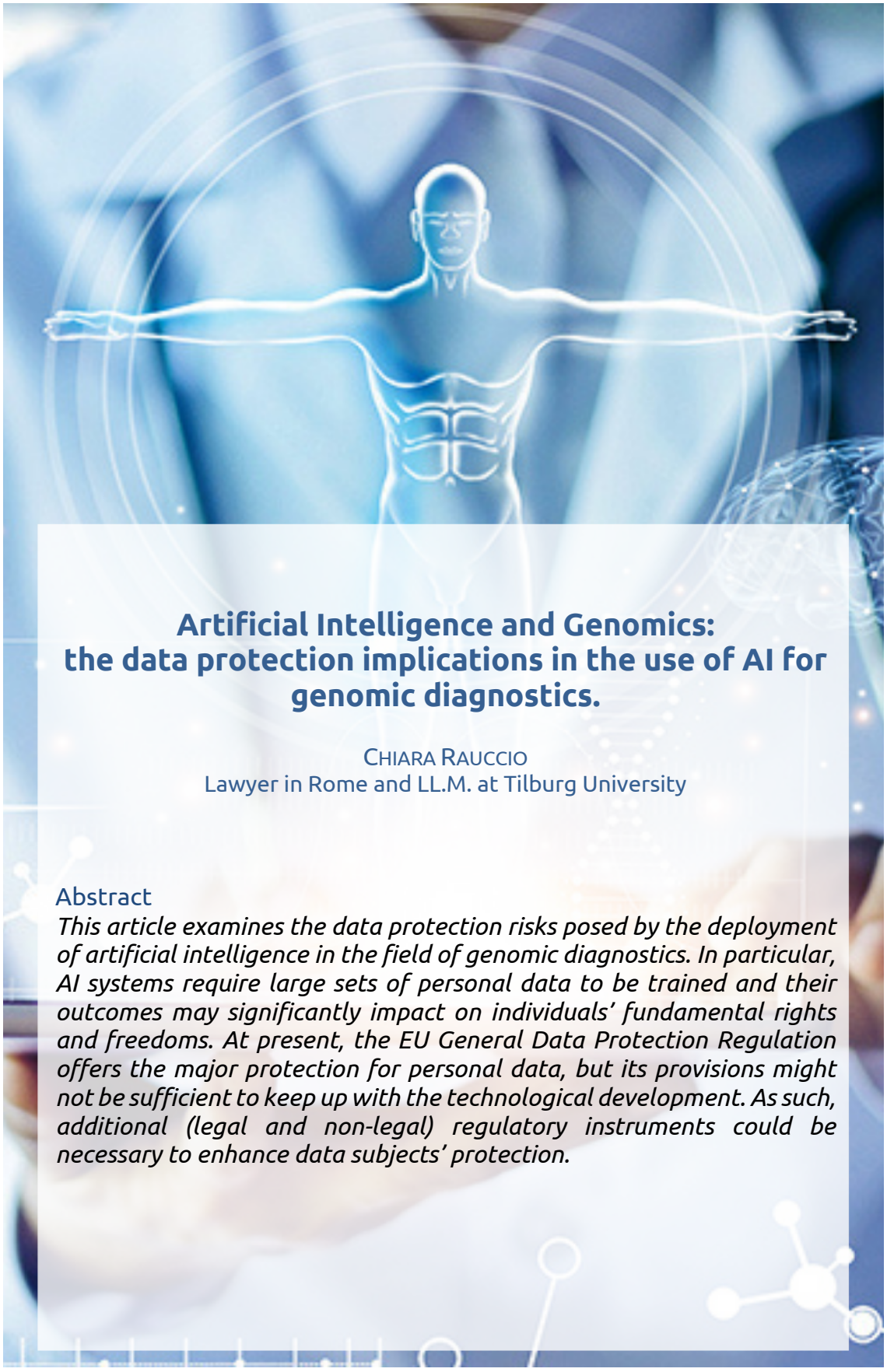
EJPLT

European Journal of Privacy Law & Technologies • 2021/1 • ISSN 2704-8012



Artificial Intelligence and Genomics: the data protection implications in the use of AI for genomic diagnostics.

CHIARA RAUCCIO
Lawyer in Rome and LL.M. at Tilburg University

Abstract

This article examines the data protection risks posed by the deployment of artificial intelligence in the field of genomic diagnostics. In particular, AI systems require large sets of personal data to be trained and their outcomes may significantly impact on individuals' fundamental rights and freedoms. At present, the EU General Data Protection Regulation offers the major protection for personal data, but its provisions might not be sufficient to keep up with the technological development. As such, additional (legal and non-legal) regulatory instruments could be necessary to enhance data subjects' protection.

Keywords: data protection; artificial intelligence; machine learning; healthcare; genomic diagnostics; GDPR; soft law; ethics.

Summary: Introduction. – 1. An overview of artificial intelligence, machine learning and deep learning. – 1.1. Application of AI in genomic diagnostics. – 1.2. Big health data and AI. – 2. The implications of AI on data protection. – 3. Application of European data protection law to AI. – 3.1. The GDPR underlying concepts. – 3.2. The GDPR fundamental principles. – 3.3. The GDPR data subjects' rights. – 4. Alternative regulatory instruments to deal with data protection issues in AI-based genomic diagnostics. – 4.1. Medical law. – 4.2. The MDR Regulation. – 4.3. Soft law. – 4.4. Digital Ethics. – Conclusions.

Introduction.

In the last few years artificial intelligence (AI)[1] has been increasingly applied in healthcare with the consequence to transform several aspects of medical practice and the potential to create new opportunities for patients' care, from more precise diagnosis and treatment applications (i.e. precision medicine)[2] to robotic assistance to caregivers, support for elderly care, and monitoring of patients' conditions.

An important field where AI will likely be deployed in the next years is genomic diagnostics, the branch of medicine which studies the interaction between disease and human genes and that finds its roots in genetic science. Genetics studies the structure of DNA and the functioning of genes (fragments of DNA that dictate the production of proteins). [3] The complete set of genes in a human being constitutes the "human genome" that encodes the genetic information and differs from individual to individual. Disorders and mutations of genes, combined with external factors like diet, lifestyle, and environment, may cause diseases. Genetic testing is used to ascertain whether a person has a genetic predisposition to developing a certain disease, or to confirm a diagnosis.[4] Yet, the development of a disease depends not only on genes, but

[1] For the purpose of this introduction the term "artificial intelligence" will be used as a generic reference to any computer systems intended to simulate human intelligence and human skills such as learning and problem-solving.

[2] D Cirillo, A Valencia, 'Big data analytics for personalized medicine' (2019) 58 Current Opinion in Biotechnology 161-167. See also the REVOLVER (Repeated Evolution of Cancer) project: https://www.healtheuropa.eu/personalised-cancer- treatment/87958/, or the Murab project which conducts more accurate biopsies, and which aims at diagnosing cancer and other illnesses faster: https://ec.europa.eu/digital-single-market/en/news/murab-eu-funded-project-success-story.

[3] It is based on the studies of genetics initiated in 1953 by Crick and Watson (see JD Watson, FHC Crick, 'Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid' (1953) Nature 737) who discovered that DNA is structured in a double helix constituted of genes that form the "human genome", a sequence of genes encoding the entirety of the genetic information stored in each cell and passed from each generation. In simple words, DNA is a string of complex molecules called nucleotides. It contains the genetic information of the individual in the form of instructions for building the molecules that make the body work (like proteins). Genes are segments of DNA that each carry a specific set of instructions for how to make a certain aspect of the individual (our genetic code contains around 23,000 genes). A genome is the complete set of the genetic material in an organism. For further details see https://scienceblog.cancerresearchuk.org/2018/05/29/science-surgery-whats-the-difference-between-the-words-genome-gene-and-chromosome/.

[4] PS Harper, 'What Do We Mean by Genetic Testing?' (1997) 34 Journal of Medical Genetics, 749.

on a variety of internal and external factors. Hence, traditional genetic tests can only provide an estimation of the risk to develop that disease.[5]

In this respect, AI may enhance diagnostic tools. Indeed, its capacity to analyse large amounts of data, identify correlations and generate inferences, as it will be explained in the next paragraphs, may enable more accurate predictions and thus earlier diagnosis and more effective treatments.

To properly work AI systems need to process large amounts of personal data related not only to health *stricto sensu,* but also to environmental and social aspects (e.g. economic status, air pollution levels, living place, job, eating habits).[6] Such data is combined into complex datasets and is analysed to create profiles to predict whether patients will likely develop a particular disease.[7] While on the one hand these technologies may produce beneficial effects for diagnosis and treatment of serious diseases, on the other hand they pose several risks in terms of protection of individuals' rights.

This paper will seek to analyse the risks associated with the application of AI technologies in the field of genomic diagnostics in relation to the protection of personal data and to assess whether the current EU legal framework can effectively face them and, if not, which further regulatory interventions would be necessary.

The paper will be structured as follows. Paragraph 1 will introduce AI with specific reference to machine learning, neural networks and deep learning. Then, it will focus on the application of AI in healthcare, specifically in genomic diagnostics, and its connection with big data. Paragraph 2 will analyse the main data protection implications and the risks that AI poses especially in relation to profiling, automated decision-making and opacity. Paragraph 3 will examine the current EU legal framework for the protection of personal data and its adequacy to regulate AI technologies for genomic diagnostics. In particular, the effectiveness of the GDPR to keep up with the recent technological developments will be assessed. Finally, Paragraph 4 will investigate alternative instruments like soft law and medical ethics that may add value to the tools currently in place and straighten the safeguard of individuals' fundamental rights.

