Equality in Healthcare AI: Did Anyone Mention Data Quality?

***Aiste Gerybaite, Sofia Palmieri, Francesco Vigna* ***

ABSTRACT: This article explores how the concept of equality is affected by the technological developments in healthcare, focusing on AI. In this regard, the article develops that in AI healthcare systems ensuring data quality is pivotal when trying to ensure the implementation of the principle of equality. Analysing the European Commission's proposed AI Act, we highlight how horizontal rules, coupled with standardisation, can achieve data quality in AI based healthcare technologies. The article concludes with a reflection on remaining grey areas that require further elaboration by the European legislator.

KEYWORDS: AI; equality; EU law; GDPR; harmonised standards

SUMMARY: 1. Introduction – 1.1. Understanding equality: A fundamental rights perspective – 1.2. Equality and data quality – 2. The proposed AI Act: An overview – 3. Data quality mechanisms in the proposed AI Act – 3.1. Data governance – 3.2. Data accuracy and data robustness – 3.3. Human oversight – 3.4. Transparency of AI systems – 4. Data quality and standardisation – 4.1. Harmonised standards in EU legal framework – 4.2. Background on harmonised standards in the proposed AI Act – 4.3. Ongoing AI standardisation activities – 4.4. Standardisation and impact on AI and digital healthcare – 5. Conclusions.

1. Introduction

Artificial Intelligence (AI) in healthcare paves ways for optimistic scenarios concerning the development and improvement of patient-tailored healthcare. Through the massive use of data, AI can improve diagnosis, treatment, decision-making processes, and can increase the accuracy and efficiency of the individual therapeutic experience.[1] However, innovation often comes with a price. Despite these encouraging prospects, the use of AI poses profound legal concerns, inter alia, regarding the respect for the right to equality and non-discrimination. The permeant use of data

* *Aiste Gerybaite: CIRSFID, Alma AI, Alma Mater Studiorum, Università di Bologna, Università di Torino, University of Luxembourg. Mail: aiste.gerybaite@unito.it; Sofia Palmieri: Metamedica, University of Ghent. Mail: sofia.palmieri@ugent.be; Francesco Vigna: CIRSFID, Alma AI, Alma Mater Studiorum, Università di Bologna, Università di Torino, University of Luxembourg University. Mail: francesco.vigna6@unibo.it.* *This contribution is a result of joint research in the fields of AI, privacy, and equality. It fills in the gap in the current state of the art. All authors equally contributed to the research of scientific articles selected, and interpreted them, selected authors to be cited, and concepted the structure of the research. SP wrote sections 1; 1.1; 1.2. AG wrote sections 2; 3; 3.1; 3.2; 3.3; 3.4. FV wrote sections 4; 4.1; 4.2; 4.3; 4.4. Section 5 was jointly written by SP, AG and FV. Finally, all authors jointly reviewed the manuscript and approved the final version submitted. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie ITN EJD "Law, Science and Technology Rights of Internet of Everything" grant agreement No 814177. The article was subject to a double-blind peer review process.*

[1] G. BRIGANTI, O. LE MOINE, *Artificial Intelligence in Medicine: Today and Tomorrow*, in *Front. Med. Frontiers* in *Medicine*, 7/27, 2020, doi: 10.3389/fmed.2020.00027.

characterising AI may cause potential problems regarding this right. For instance, data used to train an AI system may be biased, infringing the right to non-discrimination, and causing issues regarding the safety and accuracy of the output processing for a given patient.

Aware of these biases and consequent possible discriminatory effects, the recent legislative proposal for an EU's AI regulation[2] (hereafter, the proposed AI Act) also considers equality, albeit in a broad sense. The proposed AI Act elaborates on the concept of equality and non-discrimination, formulating requirements to favour the elimination of biases and, at the same time, referring to standardisation as possible means to implement on a practical level non-discrimination. In this sense, the references made by the proposed AI Act to the General Data Protection Regulation[3] (GDPR) and standardisation rules may offer indirect and partial protection for this fundamental right.

This article, therefore, aims to shed light on how the right to equality and non-discrimination is enshrined in the proposed AI Act from a *data quality* perspective. Firstly, the article discusses the source of the right to equality and non-discrimination in EU law and briefly describes the inaptitude of those sources to protect this right when we are moving at the intersection between AI and healthcare. Secondly, it explores the connection between equality and AI through a data quality perspective. Following, we establish how equality is embedded in the proposed AI Act, scrutinising how the requirements elaborated by the proposed AI Act ensure data quality. Furthermore, we analyse how the proposed AI Act collaborates with the existing data protection legislation, in particular the GDPR, in order to realise data quality via protection against several discriminatory effects. The work then addresses how harmonised standards could give practical implementation to data equality. Finally, the article concludes with several thoughts on desirable regulatory and interpretative interventions in the discussed areas.

1.1. Understanding equality: A fundamental rights perspective

The principle of equality and non-discrimination resonates in the European acquis as a value, objective, and fundamental right within the European Union.[4] Exploring the multiple sources present in the European acquis, Article 2 of the Treaty on European Union (hereafter, TEU) enshrines the respect for human dignity, freedom, democracy, equality, the rule of law, and respect for human rights, including the rights of persons belonging to minorities, as founding values of the European Union. The Union not only recognises these values as founding values but, under Article 3(3) of the TEU, recognises equality as an objective for the Union, stating that it shall work proactively to combat "social exclusion and discrimination" by promoting "social justice and protection, equality between women and men, solidarity between generations, and protection of the rights of the child".

Still moving among the treaties of the EU, the Treaty on the Functioning of the European Union (hereafter, TFEU) also presents an important basis for understanding the importance of equality within the

[2] European Commission, 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.

[3] 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).

[4] G. ZACCARONI, *Equality and Non-Discrimination in the EU: The Foundations of the EU Legal Order*, Cheltenham, 2021, 5.

ISSN 2284-4503

Union. Article 8 TFEU, which commits the Union to eliminate inequalities and promote equality between men and women, is counterbalanced by Article 10. The latter requires the Union, in defining and implementing its policies and activities, to pursue the objective of combating all forms of discrimination, be it on the grounds of sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation. Completing this framework of normative references provided by the TFEU, again with a provision of a general nature, Article 18 of the TFEU reiterates the prohibition of discrimination based on nationality.

Furthermore, it could be observed that the principle of equality has a "constitutional" place in the Charter of Fundamental Rights of the European Union (hereafter: the Charter or CFR).[5] Under the heading "Equality", the Charter provides protection to peculiar categories of individuals because of their fragile position in society. This is the case of Article 20, which enshrines the principle that "All persons are equal before the law". Following, Article 21 of the Charter, entitled "Non-discrimination", it`s an open-ended disposition that prohibits any form of discrimination "based, in particular, on sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability or sexual orientation". Respect for any cultural, religious, and linguistic diversity is addressed in Article 22, while the principle of equal treatment between women and men is enshrined in Article 23, which states that "equality between women and men shall be ensured in all areas, including employment, work and pay. [...]". Placed in the context of the provisions on equality, which are often aimed at protecting weaker persons whether for reasons of age or disability, the following three provisions of the Charter aim respectively at ensuring protection for minors (Article 24), the elderly (Article 25) and the disabled (Article 26).

Finally, particularly relevant to the issue of equality in the field of healthcare is Article 35 CFR, according to which, "everyone has the right of access to preventive health care and the right to benefit from medical treatment under the conditions established by national laws and practices".[6] Furthermore, together with the sources coming from the Charter, the EU regulatory framework is equipped with antidiscrimination directives that offer protection to the right of non-discrimination in specific contexts. Despite the many sources available in the European context to protect the right to equality and non-discrimination, they have limited meaning when moving at the intersection of healthcare and Artificial Intelligence. This is mainly for three reasons. Firstly, as already explored by relevant literature,[7] when in the specific context of healthcare, the antidiscrimination directives offer only three grounds for protection, namely race, ethnic origin[8] and gender.[9] Secondly, both the aforementioned directives and the open-ended provisions of the EU Charter of Fundamental Rights only apply when a matter falls within the scope of the EU law.[10]

[5] P. Craig, G. De Burca, *EU Law – Text, cases, and materials*, Oxford, 2020, 18-21.
[6] G. Zaccaroni, *Equality and Non-Discrimination in the EU: The Foundations of the EU Legal Order*, cit., 111.
[7] G. Di Federico, *Access to Healthcare in the European Union: Are EU Patients (Effectively) Protected Against Discriminatory Practices?*, in L. Rossi, F. Casolari (eds.), *The Principle of Equality in EU Law*, Cham ,2017.
[8] Council Directive 2000/43/EC of June 29, 2000, Implementing the principle of equal treatment between persons irrespective of racial or ethnic origin.
[9] Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of good and services.
[10] T. Hervey, J. McHale, *European Union Health Law: Themes and Implications*, Cambridge, 2015, 156-183.

