of AI-driven PRSs. However, regulatory measures can be a burden for people with access to PRS technology (Knoppers et al., 2021). In this sense, the challenge of creating a balance between sufficient regulation and rapid scientific advancement in the application of AI for PRSs must be considered in the development of AI-driven PRSs.

Beyond the ethical concerns mentioned above, further ethical challenges of AI-driven PRSs, such as informed consent procedures for AI-driven PRSs in absentia of explainability could become even more relevant in the future. In addition, the importance of AI for ethics committees has to be emphasised as does the need to involve research ethics committees and clinical ethics committees in the translation and implementation of AI-driven PRSs.

6 Conclusion

Our article has delineated the multiple layers of ethical and social concerns associated with PRSs, AI for PRSs and AI-driven PRSs in medicine. A clear limitation of most ML-based approaches compared with the *classic PRS* method is the requirement for individual level data to train the models, whereas the latter uses publicly available summary statistics about estimated effect sizes. Hence, there is room for development of new ways to leverage published summary statistics in training of more flexible ML-based PRS methods. Another limitation and future challenge common to all PRS methods is the poor generalisability of the scores in populations with different ancestries, which also stems from different allele frequencies, linkage disequilibrium and genetic effect sizes in different populations (Wang et al., 2022). Regarding the use of AI in PRS, there is great potential for improvement by developing models that integrate a

variety of health data types and risk factors into comprehensive predictors of disease risk (Dias and Torkamani, 2019). The clinical utility of PRSs is currently hotly debated; thus, more research is warranted on the best ways to implement PRSs as part of clinical practice, either to improve diagnoses, personalise treatments, or as part of preventive medicine (Torkamani et al., 2018; Choi et al., 2020). In particular, the additional challenges for the clinical implementation posed by the AI based PRS methods remain to be addressed. Furthermore, our discussion of some of the ethical issues that need to be considered in AI-driven PRS is in no way exhaustive. Rather, this article can serve as a basis for further discussions of the ethical challenges that could arise from the future application of AI-driven PRSs.

Where PRSs, ML and big data are part of the picture, we have teased out the more complex ethical challenges emerging from the relation between them, as well as pertaining to them individually. Based on a comprehensive review of the existing literature, we argue that this stands in need of urgent consideration for research and translation into the clinical setting. Different layers of ethical implications could lead to more challenges for explainability of AI-driven PRSs, more complexity of fairness with biases in data (sets) and ML for PRSs and biased outputs, more challenges in building trust, communication and education as well as regulatory uncertainties for and challenges in privacy and autonomy of AI-driven PRSs. Among these, we would especially like to highlight a lack of specificity in the literature on issues of trust in the more recent instantiations of AI-driven PRSs. We maintain that this is a future challenge that will need to be addressed in interdisciplinary, multi-stakeholder teams. The fact that the lack of explainability seems to be an inherent problem of certain ML techniques, which may never be fully solved, should not hinder efforts to make ML for PRSs more explainable and trustworthy for all stakeholders involved in the healthcare system. It has become clear that much of the more explainable PRSs depends not only on more explainable ML techniques, but also on awareness, context- and user-specific communication and engagement, education and training for all stakeholders. In addition, there are limitations to the influence of explainable ML that relate to ethical and social aspects associated with large amounts of data, such as EHRs, genomic and other health data fed into ML models. Apart from more technical research on e.g. techniques of explainable ML for PRSs, more ethical analyses are needed, covering epistemic and normative aspects of AI-driven PRSs including methods of normative and empirical ethics. We have also pointed out that hitherto there are few to no regulatory guidelines, and a lack of commensurate up-to-date research, let alone clear advice on how to communicate the potential implications, costs or benefits of these technological advances to and between the various stakeholders involved. For this, technical and bioethical content as well as discussions on the larger societal implications and public health aspects should also be included in the training for students and healthcare professionals. Although there are efforts to address the ethical and regulatory challenges of AI-driven PRSs, more work is required when AI tools are used with more complex health data such as EHRs and medical images or real world data. This should be an important item on the agenda of citizens, policymakers, scientists and funders of AI-driven PRS development as a co-production. This approach would make an important contribution to the clinical utility of PRSs in terms of transparency, responsibility and finally trustworthiness.

