TCRS
ISSN: 1970 - 5476

Laura Palazzani[*]

AI and health: ethical aspects for regulation

ABSTRACT
The article will focus on the ethical requirements for a regulation of AI in health. More precisely: meaningful human control or human-centrich approach and huma oversight; reliability of AI thorugh safety and validation of software; transparency and explainability overcoming, whern possible, the problem of opacity; equality, fairness and nondiscrimination avoiding the problem of possible bias (related to gender, age, ethnicity); data quality/accuracy in collection and data sharing, compatible with privacy; shared responsibility; information/education promoting AI literacy; sustainability.

KEYWORDS
Human-centric AI, transparency, fairness, sharing data, responsibility.

1. Definition of AI and applications to health

AI includes all machines which imitate certain aspects of human intelligence using information technologies, being able both to interact and learn from the external environment, and to make decisions with increasing degrees of automation/autonomy[1].

The rapid evolution of AI technologies, in recent years, is characterized by a ‘disruptive’ progress, because of the complexity, broadness of applications and velocity: the progress of AI is mainly due to the increase in computing power, the availability of huge amounts of data and information (which

[*] Professore ordinario di Filosofia del Diritto presso l'Università Lumsa, vicepresidente vicario Comitato Nazionale per la Bioetica, componente del comitato di bioetica presso Commissione Europea, Consiglio d’Europa, Unesco: palazzani@lumsa.it

[1] John McCarthy *et al.*, who coined the term artificial intelligence in 1955, defined AI as “the science and engineering of making intelligent machines”. Calo 2017: 406.

constitute the "examples" for the machine), the development of algorithms, the "learning" capacity of the machine, on the basis of collected and stored information (data), to identify hidden relationships between data (correlations and predictions) through algorithms.

There are various types of AI so also the reflection on AI should be differentiated, according to the difference in technologies. The distinction between ‘weak’ AI and ‘strong’ AI makes a difference also on an ethical and legal level: weak AI is limited to the mechanical application of the human instructions of the programmer; strong AI (that includes machine learning and deep machine learning) uses algorithms that enable an AI system to “learn” to perform a task based on data without having been given specific instructions for that precise task. The difference may be identified in the possible ‘autonomy’ of machines from humans. AI devices are very different depending on their structure, funtioning and logic, their degree of autonomy, their narrow or broad scope. It is important to stress the variability and plurality of AI, which may require specific reflections.

Research and development on applications for AI (both weak and strong) in healthcare is currently being conducted in a large number of fields.

The main applications, bioethically relevant, are AI or cognitive assistence in classifying and stratifying patients in diagnosis[2]; in understanding why and how patients develop diseases in clinical evaluation; in considering which treatment will be most appropriate for them in therapy; in predicting their recovery in prognosis. AI can update appropriate scientific review and guidelines and compare huge amounts of available data. Many applications are also in the field of research and precision medicine, in the analysis of big data in genomics, tailoring optimal treatment for every patient based on genetic factors, medical history, lifestyle, environmental factors, etc.

2. The role of ethics as a critical reflection to inspire regulation

There is a huge ethical debate in the framework of pluralism between on the one hand the technophilic attitude based on libertarian and utilitarian theories that is open and optimistic to every kind of development and use of AI, hoping in a post-humanistic or trans-humanistic future, with the replacement of humans by machines, and on the other hand the technophobic attitude based on the principle of precaution understood as abstention from any technology that in principle may harm a human being, fearing the threat of certain developments and applications ‘beyond’ humans, as a de-humanization and dis-humanization.

In between, there is a balanced and prudent approach - a sort of minimum common shareable ethics – elaborated through interdisciplinary and dialectic reflections, that tries to avoid excessive hopes and hypes, but also excessive fears, adopting an attitude of caution, in order not to hinder techno-scientific progress and at the same time to guarantee a development that is respectful of fundamental human values and rights, such as human dignity, freedom, responsibility, justice as equality and non discrimination.

The elaboration of minimum ethical elements for regulating techno-science draws inspiration from the horizon of fundamental human rights as a conceptual framework, which form a crucial part of national constitutions and international documents. These documents have undergone, in recent decades, a process of explicit specification and interpretation, in light of emerging issues stemming from scientific and technological development, through declarations issued by international organizations (UNESCO, WHO), conventions, resolutions, recommendations, directives, regional

[2] Coeckelbergh 2019; Morley, Machado, Burr *et al*. 2020, 260, pp. 113-172; Pagallo, Aurucci, Casanovas, Chatila, Chazerand, Ignum, Luetge, Madelin, Schafer, Valcke 2019.

regulations (*i.e.* in Europe, the European Group on Ethics in Science and New Technologies at the European Commission, the Committee on Bioethics DH-BIO of the Council of Europe, or ad-hoc groups on AI at WHO).

Within this perspective, ethics plays the role of critical reflection for an understanding and evaluation of AI, which justifies the requirements for regulation trying on the one hand to open innovative technological opportunities 'for' human beings and humankind in the field of medicine and on the other hand to avoid or at least control and manage risks.