Third, the bases of discrimination we encounter when discussing AI bias and healthcare are often not protected under EU law. The problem of inadequate protection from intersectional discrimination is already well known in the literature when discussing discriminations in the context of health. This means that in the healthcare context patients face disadvantages and discriminations based on a unique combination of protected grounds which combination do not find protection within the EU law.[11] When AI systems make their way into the healthcare context certain complex patterns of disadvantage are "woven into the data fabric"[12] and algorithmic decisions can amplify intersectional discrimination.[13] In addition, AI systems often process results based on artificial groups, so-called "algorithmic groups".[14]

Differently to protected groups, algorithmic groups frequently do not depend on fixed traits (such as age or ethnicity), their attribution is frequently not entirely arbitrary (think credit scores), members are not always the targets of historical oppression (think sad teens), and these groups are not always socially salient (such as "people who scroll slowly").[15] Currently, algorithmic groupings are not legally protected unless they map onto an already-protected group, which is rarely the case.[16]

The proposed AI Act fits into this complex context by bringing two mild solutions to these issues. Firstly, the proposed AI Act, will expand the applicability of the two directives mentioned above and of the EU Charter. Being an instrument of EU law, the proposed AI act will subject the use of AI systems to respect the right to equality and non-discrimination as protected from the named sources. Secondly, in addition to extending legal protections to the right to non-discrimination, the proposed AI Act offers *ex ante* protection for this right. The proposed AI Act's data governance requirements concur to ensure equality thanks to the maximization of data quality. In the following paragraphs, we explore this second assumption. We introduce the concept of data quality as protection of the right to equality and non-discrimination and analyse how the provisions of the proposed AI Act contribute to achieving data quality and thereby protecting the right in question.

[11] EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, *Inequalities and multiple discrimination in access to and quality of healthcare*, 2013, doi:10.2811/17523.

[12] M. A. WOJCIK, *Combating algorithmic bias in healthcare: towards a European regulatory framework based on the right to health and the right to science*, essay published on the website of Centre for Legal & Court Technology https://legaltechcenter.net/a-i/commentary/ (last visited 30/08/2022).

[13] R. XENIDIS, *Tuning EU equality law to algorithmic discrimination: Three pathways to resilience*, in *Maastricht Journal of European and Comparative Law*, 27, 6, 2020, 736-758.

[14] S. WACHTER, D. SUTCLIFFE, R. GILLIS, *How can we protect members of algorithmic groups from AI-generated unfair outcomes?*, available at https://www.oii.ox.ac.uk/news-events/news/how-can-we-protect-members-of-algorithmic-groups-from-ai-generated-unfair-outcomes/ (last visited 30/08/2022).

[15] K. LOBOSCO, *Facebook friends could change your credit score*, CNNMONEY (2013), https://money.cnn.com/2013/08/26/technology/social/facebook-credit-score/index.html (last visited 29/08/2022).

[16] S. WACHTER, *The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law*, in *Tulane Law Review*, Forthcoming, available at SSRN: https://ssrn.com/abstract=4099100 or http://dx.doi.org/10.2139/ssrn.4099100.

1.2. Equality and data quality

"Algorithms are like drugs"[17] they can have a profound impact on human life, their performance varies according to geographics and ethnicity and, even more significantly, they are prone to cause potential side and harmful effects.[18] The fact that AI technology can reflect and exacerbate existing or even create new discrimination patterns is already extensively discussed in the literature.[19] Data quality for building AI algorithms is one of the main concerns to stem potential breaches of fundamental rights, inter alia, the right to equality and non-discrimination. Following the well-known principle, "garbage-in garbage-out", poor quality of data can only lead to poor quality outcomes that are prone to cause discrimination and in general, violations of fundamental rights. In a nutshell: AI can only be as good as the data it uses.

Data quality is a broad and multifaceted concept. For the purposes of our discourse, we will relate data quality to the two main aspects highlighted by the European Union Agency for Fundamental Rights, namely, errors of representation and measurement errors.[20] Briefly, representation error generally refers to the bias created if the population to be covered by an AI application is not included in the input data used for training the application.[21] Differently, measurement errors refer to how accurately the data is used to indicate or reflect what is intended to be measured.[22] When these errors are reflected in the data used to train AI systems, the use of these systems might support or exacerbate existing inequalities. Such prejudices are implanted in AI and can cause AI systems to operate unfairly or harmfully. Let's elaborate on an example. Today, robotic surgeons can assist human surgeons to perform surgeries, aiding in activities such as cutting or suturing tissues.[23] Nevertheless, endeavours

[17] A. COVAROS, I. CHEN, A. GORDHANDAS, A. DORA, *We should treat algorithms like prescription drugs*, available at https://qz.com/1540594/treating-algorithms-like-prescription-drugs-could-reduce-ai-bias/ (last visited 29/08/2022).

[18] M. A. WOJCIK, *op. cit*.

[19] Among many: WHO, *Ethics and governance of Artificial Intelligence for Health*, 2021; P. HACKER, *A legal framework for AI training data-from first principles to the Artificial Intelligence Act*, in *Law Innovation and Technology*, 13, 2, 2021; D. SCHONBERGER, *Artificial Intelligence in healthcare: A critical Analysis of the Legal and Ethical Implications*, in *International Journal of Law and Information Technology*, 27, 2, 2019, 171-203.

[20] EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, Data Quality and Artificial Intelligence, *Mitigating bias and error to protect fundamental rights*, 2019, available at https://fra.europa.eu/en/publication/2019/data-quality-and-artificial-intelligence-mitigating-bias-and-error-protect (last visited 30/08/2022).

[21] As exemplified, inter alia, in S. AKGÜN, et al., *Estimating mortality and causes of death in Turkey: methods, results and policy implications*, in *European Journal of Public Health*, 17, 6, 2007 593-9; R. BASTIAN, *Why Representation Matters When Building AI*, 2021, available at https://www.forbes.com/sites/rebekahbastian/2021/03/28/why-representation-matters-when-building-ai/ (last visited 30/08/2022).

[22] EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, *Data Quality and Artificial Intelligence, mitigating bias and error to protect fundamental rights*, cit.

[23] See, among the vast literature, for example: G.P. MOUSTRIS, et al., *Evolution of autonomous and semi-autonomous robotic surgical systems: A review of the literature*, in *International Journal of Medical Robotics and Computer Assisted Surgery*, 7, 4, 2011, 375-392; Y. KASSAHUN, et al., *Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions*, in *International Journal of Computer Assisted Radiology and Surgery*, 2016, 11, 4, 553-568; M. DOHLER, et al., *Internet of Skills, Where Robotics Meets AI, 5g and the Tactile Internet*, in *EuCNC 2017 European Conference on Networks and Communications*, 2017; S. KIM, et al., *The Internet of Skills: use of fifth-generation telecommunications, haptics and artificial intelligence in robotic surgery*, in *BJU International*, 122, 3, 2018, 356-358; S. BEYAZ,

to make such robots more autonomous are spreading. In such a scenario, robots would learn from their own experience by applying AI techniques. The final goal would be to realise fully self-reliant, automated robot surgeons that can recognise organs and tissues and autonomously perform surgeries. Massive datasets of high-quality labelled data are needed to train such AI-driven robotic surgeons, which should follow uniform standards during the training to avoid the creation of dangerous biases.[24] Different types of biases (gender, race, status, religion and so on)[25] could be introduced in several ways into an AI system,[26] often non-intentionally, by the trainer or the human surgeon.[27]

As an example of the peculiarities of discrimination that AI could bring in healthcare, some scholars have noted that a particular kind of gender bias in autonomous surgical robots could be introduced by the surgeon who trains the AI system.[28] In terms of representation errors, since a robot is supposed to learn from its own experience, and since some surveys have shown that male surgeons tend to perform riskier surgeries than female surgeons, the gender of the trainer could indirectly affect the robot's final performance. Therefore, robots trained by male surgeons could be more prone to engage in risky operations that could endanger patients' life. That would result from using biased data about surgeons' behaviour to feed the AI system. Similarly, training the robot with data from a limited geographical area and underrepresenting or misrepresenting certain population groups could lead to inaccurate operations and results of non-generalisable quality. In terms of error measurements, we can briefly refer to data that has been incorrectly collected. Indeed, it has been found that many electronic health records are sketchy, with inaccurate or even erroneous information.[29] These types of errors or inaccuracies are present in greater numbers for elderly patients or patients with comorbidities.[30] Such errors in the measurement or labelling of certain symptoms or diseases will negatively influence the data with which the AI will be trained. The Robotic surgeon trained on these data would be then act based on wrongful information, particularly endangering those patients who are inaccurately represented by the data.

In order to mitigate possible discrimination issues and emphasise the positive aspects of AI, the regulatory framework for AI must be able to ensure that the data with which AI is trained are complete, representative and with minimum bias, in line with the concept of data quality.