If we fail to address these challenges, the danger is that not only will advances in AI and/or the applications of PRSs outstrip our ability to understand or regulate them, but that the potential for overreliance and indeed misapplication or misuse from an ethical and social standpoint may create further and insurmountable complexities in the future.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

Where authors are identified as personnel of the Biobanking and BioMolecular resources Research Infrastructure (BBMRI-ERIC), the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of BBMRI-ERIC.

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., et al. (2016). *Deep learning with differential privacy. Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. Vienna, Austria: Association for Computing Machinery, 308–318.

Adeyemo, A., Balaconis, M. K., Darnes, D. R., Fatumo, S., Granados Moreno, P., Hodonsky, C. J., et al. (2021). Responsible use of polygenic risk scores in the clinic: Potential benefits, risks and gaps. *Nat. Med.* 27 (11), 1876–1884. doi:10.1038/s41591-021-01549-6

Agerbo, E., Sullivan, P. F., Vilhjálmsson, B. J., Pedersen, C. B., Mors, O., Børglum, A. D., et al. (2015). Polygenic risk score, parental socioeconomic status, family history of psychiatric disorders, and the risk for schizophrenia: A Danish population-based study and meta-analysis. *JAMA Psychiatry* 72 (7), 635–641. doi:10.1001/jamapsychiatry.2015.0346

Akyüz, K., Chassang, G., Goisauf, M., Kozera, Ł., Mezinska, S., Tzortzatou, O., et al. (2021). Biobanking and risk assessment: A comprehensive typology of risks for an adaptive risk governance. *Life Sci. Soc. Policy* 17 (1), 10. doi:10.1186/s40504-021-00117-7

Akyüz, K., Goisauf, M., Chassang, G., Kozera, Ł., Mežinska, S., Tzortzatou-Nanopoulou, O., et al. (2023). Post-identifiability in changing sociotechnological genomic data environments. *BioSocieties*. (Accepted for publication).

Amann, J., Blasimme, A., Vayena, E., and Frey, D. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Med. Inf. Decis. Mak.* 20 (1), 310. doi:10.1186/s12911-020-01332-6

American Medical Association (2018). *AMA adopts new policies at 2018 interim meeting*. National Harbor: Press Release.

Azodi, C. B., Tang, J., and Shiu, S-H. (2020). Opening the black box: Interpretable machine learning for geneticists. *Trends Genet.* 36 (6), 442–455. doi:10.1016/j.tig.2020.03.005

Babb de Villiers, C., Kroese, M., and Moorthie, S. (2020). Understanding polygenic models, their development and the potential application of polygenic scores in healthcare. *J. Med. Genet.* 57 (11), 725–732. doi:10.1136/jmedgenet-2019-106763

Badré, A., Zhang, L., Muchero, W., Reynolds, J. C., and Pan, C. (2021). Deep neural network improves the estimation of polygenic risk scores for breast cancer. *J. Hum. Genet.* 66 (4), 359–369. doi:10.1038/s10038-020-00832-7

Beaulieu-Jones, B. K., Wu, Z. S., Williams, C., Lee, R., Bhavnani, S. P., Byrd, J. B., et al. (2019). Privacy-preserving generative deep neural networks support clinical data sharing. *Circulation Cardiovasc. Qual. Outcomes* 12 (7), e005122. doi:10.1161/CIRCOUTCOMES.118.005122

Bell, A. R., Davies, S. R., and Mellor, F. (2008). *Science and its publics*. Newcastle: Cambridge Scholars.

Bolukbasi, T., Chang, K-W., Zou, J. Y., Saligrama, V., and Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Adv. Neural Inf. Process. Syst.* 29, 4349–4357.

Border, R., Athanasiadis, G., Buil, A., Schork, A., Cai, N., Young, A., et al. (2022). Cross-trait assortative mating is widespread and inflates genetic correlation estimates. *Science* 378, 754–761. doi:10.1126/science.abo2059

Broad Institute (2021). Polygenic scores. Available at: https://polygenicscores.org/explained/ (Accessed December 20, 2022).