The regulation of the new emerging technologies, characterized by speed of development, uncertainties and unpredictability, is based on some criteria: anticipation, proactive imagination and identification of the potential scenario, possible or probable negative features and outcomes of new technologies. This new kind of governance is oriented towards soft instruments, more easily allowing changes, adaptations and reviewing, and does not require setting up a formally complete and timely regulatory framework, which may or may not eventually take place. The inefficiencies of the law, constantly chronologically "lagging behind" (the so called 'law lag') techno-scientific innovation and giving space to diversification due to pluralistic ethical approach, is changed by the methodology of anticipation that requires not only accelerating the pace of normativity but even being ahead of its objects as in the application of the so called 'ethics-by/in/for-design'.

AI systems do not operate in a lawless world or 'legal empty space'. A number of legally binding rules at national, European, and international level already apply or are relevant to the development of AI.

Legal sources are: the Treaties of the European Union (Treaty of Rome 1957; Treaty of Maastricht 1992; Treaty of Amsterdam, 1997; Treaty of Maastricht, 2001; Treaty of Lisbon, 2007), the Charter of Fundamental Rights (Nice, 2000), the General Data Protection Regulation (2016), the Product Liability Directive (1985), the Regulation on the Free Flow of Non-Personal Data (May 2019), the Medical Device Regulation (2017), Anti-discrimination Directives (2009), Consumer Law and Safety and Health at Work Directives (2014-2020), the UN Human Rights treaties and the Council of Europe conventions (such as the Convention on Human Rights and Biomedicine, 1997), and numerous EU Member State laws[3].

But the speed of the development and deployment of these technological developments is much faster than that of the legal framework regulating them, which requires the close attention of policy makers and politicians[4].

The point of departure of regulation is the anthropological view: that AI is a tool designed and produced by a human being (who is the subject); that human beings are not replacable and should not be replaced by machines (as is the post-humanists/transhumanist dream); that decisions are human and not replaceable/should not be replaced with artificial and mechanical decisions (above all in healthcare). Within this general 'human-centric' framework, the most convincing approach in the introduction of AI in medicine is the non complete replacement of human intelligence by AI, but its assistance or cognitive assistance.

For an appropriate regulation of AI in medicine[5], there are a number of new (or renewed) rights, in order to achieve a balanced framework for humans and AI with regard to health: there is a perceived need for interdisciplinary discussion.

[3] The Council of Europe is preparing a binding legal instrument on AI, such as a convention open to non-member States, with an emphasis on the human rights implications of AI in general and on the right to health in particular (see *Report on Artificial Intelligence in Health Care: Medical, Legal and Ethical Challenges Ahead*, 1 October 2020).

[4] Brownsword 2020: 26; Santosuosso 2015; Winfield *et al.* 2019; Coeckelbergh 2020.

[5] The European Commission's white paper *On AI – A European Approach to Excellence and Trust*, published in Brussels on February 19, 2020; the *Proposals for Ensuring Appropriate Regulation of AI*, issued by the Office of the Privacy Commissioner of Canada published on March 13, 2020.

3. Main emerging ethical requirements for regulation of AI

There are some ethical requirements that need to be taken into consideration in a sort of ‘anticipatory ethics’ for a future regulation of AI, on a European level[6].

3.1. *Meaningful human (physician) control*

One of the main problems debated in bioethics, in the field of human-machine interaction, is the possible complete "replacement" of the human decision-making capacity, or the "autonomisation" of machines that could escape human control in a possibile future scenario. When systems can learn to perform tasks without human direction or without supervision they are now often called ‘autonomous’, as they develop and can perform tasks independently from human operators, and for that reason unpredictable and without human control.

This is considered a threat to human dignity, as it may open possible applications/decisions against humans or provoking harm to humans. This position is argued on the basis of the recognition of the principle of human dignity, in a human-centric approach, and principle of non maleficence (do no harm to humans) and beneficence (do good to humans) in bioethics. Even if humans construct AI, select data, elaborate algorithms, train machines, they need to keep control and oversight over what they design, program, apply; machines should remain a ‘support’ to human decision, that cognitively ‘assist’ human decisions, but do not ‘substitute’. Machines should not ‘compete’, but comp*l*ete’ human actions.

In this sense also the language should be kept anthropocentric, considering AI machines as ‘automatic’ rather than ‘autonomous’ in learning. Autonomy in its ethical original meaning can be attributed only to human beings: autonomy means the capacity of an agent to act in accordance with values. The term ‘autonomy’ cannot be applied to artefacts, even if very advanced complex or even ‘intelligent’ systems. The terminology of ‘autonomous’ systems is however widely used in scientific literature and in public debate to refer to the highest degree of automation and the highest degree of independence from human beings in terms of operations and decisions. But autonomy in its original sense is an important aspect of human dignity that ought not to be relativised, but referred primarily to human beings. In the same way AI can not be considered as ‘electronic persons’ (as in the *European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics*), as persons or moral and juridical subjects are only humans and not machines. Machines may also reduce cognitive errors and achieve possible superior performance results above all on a quantitative level (collection of data and correlations of information), if compared to humans, but there are some functions which remain specifically human and irreplaceable, such as intuition, imagination, creativity, interpretation, empathy, self awareness, self-consciousness, self-authorship. All of which play a role in decision-making.