A brief history of artificial intelligence and robotic surgery in orthopedics & traumatology and future expectations, in *Joint Diseases and Related Surgery*, 31, 3, 2020, 653-655.

[24] M. BHANDARI, *Artificial intelligence and robotic surgery: Current perspective and future directions*, in *Current Opinion in Urology*, 30, 1, 2020, 48-54.

[25] P. HACKER, *Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination under EU Law*, in *Common Market Law Review*, 55, 2018, 1146.

[26] A. RAJKOMAR, et al., *Ensuring fairness in machine learning to advance health equality*, in *Annals of Internal Medicine*, 169, 12, 2018, 886-873.

[27] S. O'SULLIVAN, et al., *Operational framework and training standard requirements for AI-empowered robotic surgery*, in *International Journal of Medical Robotics and Computer Assisted Surgery*, 16, 5, 2020, 1-13.

[28] S. O'SULLIVAN, et al., *op. cit.*

[29] H. AERTS, et al., *Quality of Hospital Electronic Health Record (EHR) Data Based on the International Consortium for Health Outcomes Measurement (ICHOM) in Heart Failure: Pilot Data Quality Assessment Study*, in *JMIR Medical Informatics*, 9, 8, 2021.

[30] S. K. BELL, et al., *Frequency and Types of Patient-Reported Errors in Electronic Health Record Ambulatory Care Notes*, in *JAMA Network Open*, 3, 6, 2020.

ISSN 2284-4503

In the following paragraph, we explore how the recent recently proposed AI Act enforces data governance requirements having a meaningful impact on the quality of data.
Analysing the proposed AI Act from this perspective, it is also possible to determine whether the proposed regulation is ultimately able to give ex-ante protection to the right to equality and non-discrimination.

2. The proposed AI Act: An overview

Behind the proposed AI Act and its requirements[31] lay two years of intensive consultation work by the HLEG tasked with advising the Commission on implementing the AI strategy. In April 2019 and later in July 2020, the HLEG released two documents that guided the proposed AI Act's realisation. The Ethics guidelines for trustworthy AI,[32] followed by the Assessment list for trustworthy AI,[33] set out key requirements reflecting an approach where EU values and fundamental rights should guide the development and use of AI. The two documents guide how trustworthy AI systems can be realised by listing seven requirements those AI systems should meet. Without claiming to be exhaustive, the guidelines indicate the following requirements: human agency and oversight; Technical robustness and safety; Privacy and data governance; Transparency; Diversity, non-discrimination, and fairness; Societal and environmental well-being; and Accountability.
The HLEG is aware that datasets used by AI systems may suffer from the inclusion of unintentional historical biases, incompleteness, and poor governance models. The group acknowledge that the persistence of such biases, intrinsic to AI, could lead to discrimination, potentially exacerbating prejudice and marginalisation. Precisely for these reasons, the ethics guidelines encourage the elimination, where possible, of such identifiable and discriminatory biases from the data collection stage. Furthermore, to avoid unfair biases inherent in operating algorithms, systems development must also be subject to oversight processes to analyse and address the system's purpose, constraints, requirements, and decisions clearly and transparently. The reference to and elaboration of the principle of non-discrimination and fairness is certainly not surprising, but rather an almost due reference.
The work of the HLEG is embodied in the proposed AI Act, published on April 21st, 2021. The proposed AI Act is a horizontal approach to regulation. It is not sector-specific but aims at laying down general rules for the use of Artificial Intelligence. It is harmonised with sector-specific EU legislation as it is mentioned in the proposed AI Act itself, such as the GDPR or the Medical Devices Regulation. Therefore, the Commission hopes that such a horizontal approach will ensure the future-proof of AI legislation that is flexible enough considering technology advancements. Considering the horizontal approach to regulation of AI, the proposed regulation would apply directly to both public and private

[31] For an overview on the proposed AI Act's requirements and their rationale, see C. CASONATO, B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, in *Bio Law Journal*, 3, 2021, 415-437.
[32] INDEPENDENT HIGH LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, Ethics Guidelines for Trustworthy AI, 2019.
[33] INDEPENDENT HIGH LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, Assessment List for Trustworthy Artificial Intelligence (ALTAI), 2020.

bodies within the EU and outside, as long as an AI system is placed on the market in the EU or its use affects EU citizens as observed in article 2(1), leading to the so-called *Brussel's effect*.[34]

The proposed AI Act classifies AI systems into minimal, limited, high, and unacceptable risk AI systems as a risk-based regulation. Low and minimal-risk AI systems have minimum compliance requirements, while the great focus is on compliance with high-risk AI systems. In this respect, the proposed AI Act establishes that high-risk AI systems must comply with certain rules. Compliance with the requirement must be done *ex-ante* through a conformity assessment to establish that high-risk AI systems meet these requirements before they are put on the market or into service. It can be argued that most of the health-related AI systems will fall under the latter high-risk category, establishing further requirements for AI-based healthcare devices producers. Further, another key innovation is a mandate for a post-market monitoring system to detect problems and mitigate safety and security related issues.

Perhaps the most debatable innovation of the proposed AI Act is the definition that the proposed AI Act provides regarding the subject matter of the regulation, Artificial Intelligence. In this respect, the proposed AI Act includes an attempt to define AI, which is a subject of great debate. The proposed draft AI definition has already been amended twice, considering Brussel's Effect on the AI regulation. Lobbying done by the big tech companies and civil society (NGOs) has uncovered the struggles in finding the right balance of the notion of AI that would allow the *big tech* to continue servicing the European continent while ensuring that all risky AI technologies are prohibited, or their use is heavily regulated. An observation should be made that lobbying by different groups on the notion of AI showcases how groups try to influence European values and set a new risk appetite for a one-of-a-kind AI regulation. Depending on the adopted definition of AI, European values may shift towards private corporation values, and conservative risk appetite may shift to the high tolerance of AI solutions. Such influences and the final scope of the proposed AI Act's subject matter would also impact equality and data equality and its protection in the final adopted regulation.

The proposed AI Act expands beyond the regulatory compliance of AI systems. It also introduces the creation of a European Artificial Intelligence Board (hereafter, the EAIB) that would facilitate the development of AI standards and introduce regulatory sandboxes and voluntary codes of conduct for certain AI systems. Besides creating the EAIB, the proposed AI Act also establishes fines of up to 6% of annual worldwide turnover for breaching the rules on high-risk systems. All other breaches would be subject to fines of up to 4% of annual worldwide turnover.

The proposed AI Act attempts to achieve several different goals. First, the proposed AI Act attempts to ensure that AI systems entering the EU market are safe and respect existing laws on fundamental rights and EU values. It proposes novel governance and enforcement mechanisms on fundamental rights and applicable safety requirements to achieve the first goal. Second, it attempts to be the cornerstone legislation in ensuring legal certainty in AI. Third, it aims at facilitating the development of a single European market for lawful, safe, and trustworthy AI, based on fundamental rights and values of the EU. In the case of the use of AI systems, these values in the proposed AI Act are extensively ensured through requirements aimed at ensuring data quality. Specifically, the proposed AI Act enshrines ex-

[34] Brussel's effect refers to externalisation of EU laws outside its borders through market mechanisms and their de facto application. For further insight into the topic, please see: A. BRADFORD, *The Brussels Effect*, in *Northwestern University Law Review*, 107, 1, 2012, 1-68.

ante requirements for high-risk AI systems (such as healthcare AI) which include requirements for data governance, data completeness, data accuracy and robustness, human oversight, and provision of information to users that, when considered in totem, ensure a higher standard of data quality. In the following paragraphs, we analyse how the proposed AI Act attempts to ensure data quality through the elaboration requirements on data for training AI systems.

3. Data quality mechanisms in the proposed AI Act

By taking up the principles suggested by the HLEG and reinterpreting European values, the Commission has committed itself to data quality. Within our work, we explore how the Commission ensures data quality through the requirements laid down in the proposed AI Act and by reference to harmonised standards, ultimately protecting the European right and value of equality.
We argue that, given the peculiar risks brought by AI technologies, the EU legislator relies on other legal instruments such as GDPR to ensure data quality, intersecting the requirements of the proposed AI Act with the dictates of the GDPR. Furthermore, an additional level of assurance is provided within the proposed AI Act through the reference to harmonised standards, moving the focus from a secondary law to co-regulation mechanisms.[35]
The following paragraphs analyse how this interplay of requirements and references to ancillary norms and standards concur to ensure data quality.