1. An overview of artificial intelligence, machine learning and deep learning.

Artificial intelligence (AI) is among the most important and debated drivers of technological development of the last decades and, especially, the last years.[8] Several different definitions of AI have been proposed (one of the most

[5] A De Paor, 'Advancing Genetic Science and New Technologies' in *Genetics, Disability and the Law: Towards an EU Legal Framework* (Cambridge University Press 2017).
[6] D Cirillo, A Valencia, 'Big data analytics for personalized medicine' (n 2).
[7] Nuffield Council on Bioethics, 'Artificial intelligence (AI) in healthcare and research Nuffield Council' [2018].
[8] Modern studies on artificial intelligence date back to the early works of Alan Turing, John McCarthy, Arthur Samuels, Alan Newell, and Frank Rosenblatt, among others. In particular, artificial intelligence was officially born in 1956 during the workshop organised by John McCarthy at the Dartmouth Summer Research Project on Artificial Intelligence with the goal to find how to make machines simulate aspects of human intelligence. In the proposal for that workshop McCarthy used for the first time the term "artificial intelligence" explaining that "the goal of AI is to develop machines that behave as though they were intelligent" (see J

recent is the one suggested by the EU Commission in the Communication on AI)[9] but, despite the lack of a universally accepted definition, the common goal of AI researchers is to create a system that acts in a rational way, namely by choosing the best action to achieve a certain goal.[10] To do so, in simple terms, AI systems through sensors perceive the external environment and collect data from it (*input*). The data coming from the sensors is transformed into information understandable by the system (*knowledge representation*) that reasons on such information (*knowledge reasoning*) to decide what the best action is; finally, it takes the action through actuators (*output*).[11] To perform this function, AI systems rely on algorithms, sequences of unambiguous, well-defined, computer-implementable instructions to execute a task, usually to solve a class of problems or to perform a computation.[12]

The main limit of early AI systems was that a formal representation of the relevant knowledge had to be provided in advance by humans since the system could only deal with known situations but was unable to address new cases out of its knowledge base.[13] A step forward was the development of machine learning (ML),[14] an AI system able to automatically improve its performance at a task by learning from experience and input data, just as humans do.[15]

A particular method of ML is "artificial neural network" (ANN), so called because it simulates the human brain made of neurons which receive and transmit information. Similarly, artificial neural networks are a net of nodes, called "neurons" or "perceptrons", arranged in multiple layers each of which connected to the layers on either side. Neurons in the first layer receive information from the outside (*input units*). The information is processed by neurons in the internal layers (*hidden units*) and the last layer generates the

---

McCarthy, ML Minsky, N Rochester, CE Shannon, 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence' [1955]).

[9] "*Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications*)", Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe, Brussels, 25.4.2018 COM (2018) 237 (final).

[10] SJ Russell, P Norvig, *Artificial Intelligence: A Modern Approach* (3rd edition, Prentice Hall, 2009).

[11] High-Level Expert Group on Artificial Intelligence, 'A definition of AI: Main capabilities and scientific disciplines' made public on 8 April 2019.

[12] D Harel, YA Feldman, *Algorithmics* (Addison Wesley, 2004).

[13] European Parliamentary Research Service, Scientific Foresight Unit (STOA), 'The impact of the General Data Protection Regulation (GDPR) on artificial intelligence' June 2020.

[14] Machine Learning has been defined as the "field of study that gives computers the ability to learn without being explicitly programmed". See AL. Samuel, 'Some Studies in Machine Learning Using the Game of Checkers', (1959) 3 IBM J. Res. Dev., 210-229.

[15] There are three main machine learning approaches. (i) Supervised learning: the machine learns through "supervision" or "teaching"; it is provided with a "training set", that is a set of labelled data, and is instructed on how this data has to be categorised. The learning algorithm (trainer) uses this set to learn how to identify specific features, infers the logic underlying the set and builds a model. Then, the model is applied to new data in order to identify and categorise unseen data based on patterns detected in the training set. (ii) Unsupervised learning: the system does not receive external instructions in the form of labelled data but finds patterns on its own. (iii) Reinforcement learning: the system learns by trial-and-error through a "reward and punishment" approach. It is not provided with training data but learns by the outcome of its own actions: if it is given a reward signal, the algorithm (learner) learns that the action is right, if no reward follows the action, this is learnt to be wrong. For a more in-depth analysis see P Natarajan, B Rogers, 'Applied Machine Learning for Healthcare" in P Natarajan, JC Frenzel and DH Smaltz (eds.), *Demystifying Big Data and Machine Learning for Healthcare* (CRC Press 2017), 29.

result (*output unit*). The connection between one layer and another is associated with a number called "weight", which defines the value of each input feature in predicting the final output (i.e. how important a certain genetic variation is in developing a disease). When the neural network is trained on the training set, it is initialised with a set of weights. During the training phase weights are adjusted based on whether the outputs are right or wrong. If the output matches the labelled data, the weight is kept. If the output is sub-optimal (i.e. doesn't match the label), the learning algorithm adjusts the weight until the network, when presented with that input, will generate the correct output.[16]

Neural networks may be more or less complex depending on the number of hidden layers they have. A subset of neural networks is "deep learning" which includes several hidden layers, and each layer learns from the layer below: it receives as data input the previous layer's output, so the deeper the layers, the more complex the features that nodes can recognise.[17] As a consequence, the network is more accurate and with less need of human guidance.

1.1 Application of AI in genomic diagnostics.

AI, in particular machine learning and deep learning, may be extensively deployed in genomic diagnostics to detect disease at an earlier stage thus improving patients' quality of life. Such technologies are not intended to replace physicians but rather to augment their capabilities and make more efficient and cost-effective routine standardised tasks.[18] Different algorithms can perform different tasks, but the main areas where AI has proven helpful are classification, image and video recognition, clustering, and prediction.[19] Often these functions are combined to achieve more effective results. In particular, image recognition is used to train neural networks to predict disease, especially in oncology.[20]

Important results in the use of AI in genomic diagnostics have already been achieved and further are expected by the next years.[21] An example is IDx-DR, a

[16] European Parliamentary Research Service, Scientific Foresight Unit (STOA), 'The impact of the General Data Protection Regulation (GDPR) on artificial intelligence' June 2020.
[17] J Schmidhuber, 'Deep Learning in Neural Networks: An Overview' (2015) 61 Neural Networks.
[18] See for instance KY Ngiam, IW Khor, 'Big Data and Machine Learning Algorithms for Health-Care Delivery' (2019) 20 The Lancet Oncology.
[19] Classification is carried out by deep learning algorithms that are trained with a sheer number of examples to analyse symptoms and classify them into labelled diseases in order to suggest possible diagnosis or identify patients with high readmission risk. Image and video recognition means that deep learning systems can detect objects in complex images, label and classify them into disease types. Clustering is the function of AI systems that identify similarities and connections among items and group them accordingly; typically, the algorithm is given a set of features for each item and a number of clusters to create, and it will combine such features and divide the items into the given number of clusters). Prediction is achieved by neural networks able to analyse large amounts of data as input, combine them and produce as output the likelihood that a patient will get a certain disease in the next future.
[20] The network is provided with raw image data used as training data to identify highly predictive features (regions of interest). Such regions of interest are combined with other relevant data – in particular, genomic data – to build a model to detect patterns in new data. This is the so-called genomic survival convolutional neural network (GSCNN model). See P Mobadersany *et al.*, 'Predicting Cancer Outcomes from Histology and Genomics Using Convolutional Networks' (2018) 115 Proceedings of the National Academy of Sciences.
[21] For instance, Google is collaborating with health delivery networks to build prediction models from big data to warn clinicians of high-risk conditions, such as sepsis and heart failure, and provide them decision support to find the best diagnosis and treatment for patients. Other firms, such as Foundation Medicine

software program that uses a ML algorithm to analyse the retinal images from patients, producing one of two possible screening results: positive (more than mild diabetic retinopathy) or negative (mild diabetic retinopathy or lower).[22]

AI has also been used in the fight against amyotrophic lateral sclerosis (ALS), a devastating neurodegenerative disease caused by mutations of RNA-binding proteins (RBPs). [23] The number of RBPs currently associated with ALS represents only a small fraction of the total RBPs present in the human genome and it has been hypothesised that further unidentified RBPs may be linked to ALS. On this basis, IBM Watson has been developed to predict new potential RBPs in ALS.[24]

IBM also developed other AI systems to support physicians in making treatment decisions for different types of cancer [25]. One is Watson for Oncology (WFO), a system trained with different data sources like biomarkers derived from the patients (e.g. sex, age, type of tumour), medical literature, national treatment standards and medical records. Based on the analysis of such data, the algorithm suggests what treatment options are available and ranks them in three categories (recommended, for consideration, and not recommended) according to their suitability for the patient.[26] Several studies have been conducted to examine the concordance between the treatment recommendation proposed by WFO and actual clinical decisions by expert oncologists.[27]

A further development is IBM Watson for Genomics (WfG), an AI system for treating cancer with a personalised approach. To date, the analysis of patient genomes is performed manually by teams of human experts and may take weeks. WfG aims at supporting oncologists by analysing large volumes of genomic data.[28] WfG is first trained with databases of genomic alterations,

---

and Flatiron Health, both owned by Roche, specifically focus on diagnosis and treatment recommendations for certain cancers based on their genetic profiles given the difficulty of human clinicians to understand all genetic variants of cancer and their response to new drugs and protocols. See T Davenport, R Kalakota, 'The potential for artificial intelligence in healthcare' (2019) 6 Future Healthcare Journal, 94-98.

[22] See MD Abràmoff *et al.*, 'Pivotal Trial of an Autonomous AI-Based Diagnostic System for Detection of Diabetic Retinopathy in Primary Care Offices' (2018) 1 npj Digital Medicine.

[23] RNA-binding proteins (RBPs) are proteins having important functions in the regulation of gene expression. See C Oliveira et al., 'RNA-binding proteins and their role in the regulation of gene expression in *Trypanosoma cruzi* and *Saccharomyces cerevisiae*' (2017) 40(1) Genetics and Molecular Biology, 22-30.

[24] To achieve this goal the system has first been trained with published literature from which the system extracted specific features to identify new connections between genes, proteins, drugs, and diseases. As a result, Watson created a model of the known set of RBPs linked to ALS in such a way to apply that model to an unseen set of other RBPs, in order to rank all the new RBPs by similarity to the known set. Finally, the top-ten-ranked RBPs are validated by medical experts. For more details see N Bakkar *et al.*, 'Artificial Intelligence in Neurodegenerative Disease Research: Use of IBM Watson to Identify Additional RNA-Binding Proteins Altered in Amyotrophic Lateral Sclerosis' (2017) 135 Acta Neuropathologica.

[25] At present these systems, now in use in some hospitals especially in Asia, still have some limitations with the risk of producing inaccurate predictions or proposing incorrect treatment recommendations. However, they demonstrate that technology and medicine are moving towards this direction and in the next future we could likely assist to relevant developments in this sense. See C Ross, I Swetlitz, 'IBM's Watson supercomputer recommended "unsafe and incorrect" cancer treatments, internal documents show' (2018) STAT.

[26] A Tupasela, E Di Nucci, 'Concordance as Evidence in the Watson for Oncology Decision-Support System' (2020) AI & Society.

