Presidenza del Consiglio dei Ministri

CNB
COMITATO NAZIONALE PER LA BIOETICA

CNBBSV
COMITATO NAZIONALE PER LA BIOSICUREZZA LE BIOTECNOLOGIE E LE SCIENZE DELLA VITA

ARTIFICIAL INTELLIGENCE AND MEDICINE: ETHICAL ASPECTS

29 May 2020

CONTENTS

Presentation

During a meeting between the President of the Council of Ministers and the Italian Committee for Bioethics (26 September 2019), a specific request was made by President Conte for a pronouncement by the Committee on the use of artificial intelligence (AI) in the field of healthcare. In the context of the speech addressed to the Italian Committee for Bioethics (ICB), President Conte underlined: "In the perspective in which we are heading, where technological innovation will further characterise our daily lives and where obviously, from a government perspective, we will push even harder for a decisive transformation in the digital sense, it is clear that artificial intelligence and robotics will play an even greater role and will challenge us to deal with moral dilemmas (...). I would therefore ask you to accompany me with your reflections on this at the very moment in which we are moving in that direction of development".

In order to provide a response to the query with reference to the applications in the field of health and medicine, the Italian Committee for Bioethics has set up a mixed group with the Italian Committee for Biosafety, Biotechnology and Life Sciences (ICBBSL) coordinated by Profs: Salvatore Amato, Carlo Casonato, Amedeo Cesta, Roberto Cingolani, Lorenzo d'Avack, Silvio Garattini, Laura Palazzani. The document was edited by Prof. Laura Palazzani.

The opinion, starting with a definition of AI, it analyzes its origins and most recent developments, with specific reference to the huge availability of data and computing power. The document highlights the opportunities and risks of AI and the main applications in medicine, including the context of the Covid-19 pandemic.

The Committees, in the context of the rapid evolution of these technologies, call on some elements of critical reflection for an understanding and evaluation of AI. As part of the doctor-patient relationship, they underline, on the one hand, the opportunities that can allow health professionals to reduce the time needed for bureaucratic, routine or dangerous activities, allowing them to be more available in the patient care relationship, on the other hand, they describe the risk that "automated cognitive assistance" could reduce the skills of doctors and healthcare workers. The document underlines the importance of tools that guarantee the reliability of AI, through validation, reducing, as far as possible, opacity, errors and possible discrimination due to technological and/or human causes. Given the enormous use of data, adequate protection of privacy is also essential, also considering the possibility of sharing data for "social good".

Informed consent remains an essential element of the doctor-patient relationship, despite certain difficulties, given by the informative process of the doctor and the not always simple and usual understanding of it by the patient. Particular attention is therefore also devoted to new training in the medical, technological and social fields. In this regard, the Committees believe it is essential to rethink the training of health professionals dynamically, with a flexible review of the study programs by interdisciplinary commissions, to combine the various competencies of AI in a transversal and interdisciplinary way and, at the same time, introduce the importance of ethics in the training courses of engineers, computer scientists, developers, with particular reference to ethics in the design and application of technologies. An important objective should also be the raising of public awareness within society regarding the opportunities and risks of new technologies, as well as a regulatory update on the profiles

concerning responsibility in the application of new technologies and the promotion of AI research in both the public and private sectors.

The identification of responsibility, from a legal point of view, requires an assessment of existing categories, given the pluralism of competencies between the designer, the software vendor, the owner, the user (the doctor) or third parties.

Auditions were organised with the internal experts of the two Committees: Dr. Amedeo Cesta, member of the Italian Committee for Bioethics (13 December 2019) and Prof. Roberto Cingolani, member of the Italian Committee for Biosafety, Biotechnology and Life Sciences (31 January 2020) as well as with invited external experts Dr. Alberto Tozzi, Head of Digital Medicine and Telemedicine Unit, Director of the Multifactorial Diseases and Complex Phenotypes Research Area at Bambino Gesù Paediatric Hospital (May 29, 2020); Prof. Carlos Romeo Casabona, Professor of Constitutional Law at the University of the Basque Country, Representative of Spain at the Council of Europe Bioethics Committee, Member of the European Group on Ethics in Science and New Technologies at the European Commission, Visiting Professor University of Rome Tor Vergata (13 December 2019).