3.1. Data Governance

To start with, article 10 of the proposed AI Act considers data and data governance of high-risk AI systems. Such high-risk systems also include medical devices and AI-based digital health technologies. It should be observed that ensuring data quality in AI training requires proper data governance management practices that allow ensuring data quality.[36] In this respect, the Commission has proposed that in the deployment of AI systems, data sets used would be required to employ appropriate data governance and management practices. Such practices include appropriate design choices, data collection, and relevant data preparation processing operations, such as annotation, labelling, cleaning, enrichment, interoperability, and aggregation. It also requires developers of AI systems to include formulation of relevant assumptions, notably concerning the information that the data are supposed to measure and represent.[37] Formulating assumptions based on which an AI system would reach its goal

[35] C. MARSDEN, *Internet co-regulation and constitutionalism: Towards European judicial review*, in *International Review of Law, in Computers and Technology*, 26, 2-3, 2012, 211-228.
[36] As observed S. O'Neil et al in their study, interoperability of data and governing data sharing can encourage individuals, organisations, and agencies to share data. For further reading please see: S. O'NEIL, S. TAYLOR, A. SIVASANKARAN, *Data equality to Advance Health and Health equality in Low- and Middle-Income Countries: A Scoping Review*, in *Digital Health*, 22, 7, 2021.
[37] As observed by Kim et al: "Narratives about ML might create an impression that a mere statement of what needs to be achieved without how that should be done can be enough for a computer to accomplish a task". Moreover: "To translate a real-life problem into something that can be processed and solved by a computer, a problem essentially needs to be expressed in an abstract way – by using a formal notation (a mathematical model, functions, logic rules, etc.) – that a computer can decipher and implement". D. KIM, et al., *Clarifying*

requires such goal to be "expressed abstractly – by using a formal notation (a mathematical model, functions, logic rules, etc.) – that a computer can decipher and implement".[38] Therefore, the formulation of assumptions in healthcare may be crucial in order for healthcare AI systems to ensure the error-free functioning of such systems.

Ensuring quality data, which also contributed to equitable AI systems, is achieved by the establishment of an a priori assessment of the availability, quantity and suitability of the data sets that are needed; examination in view of possible biases; the identification of any possible data gaps or shortcomings, and how those gaps and shortcomings can be addressed. Established minimum standards for data and data governance encompassing training, validation, and testing data of high-risk AI systems, ensuring that appropriate standards are maintained at all stages. These data governance standards essentially have been created to ensure that data sets used in the training of AI systems are of the highest quality, not biased, and do not lead to discriminatory outcomes.

Further, Article 10(4) enshrines a specific requirement that all used data sets in high-risk AI systems consider the intended purpose of an AI system and the associated characteristics and elements particular to the specific geographical, behavioural, or functional setting within which a high-risk AI system is intended to be used. For example, the training of AI software for healthcare insurance purposes in hospitals in less developed areas where might lead to misleading results vis-a-vis training of same software in a more developed area. The proposed AI Act here sets a specific requirement for the developers of AI systems requiring them to consider how to ensure data quality a priori and prevent the potentially harmful or discriminatory outcomes of AI systems developed and employ measures to avoid such outcomes. Specifically, with data related to healthcare, the proposed AI Act allows the providers of AI systems to process special categories of personal data referred to in Article 9(1) of the GDPR, subject to proper safeguards for fundamental rights and freedoms of natural persons. These include technical limitations on the reuse and use of state-of-the-art security and privacy-preserving measures. Nevertheless, the processing of special categories of personal data has its limitations. The limitation lies in art. 10(5) where special categories of data may only be processed to ensure bias monitoring, detection, and correction of high-risk AI systems to avoid harmful or discriminatory outcomes, again referencing the need to ensure data quality in health-related AI systems. Consequently, the Commission requires health AI system developers to consider not only data quality within the remits of the proposed AI Act but also in coordination with other applicable regulations such as the GDPR.

In summary, since data is the backbone of all AI systems, the Commission sets high standards in order to ensure quality data in all AI systems, including healthcare. In this respect, the proposed AI Act requirements add a layer of compliance to the providers of AI systems to ensure that quality is maintained at all stages of AI systems' deployment, starting at the collection of data for the training of AI systems.

Assumptions About Artificial Intelligence Before Revolutionising Patent Law, in *GRUR International*, 71, 4, 2022, 295–321.

[38] D. KIM, et al., *op. cit*.

ISSN 2284-4503

3.2. Data accuracy and data robustness

Ensuring data quality in the proposed AI Act is also achieved by transposing principles that already exist in part in other regulations, especially the GDPR. Indeed, the proposed AI Act introduces three already existing requirements. Namely, the duty to guarantee an appropriate level of data accuracy, data robustness, and cybersecurity of the high-risk AI system considering the purpose pursued by the AI system itself. The appropriate levels according to the proposed Act shall be ensured from the design to the development of the system and kept throughout the whole AI system's life cycle. Following data governance practices discussed above, ensuring data accuracy and data robustness in AI systems is crucial in order to avoid measurement errors or poor quality of data in healthcare AI systems which can lead to discriminatory outcomes.

However, what is meant by the accuracy is not explicitly defined in the proposed AI Act's text.[39] It could be envisaged that AI providers will be asked to evaluate both the level of accuracy of AI systems, data accuracy, and the metrics of accuracy, i.e., the standards, against which the level of accuracy is evaluated.[40] In the GDPR the concept of accuracy has a clearer meaning. Article 5(1)(d) of the GDPR as data accuracy means that personal data must be kept up to date and be accurate, i.e., without errors vis-a-vis the purpose of the personal data processing. Therefore, the accuracy of an AI system depends for sure on the accuracy of data (as meant in the GDPR) and other features of the AI system, such as the algorithm or the interaction with external agents. Regarding this, setting standards for evaluating the level of accuracy will be crucial. In order to ensure levels of accuracy regarding data which would leave to data quality one may consider using 'dimensions' of data equality as a starting point for ensuring data accuracy and in turn the accuracy of AI systems. In this respect, research has addressed 4 distinct data equality dimensions: 1. representation equality, i.e. ensuring the increased visibility of underrepresented groups in data records; 2. feature equality, i.e., facilitating linkage across datasets that would allow to expose any quantify potential inequities; 3. access equality, i.e. providing equitable access to data at all levels; and 4. outcome equality, i.e. monitoring and mitigating discriminatory consequences for affected groups.[41] In this regard, data quality in healthcare AI systems can be ensured by guaranteeing that underrepresented groups are present in data collected for AI training; by ensuring data linkages between different types of data collected; by providing transparency to data; and finally, by monitoring potential discriminatory outcomes. If the above mentioned 4 aspects are considered by developers of healthcare AI systems before training AI on data collected, it could lead to a higher quality of data and more equitable AI systems' decisions.

Moreover, the proposed AI Act provides that AI systems shall be "resilient" against errors, faults, or inconsistencies. The first observation here is that it could be hard to understand the difference

[39] However, the accuracy of the AI system is a requirement mentioned several times both in the explanatory memorandum and in the text of the proposed AI Act and it seems to be considered an essential step to achieve an unbiased and fair AI.

[40] As observed S. O'Neil et al in their study, data accuracy, affects the AI's ability to make well-informed, data-driven decisions regarding, for example, health policy and resource allocation. For further reading please see: S. O'NEIL S, S. TAYLOR S, A. SIVASANKARAN, *op. cit*.

[41] For further insight on this topic please refer to: H. V. JAGADISH, J. STOYANOVICH, B. HOWE, *COVID-19 Brings Data equality Challenges to the Fore*, in *Digital Government: Research and Practice*, 2, 2, 2021.

between the requirement to make an AI system resilient to errors and faults and the accuracy requirements. The need to keep AI system consistent to the intended purposes could in part match with the data minimisation principle in GDPR, which asks for the coherence between the amount and the kind of data elaborated and the purpose of the data processing (but it could also be linked to the purpose limitation principle of the GDPR[42] or again to the accuracy principle). However, such requirements have a broader scope than GDPR's requirements. They are not limited to data elaboration but apply to the whole environment within which the AI system is deployed and to the interaction between an AI system and a natural person or other AI systems. While this requirement expands beyond data quality as may be understood in its strictest sense, it is still relevant. In particular, consider an example of an AI healthcare system that automatically assigns medication to patients in the hospital based on the data of the patient. It is not unheard that databases can be hacked, and data can be changes or deleted. in the following example, should such an event occur, an AI system could become compromised as a hacker could modify the patient's data set to increase or decrease the dose of the prescribed medication. Thus, ensuring the overall AI systems resilience is crucial in ensuring not only data quality but also an overall safety of the end-users of such system.