Brockman, D. G., Petronio, L., Dron, J. S., Kwon, B. C., Vosburg, T., Nip, L., et al. (2021). Design and user experience testing of a polygenic score report: A qualitative study of prospective users. *BMC Med. Genomics* 14 (1), 238. doi:10.1186/s12920-021-01056-0

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2018). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47 (D1), D1005–D12. doi:10.1093/nar/gky1120

Burgess, S. N. (2022). Understudied, under-recognized, underdiagnosed, andundertreated: Sex-based disparities incardiovascular medicine. *Circ. Cardiovasc. Interv.* 15, e011714. doi:10.1161/CIRCINTERVENTIONS.121.011714

Char, D. S., Shah, N. H., and Magnus, D. (2018). Implementing machine learning in health care — addressing ethical challenges. *N. Engl. J. Med.* 378 (11), 981–983. doi:10.1056/NEJMp1714229

Chatterjee, N., Shi, J., and Garcia-Closas, M. (2016). Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat. Rev. Genet.* 17 (7), 392–406. doi:10.1038/nrg.2016.27

Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., et al. (2018). Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface* 15 (141), 20170387. doi:10.1098/rsif.2017.0387

Choi, S. W., Mak, T. S., and O'Reilly, P. F. (2020). Tutorial: A guide to performing polygenic risk score analyses. *Nat. Protoc.* 15 (9), 2759–2772. doi:10.1038/s41596-020-0353-1

Chowdhury, S., Dent, T., Pashayan, N., Hall, A., Lyratzopoulos, G., Hallowell, N., et al. (2013). Incorporating genomics into breast and prostate cancer screening: Assessing the implications. *Genet. Med.* 15 (6), 423–432. doi:10.1038/gim.2012.167

Cowls, J., and Floridi, L. (2018). Prolegomena to a white paper on an ethical framework for a good AI society. SSRN: https://ssrncom/abstract=3198732 2018.

Cui, T., Mekkaoui, K. E., Havulinna, A., Marttinen, P., and Kaski, S. (2021). "Improving neural networks for genotype-phenotype prediction using published summary statistics,". bioRxiv.

Daitch, V., Turjeman, A., Poran, I., Tau, N., Ayalon-Dangur, I., Nashashibi, J., et al. (2022). Underrepresentation of women in randomized controlled trials: A systematic review and meta-analysis. *Trials* 23 (1), 1038. doi:10.1186/s13063-022-07004-2

Diao, J. A., Kohane, I. S., and Manrai, A. K. (2018). Biomedical informatics and machine learning for clinical genomics. *Hum. Mol. Genet.* 27 (R1), R29–R34. doi:10.1093/hmg/ddy088

Dias, R., and Torkamani, A. (2019). Artificial intelligence in clinical and genomic diagnostics. *Genome Med.* 11 (1), 70. doi:10.1186/s13073-019-0689-8

Docherty, A., Kious, B., Brown, T., Francis, L., Stark, L., Keeshin, B., et al. (2021). Ethical concerns relating to genetic risk scores for suicide. *Am. J. Med. Genet. Part B Neuropsychiatric Genet.* 186 (8), 433–444. doi:10.1002/ajmg.b.32871

Domingue, B. W., Trejo, S., Armstrong-Carter, E., and Tucker-Drob, E. M. (2020). Interactions between polygenic scores and environments: Methodological and conceptual challenges. *Sociol. Sci.* 7 (19), 465–486. doi:10.15195/v7.a19

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLOS Genet.* 9 (3), e1003348. doi:10.1371/journal.pgen.1003348

Elgart, M., Lyons, G., Romero-Brufau, S., Kurniansyah, N., Brody, J. A., Guo, X., et al. (2022). Non-linear machine learning models incorporating SNPs and PRS improve polygenic prediction in diverse human populations. *Commun. Biol.* 5 (1), 856. doi:10.1038/s42003-022-03812-z

European Commission (2021). *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts COM/2021/206 final*. European: European Commission.

European Parliament, Council of the European Union (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official J. L* 119, 1–88.

Felt, U., Gugglberger, L., and Mager, A. (2009). Shaping the future e-patient: The citizen-patient in public discourse on e-health. *Sci. Technol. Stud.* 22 (1), 24–43. doi:10.23987/sts.55244

Fiske, A., Buyx, A., and Prainsack, B. (2019). Health information counselors: A new profession for the age of big data. *Acad. Med.* 94 (1), 37–41. doi:10.1097/ACM.0000000000002395

Floridi, L., and Cowls, J. (2021). "A unified framework of five principles for AI in society," in *Ethics, governance, and policies in artificial intelligence*. Editor L. Floridi (Cham: Springer International Publishing), 5–17.

Future AI (2022). FUTURE-AI: Best practices for trustworthy AI in medicine. Available at https://future-ai.eu (Accessed December 20, 2022).

