Data donation and data altruism to face algorithmic bias for an inclusive digital healthcare

Giulia Re Ferrè*

ABSTRACT: This paper addresses the issue of algorithmic bias in the medical field and the need to regulate data donation. Although bias and discrimination in health care are not only related to the use of algorithms but have long-standing origins and heterogeneous causes, the use of artificial intelligence could exacerbate such biases by making them structural and difficult to identify. In many cases, the presence of algorithmic bias is due to incomplete or unrepresentative datasets and the difficulty for researchers to access the data. Data donation could be a useful tool to face this phenomenon, increasing the individual's sovereignty over his or her personal data sphere and enabling an individual's participation in scientific and technological progress from a solidarity perspective.

KEYWORDS: Artificial intelligence; algorithmic bias; data donation; data altruism; healthcare

SUMMARY: 1. Introduction – 2. Bias in the medical activity: *nihil novi sub sole* – 2.1. Algorithmic bias – 3. Health data in the GDPR – 3.1. The research exemption – 4. Data donation as a tool to foster scientific progress – 4.1. What is data donation? – 4.2. Three good reasons to regulate and allow data donation – 4.3. Three main challenges of PMDD – 5. The data altruism mechanism – 6. Conclusions.

1. Introduction

The extent and the impact of the artificial intelligence (AI) revolution have emphasized the need for regulation of new tools and technology and, as stated in the explanatory memorandum to the EU Commission proposal for a Regulation on AI[1], health is one of the sectors that deserves strong attention because of the significant impact that AI could have, and the sensitive nature of the interests involved.

Machine Learning (ML) algorithms have shown to be trustful tools in the medical field, for instance diagnosing several medical conditions and treating chronic diseases. On the other side, it has been noticed that such systems could be subjected to biases related to the training dataset that could lead to discrimination of already marginalized social categories. An emblematic case is AI applied to detect skin cancer; it was found that the algorithms designed for fair skin tone color misdiagnoses for the

* *PhD student at the University of Milan in cotutelle with the German University for Administrative Sciences Speyer. Mail: giulia.referre@unimi.it. The article was subject to a double-blind peer review process.*

[1] 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 2021/0106(cod).

dark-colored skin tone and the cutaneous infections shown diversely for the dark skin.[2] Therefore, it is self-evident how these biases could result in a non-inclusive healthcare system, a lack of trust from patients towards new tools, and in the end to a reduction of the benefits that the application of AI to the healthcare sector could bring.

The Commission Proposal on AI tries to face the problem of algorithmic bias, but it has been underlined that the root of the problem is connected to the difficulties in obtaining large, diverse medical datasets. Even if the GDPR provides a research exemption, it is controversial whether its legal provisions are appropriate and sufficient to guarantee a fair trade-off between privacy and research interests. Not even the use of anonymizing procedures alone could be a satisfactory answer since anonymization could actually reduce the quantity and quality of data and moreover, makes it impossible to verify them and prevent linking of different datasets on relevant issues such as socioeconomic indicators.

In this context, the opportunity to regulate data donation, that is, the donation of data after one's death for research purposes, seems to become increasingly important; in fact, as it has been underlined, it is easier to donate our bodies to science rather than our data, with the result of a massive amount of wasted precious medical data.

In recent years, scholars have outlined the characteristics and issues underlying data donation, but in adopting the Data Governance Act, the European Union instead introduced a so-called data altruism mechanism, i.e., the possibility of sharing one's data while still alive through bodies designated for this purpose, thus not accepting the proposals of that part of the doctrine that suggested that possibility of *post-mortem* donation should be provided for in the first place.

The paper aims to give an overview starting from the problematics related to algorithmic bias in the healthcare sector and to the difficulties accessing medical datasets and focuses on the possibility of regulating data donation. Regulating the posthumous donation of personal data clearly raises not only legal and technical challenges, but also ethical questions that need to be addressed. Is in the first place the word "donation" the correct term to use or does it imply the ownership of personal data and the transfer of a fundamental right? Which are the similarities to organ donation and to what extent can we compare these two hypotheses? Is the will of the patient-donor the only one that matters? What is the role and the relevance of the needs of other patients that could directly benefit from large-scale data collection? In fact, to outline an effective regulation it is first necessary to understand the nature of such a donation, the motivation behind that will, the concerns and the interests at stake and ultimately the role of individual sovereignty on data in relation to the public interest of developing an inclusive healthcare system. The last part of the article is focused on the main features of data altruism as recently introduced by the Data Governance Act and that will be detailed more specifically by the European Health Data Space, if it is approved.

[2] S. Naresh Kumar, B. Mohammed Ismail, *Systematic investigation on Multi-Class skin cancer categorization using machine learning approach* in *Materials Today: Proceedings*, 2020.

ISSN 2284-4503

2. Bias in the medical activity: *nihil novi sub sole*

The presence of biases in the medical field is nothing new and it is not only linked to the use of artificial intelligence. In fact, discrimination against several categories of patients has ancient roots and heterogenous causes.

First, it is necessary to define what discrimination in the healthcare sector means. In a recent study, discrimination in the healthcare setting is defined as "negative actions or lack of consideration given to an individual or group that occurs because of a preconceived and unjustified opinion".[3] The definition could be broadened in order to include also algorithmic discrimination and therefore we could identify the presence of a bias when the same inputs (e.g. symptoms) lead to different and unfair outputs (e.g. diagnosis) just on the ground of the belonging of the patient to a minority group.