3.3. Human oversight

Ensuring data quality and an overall right to equitable AI within the proposed AI Act is not only facilitated by introducing requirements to technical elements related to data as the ones mentioned previously in the chapter, but by introducing human-in-the-loop requirements. In this respect, the proposed AI Act introduces human oversight of high-risk AI systems and the right to explanation contained in the GDPR, which are intrinsically linked. The proposed AI Act enshrines an important level of protective transparency, that is, that AI applications ought to be equipped with interface tools allowing effective oversight by humans to minimise risks.[43] The usage of personal data for AI learning may lead to biased decision outcomes and biased decision-making patterns that form the core of AI systems. Thus, including safeguards to avoid such outcomes is necessary. Human oversight is one of such safeguards. In terms of the protection of personal data, data subjects should be informed of when their data is used for AI training, the legal basis for such processing, general explanation of the logic and scope of the AI system. It could be argued that irrespective of all the steps taken to ensure equitable outcomes in AI-based healthcare systems, such systems may still, intrinsically, lead to inequitable health outcomes for patients. Thus, human oversight of such systems is the *last resort* within AI systems. It should be observed, however, that it is challenging to maintain human oversight in high-risk AI systems which make life decisions, such as the ones used in health sector for the delivery of emergency medical aid, as the requirement reinforces the focus of the proposed AI Act on transparency vis-a-vis professional operators, not affected persons.[44] Again, here the Commission stems away from technical requirements

[42] Article 5(1)(b) GDPR requires that personal data shall be processed only in a compatible way to an explicit purpose, and not further processed in an incompatible way.

[43] P. HACKER, J. PASSOTH, *Varieties of AI Explanations under the Law. From the GDPR to the proposed AI Act, and beyond*, in A. HOLZINGER, R. GOEBEL, R. FONG, T. MOON, K. MÜLLER, W. SAMEK (eds.), *Lecture Notes on Artificial Intelligence 13200: xxAI beyond explainable AI*, 2022, Available at SSRN: https://ssrn.com/abstract=3911324.

[44] P. HACKER, J. PASSOTH, *Varieties of AI Explanations under the Law. From the GDPR to the proposed AI Act, and beyond*, cit.

regarding datasets, but instead focuses on the broader requirement to ensure that any AI system's decisions can be questioned by a human, who would also have the right to 'audit' the AI system in order to uncover any errors in decision that could be based on poor-quality data.

3.4. Transparency of AI systems

Linked to human oversight are the requirements that are enshrined in the proposed AI Act regarding the provision of transparency and provision of information to users, contained in article 13 of the proposed AI Act. These requirements, although referring only to explainability/interpretability of the AI system itself somehow recall the right to explanation enshrined in recital 71 and art. 22(3) of the GDPR. Interestingly, however, the requirement's scope prescribes that high-risk AI systems are designed in a way that is sufficiently transparent "to enable users to interpret the system's output and use it appropriately".[45] The requirement for explainability includes a requirement for an "appropriate type and degree of transparency". It can be observed that the proposed AI Act interplays with the GDPR article 22 considering the right to explanation. Some authors have noted that article 13 of the proposed AI Act provides specific requirements for high-risk AI system's transparency and explainability and acknowledges the different varieties of explanations that could be provided, such as local, global, or counterfactual explanations, or granular information on feature weights.[46] It should be observed however, that the scopes of article 13 of the proposed AI Act and the GDPR article 22 are different. Article 13 of the proposed AI Act considers the provision of information and explanation of the AI decisions to users, who are not the end-user as defined by the proposed AI Act. Requirements of Article 13 are directed towards the users of the AI system itself. It is important to note that 'users' under the proposed AI Act include anyone using the system, except the consumer.[47] The proposed AI Act's right to explanation is a top-down approached, transposed throughout the whole process of the use of AI software, for example, in cases where AI makes decision affecting an individual.

On the other hand, the right to explanation in the GDPR takes a bottom-up approach, the right to explanation arises only after an automated decision (often based on AI) is taken, i.e., at the end of the cycle. Nonetheless, the GDPR right to explanation is directed towards the data subject, i.e., the natural individual whose rights are affected by the processing of personal data, which may also include a consumer of AI systems. This creates a great discrepancy between the two regulations regarding what information could be provided to the data subject in terms of GDPR. Thus, the proposed AI Act's articles may fall short in reconciling the GDPR provisions on the right to explanation to data subjects with article 13 of the proposed AI Act concerning requirements for transparency and provision of information to users (not being data subjects under the scope of the proposed AI Act).

Providing transparency and information to users is also indirectly linked with equitable AI. The introduction of such provision in the proposed AI Act may ensure that users who receive information about

[45] European Commission, 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, cit.

[46] P. HACKER, J. PASSOTH, *Varieties of AI Explanations under the Law. From the GDPR to the proposed AI Act, and beyond*, cit.

[47] P. HACKER, J. PASSOTH, *Varieties of AI Explanations under the Law. From the GDPR to the proposed AI Act, and beyond*, cit.

AI's logic, the way it decides on an outcome, can receive such information in a transparent manner to interpret the system's output appropriately. This facilitates the possibility of holding the providers of AI systems accountable for any discriminatory or biased outcomes. Nevertheless, it has several shortcomings. Firstly, this indirect accountability measure is only limited to the users of AI systems and excludes the consumer (potentially protected under consumer legislation). Secondly, the provision falls short of reconciling GDPR provisions on the right to explanation since article 13 does not indicate what meaningful information should be provided to affected data subjects under the GDPR.

4. Data quality and standardisation

Equality in the health care sector needs pragmatic ways to be realised. This pragmatisation might occur through data quality. From a policy and regulatory perspective, data quality can be partially ensured through the interplay between the proposed AI Act requirements and the data protection requirements enshrined in the GDPR. The interaction between the rules of these two regulations provides a legal framework to ensure that the training of AI systems is not affected by bias that could hinder equality goals.

Harmonised standards, on the other hand, operate at the lowest level of abstraction of the data quality implementation process, i.e., the definition of procedures and techniques to be adopted during the development of the system. By adopting these technical and organisational specifications, AI developers can reach and demonstrate compliance with data equality-related requirements enshrined in the proposed AI Act, namely those in Title III, Chapter 2. In other words, standardisation of AI training procedures could contribute to reducing biases and possible discriminations and therefore increase data quality.

4.1. Harmonised standards in EU legal framework

Standardisation has been incorporated into EU legislation by the 1985 *New Approach.*[48] Since then, the directives adhering to this approach started including only essential requirements. These requirements were then to be integrated by non-legally binding technical standards developed by standardisation bodies (usually called Standard Development Organisations or SDOs). Although following these standards is not mandatory, the adoption of harmonised standards provides a "presumption of conformity" with a set of given requirements.

Today, there is an ongoing debate, both in literature and at the court level about critical features of harmonised standards. For example, whether they are actually voluntary measures, or on their nature

[48] Council Resolution of May 7th, 1985, on a new approach to technical harmonisation and standards (85/C 136/01).

as a source of the law,[49] or on the liability aspects of notified bodies[50] in case of damages due to the violation of standards by organisations.[51]

Harmonised standards are part of the governance strategy that aims at strengthening the internal market through common rules for safety controls on certain products before commercialisation. In 2008, the package of measures called *New Legislative Framework* (NLF), built upon the already mentioned *New Approach*, set up rules for market surveillance and conformity assessment of products that could endanger health and environment. Namely, these measures are meant to regulate the accreditation of notified bodies, conformity assessment procedures e the release of the CE mark. The goal is to protect consumers from risky products introduced in the internal market, e.g., medical devices or, as foreseen in the proposed AI Act, some types of high-risk AI systems.

The proposed AI Act adopts the conformity assessment procedure through harmonised standards in order to achieve in the first place an internal market result. This is also clear when looking at the legal basis of the proposed AI Act, which is, first and foremost, art. 114 TFEU (though supported by art. 16 TFEU). However, as it is possible to infer from the explanatory memorandum of the proposed AI Act, harmonised standards and conformity assessment procedures could also help reach other objectives, besides strict internal market functioning goals.[52] Enhancing compliance with data quality related requirements, and therefore reducing biases and related discrimination, is one of the aims this paper argues standardisation help to reach.

4.2. Background on harmonised standards in the proposed AI Act

Title III, Chapter 5 of the proposed AI Act, namely "Standards, Conformity Assessment, Certificates, Registration" of the proposed AI Act ascertains the role played by standardisation mechanisms. To understand the specific role played by harmonised standards in the proposed AI Act and their function in light of data equality, it is necessary to take a step back and briefly analyse the conformity assessment procedure in the proposed AI Act.

Article 19 establishes that providers shall undergo the conformity assessment procedure prior to putting into service or placing on the market a high-risk AI system. This conformity assessment is a procedure aimed at verifying the compliance of AI systems with requirements enshrined in Title III, Chapter 2, which are dedicated in large part to the data governance and transparency requirements of the AI training.

[49] See on this argument R. GESTEL, H. MICKLITZ, *European Integration Through Standardization: How Judicial Review is Breaking the Club House of Private Standardization Bodies*, in *Common Market Law Review*, 50, 2013, 145-182; J. GALLAND, *The Difficulties of Regulating Markets and Risks in Europe through Notified Bodies*, in *European Journal of Risk Regulation*, 4, 3, 2013, 365-374.