Ganna, A., Verweij, K. J. H., Nivard, M. G., Maier, R., Wedow, R., Busch, A. S., et al. (2019). Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior. *Science* 365 (6456), eaat7693. doi:10.1126/science.aat7693

Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A., and Smoller, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* 10 (1), 1776. doi:10.1038/s41467-019-09718-5

Gershman, S. J., Horvitz, E. J., and Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science* 349 (6245), 273–278. doi:10.1126/science.aac6076

Gianfrancesco, M. A., Tamang, S., Yazdany, J., and Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern. Med.* 178 (11), 1544–1547. doi:10.1001/jamainternmed.2018.3763

Goodfellow, I., and Yoshua Bengio, A. C. (2016). *Deep learning*. Massachusetts: MIT Press.

Goodman, B., and Flaxman, S. (2017). European union regulations on algorithmic decision-making and a "right to explanation". *AI Mag.* 38 (3), 50–57. doi:10.1609/aimag.v38i3.2741

High-Level Expert Group on AI (2019). *Ethics guidelines for trustworthy AI*. Brussels: European Commission.

Ho, D. S. W., Schierding, W., Wake, M., Saffery, R., and O'Sullivan, J. (2019). Machine learning SNP based prediction for precision medicine. *Front. Genet.* 10, 267. doi:10.3389/fgene.2019.00267

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., and Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *WIREs Data Min. Knowl. Discov.* 9 (4), e1312. doi:10.1002/widm.1312

Horst, M., Davies, S. R., and Irwin, A. (2017). "Reframing science communication," in *Handbook of science and technology studies*. Editors U. Felt, R. Fouché, C. A. Miller, and L. Smith-Doerr Fourth Edition (Cambridge/London: MIT Press), 881–907.

Hudec, M., Bednárová, E., and Holzinger, A. (2018). Augmenting statistical data dissemination by short quantified sentences of natural language. *J. Official Statistics* 34 (4), 981–1010. doi:10.2478/jos-2018-0048

Inouye, M., Abraham, G., Nelson, C. P., Wood, A. M., Sweeting, M. J., Dudbridge, F., et al. (2018). Genomic risk prediction of coronary artery disease in 480, 000 adults: Implications for primary prevention. *J. Am. Coll. Cardiol.* 72 (16), 1883–1893. doi:10.1016/j.jacc.2018.07.079

Janssens, A. C. (2019). Validity of polygenic risk scores: Are we measuring what we think we are? . *Hum. Mol. Genet.* 28 (R2), R143–50.

Kerminen, S., Martin, A. R., Koskela, J., Ruotsalainen, S. E., Havulinna, A. S., Surakka, I., et al. (2019). Geographic variation and bias in the polygenic scores of complex diseases and traits in Finland. *Am. J. Hum. Genet.* 104 (6), 1169–1181. doi:10.1016/j.ajhg.2019.05.001

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., et al. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* 50 (9), 1219–1224. doi:10.1038/s41588-018-0183-z

Khera, A. V., Chaffin, M., Wade, K. H., Zahid, S., Brancale, J., Xia, R., et al. (2019). Polygenic prediction of weight and obesity trajectories from birth to adulthood. *Cell* 177 (3), 587–596. e9. doi:10.1016/j.cell.2019.03.028

Khera, A. V., Emdin, C. A., Drake, I., Natarajan, P., Bick, A. G., Cook, N. R., et al. (2016). Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N. Engl. J. Med.* 375 (24), 2349–2358. doi:10.1056/NEJMoa1605086

Knoppers, B. M., Bernier, A., Granados Moreno, P., and Pashayan, N. (2021). Of screening, stratification, and scores. *J. Personalized Med.* 11 (8), 736. doi:10.3390/jpm11080736

Kopp, W., Monti, R., Tamburrini, A., Ohler, U., and Akalin, A. (2020). Deep learning for genomics using Janggu. *Nat. Commun.* 11 (1), 3488. doi:10.1038/s41467-020-17155-y

Kozlov, M. (2022). The controversial embryo tests that promise a better baby. *Nature* 609.