Medical biases have been observed against women[4], ethnical groups,[5] low-income subjects,[6] transgender and queers.[7] It has, for example, been argued that even if men and women experience pain in different ways, with women reporting more frequent and greater level of pain, female patients are more likely to be undertreated because health-care providers tend to discount women self-reports of pain at least until there is objective evidence of the cause;[8] moreover several studies showed that woman with chronic pain are rather assigned psychological than somatic causes for their pain and are perceived as "hysterical", "emotional", "complaining", "not wanting to get better", "malingerers" and "fabricating the pain".[9]

Even though medical discrimination has a long history, there is now a pressing need to specifically address the topic of algorithmic bias, because if the discrimination carried out by a human being, albeit harmful, can be identified as a pathological moment of relationship between the patient and the healthcare provider, the algorithmic discrimination may amplify and exacerbate inequalities,[10] making

[3] BM. Togioka, D. Duvivier, E. Young *Diversity and Discrimination In Healthcare* in *StatPearls*, *Treasure Island (FL)*, January 2022.

[4] J.A. Marcum, *Clinical Decision-Making, Gender Bias, Virtue Epistemology, and Quality Healthcare,* in *Topoi,* 36**,** 2017, 501–508.

[5] D.R. Williams, R. Wyatt, *Racial Bias in Health Care and Health: Challenges and Opportunities*, in *JAMA,* September 2015, 314.

[6] L. Hoyt D'Anna, M. Hansen, B. Mull, C. Canjura, E. Lee, S. Sumstine, *Social Discrimination and Health Care: A Multidimensional Framework of Experiences among a Low-Income Multiethnic Sample* in *Social Work in Public Health*, 2018; O. N. Okoro, L. A Hillman, A. Cernasev, *"We get double slammed!": Healthcare experiences of perceived discrimination among low-income African-American women*, in *Womens Health* (Lond), 2020.

[7] K.D. Jaffee, D. Shires, D. Stroumsa, *Discrimination and Delayed Health Care Among Transgender Women and Men,*: *implications for Improving Medical Education and Health Care Delivery,* in *Medical Care*, 54, 2016; A. S. Deirdre, K. Jaffee, *Factors Associated with Health Care Discrimination Experiences among a National Sample of Female-to-Male Transgender Individuals,* in *Health & Social Work*, 40, 2015, 134–141.

[8] D. Hoffmann, A. Tarzian, *The Girl Who Cried Pain: A Bias against Women in the Treatment of Pain,* in *Journal of Law*, *Medicine & Ethics*, 2001

[9] A. Samulowitz et al. *'Brave Men' and 'Emotional Women': A Theory-Guided Literature Review on Gender Bias in Health Care and Gendered Norms towards Patients with Chronic Pain* in *Pain Research and Management*, 2018

[10] I. Straw, *The automation of bias in medical Artificial Intelligence (AI): Decoding the past to create a better future,* in *Artificial Intelligence in Medicine*, 110, 2020.

them structural and more difficult to detect. In fact, as underlined by the WHO in its guidance on ethics and governance on AI for health,[11] societal bias and discrimination are often replicated by AI systems. Adjusting the above given definition, one could refer to algorithmic bias as a systematic error caused by the training dataset, that create unfair outcomes. Several classifications of such biases have been proposed, based on the source of the discrimination. We may adopt a macro division between human-based biases, which are a replication of human prejudices that causes algorithms to mirror historical inequalities or a mistake in the algorithm design[12] and dataset-based biases, which are caused by imbalanced or misrepresentative training data.[13] For the purpose of this contribution only the second type will be taken into consideration. In fact, most of the AI bias reported so far are imputable to the lack of some patient-groups data: the underrepresentation of certain groups results in low performance of the tools. Considering that the two main areas of machine learning application in healthcare are medical diagnosis and prediction of health risks, to use a biased algorithm means to obtain misdiagnosis and misprediction for those patients' categories that are also statistically more subject to human discrimination.[14] Moreover, it is necessary to underline that a poor functioning of ML algorithms does not only concern social marginalized people but could strongly affect also patient with rare disease, that are considerably vulnerable to paucity of data.[15] Consequently, while the artificial intelligence potential could provide personalized, more accurate and effective care for the entire society, the inefficient use of technological tools could result in a non-inclusive healthcare system, where many vulnerable categories would experience distrust[16] and therefore in a denial of the right to get proper care.[17]

[11] *Ethics and governance of artificial intelligence for health: WHO guidance*, 2021.

[12] Z. Obermeyer et al., *Dissecting racial bias in an algorithm used to manage the health of populations*, in *Science* 2019, 447-453: the study find evidence of racial bias in one widely used algorithm in the U.S. health care system. "The bias arises because the algorithm predicts health care costs rather than illness, but unequal access to care means that we spend less money caring for Black patients than for White patients. Thus, despite health care cost appearing to be an effective proxy for health by some measures of predictive accuracy, large racial biases arise."

[13] N. Norori, Q. Hu, F. M. Aellen, F. D. Faraci, A. Tzovara, *Addressing bias in big data and AI for health care: A call for open science* in *Patterns*, October 2021.