The need to keep the human oversight remains essential also in order to avoid the possible problem of technological delegation. An expert system that becomes optimal in suggesting "decisions" to humans (also in medicine) poses the risk of decreasing human attention with the possible consequence of reducing human skills (the so called phenomenon of de-skilling or de-professionalization),

[6] Nuffield Council on Bioethics, *Artificial Intelligence in Healthcare and Research*, 2020; the Swedish National Council on Medical Ethics, *Artificial Intelligence in Healthcare*, 2020; the Italian Committee for Bioethics, together with the Italian Committee for Biosafety, Biotechnology and Sciences of Life, *AI and Medicine: Ethical Aspects*, 2020.

reducing responsability (de-responsabilization) going towards the artificialization of choices that in medicine can impoverish and even cancel the patient/physician concrete relationship (or de-humanization). In this sense, it is important to discipline the productive synergy as a complementarity between human beings and machine, searching for ways of intelligent "support" that allows humans to have "significant or meaningful human control"[7] in terms of attention, contribution, supervision, control and responsibility.

In the physician-patient relationship**,** AI may be more efficient, precise, rapid and less expensive: it may be desirable if we consider the automation of certain tasks involving repetitive, boring, dangerous, or strenuous activities. If properly used, AI could reduce the time that professionals have to devote to merely routine or bureaucratic incumbencies, or activities which expose them to avoidable dangers, allowing them to have fewer risks and more time for the patient.

AI should be considered exclusively as an aid to the physician's decisions, which remain controlled and supervised by humans. It is for the physician in any case to make the final decision, as the machine solely and exclusively provides support for data collection and analysis, of a consultative nature. An "automated cognitive assistance" system in diagnostic, therapeutic and prognostic activity is not an "autonomous decision-making system". It collects clinical and documentary data, compares them with statistics relating to similar patients, speeding up the analysis process of the physician. In this sense, the machine cannot replace the human being in a relationship that is built on the meeting of complementary areas of autonomy and responsibility. Personal contact is the essential element of every diagnosis, prognosis and therapy. Delegating complex tasks to intelligent systems can lead to the loss of human and professional qualities, and the impoverishment of the patient-physician relationship.

Above all, difficult medical decisions such as priority in surgical operations, access to intensive care or triage, end of life decisions, need to be based on solid and transparent human reasoning, that cannot be replaced – because of the complex ethical issues – by machines[8].

A relational conception of human dignity which is characterised by our social relations, requires that we should be aware of whether and when we are interacting with a machine or another human being, and that we reserve the right to vest certain tasks to the human or the machine[9]. In this ethical framework, the ethics of AI is the ethics of human beings: the machine cannot obscure the agency, which is human. Humans conceive, design, use AI and humans should be kept at the centre (human-centric approach).

[7] The principle of Meaningful Human Control was first suggested in the field of weapon systems. This means that humans - and not computers and their algorithms - should ultimately remain in control, and thus be morally responsible. But now it is also used with reference to human oversight.

[8] The High-Level Expert Group on Artificial Intelligence, in *Ethics Guidelines for Trustworthy AI* (April 2019) underlines the need to preserve the "*human-centric*" dimension of the new technologies. The European Group on Ethics in Science and New Technologies in *Artificial Intelligence, Robotics and so called 'Autonomous' Systems* (March 2018) affirms the importance "that humans - and not computers and their algorithms - should ultimately remain in control, and thus be morally responsible" and "The principle of human dignity, understood as the recognition of the inherent human state of being worthy of respect, must not be violated by 'autonomous' technologies". The European Commission *White Paper On Artificial Intelligence - A European approach to excellence and trust* (19 February 2020, p. 12) admits that "the specific characteristics of many AI technologies, including (…) partially autonomous behaviour, may make it hard to verify compliance with, and may hamper the effective enforcement of rules of existing EU law meant to protect fundamental right". See also Recommendation of the Council on OECD, *Legal Instruments Artificial Intelligence*, 2020.

[9] In the *General Data Protection Regulation* (2016) there is a mention to guarantee that AI-driven health applications do not replace human judgement completely and that thus enabled decisions in professional health care are always validated by adequately trained health professionals. But it could be implemented.

Given the technical possibility of creating artificial systems that can be confused with humans[10], there should be a right of knowing the human or artificial nature of the interlocutor[11]. Ignorance or not clear understanding about the nature of the interlocutor could lead to misunderstandings, and betray the expectation of an empathic understanding[12] and could affect human dignity[13], from the side of the patient. The right to a non-fully automated decision is based on the need to identify who is to be considered in charge of the function and the related responsibility.

The European Union has expressly considered in art. 22 of the GDPR (*General Data Protection Regulation* 2016/679) that "The data subject 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"[14]. The effectiveness of this section is weakened, by the exceptions provided for in the same article, when the decision "is based on the data subject's explicit consent"[15]. This clause risks impoverishing the right to a non-fully automated decision.· There is the risk that people consider it more convenient to technological delegation[16]. The decision would be substantially 'captured' by the machine, the so called 'sheep effect' (*effet moutonnier*), the human's role would 'vanish/disappear' (the so called 'human in the loop' or 'human on the loop' or 'human-incommand').

Under this perspective it could be necessary to explicitly regulate the necessity of the 'meaningful human/physician control' in the application of AI to medicine, from and by design, in order not to leave alone neither the physician nor the patient.

3.2. *Reliability of AI: safety and validation of software applied to health*

For every machine or technology the design needs to be safe; safety is an ethical requirement for every machine/technology just as it is for pharmaceuticals, food, transportation etc. This needs to be applied also in AI.