[27] Among others N Zhou *et al.*, 'Concordance Study between IBM Watson for Oncology and Clinical Practice for Patients with Cancer in China' (2018) 24 The Oncologist.

[28] K Rhrissorrakrai, T Koyama, L Parida, 'Watson for Genomics: Moving Personalized Medicine Forward' (2016) 2 Trends in Cancer.

published literature and clinical studies; then, it receives as input data the patient's cancer genetic variants and evaluates each variant using advanced cognitive analytics, an analytical technique that analyses large datasets by using AI.[29] As an output, it presents a report of relevant therapies along with links to the relevant medical literature. Doctors will then review the report together with additional clinical evidence to make an informed treatment decision.[30]

1.2 Big Health Data and AI.

To efficiently operate ML systems need to process large amounts of data. The more numerous and diverse the data they are trained on, the more accurate and correct the outputs, especially when AI systems are used to make diagnosis predictions or suggest treatment recommendations. [31] However, having more data also gives rise to more problems, as it will be extensively illustrated in the following paragraphs.

In the past, collecting and analysing data was time-consuming and expensive; thus, physicians only collected strictly necessary data. Not surprisingly AI has reached a turning point over the last years, when the amount of data gathered and processed around the world has exponentially grown and data analysis has become much cheaper and faster.[32]

Today data collection has evolved under two aspects: (i) the number of data available is much higher; and (ii) the content of such data is much broader and diverse. This phenomenon is commonly referred to as "big data".[33] Big data has revolutionised healthcare: while in the past health-related data only included data collected in the healthcare environment by health actors, with the advent of big data this category has included any types of data from which it is possible to infer information about health (data related to physical, environmental, or biological aspects, as well as social, economic, or individual *status*, lifestyle, commercial preferences, so-called "big health data"). Furthermore, data may be collected from an array of (non-medical) sources like social networks that aggregate information about people's preferences, interests, contacts, etc.;[34] smartphones, able to track people's movements, activities and social interactions; "mHealth apps" developed to store health-related data and keep

[29] See 'Cognitive Analytics — Combining Artificial Intelligence (AI) and Data Analytics', Ulster University, Cognitive Analytics Research Lab at https://www.ulster.ac.uk/cognitive-analytics-research/cognitive-analytics.

[30] K Itahashi *et al.*, 'Evaluating Clinical Genome Sequence Analysis by Watson for Genomics' (2018) 5 Frontiers in medicine.

[31] WN Price, IG Cohen, 'Privacy in the Age of Medical Big Data' (2019) 25 Nature Medicine.

[32] It is estimated that between 1987 and 2007 alone the amount of data in the world grew one hundred times. See M Hilbert, P Lopez, 'The World's Technological Capacity to Store, Communicate, and Compute Information' (2011) 332 Science.

[33] No unanimous definition of big data exists, but according to a generally accepted opinion, it may be described by the "three Vs": volume (large amounts of data), velocity (high speed of access and analysis), and variety (substantial data heterogeneity across individuals and data types). See Executive Office of the President. Big data: seizing opportunities, preserving values https://bigdatawg.nist.gov/pdf/big_data_privacy_report_may_1_2014.pdf (2014).

[34] S Hoffman, 'Big Data's New Discrimination Threats Amending the Americans with Disabilities Act to Cover Discrimination Based on Data-Driven Predictions of Future Disease' in G Cohen, A Hoffman and W Sage (eds.), *Big Data, Health Law, and Bioethics* (Cambridge University Press, 2017).

track of users' health and physical conditions; and Internet of Things (IoT) devices that, through their sensors, perceive external inputs, process them and interact with each other.[35]

Such an amount of data requires adequate data analysis methods to process it, and these are AI technologies, in particular ML that is able to continuously learn from training data and apply the learnt model to new and unseen data. It follows that algorithms trained on big health data not only may confirm existing hypothesis and support existing diagnosis but can also suggest new hypothesis and make predictions before symptoms are being experienced by the patient.[36]

Such predictions are mainly based on correlation as opposed to causation that, on the contrary, has always been the basis for scientific research. Traditionally, indeed, any scientific position could only be claimed by understanding causes, thus after having found and proved linkages between causes and effects. By contrast, the typical big data approach is correlation, meaning that algorithms produce outcomes based on statistical analysis. Hence, not necessarily a cause-effect relation is demonstrated, but rather data are associated according to patterns identified in the training set in such a way to predict the occurrence of health-related events.[37] Correlation is particularly significant in genomics where identification of the causative relation between genetic mutations and disease has long been limited. Correlation has several benefits against causation: it is quick, automatic, and inexpensive, as opposed to costly and time-consuming causation approaches. Furthermore, links between variants and factors are more likely to emerge, also considering that it is not uncommon that associations are clearly experienced but difficult to prove in theory. Yet, there are some side-effects. The main problem is demonstrating to what extent a genetic mutation impacts on the risk to develop a particular disease.[38] In addition, without a theory to understand the prediction, the outcome achieved cannot be generalised and can only be used in the specific context. Finally, in the absence of a solid basis the risk of error and inaccuracy is much higher.[39]

2. The implications of AI on data protection.

As it emerges from the previous paragraph, the key to have efficient AI systems is training algorithms with huge amounts of personal data.[40] Given that, it is not surprising that several issues arise when AI is deployed in genomic

[35] An example are "wearables", tools that collect relevant health data such as blood pressure, glucose, sleep apnea, cardiac and other monitors and interact with smartphones or directly with third parties (e.g. GPs). Several other IoT devices surround people, able to collect enormous amounts of data of any kind (for instance domotic assistants know habits, preferences and behaviours of the residents of the house).
[36] V Mayer-Schönberger, E Ingelsson, 'Big Data and Medicine - a Big Deal?' (2018) 283 Journal of Internal Medicine.
[37] TZ Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (2017) 47 Seton Hall Law Review.
[38] MKK Leung *et al.*, 'Machine Learning in Genomic Medicine: A Review of Computational Problems and Data Sets' (2016) 104(1) Proceedings of the IEEE, 176-197.
[39] *Ibid.*
[40] Under Art. 4 (1)(1) GDPR personal data are defined as "namely data related to an identified or identifiable subject".

diagnostics. Such issues relate both to the collection of personal data as input and to the processing of data resulting in output.

As for the collection, data not directly related to health (obtained from different sources) are also necessary to properly train algorithms. Herein, people providing their personal data may not know how and for what purposes such data will be used because at the time of data collection it may be difficult to exactly foresee the future uses of data.[41] Besides, data can be cross-context, meaning that they can be used for multiple different purposes, also in different fields and by different data controllers.[42] Hence, data subjects might not be adequately informed and, thus, cannot maintain an effective control over their data.

Another aspect concerns the quality of input data: if data is incorrect or altered, the outcome will be inaccurate. Input data may be altered not only when it is false or wrong, but also when it is biased. If datasets reflect existing biases against minorities or other vulnerable groups (e.g. entrenched overdiagnosis of schizophrenia in African Americans),[43] such biases will be reproduced in the outcomes.[44] Biases may also be unintentional, because they are always latent in society, reflecting cultural or organisational values.[45] Therefore, even if clear sensitive data like race, age or gender were excluded, other apparently neutral data (e.g. postal code) could be associated in such a way to result in biased outcomes.[46] In addition, biases can result from non-representation of marginalised groups: datasets used to train algorithms might not adequately represent the whole population because there are not enough data related to certain vulnerable groups like racial minorities, immigrants, or people with low socioeconomic *status*.[47] One of the reasons is that such groups have no access to the most common data sources like social media, smartphones, wearables and often even the Internet and healthcare itself, thus not being represented in clinical data.[48] Another reason is the absence of studies on certain segments of the population. Genomic diagnostics, in particular, has a history of under-representation of ethnically diverse populations since genetic studies of human disease have always been predominantly based on populations of European ancestry.[49] These exclusions

[41] IG Cohen *et al.*, 'Introduction' in G Cohen, A Hoffman and W Sage (eds.), *Big Data, Health Law, and Bioethics* (Cambridge University Press, 2017).
[42] WN Price, IG Cohen, 'Privacy in the Age of Medical Big Data' (n 31).
[43] HW Neighbors *et al.*, 'The Influence of Racial Factors on Psychiatric Diagnosis: A Review and Suggestions for Research' (1989) 25 Community Ment Health J, 301–11.
[44] D Schönberger, 'Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications' (2019) 27 International Journal of Law and Information Technology, 171–203.
[45] M Hardt, 'How Big Data is Unfair', 26 September 2014, <https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de>.
[46] A Romei, S Ruggieri, 'A Multidisciplinary Survey on Discrimination Analysis' (2014) 29 The Knowledge Engineering Review, 582.
[47] D Schönberger, 'Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications' (n 44).
[48] SE Malanga *et al.*, 'Who's Left Out of Big Data? How Big Data Collection, Analysis, and Use Neglect Populations Most in Need of Medical and Public Health Research and Interventions' in G Cohen, A Hoffman and W Sage (eds.), *Big Data, Health Law, and Bioethics* (Cambridge University Press, 2017).
[49] Recent studies indicate that the proportion of individuals included in GWAS who are not of European descent is less than 20%. In particular, individuals of African and Latin American ancestry, Hispanic people and native or indigenous peoples represent together less than 4%. See AB Popejoy, SM Fullerton, 'Genomics

reflect pre-existing health disparities and amplify them, thus bearing the risk of causing serious diagnosis mistakes.[50]

When it comes to the generation of outcomes one of the main issues is the "opacity" of AI systems (more precisely, of algorithms).[51] Given that it combines large-scale high-quality datasets with sophisticated predictive algorithms and use implicit, complex connections between multiple variables, ML may identify patterns and make associations within such a high number of variables and with such an elaborate computation that it is extremely hard for the human mind to grasp the underlying logic. As a consequence, neither doctors nor AI experts would be able to fully understand how the data is processed and how the output is obtained.