[14] T. Grote, G. Keeling, *On Algorithmic Fairness in Medical Practice* in *Cambridge Quarterly of Healthcare Ethics*, January 2022.

[15] N. Hasani, F. Farhadi, M. A. Morris, M. Nikpanah, A. Rahmim, Y. Xu, A. Pariser, M. T. Collins, R. M. Summers, E. Jones, E. Siegel, B. Saboury, *Artificial Intelligence in Medical Imaging and its Impact on the Rare Disease Community: Threats, Challenges and Opportunities,* in *PET Clinics*, 17, 2022, 13-29.

[16] Several studies have been carried out on the topic of trust and distrust in the healthcare personnel-patient relationship; this problem should be addressed also when the use of AI is involved, since distrust could be experienced also against technologies. S.D. Goold, *Trust, distrust and trustworthiness,* in *Journal of general internal medicine*, January 2002, 79-81; YY. Lee, JL Lin*. Linking patients' trust in physicians to health outcomes,* in *British Journal of Hospital Medicine,* 2008.

[17] On issues related to equity in benefiting from data sharing and open science see C. Staunton, C.A. Barragán, S. Canali, C. Ho, S. Leonelli, M. Mayernik, B. Prainsack, A. Wonkham *Open Science, Data Sharing and Solidarity: Who Benefits?* in *History and Philosophy of the Life Sciences*, 2021; on the risk of underutilization of artificial intelligence in the medical field, see U. Pagallo*, Il dovere alla salute: Sul rischio di sottoutilizzo dell'intelligenza artificiale in ambito sanitario*, 2022.

2.1. Algorithmic bias

As mentioned, since most of the bias derive from the training dataset, AI model should be developed from a wide variety of data, able to reflect as much as possible all the different social categories. The implementation of this principle meets the obstacle of the scarcity of available data related to some socioeconomics groups, that results in a lack of data diversity.

An emblematic case is AI applied to detect skin cancer; it was found that the algorithms designed for fair skin tone colour misdiagnoses for the dark-coloured skin tone and the cutaneous infections shown diversely for the dark skin.[18] Some researchers have underlined that the root of the problem is connected to the difficulties in obtaining large, diverse medical datasets. Other examples can be found in cardiovascular medicine,[19] since women and minorities are historically under-represented[20] as well as in oncology, where clinical trial data traditionally underrepresent subgroups such as adolescents and young adults, women, ethnic minorities and elderly.[21]

Considering that in the nearest future the use of artificial intelligence in medicine will find more and more space, the problematic of algorithmic biases needs to be addressed not only from a technical point of view, but also from a legal perspective. The provisions of the new EU Regulation Proposal on AI seem to head in the right direction but could be insufficient. In fact, art. 10 (par. 3), entitled "Data and data governance", states that training, validation and testing data sets shall be relevant, representative, free of errors and complete. Moreover, they shall have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons on which the high-risk AI system is intended to be used.

The request of complete and representative training datasets, however valuable, could stay unimplemented unless accompanied by a regulation that guarantees the availability of a wide data variety. In fact, while it is true that representative datasets are preconditions for non-discriminatory AI systems, the lack of data from certain subgroups in clinical trials,[22] research and studies represent a substantial impediment. It should also be considered that another obstacle could be represented by the privacy legislation; therefore, a brief analysis of the GDPR provisions seems needed.

---

[18] S. Naresh Kumar, B. Mohammed Ismail, *Systematic investigation on Multi-Class skin cancer categorization using machine learning approach* in *Materials Today: Proceedings*, 2020.

[19] E. Tat, DL. Bhatt, MG. Rabbat, *Addressing bias: artificial intelligence in cardiovascular medicine* in *Lancet Digit Health*, 2020.

[20] Enrollment of women in randomized clinical trials has increased over time but remains low: C. Melloni, JS. Berger, TY. Wang, F. Gunes, A. Stebbins, KS. Pieper, RJ. Dolor, PS. Douglas, DB. Mark, LK. Newby, *Representation of women in randomized clinical trials of cardiovascular disease prevention*, in *Circulation Cardiovascular Quality Outcomes*, 2010.

[21] I.S. Chua, M. Gaziel-Yablowitz, Z.T. Korach, K.L. Kehl, N.A. Levitan, Y.E. Arriaga, G.P. Jackson, D.W. Bates, M. Hassett, *Artificial intelligence in oncology: Path to implementation,* in *Cancer Medicine*, 10, 2021, 4138-4149.

[22] On the topic of lack of representation of certain groups in clinical trials: T. Zhang, W. Tsang, HC. Wijeysundera, DT Ko, *Reporting and representation of ethnic minorities in cardiovascular trials: a systematic review* in *American Heart Journal*, 2013; K. Kwiatkowski, K. Coe, JC. Bailar, GM. Swanson, *Inclusion of minorities and women in cancer clinical trials, a decade later: Have we improved?,* in *Cancer*, 2013. On possible causes and solution to overcome the lack of diversity: LT. Clark, L. Watkins, IL. Piña, M. Elmer, O. Akinboboye, M. Gorham, B. Jamerson, C. McCullough, C. Pierre, AB. Polis, G. Puckrein, JM. Regnante *Increasing Diversity in Clinical Trials: Overcoming Critical Barriers,* in *Current Problems in Cardiology*, 2019.