Data may be considered 'reliable' as they are collected from reality. Algorithms are generally considered 'reliable' in themselves, only for the fact that their methods are represented through measurable and mathematical systems. But accurate controls should be made, of both data (the accuracy of collection) and algorithms (the validation of softwares), in order to obtain the most

[10] As is well known, this is at the center of the Turing test: Turing 1950: 433.

[11] See the mentioned *Statement on Artificial Intelligence, Robotics and 'Autonomous Systems'*, issued by EGE, 11: "we may ask whether people have a right to know whether they are dealing with a human being or with an AI artefact".

[12] In a few specific areas, this 'distraction' can have beneficial results: Huijnen *et al.* 2019: 11.

[13] Report of UNESCO, COMEST on *Robotics Ethics*, 2018. As stated in a UNESCO report, "Dignity is inherent to human beings, not to machines or robots. Therefore, robots and humans are not to be confused even if an android robot has the seductive appearance of a human, or if a powerful cognitive robot has learning capacity that exceeds individual human cognition".

[14] In recital 71, the GDPR states as follows: "The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online credit application or e-recruiting practices without any human intervention. Such processing includes 'profiling' that consists of any form of automated processing of personal data evaluating the personal aspects relating to a natural person, in particular to analyse or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, where it produces legal effects concerning him or her or similarly significantly affects him or her".

[15] A commentary in Bygrave 2017, quoted.

[16] This risk has been reported both in medicine and in justice. In medicine: "The collective medical mind is becoming the combination of published literature and the data captured in health care systems, as opposed to individual clinical experience": Char, Shah, Magnus 2018: 981.

probable certainty that the introduction of various forms of AI are beneficial (and not maleficial), above all when applied to health. In the same way as we validate pharmaceuticals or devices in clinical trials.

All the "products" of AI should be compared, through studies conducted according to the rules of controlled clinical trials (comparing results regarding the health of the patient with and without the use of AI), with decisions that are made independently of AI by groups of competent physicians, together with informatics experts and engineers. Controlled clinical studies remain the "gold standard" for the demonstration of the safety and efficacy of treatments. A new methodology to control the software applied to medicine is required, including the problem of the mechanism changing over time and the validation requiring monitoring and further checks.

It will therefore be necessary to demonstrate AI safety to ensure that also unintentional harm can be minimized and prevented and to ensure technical robustness on the basis of control starting from the data base (quality, accuracy, interoperability of clinical data, both collected and compared), the algorithms applied, the advantage in terms of benefits and risks in the application to the patients. Only in this way will it be possible to demonstrate the reliability of these systems through certifications/ validations/ monitoring that guarantee their usability in clinical practice. Only in this way can there be the entrusting of complex tasks in order to support the trust relationship between patients and AI[17].

In this direction there could be implementation of the regulation on clinical trials (*Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on Clinical Trials on Medicinal Products for Human Use, and Repealing Directive 2001/20/EC* and *Regulation on Medical Device*, 2017) including a specification of AI trials and protocols to be considered as scientific, ethical and legal requirements for validation. In this sense it could be necessary to integrate Ethical Committees (or Ethical Review Boards) with informatics/engineers, by reason of their specific competence which is necessary in the interaction with physicians and ethicists. Also deontological codes could improve the duty of the operator in terms of robustness, security and safety in order to ensure accountability of the technologies and their applications.

3.3. *Transparency and explainability: the problem of opacity*

"Opacity" refers to the non explainability or limit of explainability of algorithms that interpret/classify data. It is in some circumstances impossible also for software programmers and informatics to explain how the system has achieved certain results (the 'black box problem', where only the input and output are known, but not what stays in between). It is practically impossible for a human being (even an expert) to analyze the huge amount of calculations made by the algorithm and find out exactly how the machine managed to decide. And this is almost impossible for the physician, who is not an informatician and is not in any case sufficiently competent to do it.

Automation can lead to 'opacity' or absence/lack of transparency on the paths followed by the machine. The machine does not provide complete information on the correlations of data and/or on the logic adopted to achieve a conclusion or propose a decision. In this sense it is not possible for the physician, that uses AI, to track or trace back the processes of the decision proposed by AI and explain it in all its various phases of argumentation to the patient.

[17] The Italian Committee for Bioetics recommends in the Opinion on *AI and Medicine: Ethical Aspects* (May 2020) the encouraging of research in technology validation and certification tools and surveillance and monitoring, as indispensable elements for creating a "social pact of trust and reliability" of technologies in the medical field. In this sense it would be advisable to integrate the figure of a computer scientist or an AI expert into ethics committees for experimentation, and also update the legislation on experimentation with reference to software in the medical field.

The opacity surrounding the essential elements and the decision-making process by which an AI system applied to medicine can reach a conclusion, involves the risk that physicians cannot confirm, or reasonably reject, the proposal made by the intelligent system applying their own decision, because the motivation is unknown. This raises problems for the physician in relation to the machine (whether or not to rely on the algorithms) and in relation to the patient, to whom the physician cannot provide transparent and complete information on the decision regarding his/her health. Non-transparent systems or non intelligible systems make it harder to identify errors and therefore can also undermine the reliability of AI, and trust of both the physician and the patient.