[50] Notified bodies are private entities in charge of monitoring the compliance of organisations with standards.

[51] The issue relates to liability profiles of notified bodies, which have the duty to monitor the actual adherence of standards by organisations, in case damages occurred because of a poor application of standards. A landmark CJEU case is Elisabeth Schmitt vs TÜV Rheinland LGA Products GmbH, case C-219/15 ECLI:EU:C:2017:128 where a medical device producer lacked to respect industrial standards for breast implants (which are Class III medical device) that forced several patients to remove their breast implants.

[52] European Commission, 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, cit., 7.

The first distinction, in this regard, shall be made between high-risk systems listed in Annex II and those listed in Annex III of the proposed AI Act. In Annex II are listed high-risk AI systems that are a product, or a component of a product, which is already regulated by a piece of legislation part of the NLF, such as the *Medical Device Regulation* (MDR).[53] On the other hand, Annex III lists the so-called *standalone* AI systems, i.e., systems that are not subject to other harmonised legislation based on the NLF.

The conformity assessment of AI systems in Annex II shall be carried out according to the procedure enshrined in the legal acts concerning the product at issue. Relevant for our discussion is the example of an AI system which is part of a medical device or that can constitute a medical device under the MDR. This situation will be extremely common in the health care once the proposed AI Act will be adopted. This *overlap* is especially likely in situations where AI systems are used to improve the provision of care services. For example, AI systems supporting surgical operations or used for diagnostic purposes. In these cases, the manufacturer of the medical device that embeds an AI system shall need to check during the conformity assessment – performed according to the MDR – also the compliance with requirements of Title III, Chapter 2 of the AI Act, to the extent that they will be applicable. In other words, the conformity assessment pertaining to the proposed AI Act provisions will be included into the MDR one.

On the other hand, the conformity assessment procedure concerning the *standalone* AI system, which are listed in Annex III, follows an *ad hoc* procedure defined by the proposed AI Act itself. In this perspective a further distinction shall be observed. Indeed, conformity assessment of *standalone* AI systems can be carried out according to two different ways: 1) "conformity assessment procedure based on internal control" (Annex VI); 2) "conformity based on an assessment of quality management system and assessment of technical documentation" (Annex VII). The former is the shorter procedure that does not entail the involvement of the notified body; the latter is the longer and more detailed procedure that is monitored and audited by the notified body. Article 43 establishes the cases where it is necessary to follow one or the other procedure.

Regarding the discussion on the role of harmonised standards in *standalone* AI systems, there is a presumption of conformity with requirements laid down in Chapter 2 of Title III, in case a high-risk AI system follows harmonised standards (or parts thereof) whose references are published in the O.J. by the Commission.[54] Therefore, the presumption of conformity helps AI providers demonstrate compliance with the procedure of conformity assessment. To be more precise, according to article 43.1, the provider who has adopted a harmonised standard may choose not to follow the conformity assessment procedure referred to in Annex VII, whilst it is allowed to follow the procedure in Annex VI. The adoption of harmonised standards is a way to avoid the longer and more complex procedure under the supervision of the notified body.

Nevertheless, article 43.1 applies only to a specific type of *standalone* AI system: "Biometric identification and categorisation of natural persons" (Annex III, point 1). Indeed, the procedure of Annex VII (with the notified body) is mandatory, in the absence of the adoption of harmonised standards, only

[53] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC.

[54] Article 40, of the proposed AI Act.

for this kind of high-risk AI system. All the other types of listed *standalone* high-risk AI systems are not subjected to this procedure, but they can follow, by default and even in absence of the adoption of harmonised standards, the procedure of Annex VI.
The decision of the Commission to limit the applicability of the conformity assessment through the notified body only to AI systems aimed at the biometric identification and categorisation of natural persons could restrict the spread of standardisation among *standalone* AI systems.
As regards AI systems as part of a product which is already part of the NLF, such as medical devices, the conformity assessment by the notified body could be required even if the AI provider has adopted harmonised standards. It will depend on the specific conformity assessment envisaged for the product under the NLF framework the AI system is part of.
It is possible to state that harmonised standards are meant to introduce a mechanism that, on the one hand, should ensure the smooth functioning of the internal market, ensuring that high-risk AI systems are safely placed on the market. On the other hand, such standards provide practical guidance to AI developers about ensuring compliance with legal requirements. Therefore, they operate on the crucial step of transposing general and abstract rules of law in practical inner procedures of organisations that develop AI systems.
In conclusion, through such standards the EU Commission provides a system of meta-rules applicable in specific sectors, such as health care or medical devices. By this way, the EU Commission creates sectorial rules (though not mandatory) whilst the main body of rules in the proposed AI Act has a horizontal approach.

4.3. Ongoing AI standardisation activities

Standards find their formal definition in article 2(1) Regulation 1025/2012[55] as "a technical specification, adopted by a recognised standardisation body, for repeated or continuous application, with which compliance is not compulsory". More specifically, harmonised standards are classified by letter (c) of the same article as "a European standard adopted on the basis of a request made by the Commission for the application of Union harmonisation legislation".
Harmonised standards must be developed by the three European Standards Development Organisations (SDOs), which are CEN, CENELEC and ETSI that are also known as European Standardisation Organisations (ESOs). ESOs can either develop their own standards from scratch or transpose standards already developed by international SDOs, such as ISO/IEC, IEEE, or ITU-T. However, if ESOs endorse standards developed at international level, they are supposed to adapt such standards to the EU market necessities and EU regulatory requirements.
Standardisation in healthcare, especially in the scenario of digital health has a significant role, for instance in the context of medical devices put on the European market. Even though many SDOs have been working on AI standardisation projects lately, AI based systems in healthcare still lack full

[55] Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation, amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC, 2009/23/EC and 2009/105/EC of the European Parliament and of the Council and repealing Council Decision 87/95/EEC and Decision No 1673/2006/EC of the European Parliament and of the Council.

harmonisation under many aspects. Two of the most active SDOs in setting standards for AI are IEEE SA and ISO/IEC.

IEEE has several working groups that are currently focusing on different aspects of AI standardisation. Many of the standards adopted or to be adopted by IEEE focus on ethical aspects of AI.[56] For example, the IEEE working group on algorithmic bias produces the P7003 "Standards Algorithmic Bias Considerations". Another standards project within the IEEE P7000 series that focuses on ethical aspects of AI systems is the P7000 standard which sets "Standards model process for addressing ethical concerns during system design". Moreover, IEEE standards also focus on personal data protection, such as through P7002 "Standard data privacy process". Data management and transparency are, on the other hand, objects of P7001 "Standards transparency of autonomous systems".[57] Relevant for the health sector are the standards IEEE 2801-2022[58] and the IEEE P2802.[59] The former is a recently adopted standard which aims at establishing quality procedures for health datasets management; it focuses on many aspects of the AI systems development, dealing with data quality, data collection and data annotation, but also the training of the professionals and the environment of training. The latter, on the other hand, is an ongoing project for the adoption of common terminology for safe AI based medical devices.

ISO/IEC have several standards either adopted, or under development on AI systems. Such standards focus on different aspects of the AI system development, from data quality to bias reduction in AI systems, or trustworthiness of AI. Most of the ISO/IEC AI standardisation activities come from the JTC-1/ SC42 AI sub-committee and its advisory groups and working group.[60]

Part of the ISO/IEC strategy in AI standards development is to link big data standards to AI standards and complement existing ISO/IEC standards with specific requirements for AI system.[61]

However, the SDOs that the EU Commission delegates to elaborate harmonised standards are the regional European SDOs, i.e., CEN/CENELEC and ETSI.

Among them, the most active in developing standards which could fit the AI context is ETSI. It has three working groups that deal with AI. Two of them, namely the ZSM-ISG and ENI-ISG, deal with networking aspects of AI and autonomous network systems of AI; however, no standards development seems to

[56] W. ZIEGLER, *A landscape analysis of standardisation in the field of artificial intelligence*, in *Journal of ICT Standardisation*, 8, 2, 2020, 151-184. See also, for an overview on standards development worldwide the StandICT observatory website: https://www.standict.eu/standards-repository/all?field_groups_target_id[]=267 (last visited 30/08/2022).

[57] For an overview on IEEE AI P7000 series standardisation projects, see: IEEE, *Ethically Aligned Design, First edition, A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, 285.

[58] https://standards.ieee.org/ieee/2801/7459/ (last visited 30/08/2022).

[59] https://standards.ieee.org/ieee/2802/7460/ (last visited 30/08/2022).

[60] See https://www.iso.org/committee/6794475.html (last visited 30/08/2022).