Call

3. Health data in the GDPR

The General Data Protection Regulation represents the cornerstone of the new data governance regime in the European Union which aim to balance the protection of individual privacy and the promotion of a thriving European data economy.[23] Its relevance in the *eHealth* and *mHealth* sector is therefore self-evident.[24]

The GDPR offers a definition of health data, stating that "data concerning health" means personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status.[25] Consequently, health data belong to the category of personal data when they reveal information relating to the past, current, or future physical or mental health status of the data subject. According to recital n. 35 "*this includes information about the natural person collected in the course of the registration for, or the provision of, health care services as referred to in Directive 2011/24/EU of the European Parliament and of the Council (1) to that natural person; a number, symbol or particular assigned to a natural person to uniquely identify the natural person for health purposes; information derived from the testing or examination of a body part or bodily substance, including from genetic data and biological samples; and any information on, for example, a disease, disability, disease risk, medical history, clinical treatment or the physiological or biomedical state of the data subject independent of its source, for example from a physician or other health professional, a hospital, a medical device or an in vitro diagnostic test.* These definitions should be read in conjunction with recital n. 51 which reproposes the notion of 'sensitive data',[26] consisting in data that because of their sensitive nature merit a specific protection *as the context of their processing could create significant risks to the fundamental rights and freedoms*." Anonymized data, on the other hand, do not fall in the field of application of the GDPR, as they do not relate to an identified or identifiable natural person; moreover, accordingly to Recital 27, the Regulation does not apply to the personal data of deceased people.

3.1. The research exemption

A specific protection is then realized by art. 9, which opens with a general prohibition of processing special categories of personal data,[27] including health data: "*Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural*

[23] L. Marelli, E. Lievevrouw, I. Van Hoyweghen, *Fit for purpose? The GDPR and the governance of European digital health,* in *Policy Studies*, 2020.

[24] On the growing relevance of mHealth in telemedicine and telemonitoring and issues related to data protection, see G. Bincoletto, *mHealth app per la televisita e il telemonitoraggio. Le nuove frontiere della telemedicina tra disciplina sui dispositivi medici e protezione dei dati personali,* in *BioLaw Journal,* 2021.

[25] GDPR, art. 4 par. 1.

[26] Nevertheless, note that the GDPR abandons the expression of sensitive data and replaces it with that of special categories of data.

[27] Art. 9 par 1 GDPR "Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited".

person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited."

Nevertheless, the provision contained in par. 1 is balanced by a list of conditions under which the processing is permitted and that could be divided into two main groups: the general scenario in which the Regulation requires an informed, free, explicit consent given for one or more specified purpose[28] and specific hypotheses that allowed the processing without the consent of the data subject.

It is controversial whether the consent to process special categories of data for research purpose should be specific, or if a broad consent could be lawfully used. Recital 33 recognizes that it is often not possible to fully identify the purpose of personal data processing in the field of scientific research at the time of data collection and therefore states that data subjects should be allowed to give their consent to certain areas of scientific research. This is particularly true for example in biobanking research where the long-run reusability of the biological resources is essential.[29] However, it seems to be excluded the possibility to give an omnibus consent for any research purpose.[30]

The regulation provides a so-called research exemption and states that special categories of personal data, can also be processed without consent either on the legal basis of public interest, or for research purpose. The exemption contained in art. 9 should be read in conjunction with art. 89 GDPR, which requires that processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards for the rights and freedoms of the data subject, through technical and organizational measures such as pseudonymization. Even if safeguards are required, derogations from the rights referred to in art. 15, 16, 18, 19, 20 and 21 are possible, both on a Union and on a national level. The possibility for member states to introduce derogations could lead to an undesirable fragmentation of the legal framework and could increase difficulties for EU cross-border health projects.

There are conflicting positions on the GDPR provisions adequacy regarding the research exemption and some scholars have argued that the data protection level could be lowered by the possibility for the EU law or for Member states to provide derogations to some rights.[31] Moreover, the exceptions have proven in practice elusive, and many aspects would need further clarification.[32]

From another point of view, it has been argued that the implementation of the GDPR has brought several obstacles to secondary research and the distinction between "pseudonymous data", which are still considered personal data and therefore fall under the GDPR, and "anonymous" data has raised some concerns. In fact, while traditionally data used in secondary research are key-coded and there-

[28] M. FARINA, *Il cloud computing in ambito sanitario tra security e privacy*, 2019, 28.

[29] E. S. DOVE, *The EU General Data Protection Regulation: Implications for International Scientific Research in the Digital Era: Currents in Contemporary Bioethics*, in *Journal of Law, Medicine and Ethics*, 2018.

[30] M. DONNELLY, M. MCDONAGH*, Health Research, Consent and the GDPR Exemption*, in *European Journal of Health Law*, 2019.

[31] K. PORMEISTER, *Genetic data and the research exemption: is the GDPR going too far?,* in *International Data Privacy Law*, 7, 2017.