The document was voted unanimously by the Italian Committee for Bioethics in the plenary session of 29 May 2020: Profs. Salvatore Amato, Luisella Battaglia, Carlo Caltagirone, Stefano Canestrari, Cinzia Caporale, Carlo Casonato, Francesco D'Agostino, Bruno Dallapiccola, Antonio Da Re, Lorenzo d'Avack, Mario De Curtis, Gianpaolo Donzelli, Silvio Garattini, Mariapia Garavaglia, Marianna Gensabella, Assunta Morresi, Laura Palazzani, Tamar Pitch, Lucio Romano, Massimo Sargiacomo, Monica Toraldo di Francia e Grazia Zuffa

Despite their not having the right to vote assent was given by: Dr. Maurizio Benato, the delegate for the President of the National Federation of MDs and Dentists Colleges, Dr. Carla Bernasconi, the delegate for the President of the National Federation of the Orders of Italian Veterinarians, Dr. Amedeo Cesta the delegate for the President of National Research Council, Dr. Paola Di Giulio the delegate for the President of the Superior Health Council, Prof. Carlo Petrini, the delegate for the President of the National Institute of Health.

Profs. Riccardo Di Segni, Luca Savarino and Lucetta Scaraffia absent from the session, subsequently assented.

It was also voted by the Italian Committee for Biosafety, Biotechnology and Life Sciences, on June 5, 2020, by Profs. Andrea Lenzi (President), Antonio Amoroso, Antonio Bergamaschi, Carlo Caltagirone, Roberto Cingolani, Fabio Fava, Paolo Gasparini, Maurizio Genuardi, Marco Gobbetti, Paola Grammatico, Mauro Magnani, Piero Angelo Morandini, Luigi Naldini, Ferdinando Nicoletti, Giuseppe Novelli, Pierfranco Pignatti, Roberta Siliquini, Paolo Visca.

1. Definitions of AI and recent developments

It is not straightforward to give a homogeneous definition of Artificial Intelligence (AI) especially in light of recent developments that have led to a widespread use of the term[1]. In general, this term identifies the sector of Information and Communication Technologies (ICT) which has the aim of imitating certain aspects of human intelligence with IT technologies, to develop "IT products or machines" able both to interact and learn from the external environment, and to make decisions with increasing degrees of autonomy.

Nevertheless, the term "intelligence", consistent with the English meaning of the word, does not designate strictly human qualities conferred on machines, but describes functions that make certain machine behaviours similar to those of human beings. The other absolutely particular aspect is that we are, for the first time, faced with technologies that systematically act as independent users of other technologies.

Developments in AI, following a period of stagnation, have gradually resumed during the last few decades[2] enabling to achieve its current significant impact. The great interest in AI has its explanations in technological progress, it has ensured that the results for years basically relegated to the field of scientific research, were attained first in industrial laboratories and from there have moved to the market.

To interpret the times we live in we must observe two aspects related to the evolution of technology: the computing power of computers, that has made possible the brilliant results of research which until only a few years ago only solved "toy instances"[3] of a problem, thanks to this power computers have now come to solve instances of real-life problems. The second aspect is linked to the increased availability of huge amounts of data[4] and information (which constitute the "examples" or basic elements for the machine), taken from ICT and the web, and to the development of algorithms[5] which have now become "executable": data and algorithms constitute the "learning" capability (machine learning), which allows the machine, on the basis of stored and archived information (data), to discover hidden relationships between data and the connection of information (algorithms)[6]. The algorithms integrate mathematical expressions, which find application in everyday problems to find associations, identify trends and identify

[1] See the Technical Report of the Joint Research Centre of the European Commission on *Defining Artificial Intelligence*: *Towards and Operational Definition and Taxonomy of AI,* 27 February 2020 (https://op.europa.eu/en/publication-detail/-/publication/6cc0f1b6-59dd-11ea-8b81-01aa75ed71a1/language-en/format-PDF/source-118380790).

[2] In classical AI one reasoned with a mathematics linked to philosophical logic, drawing deductions on the basis of causality. Today, however, the problem is that knowledge can be abstracted directly from data.

[3] One of the first was the chess game between the *Deep Blue* computer and the then world champion Kasparov.