[61] D. LEWIS, et al., *Standardization and the Governance of Artificial Intelligence Standards*, in D. C. POFF, A.C. MICHALOS (eds), *Encyclopedia of Business and Professional Ethics*, Cham, 2021.

be expected by these working groups.[62] On the other hand, SAI-ISG aims at setting standards for security and robustness aspects of AI.[63]

The EU Commission's formal request for harmonised standards elaboration will probably be addressed just to CEN/CENELEC. It shall be noted that, to date, CEN/CENELEC have sparse projects on AI. CEN/CENELEC have published a road map for standardisation of AI after the adoption by the EU Commission of the White Paper for AI in 2020. From this document comes up that the strategy of CEN/CENELEC will be to transpose and, if the case, adjust AI standards developed by other SDOs, namely ISO/IEC. However, CEN/CENELEC do not close the door to the possibility to elaborate ad hoc standards where specific necessity for the EU market would come up.[64]

Despite the fact that the ESOs, especially CEN/CENELEC, have few ongoing standardisation projects about AI, this does not automatically mean that European interests will be necessarily under-represented in the AI standardisation scenario. Indeed, international SDOs, such as ISO/IEC, are constituted by representatives of national standardisation bodies.[65] In this respect, data shows that the distribution of chairs in SDOs' working groups on AI see a majority of representatives from Europe.[66] This consideration could indicate a strategy by European national standardisation bodies to develop standards at international level (e.g., through ISO/IEC) and then adapt these standards to the EU Commission request for harmonised standards at the EU level. It should be done using some special agreements between some SDOs and ESOs. Indeed, ISO/IEC and CEN/CENELEC have a strong relationship that is regulated by two cooperation agreement.[67] This approach is meant to avoid competing or overlapping standards being developed between SDOs and ESOs.

To sum up, there are a large number of aspects of AI systems that are in the agenda of SDOs or are already part of published standards. Such aspects can be categorised in different ways. The EU Commission's AI watch service, in its report on "AI Standardisation Landscape",[68] has classified the proposed AI Act requirements object of standardisation as follows: data and data governance; risk management system; technical data and record keeping; transparency and information to users; human oversight; accuracy robustness and cyber security; quality management system.[69]

From the analysis carried out in the report, it seems that IEEE is the SDO which is more active in developing standards in the domain of AI transparency and data governance aspect, as well as in ensuring

[62] ZSM-ISG (Zero Touch Network & Service Management) https://www.etsi.org/technologies/zero-touch-network-service-management (last visited 30/08/2022) and ENI-ISG (Experiential Networked Intelligence) https://www.etsi.org/technologies/experiential-networked-intelligence (last visited 30/08/2022).

[63] See SAI-ISG (Secure Industry Specification Group) https://www.etsi.org/committee/1640-sai (last visited 30/08/2022).

[64] CEN-CENELEC, *Focus Group Report: Road Map on Artificial Intelligence (AI)*, available at https://www.cencenelec.eu/areas-of-work/cen-cenelec-topics/artificial-intelligence/ (last visited 30/08/2022).

[65] ESOs' composition is made of representatives of national standardisation bodies as well.

[66] W. ZIEGLER, *op cit*.

[67] Namely, the Vienna agreement regulates relationship between ISO and CEN, while the Frankfurt agreement regulates relationship between IEC and CENELEC, see https://www.cencenelec.eu/european-standardization/international-cooperation/iso-and-iec/ (last visited 30/08/2022).

[68] S. NATIVI, S. DE NIGRIS, *AI Standardisation Landscape: state of play and link to the EC proposed AI Act for an AI regulatory framework*, 2021, doi:10.2760/376602, JRC125952.

[69] S. NATIVI, S. DE NIGRIS, *op. cit.*, 18-19.

proper human oversight on AI systems. Such aspects are those we argued can influence the most data quality and therefore also equality in the healthcare domain. On the contrary, other SDOs, such as ISO/IEC or ETSI or ITU-T, despite being active also in these domains, seem to focus more on accuracy, robustness and cybersecurity or technical data or record keeping aspects.[70]

4.4. Standardisation and impact on AI and digital healthcare

The proposed AI Act allows AI providers adopting harmonised standards in relation to Title III, Chapter II. It means that harmonised standards could be used to reach and demonstrate compliance with whichever requirement set in this chapter. Despite Chapter II enshrines a broad number of requirements, only some of them seem to have a relevant impact on equality.

As we have argued in the previous sections of the paper, data management and data transparency requirements are introduced by the proposed AI Act in order to prevent discrimination when high-risk AI systems are deployed. In this perspective, some scholars have already supported the added value of standardisation of AI training procedures for achieving higher objectives than pure market regulation.[71]

Defining standardised procedures would help health care providers and organisations implement general rules defined at policy and regulatory level. Indeed, industry often struggles in transposing such high-level tenets into actionable and auditable criteria.[72]

In support of this reasoning, even before the proposed AI Act, scholars have highlighted the necessity to develop a strong and harmonised system of audits and organisational practices in order to translate AI high-level principles into day-by-day organisations' operations.[73]

Moreover, the critical features of developing AI standards stand also in the necessity to transpose socio technical requirements into standardisation processes. AI systems indeed are different to other technologies that are traditionally subject to standardisation, such as classic medical devices, to the extent that their outcomes are influenced also by humans, ethical and environmental features. For example, AI-based medical devices may reinforce social biases because of poor ethical and socio awareness during the training activities.[74]

The proposed AI Act seems to be aware of such criticalities. Therefore, it has included a conformity assessment based on a notified body and harmonised standards and has connected it to the already existing NLF framework.

Moving the discussion to the content of AI standards, part of the literature criticises the current approach adopted in AI context by the main standardisation bodies. The criticism mainly stems from the

[70] S. NATIVI, S. DE NIGRIS, *op. cit.*, 20-21.

[71] K. MATUS, M. VEALE, *Certification systems for machine learning: Lessons from sustainability*, in *Regulation and Governance*, 16, 1, 2022, 177-196.

[72] J. MÖKANDER, L. FLORIDI, *Ethics-Based Auditing to Develop Trustworthy AI*, in *Minds and Machines*, 31, 2, 2021, 323-327.

[73] M. BRUNDAGE, et al., *Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims*, available at https://doi.org/10.48550/arXiv.2004.07213 (last visited 30/08/2022); I. RAJI, et al., *Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing*, in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, 33-34; J. MÖKANDER, L. FLORIDI, *op. cit*.

[74] See above section 1.2.

fact that current standards (or proposed standards) are *meta-standards* which focus on common terminology, networking aspects, technology interoperability, and safety. On the contrary, a more comprehensive and holistic approach should be adopted, which should address also social issues related to AI systems and include also environment-based aspects.[75]

The process to develop AI systems might be seen as a *sociotechnical system*. This means that AI development procedures embed complex relationships between non-human and human systems.[76] AI systems development presents more challenges in performing auditing activities than other products or processes. This is because of the network of interaction between humans, data sources, modalities of data collection and model training.[77] For third parties, it is not trivial to assess whether AI developers' behaviours and processes are consistent with norms and principles.

IEEE 2801 standard, for example, tries to adopt a more comprehensive approach when shaping standard procedures for medical AI dataset management. Indeed, such a standard deal with the full cycle of dataset management, including also possible environment and personnel interactions. On the other hand, other standards, such as ETSI DES/eHEALTH-008 data recording requirements for eHealth, focus only on few specific steps of the AI system development.[78]

Therefore, there is a recognised necessity to find operational mechanisms that demonstrate adherence of the AI developers to ethical and legal requirements.[79]

It could be noted that some AI standards can build on existing standards that deal with software development, testing and safety, while other specific aspects of AI systems need to be addressed through a new approach.[80]

Bias avoidance through data quality is one of those aspects that require a more holistic approach. Standards dealing with such features of the AI system shall therefore address both dataset quality and behaviour models within the organization.[81]

ISO/IEC has adopted several standards which are meant to reduce the risk of bias when deploying AI systems and ensuring data quality during data management procedures. On the other hand, IEEE projects for developing ethics AI systems have already led to the adoption of several standards that consider ethical and social aspects of AI deployment.

In other words, AI standardisation should not focus only on software and hardware mechanisms, but also on institutional mechanisms in order to incentivise the proper behaviours in people involved in the AI development. Only by this way, AI developers will be able to demonstrate the trustworthiness of AI systems to external parties.

---

[75] K. MATUS, M. VEALE, *Certification systems for machine learning: Lessons from sustainability*, cit., 184; G. ADAMSON, et al., *Designing a Value-Driven Future for Ethical Autonomous and Intelligent Systems*, in *Proceedings of the IEEE*, 107, 3, 2019, 218-525; I. RAJI, et al., *op. cit*.

[76] M. BRUNDAGE et al., *op. cit*.

[77] I. RAJI, et al., *op. cit*.

[78] See https://portal.etsi.org/webapp/WorkProgram/Report_WorkItem.asp?WKI_ID=56908 (last visited 30/08/2022).

[79] J. MÖKANDER, L. FLORIDI, *op cit*; M. BRUNDAGE, et al., *op. cit*.; D. LEWIS, et al., *op. cit*; G. ADAMSON, et al., *op. cit*.