[32] J. MÉSZÁROS, C. HO, *Big Data and Scientific Research: The Secondary Use of Personal Data under the Research Exemption in the GDPR,* in *Hungarian Journal of Legal Studies*, 4, 2018, 403–419.

fore pseudonymized, the obligation to carry out an anonymization process could influence the scientific quality of the data.[33] In fact, in some areas of health research, anonymization cannot be achieved without fundamentally undermining the quality and contribution of the research.[34]
Moreover, it has been pointed out that two key features of the GDPR, such as data minimization and purpose limitation do not fit the peculiarities of Big Data technologies and fail to take into account several aspects. It should, in fact, be considered that AI is used also to discover new and unpredictable patterns and correlations among the gathered data, that consequently cannot be collected only for specific and limited purpose.[35] Also the use of anonymized data could reduce the potentiality of the use of new technology tools, since it prevents linking of different dataset on relevant issues such as socioeconomic indicators.[36]
Ultimately the need for diverse, representative datasets in order to avoid algorithmic bias and the scarcity of health data related to certain categories of patients, together with the huge amount of data gathered not only by healthcare structures, but also through apps and wearable devices could give the chance to start a discussion on the opportunity to regulate data donation.

4. Data donation as a tool to foster scientific progress

The topic of data donation is for sure very complex, and the aim of this paper is not to provide a full analysis, rather to identify some of the main challenges and to offer a few general reflections. Although the matter is relatively new, some studies have been conducted, the most complete and critical, both from an ethical, philosophical and legal point of view, being the project developed at the Digital Ethics Lab at the Oxford Internet Institute, University of Oxford, and funded by Microsoft Research.[37]

4.1 What is data donation?

It is first of all necessary to define what is meant by "data donation": it mainly refers to the donation of personal data after death and describes the possibility for a subject to allow access to health data for research purpose (also known as PMDD – posthumous medical data donation).
While it is quite easy to choose to donate blood, tissues, organs and even the entire corpse for scientific purpose, there is yet no legal framework that allows the donation of medical data and this basically results in the loss of a huge amount of precious health data that people gather during their entire existence.
The topic has started to be discussed very recently and an interesting debate emerged, starting from the chosen term of donation itself. In fact, it has been argued that it would be preferable to talk about

[33] D. Peloquin, M. DiMaio, B. Bierer, M. Barnes, *Disruptive and avoidable: GDPR challenges to secondary research uses of data,* in *European Journal of Human Genetics,* 2020.
[34] M.Donnelly, M. McDonagh *Health Research, Consent and the GDPR Exemption,* in *European Journal of Health Law*, 2019.
[35] M. Favaretto et al., *Big Data and discrimination: perils, promises and solutions. A systematic review,* in *Journal of Big Data*, 2019.
[36] J. Rumbold, B. Pierscionek, *Contextual Anonymization for Secondary Use of Big Data in Biomedical Research: Proposal for an Anonymization Matrix,* in *JMIR Medical Informatics*, 2018.
[37] J. Krutzinna, L. Floridi (ed.), *The ethics of data donation*, 2019.

"data sharing", "data altruism" or "data solidarity",[38] since the concept of donation implies the ownership of the personal data and consequently the ownership transfer, which is not in line with the European vision of data protection and privacy as fundamental right and not as property asset. The institution of data donation also has points in common with the so called "data philanthropy", that is the donation of data by private companies for altruistic purposes, although the two hypotheses do not overlap.[39] The nature of rights on data is still controversial but most authors suggest that the general principles of European law, the principles deriving from the member states constitutions and the current legal provisions are incompatible with the concept of data ownership.[40]

The legal instrument of donation, rooted back in the Roman law, involves the enrichment of a subject with the correspondent impoverishment of the donor, who acts with spirit of liberality (*animus donandi*), but not obligatorily for solidarity reasons, and hence can also recall financial interests. Even if this is true, it should be pointed out that the term donation is commonly accepted with regards to the disposal of *res extra commercium* such as body parts, organs, tissues, blood, both *ante*- and *post-mortem*, even if in a non-technical sense.[41] Therefore, to coin new terms, with broader meaning but foreign to the common lexicon, seems useless and would probably, at least at the beginning, lack the evocative power of the word donation.

4.2. Three good reasons to regulate and allow data donation

Some studies have tried to investigate the existence of a willingness to donate (and more generally to share) health data and the reasons behind that will. The papers analyzed so far show a general interest and a good and proactive attitude of patients towards the possibility of sharing personal data for research intended to lead to public good;[42] moreover, it has been underlined that participants associated their motivations to donate their personal data with self-benefit and concern for others.[43]

Even in very sensitive field, like mental health, where the social stigma is still very high, it has been shown that, if adequate safeguards and transparency measures are guaranteed, patients could have a positive perception about sharing their data in hopes to contribute in better policy and care.[44] Similarly, rare disease patients, regardless of the severity of their disease and their socio-demographic profile, are clearly supportive towards data sharing to foster research and improve healthcare.[45]

[38] Health and Food Safety Directorate-General, *Assessment of the EU Member states' rules on health data in the light of GDPR*, 2021, 113.

[39] M. Taddeo, *Data philanthropy and individual rights* in *Mind and Machines,* 2017.

[40] P. Hummel, M. Braun, P. Dabrock, *Own Data? Ethical Reflections on Data Ownership,* in *Philosophy & Technology*, 2021.

[41] S. Ruscica, *I diritti della personalità*, 2013.

[42] M.A.R. Bak, M.C. Ploem, H. Ateşyürek, et al. *Stakeholders' perspectives on the post-mortem use of genetic and health-related data for research: a systematic review,* in *European Journal of Human Genetics*, 2020.

[43] A. Skatova, E. Ng, J. Goulding, *Data Donation: Sharing Personal Data for Public Good?,* in *Conference paper of the Digital Economy All Hands Meeting*, 2014.