In this sense, it would be necessary to regulate the right to explainability to the extent possibile, meaning the right to receive an explanation of the decision of AI (for the patients). The right to an explanation of the steps through which AI has produced the result, as an understandable description (not technical) of the logic, necessary above all in healthcare[18], making evident bias. This right should impose an informational burden which some experts consider impossible or unsustainable, in any case this aim of the technological feasibility should be explicit[19] in order to realize ‘trustworthy AI’. This entails the need to develop technology that is able to explain every step of the decision or at least to inform the users (both physicians and patients) of the risk of opacity, in order to acquire a critical awareness. Humans need to be aware that they are interacting with an AI system, and must be informed of the system’s capabilities and limits.

3.4. *Equality, fairness and non discrimination: the problem of bias*

Data collected from patients (clinical data, personal data, data on life style gathered from different sources) are classified according to categories, into groups and subgroups and algorithms correlate them. Classification means profiling individuals under clusters with certain schemes. Such classifications, profiling, clustering may be discriminatory. The risk arises, by stratifying patients into groups or subgroups on the basis of personal profiles, as classification may be based on various criteria (decided by the data collector) or purposes defined by different subjects, the informatics or expert/computer scientist who designs the algorithms or the physician who defines the clusters.

The identification of groups or subgroups may be based on exclusion criteria (i.e. age, gender, ethnicity), unintentionally or not intentionally: i.e. data may be collected only from individuals of a certain age, gender o ethnical origin, and the system does not recognise individuals of different age, gender, ethnicity.

Algorithmic discrimination is possible, even in the medical field, with an impact on equality and inclusiveness. Inequalities already exist in the health sector, but AI could accentuate and worsen them by creating and/or increasing the "gap", discriminations and inequalities. If a healthcare algorithm learns from a training database in which certain groups of patients are under-represented or excluded, it can lead to these groups running a greater risk of misdiagnosis, misprognosis or mistreatment.

Medical decisions may be made exclusively on the basis of these profiles or on the basis of considerations not related to healthcare (also indirectly linked, for example, to age, gender and ethnic origin) and without taking into consideration the variations that a particular patient may present. A care/cure decision based exclusively on the profiles of patients in an automated way (through algorithms) can lead to the exclusion of treatment without offering an alternative, albeit presumed less effective, but nevertheless an indicated alternative.

[18] Floridi, Cowls 2019; Wachter, Mittelstadt 2019: 494.

[19] Rudin 2019: 206.

It is possible to avoid this problem with the adoption of a broad and inclusive approach, which could be representative of all categories, groups and subgroups, to be continuously updated and placing attention on inclusion, for the collection /selection of data and the development of algorithms.

This implies a reflection on the "data ethics" supporting AI (which require accuracy in collecting and selecting) and on "algorithm ethics" (also called "algor-ethics"), which should be based on data that are not selected, or, alternatively on inclusive and non-discriminatory selections. These ethical requirements in the regulation of AI in medicine call for the implementation of technologies that can prevent and detect discriminatory biases in data sets used to train and run AI systems and at least report and neutralise at the earliest stage possible[20]. Unfair bias should be avoided, as it introduces forms of marginalization of vulnerable groups, exacerbation of prejudice and discrimination. AI systems should be accessible to all (design-for-all), regardless of any difference.

3.5. *Data quality and data sharing: the problem of privacy*

AI is based on data. The availability of data (clinical data, genetic data, personal data, data on lifestyle and environment, etc.) is not enough. In order to have reliable and safe AI, it is necessary to have 'quality of the data' (beyond the quantity of 'big data')[21], that means accuracy in collecting (verifying the authenticity and veracity of the variety of data sources and the velocity of collection), the interoperability of the data (through standardization and classification criteria) as conditions for the developments of AI and their correct applications in the medical field. Since every AI system is based on data, the problem of preparing and supervising data from human beings emerges, the avoidance of errors in data collection and classification, as well as providing AI mechanisms for checking and verifying correctness.

The protection of privacy and confidentiality is often underlined as an obstacle to the development of AI, which is based on big data. In an age of ubiquitous, pervasive and massive collection of data through digital communication technologies, the right to protection of personal information and the right to respect for privacy are crucially challenged.

AI needs to dispose of data in a broad field, on a global level (with transfer of data to other countries) and storage of data over time. Full anonymization of data, that is the use of data without any relation to personal data, are not useful for medical research. Pseudonymisation (or codification) is accepted, as an intermediate solution between anonymization and full identification, as it allows traceability, with identification in cases where it is important to communicate the results. But appropriate conditions to prevent improper disclosures are needed, in case of the use of data by health insurances or in the workplace.

The huge collection of data, necessary for AI, and pseydonymized as required by research, also highlights the risk, related to the crossing of data, of intentional and/or accidental re-identification, raising the problem of privacy, which in this context tends to "end" or "evaporate"[22]. In this sense

[20] Also European Group in Ethics in Science and New Technologies, *Ethical Issues of Healthcare in the Information Society*, 1999; *Ethics of Information and Communication Technologies*, 2012; *New Health Technologies and Citizen Participation*, 2015; *Statement on Artificial Intelligence, Robotics and 'Autonomous Systems'*, 2018; *Future of Work. Future of Society*, 2018. See European Commission, *Communication on AI for Europe*, 2018; Council of Europe, Committee of experts on *Human Rights Dimensions of Automated Data Processing and Different Forms of Artificial Intelligence*.

[21] Deutscher Ethikrat, *Big Data and Health: Data Sovereignity as the Shaping of Informational Freedom*, 2018; UNESCO, International Bioethics Committee, *Big Data and Health*, 2017.