Because of this opacity AI systems have been defined as "black boxes".[52] The use of black boxes in genomic diagnostics may lead to enormous benefits. Indeed, genomic datasets are so vast and the relationships among them and with other variables are so complex that they are not fully understood in medical literature yet. Hence, associations made by algorithms, although opaque, may suggest new treatment options and make more specific predictions.[53] On the other hand, black boxes could negatively affect the patients' right to be informed on how their data is processed and how decisions concerning themselves are taken.[54]

As such, the opacity of algorithms limits individuals' autonomy of identity, namely the ability of individuals to build their own identity without external influences. The reason is that based on the machine outcome patients are classified and grouped into clusters (e.g. based on the presence of a tumour, the likelihood of success of a treatment, the prediction of risk and survival) and subsequent decisions will be taken in relation to the cluster as a group, regardless of the identity of the group members as individuals. Yet, groups do not perfectly reflect individuals because they represent only some aspects of their identity.[55] It follows that decisions might be taken for all the members of the group, based on their common features, without taking into account the differences existing among them and the further characteristics that each member has and that contribute to delineate their identity. In the field of genomic diagnostics the consequences may be particularly dangerous because, for instance, all patients with the same type of cancer could be medically treated in the same way without considering that the treatment could have different consequences depending on their lifestyle or the existence of further genetic variants.

---

is failing on diversity' (2016) 538 Nature, 161–164; G Sirugo, SM Williams, SA Tishkoff, 'The Missing Diversity in Human Genetic Studies' (2019) 177(1) Cell., 26-31.

[50] An example is a study conducted to detect skin cancer by using ML, where fewer than 5 per cent of the images the model was trained on were from individuals with dark skin. See J Zou, L Schiebinger, 'AI can be Sexist and Racist — it's time to make it Fair' (2018) 559 Nature, 324–6.

[51] WN Price II, 'Black-Box Medicine' (2014) 28 Harvard Journal of Law & Technology, 419.

[52] One of the first authors to propose the idea of black boxes was Frank Pasquale in 2015. See F Pasquale, *The black box society: The secret algorithms that control money and information* (Harvard University Press 2015). See also WN Price II, 'Black-Box Medicine' (n 51).

[53] WN Price II, 'Black-Box Medicine' (n 51).

[54] BD Mittelstadt, P Allo *et al.*, 'The Ethics of Algorithms: Mapping the Debate' (2016) 3 Big Data & Society.

[55] B Mittelstadt, 'From Individual to Group Privacy in Biomedical Big' in G Cohen, A Hoffman and W Sage (eds.), *Big Data, Health Law, and Bioethics* (Cambridge University Press, 2017).

Finally, ML outcomes could lead to discriminatory treatments of data subjects. In particular, within the healthcare environment medical predictions and diagnosis could lead to discriminatory treatment plans if relied on biased data. In addition, should third parties (e.g. employers, health insurances, financial services providers) access them, they could take advantage of such outcomes to discriminate vulnerable data subjects.[56] Besides, the opacity of algorithms makes it hard to prove that discrimination has actually occurred.[57]

3. Application of European Data Protection Law to AI.

This paragraph will examine the European General Data Protection Regulation (GDPR) [58] to assess how it deals with the data protection issues raised by the use of AI in genomic diagnostics and evaluate whether it is sufficient to ensure adequate protection of patients' personal data. In particular, three aspects will be taken into consideration: the GDPR underlying concepts, its fundamental principles and the data subjects' rights.

3.1 The GDPR underlying concepts.

First of all, the material scope of the GDPR covers "personal data", defined as "*any information relating to an identified or identifiable natural person*".[59] This notion is based on the concept of identifiability, so that anonymous data fall outside the scope of the Regulation. This choice was justified when processing focused on individuals, but it seems no longer meaningful in a world dominated by big data and algorithms that process large amounts of data not related to any specific person. [60] In particular, the deployment of ML in genomic diagnostics requires the collection of huge datasets to train the algorithm and make it as much accurate as possible, but it is not necessary that such data is linked to identified data subjects since – in the training phase – it is only their aggregation that matters. Therefore, data may be anonymised to easily avoid the GDPR obligations with the consequence that no protection will be ensured to people whose data is collected and used to identify patterns and create clusters. In this perspective identifiability represents a limit to data protection since under the GDPR an effective protection is ensured only when data can be linked to a precise person. However, as noted before, AI systems are able to make inferences from input data based on statistical associations, so the mere fact that a person is included in a group allows to collect a lot of data about him/her because, even though his/her name would not be known, the group characteristics would describe that person with sufficient accuracy. As a

[56] S Hoffman, 'Big Data's New Discrimination Threats Amending the Americans with Disabilities Act to Cover Discrimination Based on Data-Driven Predictions of Future Disease' (n 34).
[57] WN Price, IG Cohen, 'Privacy in the Age of Medical Big Data' (n 31).
[58] Regulation 2016/679 of the European Parliament and of the Council on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Advancement of Such Data, and repealing Directive 95/46/EC, 2016 O.J. L 119/1.
[59] Art. 4(1) GDPR.
[60] EG González, P de Hert, 'Understanding the Legal Provisions That Allow Processing and Profiling of Personal Data—an Analysis of GDPR Provisions and Principles' (2019) 19 ERA Forum, 597–621.

consequence, it should not be necessary to identify a person with absolute certainty to provide for data protection; rather, it should be enough to know that a person is included in a certain group.[61]

A second aspect, strictly related to the previous one, is the individual dimension: the GDPR regulates data protection as an individual right and focuses on individuals' protection. However, data processing in genomic diagnostics mainly concerns group data: algorithms manage to discover unknown patterns among individuals with certain characteristics thus creating groups based, for instance, on the likelihood to develop a disease or the chance to survive. In this case, risks do not concern directly members of the group as such, but rather the group itself. [62] Indeed, by processing group data, controllers are able to create "inferences", namely non-verifiable information, opinions, or assessments[63] that will be used to take decisions towards any patients included in that group. Patients have no rights on such inferences, given that the GDPR focuses primarily on mechanisms related to the input side (collection and processing), while mechanisms addressing outputs (i.e. inferred data, profiles, decisions) like, for example, the right to explanation, are far weaker (as it will be better explained below). Yet, outputs are the ones that pose major risks, given that problems do not lie much with data collection, but rather with what can be inferred from such data and the decisions that can be taken based on this knowledge (e.g. the decision to apply a therapeutic plan or make surgery based on the likelihood to develop a disease). [64]

Another cornerstone of the GDPR is the distinction between sensitive and non-sensitive data. [65] The *ratio* of this distinction is ensuring stronger protection for sensitive data by introducing further limitations for processing and providing data subjects with major safeguards. However, such a distinction is challenged by AI and big data analytics because algorithms can infer sensitive data even from non-sensitive data, thus shifting from one category to another. In particular, a lot of non-medical data collected from different sources can be used to create medical profiles and make inferences about health. This fluidity makes the abovementioned distinction fundamentally flawed and almost useless because any data could potentially become sensitive if used to infer information about health.[66] For this reason, it may become difficult to set the point when non-sensitive data should start being treated (and protected) as

[61] On the distinction between identifiable and non-identifiable data in the Big Data era see CJ Bennett, RM Bayley, 'Privacy Protection in the Era of "Big Data": Regulatory Challenges and Social Assessments', in B van der Sloot, D Broeders and E Schrijvers (eds.), *Exploring the Boundaries of Big Data* (WRR/Amsterdam University Press, 2016), 205.

[62] EG González, P de Hert, 'Understanding the Legal Provisions That Allow Processing and Profiling of Personal Data—an Analysis of GDPR Provisions and Principles' (n 60).

[63] Article 29 Data Prot. Working Party, Opinion 4/2007 on the Concept of Personal Data, 01248/07/EN WP136, at 8 (June 20, 2007).

[64] S Wachter, B Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) 1 Columbia Business Law Review.

[65] "Sensitive data" are so called because they relate to the inner – and more sensitive – sphere of individuals (e.g. health, religion, sexual orientation) and thus unlawful processing would produce particularly serious consequences on data subjects. The terminology "sensitive data" was used under the Data Protection Directive. Art. 9 GDPR defines such data as "*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*".

[66] TZ Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (n 37).

sensitive data.

Finally, the GDPR is based on the idea that personal data can be processed provided that data subjects maintain control over data related to them. In this perspective the GDPR grants individuals several rights (e.g. right of access, right to rectify, erasure, object) and includes among the lawful basis for processing data subjects' consent [67] that must be freely given, specific, informed, and unambiguous.[68] This is referred to as the "notice-and-consent" model according to which patients must be adequately informed about the processing of their personal data in order to take meaningful decisions about it. Yet, this model has been weakened by AI-based technologies that make the explanation and understanding of data processing much more complicated especially to non-expert data subjects. Herein, even though the GDPR imposes on controllers to provide data subjects with all the relevant information about the processing, patients could hardly understand it and make fully conscious choices. Hence, behind the façade of self-determination data subjects risk not to be actually protected. This should lead regulators to reconsider the role of data subjects in the processing of personal data, at least when the processing is extremely complex as the AI-based processing for diagnostic purposes.[69] Rather than placing on patients the burden to understand how the processing is carried out and what are the best decisions to protect their rights, a stronger obligation should be imposed on data controllers to take all the measures necessary to protect individuals even out of their control. Indeed, it would require a huge effort to patients to fully understand what is behind data processing and not all of them might have time, resources, and capability to understand it. On the contrary, developers and medical structures deploying AI systems are in a better position to assess and guarantee that the processing is safe.[70] This approach is not new but is already adopted in other regulatory fields involving information too complex to be understood without expert knowledge like, for instance, food industry, building engineering and car safety.[71] It would mean a shift from individual consent to *ex ante* assessments (carried out by controllers and external independent expert entities) aimed to certify the security of the processing and the absence of potential risks and harms for individuals.

3.2 The GDPR fundamental principles.

Art. 5 GDPR sets the fundamental principles to be respected when processing personal data. In particular, Art. 5(1)(a) sets the principles of lawfulness, fairness and transparency, meaning that the processing must have

[67] Art. 6(1)(a) GDPR.
[68] Art. 4(11) GDPR.
[69] A Mantelero, 'Regulating Big Data. The Guidelines of the Council of Europe in the Context of the European Data Protection Framework' (2017) 33 Computer Law & Security Review, 584–602.
[70] Developers – being the ones who design the system – can more easily understand how it works and guarantee its safety and have the competences and the resources to certify, even with external bodies, that the system respects safety standards. At the same time medical structures have a direct relationship with the developer to understand the system functioning and have the technical resources to test the safety of the system before putting it into use.
[71] V Mayer-Schönberger, E Ingelsson, 'Big Data and Medicine - a Big Deal?' (n 36).

a lawful basis[72] and must be carried out in a fair and transparent manner. Transparency is extremely important because it allows data subjects to know how their data is processed and maintain control over it, thus it is strictly linked to the right to be informed and the right of access (that will be analysed in-depth in the next paragraph). Yet, transparency is threatened by the opacity of algorithms that makes it difficult to explain on what basis data is aggregated, how clusters are created, and how outputs are generated.