[44] E. Satinsky, C. Driessens, D. Crepaz-Keay, AA. Kousoulis, *Mental health service users' perceptions of data sharing and data protection: a short qualitative report,* in *BMJ Health & Care Informatics*, 2018.

[45] S. Courbier, R. Dimond, V. Bros-Facer, *Share and protect our health data: an evidence-based approach to rare disease patients' perspectives on data sharing and data protection - quantitative survey and recommendations*, in *Journal of Rare Diseases,* 2019.

The above-mentioned studies referred to data donation during the life of the data subject, but it should be noted that another research has highlighted that patients could be more willingly to share digital data after death.[46] 40% of the participants agreed to share data from wearable devices,[47] but the percentage has risen to almost 80% in case of *post-mortem* donation; the donation of genetic data followed the same trend.

Consequently, the first reason to support the enabling of PMDD is linked to the altruistic attitude underlying the donation will, that should not be frustrated, because it can be recognize as the expression of the fundamental principle of solidarity, which is one of the cornerstones of European law and is set out in many of the member states Constitutions.

Secondly, data donation could contribute to the development of participation of citizens in scientific progress, a right recognized in art. 27 par. 1 of the Universal Declaration of Human Rights, which states that ‹‹*everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits*››. On this aspect, it has been argued that data donation, not only *post-mortem* but also during the lifetime, can be a model for health-focused citizen science, which transform the study of human health and behavior, if not only data gathered through Electronic Health Record (EHR) and in clinical settings, but also those generated by wearables, smartphones and the Internet of Things (IOT) are made available for research.[48]

Third reason that stands for the necessity of a regulation for PMDD involves sovereignty over personal data. The concept of sovereignty should not be intended as absolute and unlimited power, nor just in a negative sense. If it is true that sovereignty implies the power to exclude others, to limit the power of other subjects over personal data, this aspect does not exhaust the whole meaning. Indeed, sovereignty consists also of a positive and active component, which is expression of substantial self-determination. As brilliantly sustained by Hummel et al. ‹‹*sharing one's personal data can constitute meaningful advances and reinforcements of the social structures in which the individual seeks to realize positive aspects of her sovereignty*››.[49] In the era of IoT and society datafication, the personality of human beings consists also of a digital identity and expresses itself also on a data level. To guarantee sovereignty especially over sensitive data and to allow people to enhance their data after their death, means in the end to promote self-determination in a solidarity view, consistently with what already happened with the possibility to decide what to do with our corpse, organs and tissues.

4.3. Three main challenges of PMDD

The first criticism highlighted against the possibility of PMDD is related to the fact that often health data cannot be considered just data of the patients but involve many other subjects, such as relatives and consequently it would be more correct to refer to these data with the expression "data about the

[46] E. Seltzer et al. *Patients' willingness to share digital health and non-health data for research: a crosssectional study,* in *BMC Medical Informatics and Decision Making,* 2019.

[47] It is interesting to notice that 60% to 80% of participants declared that data gathered from wearables devices were more health-related than genetic data.

[48] M. Bietz, K. Patrick, C. Bloss, *Data Donation as a Model for Citizen Science Health Research,* in *Citizen Science: Theory and Practice*, 2019.

[49] P. Hummel, M. Braun, P. Dabrock, *Data Donations as Exercises of Sovereignty*, in J. Krutzinna, L. Floridi (eds.), *The Ethics of Medical Data Donation.*

patient" and most importantly family members should be granted the right to oppose to the donation, if this could lead to any harm to their privacy or anyhow to their personal sphere. Indeed, some kind of data are not related just to the data subject but also to relatives; this is particularly true for genetic data for examples, which are also almost impossible to anonymize.[50] The point deserves for sure some investigation, but it should also be considered that the possibility to donate blood and tissues to biobanks already exists, and these biological materials contains genetic data. Therefore, if a privacy problem exists in relation to genetic data donation, the same issue should be relevant with regards to biobanks too. Moreover, safeguards for privacy of the donor and of the family will clearly be set out, at least through pseudonymization and confidentiality duties, which do not expire with the death of the patient.

This consideration gives also the chance to underline that, even if the traditional parallelism that is so far being proposed in most of the studies is between data donation and organs donation, actually, for at least two reasons, it seems more correct to compare data donation to corpse and tissue donation to science: first, data would be donated to research to foster scientific progress and not to an individual; secondly while organs donation is possible only under very strict conditions and circumstances, data donation could, at least theoretically, be possible for everyone and also the re-use of the same data is also desirable. However, the existing framework for organs, corpse and specimen donation could be a useful guideline.