[22] On the subject of privacy, the Italian Bioethics Committee intervened in the document *Information and Communication Technologies and Big Data: Bioethical Issues*" (2016) underlining that as part of the "data processing

technologies are becoming increasingly "opaque" and the users increasingly "transparent". In the AI era and the need for the use of data for medical research, the possibility/opportunity of "sharing" data arises, as a "social/common good" for the advancement of scientific knowledge. But this requires specific regulation, in order to protect the advancement of science and at the same time to protect the patients. Besides ensuring full respect for privacy and data protection, adequate data governance mechanisms should also be ensured, taking into account the quality and integrity of the data.

There is wide debate, even on a regulatory level, on the applicability of the Regulation (EU) 2016/679 (GDPR) to AI scenarios, where it is unrealistic to protect privacy and guarantee data control, in the global research area (ICT) and in times that cannot be defined a priori. There are methods and technologies for performing data transactions while preserving data security (one of the technologies is the family of block-chain applications). The sharing needs to be guaranteed by the exclusive use for research purposes (and not for commercialization), which enables a return of information and sharing of clinically relevant results (*benefit sharing*), with specific controls on data breaches and abuse, by implementing GDPR.

3.6. *Informed consent in AI-driven medical research: autonomy*

The traditional way to conceive informed consent proves to be ineffective in medical research driven by AI. Especially when using genetic data or data associated with biological samples stored in biobanks (as in rare diseases or currently for Coronavirus), medicine requires the processing of an ever-increasing amount of health data together with personal data and data on lifestyle. In this perspective, for instance, ‘precision medicine’ is based on the possibility of collecting, processing and comparing as much clinical information belonging to as many patients as possible, in order to diagnose a disease and define a possible therapy.

In this field, data collected wherever (without borders) are a fundamental asset that should be maintained whenever (over time), also for future research and follow-ups, shared worldwide among all the researchers involved, and used for studies whose purposes may not be directly connected with the project initially planned and subscribed to or, as in the case of the Coronavirus, which may not have been even imaginable at the time of the acquisition of the original consent. Given these characteristics, informed consent, as originally designed for clinical trials, constitutes an obstacle to achieving reliable results and tends to hinder the research done and possible with AI.

Informed consent cannot just be cancelled, as it remains a relevant instrument and process that guarantees the patient’s information, awareness and autonomy. But it has to be replaced by other tools. There is ongoing discussion on the possibility/feasibility of ‘relaxing’ the informed consent requirement, which needs to become more broad, flexible and dynamic. In this perspective the informed consent, delineated for each clinical study in detailed, restricted and rigid consent, needs to become open and variable consent. Within these conditions patients can grant the first consent, admitting at the same time the possibility to extend the original consent to all clinical trials with the same outcomes or related to the same disease, or even to any clinical trials, in this case, subject to review by an ethical committee. In this case, unless otherwise noted, the consent is presumed to be

when requesting information, it must always be accompanied by an explicit informed consent", in a transparent, complete and simple way, specifying" who collects and who will use the data, what data, how it is collected, where it will be stored and for how long, for what reason and for what purpose", specifying revocability. In the opinion *Mobile-health Applications: Bioethical Aspects* (2015), the ICB expresses awareness "of the difficulty of achieving an informed consent and of protecting the privacy of users in this new field of application".

valid for all clinical trials, also in the future. Informed consent becomes a sort of awareness to remain open to all (potential) possibilities of research.

A human-centered AI needs to optimize the huge amount of clinical data in order to increase the possibility to reach therapeutic benefits for both present and generations to come in the next or remote future[23]. It is necessary, in this perspective, to move from informed consent based on the protection of privacy to a means that leads to the possible disclosure of any kind of personal data useful for research. Informed consent, in the era of ICT and AI, becomes/develops a critical awareness of the limit of privacy in the sharing of data.

Regulation (EU) 2016/679 still considers informed consent as the protection of personal data, but does not deal with the question of AI and medical research[24]. It is therefore urgent to further develop a debate in order to find more adequate tools for the protection of personal data avoiding abuse/ misuse, and to reconcile the realm of AI with biomedical research, above all in front of the development of AI in precision medicine. The right/duty to flexible privacy in medical research.

3.7. *Shared responsibility: human responsibility*

Automation in medicine can contribute to an improvement in medicine, but it is not without risks. AI can be poorly designed and applied, with negative consequences for the patient.

The issue of liability is one of the most delicate and complex problems that arise with the use and development of new AI systems also in medicine. In the moral and juridical field, it is necessary to clarify whether accountability for certain decisions made through an intelligent system should be attributed to the software designer (or designers), the manifacturer, the vendor, the owner, the user (the physician) or third parties (patients). The possible occurrence of errors should be traced and analyzed as is the case for any medical error.

The regulation needs to innovate the guarantees towards new categories of risks both for the patient and for the physician. AI has an "author" who creates it (programmer, validator) and who may not coincide with the "producer" and manufacturer of the product that incorporates it, the "seller", the “owner” and the “user” (the physician). The fact there are different subjects involved, with the consequence that there is a ‘shared responsibility’ which should be able to be directly asserted by the end user of that product, the patient, not only through the traditional “contract” that binds the physician and the healthcare facility, but also a "shared and articulated contract".