Art. 5(1)(b) sets forth the principle of "purpose limitation"[73] whose *ratio* is to prevent data controllers from using data for other purposes than those for which it was initially collected. Yet, such a principle may hinder the development of AI for genomic diagnostics since ML algorithms – to make precise inferences and generate accurate outputs – must be trained on large volumes of data that could be collected from several sources even not directly related to health (e.g. social networks, IoTs, etc.). When this is the case, data is initially collected for purposes different from training algorithms for genomic diagnostics and, at the moment it is collected, neither the controller nor the data subject might specifically know that it could also be used for diagnostic purposes. Thus, a strict application of the purpose limitation principle, albeit meant to protect data subjects, could end up posing a limit to training algorithms.[74]

Other principles are "data minimisation"[75] and "storage limitation"[76] that require respectively to collect only data adequate, relevant and necessary to achieve the purposes for which it is processed and keep data for no longer than necessary to achieve these purposes. These principles aim to prevent data controllers from collecting and retaining data not necessary for the processing, thus exploiting data for further unspecified purposes. On the other hand, such a limitation may curtail the deployment of AI systems for genomic diagnostics because, as already mentioned, algorithms must be trained on large amounts of data to identify relevant patterns and generate reliable outputs. Yet, data that could potentially be used to train algorithms but are not or no longer necessary to achieve the initial purpose (i.e. the purpose for which it is processed in the first place) cannot be collected. This could reduce the amount

[72] Among those listed in Articles 6 and 9 GDPR.
[73] According to such a principle personal data shall be "*collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes*". See Art. 5(1)(b) GDPR.
[74] A Mantelero, 'Regulating Big Data. The Guidelines of the Council of Europe in the Context of the European Data Protection Framework' (n 69). Art. 5 partly derogates this principle when further processing is not incompatible with the original purposes. Nevertheless, doubts remain on when processing may be considered "compatible". On this regard, Art. 5(1)(b) specifies that processing for "statistical purposes" is not incompatible, but according to Recital 162 statistical purposes imply that the result or processing "are not used in support of measures or decisions regarding any particular natural person". This cannot be the case for data used to train algorithms for diagnostic purposes, given that the results will likely ground medical decisions concerning patients. Indeed, training data are analysed by the algorithm to build the model that will be applied to new cases to suggest medical decisions (e.g. therapeutic plans). This means that, based on the inferences and the correlations found in the training datasets, the algorithm identifies the best medical decisions related to each patient's personal circumstances. Secondly, Art. 6(4) lists the criteria to be taken into account to assess the compatibility of further processing and mentions, inter alia, the link with the original purpose, the context in which data have been collected, and the nature of data. Under these terms, collection of personal data based for instance on the preferences expressed by the user on a social network could hardly be linked with the purpose of training algorithms to diagnose diseases.
[75] Art. 5(1)(c) GDPR.
[76] Art. 5(1)(e) GDPR.

of data available to train algorithms thus undermining their effectiveness.[77]

3.3 The GDPR data subjects' rights.

Chapter III of the GDPR grants data subjects a series of rights related to the processing of their personal data.[78]

Arts. 13 and 14 impose on the controller to provide data subjects with relevant information about the processing.[79] Such information has to be provided either at the time of data collection (Art. 13) or at the first communication with the data subject (Art. 14), thus only covering *ex ante* explanation about system functionality. Besides, the expression "*envisaged consequences of such processing*" refers to future consequences not occurred yet,[80] while no information on the reasons and the effects of specific decisions can be released, since these require that a decision has already been taken.[81]

Similarly, Art. 15 confers on data subjects the right to access information about the processing. Although in this case the right can be exercised in any moment, without the time limit set in Arts. 13-14, reference remains to the "*envisaged consequences*", meaning that the relevant information is supposed to be provided before a decision has occurred. As in the previous case the right only ensures *ex ante* explanation about the functioning of the automated decision system, but no *ex post* explanation about the grounds and the implications of a specific decision is required.[82]

A fundamental right for AI-based diagnostic data processing is set forth in Art. 22(1), whose scope is widely debated in literature. It establishes as a general rule the right of data subjects not to be subject to decisions based solely on automated processing including profiling.[83] Actually, as the Article 29 Working Party clarified, such a provision should be interpreted as a prohibition to data controllers rather than a right, with the consequence that controllers must refrain from automated decision-making without data subjects having to actively require it.[84] Art. 22(2) admits some exceptions to this general rule, allowing automated decision-making under certain conditions,[85] but even when these conditions are fulfilled, special categories of personal data can be used for automated decision-making only with data subjects' consent (or for

[77] TZ Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (n 37).
[78] These are the right to be informed (Arts. 13-14), right of access (Art. 15), right to rectification (Art. 16), erasure (Art. 17), restriction of processing (Art. 18), data portability (Art. 20), right to object (Art. 21), and right not to be subject to automated decision-making (Art. 22).
[79] In particular, information shall be provided about "*the existence of automated decision-making, including profiling, […] and, […], meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject*".
[80] EG González, P de Hert, 'Understanding the Legal Provisions That Allow Processing and Profiling of Personal Data—an Analysis of GDPR Provisions and Principles' (n 60).
[81] S Wachter *et al.*, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) International Data Privacy Law.
[82] *Ibid.*
[83] Art. 22(1) says "*The data subjects shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her*".
[84] Article 29 Working Party (A29WP), 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP 251, 3 October 2017).
[85] In particular, when it is necessary for performing a contract between the data subject and the data controller or when it is based on the data subject's consent.

reasons of substantial public interest).[86] Furthermore, when the abovementioned exceptions apply, the controller has to inform the data subject and adopt adequate safeguards, in particular ensure the right "*to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision*",[87] the so-called "human in the loop".[88] Instead, no explanation of how the processing is carried out or how the results are achieved is imposed, so that some scholars, in particular Wachter et al., have argued that "the GDPR does not […], implement a right to explanation, but rather [a] right to be informed".[89] Such a conclusion has been rebutted by Selbst and Powles who suggest that Art. 22, although does not directly set forth a right to explanation, supports the existence of that right which is derived from Articles 13-15, in particular in the part where they acknowledge data subjects' right to "*meaningful information about the logic involved*" in automated decisions.[90]

In any case, the right of explanation should not be regarded as a panacea for all the problems posed by algorithms.[91] Indeed, the way in which algorithms work is not always interpretable: the "black box" model makes it difficult to understand the correlations between input data made by algorithms to find patterns and create clusters. Therefore, the "logic" behind automated systems might not be fully explainable to humans and simplifying such systems in order to make them human interpretable could result in reducing their predictive performance.[92] Secondly, even when experts would manage to interpret the systems, these could be hardly explicable to non-expert data subjects (e.g. patients). The result would be a "transparency fallacy",[93] that is, an illusion to provide data subjects with enough information to allow them to maintain full control over their personal data, while, in fact, such information is not meaningful if cannot be really understood.[94]

---

[86] Art. 22(4) GDPR.
[87] Art. 22(3) GDPR.
[88] L Edwards, M Veale, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review, 18.
[89] S Wachter *et al.*, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (n 81).
[90] AD Selbst, J Powles, 'Meaningful Information and the Right to Explanation' (2017) 7 Int'l Data Privacy L., 233.
[91] B Casey *et al.*, 'Rethinking Explainable Machines: The GDPR's 'Right to Explanation' Debate and the Rise of Algorithmic Audits in Enterprise' (2019) 34 Berkeley Tech LJ, 143.
[92] L Edwards, M Veale, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (n 88).
[93] This notion has been first introduced in relation to consent, as giving data subjects the illusion to take control of the use of their personal data. See DA Heald, 'Varieties of Transparency' in C Hood and D Heald (eds.) *Transparency: The Key to Better Governance?: Proceedings of the British Academy* (Oxford University Press, 2006), 135 .
[94] Apart from the right to explanation, the scope of Art. 22 is narrowed under two aspects. Firstly, para. 1 only applies to decisions based "*solely*" on automated processing, thus implying that as long as there is a human intervention the right should be excluded. Current automated systems are mostly used to support rather than replace human activity, so that just a few of them are entirely autonomous and this would strongly reduce Art. 22 practical relevance. The A29WP clarified that, to be regarded as human involvement, "*any oversight of the decision [should be] meaningful, rather than just a token gesture. It should be carried out by someone who has the authority and competence to change the decision*". Doubts still remain concerning the degree of human intervention required to exclude the application of Art. 22(1). In particular, in the field of genomic diagnostics physicians still maintain a certain discretion in deciding whether or not to follow the machine's advice (there is usually a team of medical experts who interpret and discuss the outcome). It may be questioned whether such intervention is sufficient to exclude patients' right not to be subject to

4. Alternative regulatory instruments to deal with data protection issues in AI-based genomic diagnostics.

As emerged in the previous section data protection law is not sufficient to deal with the data protection issues posed by AI. There are in particular two reasons for this. Firstly, there are some misalignments between the GDPR and the development of AI. In particular, the GDPR is still focused on individual protection and grants no rights to groups, while algorithms create clusters where patients are grouped based on aggregated personal data. In addition, the GDPR is based on data subjects' self-determinism (the "notice-and-consent" model) that seems incompatible with the complexity and the opacity of ML. Secondly, data protection law only covers part of the issues posed by AI. Yet, AI-based data processing for diagnostic purposes gives rise to further issues (like discrimination and protection of aggregated data) that are not specifically addressed. Therefore, data protection law needs to be included in a broader framework and combined with other regulatory instruments to ensure adequate protection of patients/data subjects' rights and, at the same time, allow them to benefit of the advantages of AI. [95]

4.1 Medical law.

A first instrument may be medical law that, although not directly addressing AI, includes some provisions which may affect the deployment of AI-based technologies in genetic diagnostics. [96]

The need to respect the privacy of patients and the confidentiality of the

automated decision-making. Secondly, Art. 22(1) only applies when the decision produces legal effects or "*similarly significantly*" affects the data subject. This last expression has been clarified by the A29WP (see A29WP note 84). According to this interpretation automated decisions in the field of genomic diagnostics may be considered as significantly affecting patients and, as such, falling under the scope of Art. 22 given that the exclusion or confirmation of disease as well as the decision to have the patient undergone surgery or the choice of a therapeutic plan affects patients' whole life and all their consequent choices and behaviour.