The second main issue that need to be addressed is the legal mechanism to allow data donation, and in particular the alternative between an opt-in and an opt-out system. The topic is still very controversial and different positions have been sustained, but an opt-in system with the possibility to exclude the donation for specific research uses seems, in the opinion of the writer, to be the most adequate option for many reasons. First of all, as highlighted before, studies have shown a strong will and positive attitude of patients to voluntarily donate health data. Therefore, the first choice should fall on a system that allows to fully express this will, also according to the proportionality principle; only where an opt-in system proves to be ineffective, a more pressing and coercive system could be justified. Moreover, a choice imposed by the legislator without a sensitization process could have the opposite effect, i. e to generate mistrust and skepticism, which in the end could lead to a failure analogous to what happened with the English program care.data.[51] It would also be more coherent with the existing European legal framework on data protection to base the regulation on explicit consent, even though a broader one. As a matter of fact, a specific consent would not suit to the peculiarity of PMDD, considering that it can often be impossible to exactly predict for which research the data will be used. Nonetheless, the possibility to exclude some kind of uses could be advisable; for instance, some studies have shown reluctance towards research use of the donated data by commercial institutions.[52]

Third criticism is related to the need to guarantee inclusion. The first paragraphs of the paper are dedicated to algorithmic bias and it has been argued that the most of these biases are due to a lack of

[50] MD. Sorani, JK. Yue, S Sharma, GT. Manley, AR. Ferguson*, Genetic data sharing and privacy,* in *Neuroinformatics,* 2015.

[51] J. Meszaros, C. Ho, B*uilding trust and transparency? Challenges of the opt-out system and the secondary use of health data,* in *Medical Law International*, 2019.

[52] G. Richter, C. Borzikowsky, W. Lesch, *et al. Secondary research use of personal medical data: attitudes from patient and population surveys* in *The Netherlands and Germany,* in *European Journal of Human Genetic,* 2021.

diversity in the training datasets and therefore allowing people to donate data related to health and gathered not only through EHR, but also thanks to IoT, wearable technologies and smartphone could help provide more representative, diverse and complete datasets. Consequently, the result cannot be reached if some social categories are excluded or limited in the access to new technologies. The so-called digital divide is a well-known problem that needs to be addressed not only to guarantee that elderly people, that are normally less confident with new tools, are not cut off from the benefit that a digital healthcare could bring, but also to overcome an economic digital divide, which results in the exclusion of those people who cannot financially afford the purchase of the newest technologies. If healthcare is becoming more and more digital and if the role of *mHealth* is getting increasingly central, then a reimbursement system for apps and wearables is necessary.[53]

5. The data altruism mechanism

An opt-in consent-based system has been adopted in the Data Governance Act (DGA) approved on the 16th of May 2022. As stated in the Recitals, the Regulation aims at promoting the availability of data and build a trustworthy environment to facilitate data use for research and innovation; the EU framework should have the objective of building trust among individuals and undertakings in relation to data access, control, sharing, use and re-use.

One of the mechanisms set up to facilitate subjects' control over their own data is so-called data altruism, for which the European legislation outlines just the basic features, leaving the technical and organizational implementation to member states.[54] The relevance of altruistic data sharing is emphasized not only in regard to the control and governance of data by data owners, but also in relation to the impact that the formation of large pools of available data can have on the development of data analytics and machine learning systems.

It is first necessary to clarify what is meant by data altruism. The regulations provide a definition of it in Art. 2 p.16, stating that it is the "voluntary sharing of data on the basis of the consent of data subjects to process personal data pertaining to them, or permissions of data holders to allow the use of their non-personal data without seeking or receiving a reward that goes beyond compensation related to the costs that they incur where they make their data available for objectives of general interest as provided for in national law, where applicable, such as healthcare, combating climate change, improving mobility, facilitating the development, production and dissemination of official statistics, improving the provision of public services, public policy making or scientific research purposes in the general interest". As noted above, the regulation does not provide a detailed framework, nor does it contain specific provisions for health data, which are instead contained in the proposal for the European Health Data Space (EHDS).[55] Despite the definition given in the regulation there remains ample room for doubt

[53] An interesting example of legislation introducing the possibility of reimbursement for apps and devices is the German Digital-versorgung-gesetz (DVG).

[54] Recital 45.

[55] The European Health Data Space (EHDS) is the first proposal for a domain-specific common European data space, which falls within the framework of the European data strategy. According to the explanatory memorandum, the EHDS will address health-specific challenges to electronic health data access and sharing and establish a specific mechanism for data altruism in the health sector.

about the nature of data altruism and in fact even the European Data Protection Board (EDPB) and the European Data Protection Supervisor (EDPS) have pointed out that "*the concept of data altruism is still not clearly and sufficiently defined. In particular, it is unclear whether the consent envisaged in the Proposal corresponds to the notion of 'consent' under the GDPR, including the conditions for the lawfulness of such consent. In addition, it is unclear the added value of data altruism, taking into account the already existing legal framework for consent under the GDPR, which provides for specific conditions for the validity of consent*".[56]

This new consent model operates through the establishment of data altruism organizations, as provided for in Chapter IV, which must, among other things, operate on a nonprofit basis and be legally independent of any entity operating for profit. It is the responsibility of the organization, pursuant to Article 21, to put in place the protections necessary to safeguard the rights of data subjects by informing them of the public interest objectives and, where appropriate, the determined, explicit and legitimate purpose for which the data are to be processed. It should therefore be noted that even in this mechanism, the possibility of general consent to data use seems to be excluded and in fact the organization cannot use the data for any other purpose other than those of general interest for which the data subjects or data owners consent to the processing (art. 21 par. 2).