Läll, K., Mägi, R., Morris, A., Metspalu, A., and Fischer, K. (2017). Personalized risk prediction for type 2 diabetes: The potential of genetic risk scores. *Genet. Med.* 19 (3), 322–329. doi:10.1038/gim.2016.103

Lambert, S. A., Abraham, G., and Inouye, M. (2019). Towards clinical utility of polygenic risk scores. *Hum. Mol. Genet.* 28 (R2), R133–R42. doi:10.1093/hmg/ddz187

Lambert, S. A., Gil, L., Jupp, S., Ritchie, S. C., Xu, Y., Buniello, A., et al. (2021). The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat. Genet.* 53 (4), 420–425. doi:10.1038/s41588-021-00783-5

Lee, A., Mavaddat, N., Wilcox, A. N., Cunningham, A. P., Carver, T., Hartley, S., et al. (2019). Boadicea: A comprehensive breast cancer risk prediction modelincorporating genetic and nongenetic risk factors. *Genet. Med.* 21 (8), 1708–1718. doi:10.1038/s41436-018-0406-9

Lewis, A. C. F., and Green, R. C. (2021). Polygenic risk scores in the clinic: New perspectives needed on familiar ethical issues. *Genome Med.* 13 (1), 14. doi:10.1186/s13073-021-00829-7

Lewis, A. C. F., Perez, E. F., Prince, A. E. R., Flaxman, H. R., Gomez, L., Brockman, D. G., et al. (2022). Patient and provider perspectives on polygenic risk scores: Implications for clinical reporting and utilization. *Genome Med.* 14 (1), 114. doi:10.1186/s13073-022-01117-8

Lewis, C. M., and Vassos, E. (2020). Polygenic risk scores: From research tools to clinical instruments. *Genome Med.* 12 (1), 44. doi:10.1186/s13073-020-00742-5

Libbrecht, M. W., and Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 16 (6), 321–332. doi:10.1038/nrg3920

Loh, P-R., Bhatia, G., Gusev, A., Finucane, H. K., Bulik-Sullivan, B. K., Pollack, S. J., et al. (2015). Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* 47 (12), 1385–1392. doi:10.1038/ng.3431

Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X., and Sham, P. C. (2017). Polygenic scores via penalized regression on summary statistics. *Genet. Epidemiol.* 41 (6), 469–480. doi:10.1002/gepi.22050

Markus, A. F., Kors, J. A., and Rijnbeek, P. R. (2021). The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *J. Biomed. Inf.* 113, 103655. doi:10.1016/j.jbi.2020.103655

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., and Daly, M. J. (2019a). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* 51 (4), 584–591. doi:10.1038/s41588-019-0379-x

Martin, A. R., Daly, M. J., Robinson, E. B., Hyman, S. E., and Neale, B. M. (2019b). Predicting polygenic risk of psychiatric disorders. *Biol. Psychiatry* 86 (2), 97–109. doi:10.1016/j.biopsych.2018.12.015

Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., et al. (2019). Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am. J. Hum. Genet.* 104, 21–34.

McInnes, G., Sharo, A. G., Koleske, M. L., Brown, J. E. H., Norstad, M., Adhikari, A. N., et al. (2021). Opportunities and challenges for the computational interpretation of rare variation in clinically important genes. *Am. J. Hum. Genet.* 108 (4), 535–548. doi:10.1016/j.ajhg.2021.03.003

McLennan, S., Fiske, A., Tigard, D., Müller, R., Haddadin, S., and Buyx, A. (2022). Embedded ethics: A proposal for integrating ethics into the development of medical AI. *BMC Med. Ethics* 23 (1), 6. doi:10.1186/s12910-022-00746-3

Meisner, A., Kundu, P., Zhang, Y. D., Lan, L. V., Kim, S., Ghandwani, D., et al. (2020). Combined utility of 25 disease and risk factor polygenic risk scores for stratifying risk of all-cause mortality. *Am. J. Hum. Genet.* 107 (3), 418–431. doi:10.1016/j.ajhg.2020.07.002

Meskó, B., Drobni, Z., Bényei, É., Gergely, B., and Győrffy, Z. (2017). Digital health is a cultural transformation of traditional healthcare. *mHealth* 3 (9), 38. doi:10.21037/mhealth.2017.08.07

Miotto, R., Li, L., Kidd, B. A., and Dudley, J. T. (2016). Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Sci. Rep.* 6, 26094. doi:10.1038/srep26094

Miotto, R., Wang, F., Wang, S., Jiang, X., and Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities and challenges. *Briefings Bioinforma.* 19 (6), 1236–1246. doi:10.1093/bib/bbx044

Mitchell, B. L., Thorp, J. G., Wu, Y., Campos, A. I., Nyholt, D. R., Gordon, S. D., et al. (2021). Polygenic risk scores derived from varying definitions of depression and risk of depression. *JAMA Psychiatry* 78 (10), 1152–1160. doi:10.1001/jamapsychiatry.2021.1988

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., and Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Soc.* 3 (2), 205395171667967. doi:10.1177/2053951716679679

Moorthie, S., Hall, A., Janus, J., Brigden, T., Babb de Villiers, C., Blackburn, L., et al. (2021). *Polygenic scores and clinical utility*. Cambridge: PHG Foundation.