[95] A piece of this framework is non-discrimination law, given that discriminatory effects are among the main risks posed by AI, especially in the field of genomics, given its history of underrepresentation of some ethnical groups. Non-discrimination has always been one of the fundamental values of the EU, thus it is included in the EU treaties (Art. 2 TEU, Art. 10 TFUE, Art. 21 EU Charter of Fundamental Rights), Member States' constitutions and several EU directives (Directive 2000/43/EC, Directive 2000/78/EC, Directive 2004/113/EC, Directive 2006/54/EC). Yet, non-discrimination law is not always able to deal with the discriminatory effects of AI given that it mainly addresses direct discrimination that, being based on belonging to protected classes (e.g. because of race or gender), is easier to detect and prohibit. However, the most common form of discrimination in AI is indirect discrimination, based on apparently neutral elements (e.g. postal code, pet ownership) that hide protected classes, thus making it difficult to ascertain the existence of discrimination. In the case of indirect discrimination the algorithm is a neutral criterion since it is not discriminatory *per se* but its application may lead to discriminatory results. As a consequence, it falls on data subjects the burden to prove that the decision is discriminatory and lacks an objective justification. In addition, current non-discrimination law mostly addresses traditional protected classes (e.g. based on ethnicity, gender), while discrimination grounded on different elements that do not seem *prima facie* discriminatory (e.g. genetic variants) could fall outside its scope. See P Hacker, 'Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law' (2018) 55 Common Market Law Review, 1143-1186; see also FJ Zuiderveen Borgesius, 'Strengthening legal protection against discrimination by algorithms and artificial intelligence' (2020) The International Journal of Human Rights.

[96] The term "medical law" will be used in the following paragraph as generally referring to the branch of law regulating, in particular, the performing of medical operations, the relations between medical operators and patients, medical devices, the prerogatives of medical professionals and the rights of patients.

information they disclose to their physicians has been a cornerstone of medical practice since the Hippocratic oath.[97] More recent is the notion of "informed consent" grounded on respect for human dignity, according to which subjects involved in medical research must be adequately informed of the aims, methods, benefits, potential risks and any other relevant aspects of the treatment, as well as the right to refuse the treatment or to withdraw consent. [98]

However, informed consent under such terms only provides patients with the right to refuse a diagnostic procedure, a specific treatment or to participate in research, but not the right to object to the use of a particular technology, or to obtain a detailed explanation about the functioning of a certain medical tool or an algorithm. Yet, to take meaningful decisions patients would need at least general information about the technology used, how the algorithm has been constructed and what type of data it uses. Besides, clinicians should explain patients why they have chosen to use that AI system, what its influence is on the diagnosis or treatment, and why they eventually decide (not) to follow its advice; in such a way, patients would be made aware of the contribution the AI had in the final decision. At the moment there are no clear indications in this sense, so that doubts arise on the extent to which clinicians have to inform patients and how detailed information they have to provide them. Therefore, a legislative intervention would be advisable to clarify what kind of information physicians should provide patients when deploying AI systems for diagnostic procedures.

4.2 The MDR Regulation.

Another relevant legislation related to AI in healthcare is the EU Medical Device Regulation (MDR) [99] amending the Medical Device Directive [100] in regulating the use of medical devices within the EU.

The MDR classifies medical devices into four categories (classes I, IIa, IIb, and III) based on their intended purpose and inherent risks. Manufacturers of medical devices shall assess the conformity of their devices prior to placing them on the market and the applicable conformity assessment procedure depends on the classification of the device: for class I devices having a low level of vulnerability the conformity assessment procedure will be carried out under the sole responsibility of the manufacturers; class IIa, IIb, and III devices having a higher risk have to be assessed by an independent accredited certification organisation appointed by the competent authorities of EU Member States.

An important innovation of the MDR is the broader definition of "medical

[97] P Balthazar *et al.*, "Protecting Your Patients' Interests in the Era of Big Data, Artificial Intelligence, and Predictive Analytics" (2018) Journal of the American College of Radiology.
[98] This notion, firstly elaborated in 1957 in a medical malpractice suit, has then been incorporated in several conventions and declarations like the Declaration of Helsinki, the best-known policy statement of ethical principles published by the World Medical Association (WMA) to guide the protection of human participants in medical research, adopted in 1964 and amended seven times, most recently in 2013.
[99] 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. The MDR entered into force on May 2017 and entered into application on 26 May 2021.
[100] Council Directive 93/42/EEC of 14 June 1993 concerning medical devices.

devices" that now also includes software used for human beings for the "*Medical purpose of prediction or prognosis of disease as a medical device*".[101] Furthermore, the Regulation specifies that "*software intended to provide information which is used to take decisions with diagnosis or therapeutic purposes is classified as class IIa*".[102] It follows that AI software may be classified under the MDR. For example, Watson for Oncology might be classified at least as a class IIa medical device since it "*provide*(s) *information which is used to take decisions with diagnosis or therapeutic purposes*".[103] This may represent an important step forward in regulating AI-based medical devices because enables medical operators and patients to know that the conformity of medical devices to certain safety standards has been assessed.[104]

4.3 Soft law.

Another important regulatory instrument is soft law, a broad category including a variety of instruments with different characteristics that pose rules of conduct without legally binding force but still producing some legal effects.[105] Two important forms of soft law are co-regulation and self-regulation, based on the intervention of private actors respectively to implement legislative provisions or to regulate private conducts in the absence of legislative provisions.[106]

The use of soft law is still debated among scholars. On the one hand, its main advantages are specificity, that allows to fill regulatory gaps by providing operators with a guidance on how to interpret and apply the legal rules in specific sectors, and flexibility, given that it can easily and rapidly be updated to face new scenarios and new issues. Furthermore, soft law is usually the result of collaboration between public and private actors with the participation of experts; as a result, the drafted provisions are more pragmatic and in line with the needs of a particular sector/processing. On the other hand, some scholars have put forward the risk of "over-saturation"[107] and fragmentation as a result of excessive reliance on soft law and the creation of new instruments for each processing or technology. This would reduce the utility and desirability of these mechanisms by complicating too much the regulatory framework and causing legal uncertainty in an already complex field like AI. In addition, soft law instruments can be effective only when widely adopted, otherwise their regulatory power would be limited lacking strong enforcement mechanisms. A further criticism is that soft law shifts the regulatory power away from the

[101] Art. 2 MDR.
[102] Rule 11 in Chapter III of Annex VIII of the MDR.
[103] S Gerke *et al.*, 'Ethical and legal challenges of artificial intelligence-driven healthcare' (2020) Artificial Intelligence in Healthcare, 295–336.
[104] Annex I of the Medical Device Regulation lists the safety and performance requirements (also including the information to be provided with the device) that medical device must achieve to be released on the market.
[105] L Senden, 'Soft Law, Self-Regulation and Co-Regulation in European Law: Where Do They Meet?' (2005) 9(1) Electronic Journal of Comparative Law.
[106] *Ibid.*
[107] S Wrigley, 'Taming Artificial Intelligence: "Bots", the GDPR and Regulatory Approaches' in M Corrales, M Fenwick and N Forgó (eds.) *Robotics, AI and the Future of Law. Perspectives in Law; Business and Innovation* (Springer, 2018).

legislature to the private sector or to secondary bodies which lack democratic legitimacy with the risk to make private or commercial interests prevail over those of data subjects.[108]

In the light of this, soft law may still be considered a valuable regulatory instrument provided that it is subject to the control of independent authorities and is used as a complementary tool to give operators a guidance to effectively comply with the (often vague and broad) legal statutory provisions.

An important form of soft law are codes of conduct, a set of rules, ethics or values that guide people/companies/organisations in daily practices, activities, and interactions. In relation to processing of personal data, codes of conduct find a legal ground in Articles 40 and 41 GDPR which analytically describe the procedure of approval and accreditation of such codes;[109] further clarifications have been provided by the EDBP.[110] Adherence to a code of conduct is not mandatory, but several benefits follow its adoption.[111] Also, the use of an approved code of conduct can be used as a basis for a transfer of personal data outside the EU.[112]

The advantages of codes of conduct, especially in a technical sector like genomic diagnostics, are even more relevant in relation to AI since data protection law does not include specific provisions concerning the use of AI-based technologies. Firstly, codes of conduct can be written with the help of practitioners who can better pinpoint the specific needs of healthcare and diagnostics, thus ensuring that data protection issues are addressed accordingly. Besides, their involvement can increase the code's acceptance.[113] Secondly, a code of conduct can formulate best practices and set standards able to harmonise data processing for the entire category of health operators in Member States, thus increasing legal certainty and enhancing patients' trust in AI.[114] Finally, codes of conduct can have a broader scope than legislation and include ethical standards as further safeguards of personal data.[115] As such, geneticists strongly ask for codes of conduct that could help them to overcome

[108] P de Hert, 'The future of privacy. Addressing singularities to identify bright-lines that speak to us' (2016) 3(4) European Data Protection Law Review, 461.

[109] Under Art. 40 GDPR codes of conduct may be drawn up by associations and other bodies representing categories of controllers or processors and then formally approved by supervisory authorities (if the code relates to only one Member State) or the Commission (if the code relates to multiple Member States). After approval codes are collated in a register and made publicly available. The code main purpose is providing data controller and processor with a guidance for processing personal data, especially in relation to the collection of data, the security measures to adopt, the information to provide to data subjects and the ways to make data subjects exercise their rights.

[110] EDPB Guidelines 1/2019 on Codes of Conduct and Monitoring Bodies under Regulation 2016/679.

[111] For example, under Recital 77 and Art. 24(3) adherence to an approved code of conduct can be used by data controller and data processor to demonstrate compliance with the GDPR and, under Art. 83, it can be evaluated by supervisory authorities to mitigate penalties for non-compliance.

[112] Article 46 (2) (e) GDPR.

[113] F Molnár-Gábor, JO Korbel, 'Genomic data sharing in Europe is stumbling—Could a code of conduct prevent its fall?' (2020) 12(3) EMBO Mol Med.

[114] EDPB Guidelines 1/2019 (n 110).