Regulations need to clarify with whom liabilities lie for damages caused by undesired behaviour of AI, as ‘autonomous’ systems. Moreover, effective harm mitigation systems should be in place.

3.9. *Medical, technological and social information/education*

The physicians and healthcare professionals are generally not trained to use the results of AI research. It is therefore very important to introduce a specific education on AI in the activities of Continuing Medical Education (ECM) in order to avoid the so called “skills polarization” of employees, i.e. “re-skilling” workers, in this case healthcare workers, in the face of developments in emerging technologies. This gives rise to the concern of the European Group on Ethics in Science and New Technologies (EGE), in the opinion *Future of Work*, *Future of Society (*2018), underlining the “skills polarization" that can hide new forms of discrimination, excluding those who are unable

[23] Lee 2018: 1. Fei-Fei Li and John Etchemendy lead the Stanford Institute for Human-Centered AI (HAI).

[24] Among others, Lambert 2017.

to secure the new required "skills". The problem of new professions, even in the medical field, remains therefore that high-level skills will be required.

It is necessary to re-design medical education programs, allocating a significant part of the training of future doctors to the problems deriving from the virtualization, digitalisation and artificialization of medicine which is the basis of AI technologies. There is a need for interdisciplinary and crossdisciplinary courses for the training of health professionals to constant adaptation to technological change and to the possible "convergence" of traditionally separated disciplinary sectors (e.g. medicine and computer science or physics or data science).

Education should also be renewed, introducing ethics and bioethics courses for engineers, computer technicians, computer scientists and data scientists, with particular reference to ethics within technologies (ethics by design/in design/for designers) and in the planning, methodology and application of technologies. This is the only way to ensure ethical awareness and comprehension from the very beginning of the technological design. The designer and programmer of AI, particularly machine and deep learning, could benefit from interdisciplinary training, on ethical, social and legal aspects of their activity. Designing AI is not only a technical activity, but it entails, intentionally or unintentionally, ethical and legal concepts.

It is also desirable to promote public debate on the developments and limits of AI in medicine, so that all individuals - present and future patients - can acquire the basics of "AI literacy", promoting an active participation in social discussion. These are the prerequisites for a possible overcoming of the "digital divide" in medicine avoiding the marginalization, stigmatization and exclusion of people without technologies and competences and motivation to use them, in the framework of inclusiveness.

The right to free education on AI should be provided[25] as a compulsory education, both for the youngest (from elementary school onwards) and for adults (in universities and professional training). There is a need for a common understanding of AI and its pros and cons, gaining/developing critical awareness, in order to overcome the AI divide, and ensure equal access to opportunities and inclusive growth.

3.10. *Sustainability*

AI systems should benefit all human beings, including future generations.

Their sustainability should be ensured, both on a social and environmental level. They should take into account access for all, considering the costs, and respect for the environment, including all living beings. There should be special attention paid to the social and environmental impact of AI.

AI technology should be based on the human responsibility to ensure the basic conditions for life on our planet, preserving a good environment for future generations. AI can contribute to well-being and help to achieve European socio-economic goals if designed and deployed wisely.

[25] US National Science and Technology Council (2016) *Preparing for the Future of Artificial Intelligence*; Royal Society (2017) *Machine Learning: The Power and Promise of Computers that Learn by Example*, Future of Life Institute (2017).

BIBLIOGRAFIA

Brownsword R. 2020, "Law, Technology, and Society: In a State of Delicate Tension", *Notizie di Politeia*, Jan 1, 36 (137): 26-58.

Bygrave L.A. 2017, *EU Data Protection Law Falls Short as Desirable Model for Algorithmic Regulation*, in L. Andrews *et al.*, *Algorithmic Regulation*, The London School of Economics and Political Science, Discussion Paper N.85, September 2017.

Calo R. 2017, "Artificial Intelligence Policy: A Primer Roadmap", *UC Davis Law Review*, 51(2): 399-435.

Char D.S., Shah N.H., Magnus D. 2018, "Implementing Machine Learning in Health Care – Addressing Ethical Challenges", *The New England Journal of Medicine*, 378(11): 981-983.

Coeckelbergh M. 2019, "Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability", *Science and Engineering Ethics*, 26: 2051-2068.

Coeckelbergh M. 2020, *AI Ethics*, MIT Press.

Council of Europe, Committee of experts on Human Rights Dimensions of Automated Data processing and Different Forms of Artificial Intelligence.

Council of Europe, *Report on Artificial Intelligence in Health Care: Medical, Legal and Ethical Challenges Ahead* (1 October 2020).

Deutscher Ethikrat 2018, *Big Data and Health: Data Sovereignity as the Shaping of Informational Freedom*. Available at: https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/englisch/opinion-big-data-and-health-summary.pdf (last access 24 March 2021).

European Commission 2018, *Communication on AI for Europe*. Available at: https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe (last access 24 March 2021).

European Commission, *White Paper On Artificial Intelligence - A European approach to excellence and trust* (19 February 2020). Available at: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (last access 24 March 2021).

European Group in Ethics in Science and New Technologies 1999, *Ethical Issues of Healthcare in the Information Society*. Available at: *https://ec.europa.eu/info/publications/ege-opinions_en* (last access 24 March 2021).

European Group in Ethics in Science and New Technologies 2012, *Ethics of Information and Communication Technologies*. Available at: *https://ec.europa.eu/info/publications/ege-opinions_en* (last access 24 March 2021).