Morley, J., Machado, C. C. V., Burr, C., Cowls, J., Joshi, I., Taddeo, M., et al. (2020). The ethics of AI in health care: A mapping review. *Soc. Sci. Med.* 260, 113172. doi:10.1016/j.socscimed.2020.113172

Murdoch, B. (2021). Privacy and artificial intelligence: Challenges for protecting health information in a new era. *BMC Med. Ethics* 22 (1), 122. doi:10.1186/s12910-021-00687-3

Murray, G. K., Lin, T., Austin, J., McGrath, J. J., Hickie, I. B., and Wray, N. R. (2021). Could polygenic risk scores be useful in psychiatry?: A review. *JAMA Psychiatry* 78 (2), 210–219. doi:10.1001/jamapsychiatry.2020.3042

Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366 (6464), 447–453. doi:10.1126/science.aax2342

Palk, A. C., Dalvie, S., de Vries, J., Martin, A. R., and Stein, D. J. (2019). Potential use of clinical polygenic risk scores in psychiatry – ethical implications and communicating high polygenic risk. *Philosophy, Ethics, Humanit. Med.* 14 (1), 4. doi:10.1186/s13010-019-0073-8

Pearl, J. (2009). *Causality: Models, reasoning, and inference*. 2 ed. Cambridge: Cambridge University Press.

Peters, J., Janzing, D., and Schölkopf, B. (2017). *Elements of causal inference: Foundations and learning algorithms*. Cambridge: The MIT Press.

Pivovarov, R., Albers, D. J., Sepulveda, J. L., and Elhadad, N. (2014). Identifying and mitigating biases in EHR laboratory tests. *J. Biomed. Inf.* 51, 24–34. doi:10.1016/j.jbi.2014.03.016

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., et al. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460 (7256), 748–752. doi:10.1038/nature08185

Ras, G., van Gerven, M., Haselager, P., et al. (2018). "Explanation methods in deep learning: Users, values, concerns and challenges," in *Explainable and interpretable models in computer vision and machine learning*. Editors H. J. Escalante, S. Escalera, I. Guyon, X. Baró, Y. Güçlütürk, and U. Güçlü (Cham: Springer International Publishing), 19–36.

Reid, N. J., Brockman, D. G., Elisabeth Leonard, C., Pelletier, R., and Khera, A. V. (2021). Concordance of a high polygenic score among relatives: Implications for genetic counseling and cascade screening. *Circ. Genom Precis. Med.* 14 (2), e003262. doi:10.1161/CIRCGEN.120.003262

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "Why should I trust you?," *Explaining the predictions of any classifier* in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, August 2016.

Rigby, M. J. (2019). Ethical dimensions of using artificial intelligence in health care. *AMA J. Ethics* 21 (2), E121–E124.

Roth, A., and Dwork, C. (2013). The algorithmic foundations of differential privacy. *Found. Trends® Theor. Comput. Sci.* 9 (3-4), 211–407. doi:10.1561/0400000042

Rothstein, M. A. (2015). Ethical issues in big data health research: Currents in contemporary bioethics. *J. Law Med. Ethics* 43 (2), 425–429. doi:10.1111/jlme.12258

Shen, J., Zhang, C. J. P., Jiang, B., Chen, J., Song, J., Liu, Z., et al. (2019). Artificial intelligence versus clinicians in disease diagnosis: Systematic review. *JMIR Med. Inf.* 7 (3), e10010. doi:10.2196/10010

Simmons, S., Sahinalp, C., and Berger, B. (2016). Enabling privacy-preserving GWASs in heterogeneous human populations. *Cell Syst.* 3 (1), 54–61. doi:10.1016/j.cels.2016.04.013

Sirugo, G., Williams, S. M., and Tishkoff, S. A. (2019). The missing diversity in human genetic studies. *Cell* 177 (1), 26–31. doi:10.1016/j.cell.2019.02.048