[80] D. LEWIS, et al., *op. cit*.

[81] D. LEWIS, et al., *op. cit.*

Another interesting part of the holistic approach to AI standardisation is the consideration that the objective of the standardisation process should be to provide additional information to the consumer, the final user, or the patient, about some aspects of the AI systems that are otherwise impossible to assess by an external agent. Such features are also called *credence qualities*.[82] Following this approach, the aspects that a standardisation mechanism should address include privacy and ethical aspects during data collection and AI training, biases in dataset building, or cyber-attacks. All these aspects are indeed considered by Title III, Chapter 2 of the proposed AI Act. Therefore, they are potential contents of harmonised standards under article 40 the proposed AI Act.

More specifically, looking at the requirements of Title III, Chapter 2 of the proposed AI Act, article 10 requires appropriate data governance and management practices. This could be an interesting opportunity to apply the reasoning about credence qualities and set standards that harmonise hidden AI processes about data quality in training dataset or during data collection. In the same vein, we also note requirements contained in article 15 related to data accuracy, robustness, and cybersecurity. Indeed, standardisation efforts could be used to ensure data quality concerning "errors, faults or inconsistencies" in the environment where an AI system is trained and deployed. Errors in the dataset could generate, in ML context, biased outputs, which could be used as inputs in future data training and generate *feedback loops*. This is the case for AI systems that end up suffering from biases due to the embedding of societal values or historical discriminations proper of the training set. Again, harmonised standards regulating the social environment during AI training could help reduce unequal treatment among patients.

Moreover, given the several connections between article 10 and the GDPR, a mechanism of coordination between harmonised standards of the proposed AI Act and GDPR certification mechanisms[83] is more than desirable. However, the current version of the proposed AI Act does not address this issue.[84]

Harmonised standards are then explicitly mentioned in article 9 of the proposed AI Act for developing the risk management system. While article 12, on the other hand, establishes to consider "recognised standards" when setting down the requirement to deploy a log system for high-risk AI. However, it is not clear whether such recognised standards shall be meant as harmonised standards according to article 2(1)(c) of Regulation 1025/2012 on standardisation.

Finally, compliance with cybersecurity requirements (art. 15.4) could also be reached through standards within the proposed AI Act even though a specific certification system is provided by article 42.2 in combination with regulation (EU) 2019/881.[85]

From a different perspective, we can note that besides the content of the standard, other elements determine the efficacy of standardisation measures, such as the capacity to modify parties' behaviour or the diffusion of the standards. However, it is difficult to directly assess the impacts of standardisation mechanisms, especially in AI and healthcare, because of the complexity of the context.

[82] K. MATUS, M. VEALE, *Certification systems for machine learning: Lessons from sustainability*, cit., 178.

[83] See articles 42 and 43 GDPR.

[84] EDPB-EDPS, Joint Opinion 5/2021 on the proposed AI Act for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), 18 June, 20.

[85] Regulation (EU) 2019/881 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act).

Nevertheless, a set of indirect outcomes could be identified as positive effects generated by standardization:[86] 1) improvement in the relationship among stakeholders while contributing to reduce the conflicts and grow the positive relationship, such as contracts and agreements; 2) increasing of experimentations and new best practices;[87] 3) introducing best practices that one day could be transposed in traditional regulation.

To conclude, harmonised standards in the proposed AI Act could address several aspects of the AI system's value chain and deal with quality-related problems. De-biasing activity is one of them. Especially in the health domain, bias in training datasets, or bias generated by the interaction between the physician and the AI system, are concerns of the utmost importance. The standards should impact the social environment in which AI is trained rather than address only technical aspects. Requirements of Title III, Chapter 2 of the proposed AI Act impose several duties, especially in article 10, that could be the object of harmonised standards.

As noted earlier, many standardisation activities are already in the agenda of the main SDOs both at international and regional level. It should be noted that most of the standards are *horizontal*, i.e., they are applicable to AI systems used for different purposes. However, some *vertical* standards, e.g., specific for the health sector have been developed. The appropriate combination of horizontal and vertical standards will determine the success of the AI standardisation process.

At the time of writing the ESOs are waiting for the EU Commission formal request to develop AI standards; these standards might be therefore ready by 2024 (while the AI act should come into force by 2025). However, the two main ESOs, i.e., CEN/CENELEC, will probably follow, and if the case, adapt the work of ISO/IEC rather than developing their own standards.

The efficacy of the standardisation system is assessed upon many other elements, such as the capacity to modify behaviours of the parties adopting the standards, the actual level of enforcement, or how widespread the standard is. We remind that harmonised standards according to articles 40 to 43 are meant to produce a presumption of conformity just for high-risk systems. The final spread of harmonised standards will also heavily depend on these considerations.

5. Conclusions

AI can have promises and perils in healthcare. One of the main perils in healthcare AI systems is bias. Therefore, to fully appreciate the benefits of AI, we ought to face some fundamental rights issues, inter alia, equality and on discrimination. In our paper we highlighted that the EU safeguards on non-discrimination are not meaningful and fully applicable to AI and healthcare. However, the proposed AI Act could play a role by, firstly, broadening the scope of application of EU law, and secondly, minimizing ex ante the presence of bias in datasets used to train AI healthcare systems that lead to generating unequal outcomes.

In this respect, in our paper we analysed how the proposed AI Act includes provisions related to ensuring data quality and more equitable outcomes in healthcare AI systems. We note that this intertwining of areas of interest and requirements developed by the proposed AI Act is of particular

[86] K. MATUS, M. VEALE, *Certification systems for machine learning: Lessons from sustainability*, cit., 191.

[87] Title V of the proposed AI Act introduces some measures to support innovation, such as regulatory sandboxes.

importance for the protection of the fundamental rights and values of the European Union. In our article we have, in fact, highlighted the particular need to ensure the principle of equality through the implementation of data quality. In this respect, data quality is of particular importance in the healthcare sector, where the use of biased AI systems can have detrimental consequences not only for the protection of fundamental rights, but also for the health and safety of a patient. These mechanisms include specific requirements for high-risk AI systems' data quality requirements on data and data governance, transparency, provision of information to users, and data accuracy, and AI systems robustness, discussed in detail in the paper.

We also have observed in our paper that the proposed AI Act on its own cannot achieve the goals of ensuring data quality. Where potential gaps in the protection of data equality by the requirements of the proposed AI Act may arise, in our paper we suggested finding alternative methods for the protection of this concept. In this regard, as a last resort, we suggest encouraging adherence to uniform standards established within different industries and sectors, even though the impact of standardisation mechanism is yet to be directly assessed due to the complexity of AI software in healthcare. In this respect, it should be noted that the European legislator has chosen to adopt a horizontal approach in the proposed AI Act. This means that the requirements in the proposed AI Act are applicable indiscriminately to every AI system falling into the material scope of the regulation, without distinction based on the sector of application. However, some sectors, such as healthcare are characterised by problems that are not common in other sectors and vice versa. Harmonised standards therefore are supposed to specify the general and horizontal rules, laid out in the proposed AI Act, into each sector of application. Yet, the use of harmonised standards raises doubts about delegating the legislative process to private organisations that elaborate those standards in lack of democratic oversight. On the other hand, the alternative solution would be to elaborate sector-oriented rules integrated directly into the proposed AI Act or into sectoral pieces of legislation. In this case the shortcomings would be related to the hypertrophy of hard law at the EU level, with negative impacts on the technological development and on the digital single market. Moreover, it seems difficult to believe that the EU legislator would have the skills and the know-how necessary to define the requirements of an over-detailed regulation on AI that deals with sector-specific requirements.

Nonetheless, harmonisation and standards can be precisely the tools that address several aspects of the AI system's value chain dealing with data quality related problems such as de-biasing. One peculiar aspect, observed with respect to standards, should be made as concluding remark. As reiterated earlier, standards tend to focus on the technical aspects of AI systems yet overlooking the impact that the social environment has on AI systems training. Indeed, in healthcare, several elements from the training environment could determine the presence of bias. Sometimes biases are even hidden in peculiarities of social relationship within a specific medical environment.[88] Such elements are difficult to detect from the outside, and for sure are hidden to the final user (the patient). Establishing standards that regulate also social dynamics and ethic-related aspects in health care during the AI training could help shed light on dynamics and procedures potentially carrier of biases and discrimination.[89]

[88] See above section 1.2.
[89] See above section 4.5.

Finally, from the analysis of the proposed AI Act, it is clear that high-risk AI systems have to pass a conformity assessment which will be highly based on harmonised certifications and standards. In this perspective, we should underline that the subsidiary tools of compliance, coupled with the proposed AI Act's enshrined data quality insurance mechanisms, will raise the bar in ensuring that high quality data is used in training healthcare AI systems.

ISSN 2284-4503