[115] Several codes of conduct have already emerged or are being drafted in the healthcare sector. An example is the Code of Conduct drawn up by the Biobanking and BioMolecular Research Infrastructure aiming to regulate data processing and fostering transparency and trust in the use of personal data for health research within the EU (http://code-of-conduct-for-health-research.eu). In addition, some guidelines are being developed by Alliance Against Cancer for the creation of a technological platform that allows the collection, sharing and analysis of health big data from each Italian research institute (https://www.alleanzacontroilcancro.it/en/commissione-acc-gdpr/).

the gaps and uncertainties left by data protection law[116]. In particular, codes of conduct should address issues like pseudonymisation and standards to ensure adequate data protection, consent and withdrawal, data portability, access, and explainability.[117]

Another accountability tool meant to give data subjects confidence about data controllers' compliance with data processing legal requirements is certification,[118] grounded on Articles 42 and 43 GDPR[119] and addressed by the EDPB in its guidelines.[120]

Although not widespread yet, this instrument may be particularly relevant for AI-based health data processing because patients do not have the knowledge and expertise necessary to fully understand how their data is processed and whether the processing is fair. Hence, the self-determination regime based on the notice-and-consent model cannot ensure an effective protection. On the contrary, certifications would give patients the certainty that the processing is trustable because its compliance to legal provisions and conformity to safety standards has been evaluated by *ad hoc* expert certification bodies.

A further co-regulatory instrument is the data processing impact assessment, a self-assessment procedure through which data controllers and processors, either by themselves or with the help of external bodies, evaluate the impact of data processing.

The rationale behind impact assessment is twofold. [121] On the one hand, the idea that data subjects often cannot fully understand the functioning and the implications of AI-based data processing and cannot take meaningful decisions about the use of their personal data. On the other hand, the acknowledged collective dimension of data protection. The impact assessment would place on

---

[116] In relation to genomic data, major problems in terms of data subjects' identifiability arise because, given the necessity to maintain a strict connection with individual personal genetic profiles, data cannot be anonymised. See M Shabani, L Marelli, 'Re-identifiability of genomic data and the GDPR: Assessing the re-identifiability of genomic data in light of the EU General Data Protection Regulation' (2019) 20(6) EMBO reports. At the same time, large genetic data sets are required to train the algorithms and, considering that genetic diseases are often rare, data needs to be collected and gathered from different data centres around the world with consequent problems of international transfers. See F Molnár-Gábor, JO Korbel, 'Genomic data sharing in Europe is stumbling—Could a code of conduct prevent its fall?' (n 113).

[117] M Philips *et al.*, 'Genomics: data sharing needs an international code of conduct' (2020) 578 Nature, 31 – 33.

[118] Certification has been generally defined by the International Standards Organisation (ISO) as "*the provision by an independent body of written assurance (a certificate) that the product, service or system in question meets specific requirements*". See https://www.iso.org/certification.html.

[119] Arts. 42-43 GDPR recognise two different certification models. The first model is managed by accredited private certification bodies with an appropriate level of expertise in relation to data protection that can set up a certification scheme and submit it for approval to the competent national supervisory authority or to the EDPB. After approval, the certification body is entitled to manage the conformity assessment and issue the certification when the candidate demonstrates its full conformity with the approved requirements. The second certification scheme is directly set up and managed by national supervisory authorities. In addition, the Regulation allows the establishment of other certification schemes that fall outside the control of supervisory authorities thus giving no certainty about their consistency. The risk is that the proliferation of such unmonitored certification schemes could make people lose trust in certification thus making ineffective the whole system. See E Lachaud, 'What GDPR tells about certification' (2020) 38 Computer Law & Security Review.

[120] EDPB, Guidelines 1/2018 on certification and identifying certification criteria in accordance with Articles 42 and 43 of the Regulation – Revised version 3 (2019).

[121] A Mantelero, 'AI and Big Data: A blueprint for a human right, social and ethical impact assessment' (2018) 34(4) Computer Law & Security Review, 754-772.

the controller/processor, rather than on data subjects, the burden to verify and assess the impact of processing. The results would be made public in order to make data subjects aware of the risks of data uses and able to decide whether or not to accept the processing.

The DPIA set by the GDPR still adopts a risk-based approach focused on data management, data quality, data security and procedural aspects like the regulation of the different stages of data processing and the definition of the powers and tasks of the subjects involved in the process[122]. Yet, there is no reference to the impact on ethical and social values nor on fundamental human rights. This lack is particularly serious for data processing conducted in genomic diagnostics with AI. Indeed, data processing in this field poses high risks for human rights (e.g. the risk of discrimination) and such risks assume a collective dimension given that decision-making is based on clusters created by algorithms. The potential negative impact of data processing is, therefore, no longer restricted to data protection but includes other potential prejudices that can be adequately taken into account only in a broader ethical and social perspective having a collective dimension and based on a human rights assessment.[123]

4.4 Digital ethics.

Finally, the analysis of complementary regulatory instruments could not overlook the social and ethical concerns raised by the use of AI in genomic diagnostics. On this regard, it is first necessary to clarify the role of ethics and how it interacts with data protection law. Albeit data protection legislation maintains a crucial role in regulating data processing and data subjects' rights, when AI comes into play it is no longer sufficient since AI generates radical and irreversible transformative effects that go far beyond data-related issues

[122] Under Art. 35 GDPR the Data Protection Impact Assessment (DPIA) must be conducted when processing "is likely to result in a high risk" to data subjects. The article lists the cases when high risk is presumed and, thus, the DPIA is mandatory. In addition, the A29WP has adopted the Guidelines on performing and evaluating DPIAs which identify nine criteria to evaluate whether the processing results in high risk and the DPIA must be conducted. See Article 29 Working Party WP 248 Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, adopted on 4 April 2017. Most of these criteria are met in AI-based data processing for genetic diagnostics thus triggering the obligation to conduct the DPIA (in particular, the following criteria are met: automated-decision making with legal or similar significant effect, the processing on large scale of sensitive data also concerning vulnerable data subjects, matching or combining datasets, use of new technological solutions).

[123] Different models have thus been proposed in this direction. The first one is the Privacy, Ethical and Social Impact Assessment (PESIA) adopted by the Council of Europe in its Guidelines on Big Data. This model of assessment has a broader scope than the DPIA in the GDPR, since it also encompasses the societal consequences of data uses and the analysis of their potential conflicts with ethical values. See Consultative Committee of Convention 108, "Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data" T-PD (2017)01, 23 January 2017. A further step forward is the proposed Human Rights, Ethical and Social Impact Assessment (HRESIA), a principle-based model that represents an evolution of the already existing HRIA (Human Rights Impact Assessment) since it combines the assessment of ethical and societal values with the evaluation of the human rights impact. See A Mantelero, 'Towards a Big Data regulation based on social and ethical values. The guidelines of the Council of Europe' (2017) Revista de Bioética y Derecho. See also P de Hert, 'A Human Rights Perspective on Privacy and Data Protection Impact Assessments' in D Wright and P de Hert (eds.) *Privacy Impact Assessment; Law, Governance and Technology* (Springer, 2012) 6.

giving rise to new and broader risks.[124] To mitigate these risks, new solutions need to be implemented with a systemic approach that cannot only rely on hard governance measures. In fact, statutory obligations only indicate what is legal and illegal but, as Floridi notes, they say "nothing about what the good and best moves could be, among those that are legal, to win the game, that is, to have a better society".[125] When it comes to genetic diagnostics, the GDPR regulates data processing in order to ensure data quality and data safety from a legal point of view, but it says nothing on how data should be processed to reach the best result for the patient (e.g. avoiding bias, taking fair therapeutic decisions, preventing false positives or false negatives). It follows that legal compliance is necessary but not sufficient to ensure an adequate balance of the interests at stake and an effective protection of patients' rights. Therefore, it is essential to enhance data protection law with ethical guidelines that support geneticists in processing patients' personal data in accordance with data subjects' expectations, needs, and rights.[126]

To this end reference should be made to "digital ethics" or "data ethics", the branch of ethics that studies moral problems related to data, algorithms and corresponding practices, in order to formulate and support morally good solutions.[127]

In this perspective in 2015 the EDPS established the Ethics Advisory Group to analyse the new ethical challenges posed by digital developments and current legislation, especially in relation to the GDPR.[128] In the same direction the European Commission set out its vision for an "ethical, secure and cutting-edge AI made in Europe"[129] and established the High-Level Expert Group on Artificial Intelligence (AI HLEG) that on 8 April 2019 published the AI Ethics Guidelines,[130] then confirmed by the Commission itself in its White Paper on Artificial Intelligence.[131] The Guidelines set out a framework for achieving Trustworthy AI "seeking to maximise the benefits of AI systems while at the same time preventing and minimising their risks".[132] In particular, Trustworthy AI should be lawful (complying with all applicable laws and regulations), ethical (ensuring adherence to ethical principles and values), and robust (causing no harm). Ethics, thus, is considered as a further element that must be added to legal compliance for the implementation of a trustworthy AI.

The AI HLEG elaborated the basic principles for a trustworthy AI on the ground of the bioethical principles originally developed by Beauchamp and Childress and then re-interpreted in relation to the use of AI.[133] Based on such

---

[124] J Morley, L Floridi, 'An ethically mindful approach to AI for health care' (2020) 395 Lancet, 254-255.
[125] L Floridi, 'Soft ethics, the governance of the digital and the General Data Protection Regulation' (2018) 376 Phil. Trans. R. Soc.
[126] *Ibid*.
[127] L Floridi, M Taddeo, 'What is data ethics?' (2016) 374 Phil. Trans. R. Soc.
[128] EDPS Ethics Advisory Group | Report 2018.
[129] European Commission, Brussels 25.4.18 COM(2018)237 and European Commission, Brussels 7.12.18 COM(2018)795.
[130] AI HLEG Guidelines (n 11).
[131] European Commission, Brussels, 19.2.2020 COM(2020) 65 final, WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust.
[132] AI HLEG Guidelines (n 11) paragraph A.
[133] TL Beauchamp, JF Childress, *Principles of biomedical ethics* (7th edn, Oxford University Press 2013). In particular, the principle of beneficence imposes that AI is developed for the common good and the benefit of humanity; the "non-maleficence" or "do no harm" principle requires avoiding any potential negative

principles, the AI HLEG identified the principles of respect for human autonomy, prevention of harm, fairness, and explicability to ensure that AI systems are developed, deployed and used in a trustworthy manner,[134] and identified seven concrete requirements to implement these principles.[135] Lastly, on 21 April 2021 the Commission published its proposal for a Regulation on a European approach for Artificial Intelligence (the "Artificial Intelligence Act"), the first EU legal framework on AI.[136]

In order to ensure an ethical use of AI, the identified ethical principles should be embedded in the design of AI systems: each element of the system should be "pro-ethically" designed to protect the values and principles of ethics.[137] This means that the whole process, from the development and training of algorithms for diagnostics to their deployment, should be designed in ways that decrease inequality, leave room for human autonomy, ensure explicability and data protection.