Another problematic aspect to be investigated is the possibility of revoking consent; in fact, it is the responsibility of the organization to provide not only tools for obtaining data subjects' consent or authorizations for data processing made available by data controllers, but also tools for the easy revocation of such consent or authorization. From a practical point of view, the revocation of consent, although in harmony with the provisions of the GDPR, poses several critical issues especially considering that the goal of the legislation is to allow broad reuse of data and for a potentially indefinite and unlimited number of researches.[57]

Finally, it is relevant to note that Article 25 of the aforementioned regulation introduces a common European model of consent for data altruism. Consistent with the exclusion of an omnibus consent, the form should follow a modular approach and allow customization for specific sectors and for different purposes. Moreover, where personal data are provided, the European data altruism consent form shall ensure that data subjects are able to give consent to and withdraw consent from a specific data processing operation in compliance with the requirements of the GDPR.

So, while data altruism has many points of contact with data donation, it differs from it primarily because it concerns the sharing of data of people who are still alive, allows for the revocation of consent, and ultimately stands simply as a new model of consent to the processing and reuse of personal data, moving away even through the choice of a different name from a proprietary view of data.

[56] EDPB-EDPS Joint Opinion 03/2021 on the Proposal for a regulation of the European Parliament and of the Council on European data governance (Data Governance Act).

[57] As noted above, data sharing (both in the form of data donation and data altruism) has several points of contact with the donation of tissue and biological material to biobanks (see G.M. VERGALLO, *Campioni biologici da vivente capace e biobanche di ricerca: raccolta, utilizzo e circolazione,* in *European Journal of Privacy Law & Technologies*, 2021); also with regard to withdrawal of consent and related issues, the legislation and doctrine that has developed since Directive 2004/23/ec could be a good starting point for the development of consistent and informed rules (see A. BERNES*, Dati e ricerca genetica.Dalla tutela individuale alla gestione procedural,* in *BioLaw Journal*, 2022).

Conflicting positions have been formed on the data altruism system, especially in the biomedical research field:[58] on one hand it gives voice to some instances that also underlie the regulation of data donation, promoting the altruistic participation of individuals in technological and scientific research progress and enhancing the individual's sovereignty over his or her sphere of data; on the other side the success of this mechanism is by no means certain and places heavy implementation burdens on member states, including the establishment of authorities responsible for maintaining registries of data altruism organizations and charged with overseeing such organizations. The adoption of minimal and non-detailed legislation, thus leaving wide room for member states, could lead to extremely uneven and fragmented implementation.

Moreover, the essential rules dictated by the DGA do not specifically address the sharing of health-related data, which would require ad hoc legislation because of their relevance and particularly sensitive nature. A more specific regulation could be introduced only if and when the proposal for the European Health Data Space is adopted. However, even in this case it will be necessary to harmonize the two regulations; in fact, Article 40 of the Proposal for the EHDS defines and provides for the data altruism in the context of health but as pointed out by the EDPB EDPS in a joint opinion "*the provision is unclear, particularly with regards to the interplay with the respective provision introduced by the DGA.*"[59]

Finally, it is also necessary to consider that a decisive role will be played by individuals' trust in the data altruism system, as it is precisely widespread distrust and skepticism that have led to the failure of some similar projects such as care data mentioned earlier.

6.Conclusions

If the existence of biases in the medical activity is nothing new, the advent of AI and use of ML algorithms amplify this issue that need to be address with no further delay. Biases in human activities are for sure detrimental, but algorithmic biases amplify them and can lead to systematic discrimination of certain categories on a massive scale. Since one of the main bias causes is the lack of diversity and completeness in training datasets, one of the instruments that could be considered for solving the problem is data donation. The goal is to develop representative and comprehensive training datasets, with the inclusion of data from those categories of patients that are traditionally underrepresented. Through data donation it would be possible to enhance the medical data of individuals not only collected through clinical examinations, but also real-world data collected through apps and wearable devices. Obviously, this would imply the need for regulation and administrative implementation to maximize the usefulness of donated data on the one hand while still protecting the privacy of the donor and family members.

The donation of medical data is, indeed, not free of challenges because of the sensitive interests at stake, but its regulation could be a way to guarantee self-determination in a solidaristic perspective while contributing to the scientific progress. It would, in fact, fulfill an altruistic motivation that many

[58] M. SHABANI, *The Data Governance Act and the EU's move towards facilitating data sharing,* in *Molecular Systems Biology,* 17, 2021.

[59] EDPB-EDPS Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space.

studies have highlighted, increase citizen participation in scientific progress as well as strengthen individuals' sovereignty over their own personal data sphere. Alongside the many positive aspects, clearly, there are issues related to privacy, not only of the donor but also of family members, and thus to the trust that such a mechanism should inspire in order to be successful. In addition, the aspect of inclusivity cannot be overlooked: since the goal is also to eliminate bias arising from incomplete datasets, it will be necessary to work to ensure equity in access to digital tools among all patient groups.

The European legislature in adopting the European Data Strategy seems to be aware of the importance of valuing the possibility of sharing one's data for altruistic purposes, and although it did not introduce a tool for posthumous data donation, it did establish a mechanism for data altruism. This mechanism helps to enhance the control and sovereignty of individuals over their own data and make large and representative datasets available; however, many doubts have been raised about the usefulness of an instrument that appears to be merely a specification of an existing form of consent and whose characteristics in any case are unclear. Also since much of the implementation is left to the member states, it could be extremely fragmented, but today it is still too early to assess the effectiveness of such a legislative intervention, which in any case would not prevent the Union from also making *post-mortem* data donation possible, ensuring that citizens have one more chance to contribute with their health data to the creation of better and more inclusive healthcare.