Slunecka, J. L., van der Zee, M. D., Beck, J. J., Johnson, B. N., Finnicum, C. T., Pool, R., et al. (2021). Implementation and implications for polygenic risk scores in healthcare. *Hum. Genomics* 15 (1), 46. doi:10.1186/s40246-021-00339-y

Sohail, M., Maier, R. M., Ganna, A., Bloemendal, A., Martin, A. R., Turchin, M. C., et al. (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife* 8, e39702. doi:10.7554/eLife.39702

Sud, A., Turnbull, C., and Houlston, R. (2021). Will polygenic risk scores for cancer ever be clinically useful? *npj Precis. Oncol.* 5 (1), 40. doi:10.1038/s41698-021-00176-1

Tellier, L. C. A. M., Eccles, J., Treff, N. R., Lello, L., Fishel, S., and Hsu, S. (2021). Embryo screening for polygenic disease risk: Recent advances and ethical considerations. *Genes* 12 (8), 1105. doi:10.3390/genes12081105

Torkamani, A., Wineinger, N. E., and Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* 19 (9), 581–590. doi:10.1038/s41576-018-0018-x

Treff, N. R., Eccles, J., Lello, L., Bechor, E., Hsu, J., Plunkett, K., et al. (2019). Utility and first clinical application of screening embryos for polygenic disease risk reduction. *Front. Endocrinol.* 10, 845. doi:10.3389/fendo.2019.00845

Trubetskoy, V., Pardinas, A. F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T. B., et al. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* 604, 502–508. doi:10.1038/s41586-022-04434-5

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., et al. (2021). *The ethics of algorithms: Key problems and solutions*. New York: AI & Society.

Turley, P., Meyer, M. N., Wang, N., Cesarini, D., Hammonds, E., Martin, A. R., et al. (2021). Problems with using polygenic scores to select embryos. *N. Engl. J. Med.* 385 (1), 78–86. doi:10.1056/NEJMsr2105065

Vayena, E., Blasimme, A., and Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLOS Med.* 15 (11), e1002689. doi:10.1371/journal.pmed.1002689

Vilhjalmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindstrom, S., Ripke, S., et al. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* 97 (4), 576–592. doi:10.1016/j.ajhg.2015.09.001

Wald, N. J., and Old, R. (2019). The illusion of polygenic disease risk prediction. *Genet. Med.* 21 (8), 1705–1707. doi:10.1038/s41436-018-0418-5

Wand, H., Lambert, S. A., Tamburro, C., Iacocca, M. A., O'Sullivan, J. W., Sillari, C., et al. (2021). Improving reporting standards for polygenic scores in risk prediction studies. *Nature* 591 (7849), 211–219. doi:10.1038/s41586-021-03243-6

Wang, Y., Tsuo, K., Kanai, M., Neale, B. M., and Martin, A. R. (2022). Challenges and opportunities for developing more generalizable polygenic risk scores. *Annu. Rev. Biomed. Data Sci.* 5, 293–320. doi:10.1146/annurev-biodatasci-111721-074830

Whittlestone, J., Nyrup, R., Alexandrova, A., and Cave, S. (2019). The role and limits of principles in AI ethics: Towards a focus on tensions, in the 2019 AAAI/ACM Conference, Honolulu, HI, USA, January 27–28, 2019.

Widen, E., Junna, N., Ruotsalainen, S., Surakka, I., Mars, N., Ripatti, P., et al. (2022). How communicating polygenic and clinical risk for atherosclerotic cardiovascular disease impacts health behavior: An observational follow-up study. *Circ. Genom Precis. Med.* 15 (2), e003459. doi:10.1161/CIRCGEN.121.003459

Wynne, B. (1993). Public uptake of science: A case for institutional reflexivity. *Public Underst. Sci.* 2 (4), 321–337. doi:10.1088/0963-6625/2/4/003

Xu, J., Yang, P., Xue, S., Sharma, B., Sanchez-Martin, M., Wang, F., et al. (2019). Translating cancer genomics into precision medicine with artificial intelligence: Applications, challenges and future perspectives. *Hum. Genet.* 138 (2), 109–124. doi:10.1007/s00439-019-01970-5

Zhang, T., He, C., Ma, T., Gao, L., Ma, M., and Avestimehr, S. (2021). "Federated learning for internet of things," in Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems, Coimbra, Portugal, 15 November 2021.