The idea of ethics embedded in machines is based on the concept of "value sensitive design" developed by Friedman and Kahn in the late 1980s and early 1990s and claiming that human principles and standards must be considered when planning technology.[138] Actors involved in developing a new technology should identify and incorporate human values into the design and development process in order to limit or eliminate potential problems once the technology has been deployed.[139]

The main issue is the selection of human values to embed in AI-based technologies for genomic diagnostics, namely how and by whom these values should be selected and to what extent they can be universally applicable.[140] Indeed, it is extremely difficult to define shared human values and principles and establish what is good or bad in a given situation, *a fortiori* considering the diverse ethical, social and cultural backgrounds spread around the world. A further criticism is the risk of behavioural manipulation and limitation of human

---

consequences of misusing AI technologies and any violations of human rights, including the right to the protection of personal data; "autonomy" implies the power to decide, meaning that the right of individuals to make decisions for themselves about the treatment to receive must not be constrained by AI; "justice" may be translated as a requirement for a correct use of AI in such a way to avoid biased data and unfair discrimination, ensure that the benefits of AI are shared equally, and prevent the creation of new harms. A fifth principle has been added with specific reference to AI that is "explicability" (in the twofold sense of intelligibility and accountability) requiring that AI systems must be made as much as possible intelligible and understandable (at least) to experts. See L Floridi *and colleagues*, 'AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations' (2018) 28 Minds and Machines, 689–707.

[134] AI HLEG Guidelines (n 11), chapter I, paragraph 2.2.

[135] The seven requirements are: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental wellbeing; and accountability. See AI HLEG Guidelines (n 11), chapter II, paragraph 1.

[136] See 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. This Proposal falls outside the scope of this paper, thus it will not be further analysed.

[137] J Morley, L Floridi, 'How to design a governable digital health ecosystem' (2019).

[138] B Friedman, 'Value-sensitive design' (1996) 3(6) Interactions, 16–23.

[139] B Friedman, PA Kahn Jr., A Borning, 'Value sensitive design and information systems' in N Doorn, D Schuurbiers, I van de Poel and ME Gorman (eds.), *Early engagement and new technologies: Opening up the laboratory* (Springer, 2013), 55-95.

[140] A Cenci, D Cawthorne, 'Refining Value Sensitive Design: A (Capability-Based) Procedural Ethics Approach to Technological Design for Well-Being' (2020) 26 Science and Engineering Ethics. See also A Gerdes, 'An Inclusive Ethical Design Perspective for a Flourishing Future with Artificial Intelligent Systems' (2018) 9 European Journal of Risk Regulation, 677–689.

free choice. Indeed, when the technological architecture embeds certain ethical values, it nudges users to act according to those values without leaving them any choice.[141]

However, as it has been argued by Vermaas et al., "technical artefacts are not morally neutral because their functions and use plans pertain to the objectives of human actions, and those actions are always morally relevant".[142] This means that technologies are never neutral but always reflect specific values and normative goals. Given that, the main point becomes what values should be reflected in technology and who should fix them. Different suggestions have been put forward.

First of all, *ad hoc* ethical committees may play a crucial role as independent organisms made up of experts in technology, medicine, ethics and bioethics who may support the regulator in identifying the ethical issues raised by AI and the ethical values to be implemented. Several ethical committees have already been instituted by national and international organizations,[143] and both in the public and private sector several AI guidelines and ethical codes of practice have been adopted.[144] In the field of genomic diagnostics mention can be made to the UK-France Genomics and Ethics Network,[145] and the Joint Committee on Genomics in Medicine that published a guide on the ethical issues arising from the use of genetic and genomic information in the clinic.[146]

Other important instruments are codes of ethics which set the ethical values and practices to follow in developing and deploying AI systems for genomic diagnostics.[147]

Finally, a way to select the values to embed in AI systems and develop a pro-ethical design would be to engage all the categories of actors involved

[141] L Floridi, 'Tolerant Paternalism: Pro-ethical Design as a Resolution of the Dilemma of Toleration' (2016) 22 Sci Eng Ethics.

[142] P Vermaas *et al.*, 'A philosophy of technology: from technical artefacts to sociotechnical systems' (2011) 6(1) Synthesis Lectures on Engineers, Technology and Society, 1–134.

[143] Some examples are the HLEG on AI appointed by the European Commission; the expert group on AI in Society of the Organisation for Economic Co-operation and Development (OECD); the Advisory Council on the Ethical Use of Artificial Intelligence and Data in Singapore; the Select Committee on Artificial Intelligence of the UK House of Lords.

[144] Examples are Google's AI Principles; IBM's everyday ethics for AI; Microsoft's guidelines for conversational bots; Intel's recommendations for public policy principles on AI.

[145] The Network has been "set up to reflect on the ethical and social issues arising from the integration of genomics into routine clinical care". See M Gaille, R Horn, The UK-FR GENE (Genetics and Ethics Network) Consortia *et al.*, 'The ethics of genomic medicine: redefining values and norms in the UK and France' (2021) Eur J Hum Genet.

[146] Royal College of Physicians, Royal College of Pathologists and British Society for Genetic Medicine, *Consent and confidentiality in genomic medicine: Guidance on the use of genetic and genomic information in the clinic* (3rd edn. Report of the Joint Committee on Genomics in Medicine. London: RCP, RCPath and BSGM, 2019).

[147] For instance, in 2019 the UK adopted the Code of Conduct for data-driven health and care technologies setting out the behaviours required to those developing, deploying and using data-driven technologies in such a way to balance the benefits deriving from data-driven health and care technologies to patients, clinicians, service users and the system as a whole with issues such as transparency, accountability, liability, explicability, fairness, justice and bias in order to make sure that the health and care system does not cause unintended harm. This Code is based on the Nuffield Council of Bioethics' principles for data initiatives (respect for persons, respect for human rights, participation and accountability). See Nuffield Council of Bioethics, 'The collection, linking and use of data in biomedical research and health care: ethical issues' (2015).

following the model of society-in-the-loop (SITL) developed by Rahwan.[148] According to Rawhan, when AI systems perform a broad function which has wide societal implications, the algorithms should embed the values of the society as a whole.[149] Indeed, when different interests and rights are at stake, trade-offs need to be negotiated and in doing so the regulator should take into account all the parties involved. This means that in regulating the use of AI for genomic diagnostics all the relevant stakeholders, such as patients and healthcare professionals, but also engineers developing algorithms, should not be regarded just as users subject to the rules imposed by the legislator with a top-down approach, but rather as part of the regulatory process. In such a way, they can put forward their interests and express their expectations and concerns around the use of AI with a collaborative approach in order to identify and combine the values expressed by the different players. As such, they should be regularly engaged throughout the whole design process in order to take more widely accepted decisions and build societal trust in AI.[150]

Conclusions.

Over the last years AI has been revolutionising healthcare under several aspects thanks to the development of machine learning and neural networks that are able to elaborate large sets of data, identify patterns and apply them to unseen data. This development has been enabled by big data, namely the possibility to collect and analyse huge amounts of data from a sheer number of individuals around the world. However, the processing of such amounts of data may pose problems in terms of protection of personal data and individuals' fundamental rights. This topic deserves significant attention because even though AI may significantly improve current diagnostic tools based on individuals' genomic profiles, it may also significantly impact on data subjects' rights given the involvement of sensitive data like genetic data. Therefore, effective regulatory instruments must be identified and put in place.

Among those instruments, the GDPR plays a primary role. However, some misalignments have been identified between the Regulation and the development of AI. The reason is that different interests are at stake, often conflicting between each other. On the one hand, there is the need to safeguard individuals against abuses of third parties that would like to gather as much data as possible to build strong algorithms (indeed, the more data algorithms are trained on, the more accurate and effective they are). To avoid any abuses, data protection law sets strict limits such as the principle of data minimisation and purpose limitation, or the requirement of a lawful basis for processing (e.g. data subjects' consent). On the other hand, it must be noted that the development of AI is not negative *per se* but, on the contrary, may have

[148] I Rahwan, 'Society-in-the-Loop: Programming the Algorithmic Social Contract' (2018) 20(1) Ethics and Information Technology, 5–14.
[149] *Ibid.*
[150] To that end, for instance, Involve and DeepMind developed a guidance on stimulating effective public engagement on the ethics of artificial intelligence. For further details see https://www.involve.org.uk/sites/default/files/field/attachemnt/How%20to%20stimulate%20effective%20public%20debate%20on%20the%20ethics%20of%20artificial%20intelligence%20.pdf.

important benefits for the society as a whole. In particular, AI may significantly improve existing diagnostic tools making it possible to detect disease at an earlier stage and intervene earlier with an adequate medical therapy, thus increasing the quality of patients' life. Developing AI in genomic diagnostics, thus, is meant to protect patients, not to exploit them, and provisions that pose too strict limits to the development of AI, impeding or slowing it down, do not necessarily protect patient's rights. Hence, it is necessary to strike a balance between these conflicting interests to allow the development of AI in a safe and respectful fashion.

Based on this conclusion, further legal and non-legal regulatory instruments should be considered to integrate data protection law. Soft law instruments (like certifications, codes of conduct, impact assessment) have been found highly effective to provide *ad hoc* provisions tailored on the specific issues of genomic diagnostics. Yet, there are some side backs as well, like the risk of over proliferation of such instruments leading to legal uncertainty, or the risk that private interests prevail over the public ones. Thus, soft law instruments may be regarded as a valid integration and specification of data protection law without replacing it.

Secondly, data protection should be enlarged with ethical considerations. The focus should be shifted from the quality and safety of data to human rights and from an individual to a group dimension of protection. The EU is already moving in this direction, as shown by the establishment of the AI High-Level Expert Group, the Commission's White Paper on AI, and the recent proposal for the AI Act.

However, this is only a starting point and further steps are expected in the next years to put in practice the principles identified so far. In particular, it is still not clear how to embed ethics in diagnostic machines – thus realising a pro-ethical design of AI – and how to select the human values to be embedded. In addition, an effective way to engage the actors involved in the processing should be found according to the society-in-the-loop model.

In the light of the above, it may be assumed that one of the main challenges for the next years will be enhancing regulatory instruments that take into account the variety of legal, ethical, and social implications of the use of AI in genomic diagnostics. Specifically, such instruments should represent a fair balance between the protection of both individual and collective rights and the development of efficient AI systems.