European Group in Ethics in Science and New Technologies 2015, *New Health Technologies and Citizen Participation*. Available at: *https://ec.europa.eu/info/publications/ege-opinions_en* (last access 24 March 2021).

European Group in Ethics in Science and New Technologies 2018, *Statement on Artificial Intelligence, Robotics and 'Autonomous Systems'*. Available at: *https://ec.europa.eu/info/publications/ege-opinions_en* (last access 24 March 2021).

European Group in Ethics in Science and New Technologies 2018, *Future of Work. Future of Society*. Available at: *https://ec.europa.eu/info/publications/ege-opinions_en* (last access 24 March 2021).

European Group on Ethics in Science and New Technologies, *Artificial Intelligence, Robotics and so called 'Autonomous' Systems* (March 2018). Available at: *https://ec.europa.eu/info/publications/ege-opinions_en* (last access 24 March 2021).

Floridi L., Cowls J. 2019, "A Unified Framework of Five Principles for AI in Society", *Harvard Data Science Review*, 1(1): 1-15.

High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI* (April 2019). Available at: https://ec.europa.eu/futurium/en/ai-alliance-consultation (last access 24 March 2021).

Huijnen C. *et al.* 2019, "Roles, Strengths and Challenges of Using Robots in Interventions for Children with Autism Spectrum Disorder (ASD)", *Journal of Autism and Developmental Disorders*, 49(1): 11-21.

Italian Bioethics Committee 2015, *Mobile-health Applications: Bioethical Aspects*. Available at: *http://bioetica.governo.it/en/opinions/opinions-responses/mobile-health-apps-bioethical-aspects/* (last access 24 March 2021).

Italian Bioethics Committee 2016, *Information and Communication Technologies and Big Data: Bioethical Issues*. Available at: *http://bioetica.governo.it/en/opinions/opinions-responses/information-and-communication-technologies-and-big-data-bioethical-issues/* (last access 24 March 2021).

Italian Committee for Bioethics, together with the Italian Committee for Biosafety, Biotechnology, and Sciences of Life 2020, *Artificial Intelligence and Medicine: Some Ethical Aspects*. Available at: *http://bioetica.governo.it/en/opinions/joint-opinions-icbicbbsl/artificial-intelligence-and-medicine-some-ethical-aspects/* (last access 24 March 2021).

Lambert P. 2017, *Understanding the New European Data Protection Rules*, Oxfordshire: Taylor and Francis Ltd.

Lee J.E. 2018, "Artificial Intelligence in the Future Biobanking: Current Issues in the Biobank and Future Possibilities of Artificial Intelligence", *Biomedical Journal of Scientific & Technical Research*, 7(3), 1: 5937-3939.

Morley J., Machado C.C.V., Burr C. *et al.* 2020, "The Ethics of AI in Healthcare: A Mapping Review", *Social Science & Medicine*, 260: 113-172.

Nuffield Council on Bioethics 2020, *Artificial Intelligence in Healthcare and Research*. Available at: *https://www.nuffieldbioethics.org/publications/ai-in-healthcare-and-research* (last access 24 March 2021).

OECD 2021, "Recommendation of the Council on Artificial Intelligence", *OECD Legal Instruments*. Available at: *https://www.fsmb.org/siteassets/artificial-intelligence/pdfs/oecd-recommendation-on-ai-en.pdf* (last access 24 March 2021).

Office of the Privacy Commissioner of Canada, *Proposals for Ensuring Appropriate Regulation of AI*, published on March 13, 2020.

Pagallo U., Aurucci P., Casanovas P., Chatila R., Chazerand P., Dignum V., Luetge C., Madelin R., Schafer B., Valcke P. 2019, *AI4People - On Good AI Governance: 14 Priority Actions, a S.M.A.R.T. Model of Governance, and a Regulatory Toolbox* (November 6, 2019). AI4PEOPLE.

Rudin C. 2019, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead", *Nature Machine Intelligence*, 1: 206-215.

Santosuosso A. 2015, "*The Human Rights of non Human Artificial Entities: an Oxymoron?"*, *Jahrbuch für Wissenschaft und Ethik*, 19(1): 203-238.

Swedish National Council on Medical Ethics 2020, *Artificial Intelligence in Healthcare*. Available at: *https://smer.se/en/2020/05/28/in-brief-artificial-intelligence-in-healthcare/* (last access 24 March 2021).

Turing A.M. 1950, "Computing Machinery and Intelligence", *Mind*, 59: 433-460.

UNESCO, International Bioethics Committee 2017, *Big Data and Health*. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000248724 (last access 24 March 2021)

UNESCO COMEST 2017, *Report of COMEST on Robotics Ethics*. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000253952 (last access 24 March 2021).

US National Science and Technology Council 2016, *Preparing for the Future of Artificial Intelligence*. Available at: *https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf* (last access 24 March 2021).

Royal Society 2017, *Machine Learning: the Power and Promise of Computers that Learn by Example*, Future of Life Institute. Available at: *https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf* (last access 24 March 2021).

Wachter S., Mittelstadt B. 2019, "A Right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI", *Columbia Business Law Review*, 2: 494-620.

Winfield A. *et al.* 2019, “Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems”, *Proceedings of the IEEE*, 107(3): 509-517.