[4] On the subject of data, see Italian Committee for Bioethics (ICB) *Information and communication technologies and big data: bioethical issues* (2016).

[5] The algorithm is a sequence of instructions that define the elementary operations that the machine must perform on the data to obtain the results. It is a systematic calculation procedure that solves a specific problem through a finite number of elementary instructions with a finite amount of data.

[6] Among other things, there are two distinct types of AI: a) weak emulation AI (*weak AI*) based on the principle that the essence of the functioning of the brain does not lie in its structure but in its performance; b) strong AI, simulation (*strong AI*) based on research to reproduce as closely as possible the physiology of the brain, possible extension of the weak one.

regularities within any set of data, at the basis of human behaviour, expressed by data and information[7].

The algorithms require a training phase starting from human-provided examples that form the basis from which to learn[8]. These two combined effects (data and algorithms) are the basis of current progress: while previously AI programs "were hand-trained" by the skill of their creators, they can now also "train themselves", with limits, and this has broken new ground. In the case of data available on the internet, automatic data acquisition can be programmed as they become available (some marketing or financial management tools use this mechanism).

Based on this "training" AI is able to predict, with varying degrees of probability. Correctness in the prediction of AI is proportional to the number and quality and accuracy of the data entered and the experiences stored on a given theme, but it could always fail should a case that has never happened before emerge. AI predictions depend on the data and algorithms with which the system is "trained": the predictions can be "wrong" due to the inaccuracy of the data provided or the use of unfounded assumptions. It is therefore necessary to take care of the "nourishment" or "training" of AI: in this context man maintains a central role. In this sense, it is not possible - to date - to speak of the "decision-making autonomy" of the machine.

This progress has affected areas of AI to different extents: many perceptual tasks (artificial vision, recognition of specific objects, interpretation of spoken and written language), to a lesser extent tasks similar to human reasoning, such as developing the ability to reason, understand intentions, elaborate arguments, as well as construct very articulated verbal speech. There is extensive discussion on the so-called "Logical capacity" of AI: what appears to be a logical-deductive process based on the typical concatenation of human reasoning (of causal association that starts from the description of reality and infers conclusions) performed by AI is, in truth, a dynamic model, based on very fast comparison and correlation with stored examples. AI, in this sense, applies a mathematical principle that highlights correlations between the data, but does not reason in a "logical" way in the proper sense. However, correlations are always probabilistic predictions and contain many limits and exceptions.

A particular area of *machine learning* is *deep learning* deriving from a process of imitation of the human brain, based on the creation of networks of artificial neurons. In *deep learning*, the machine extracts meanings by reasoning on large amounts of data: these are "automatic" (rather than "autonomous") learning methods, which, at the moment, generate "opaque" results that are not easily explained (the problem of the black box). These lead to superior performance results, which are sectorally close to human ones, albeit with some limitations.

---

[7] AI is not a new scientific-technological discovery but an area of research whose birth generally dates back to 1956 (year of the Dartmouth Summer Research Project on Artificial Intelligence to which the AI coinage certainly dates back), but which still dates back before the studies on the artificial neuron by J. McCULLOCH (*A Logical Calculus of the Ideas Immanent in Nervous Activity*, in "Bulletin of Mathematical Biophysics", 1943, 5 (4), pp. 115–133) and the studies by Alan Turing (*Computing Machinery and Intelligence,* in "Mind", 1950). The research area has gone through various periods: a first moment of vogue occurred in the eighties when there was a first marketing attempt (linked to the so-called "expert systems") which proved to be a limited and soon disappointing success. There followed a long period of slowdown in its advancements (also identified as "AI winter") basically dedicated to a return to the laboratories to resume basic studies, also strongly theoretical (A. VESPIGNANI, *L'algoritmo e l'oracolo*, Milano 2020).

[8] An equation in which a number of examples are accumulated to the left of the equal.

Much of the current fears and uncertainties concerning AI are based on the assumption of situations that are still unrealistic today, such as the "replacement" of the human decision-making capacity, or the "autonomisation" of machines that could escape human control[9]. Although theoretically possible, we are still far from this scenario. AI is a powerful tool, but it is an accessory to human decision. The problem today therefore is not so much that - feared by several parties – of misgivings regarding the "autonomy" of AI, but if anything, the problem concerns the fact that an expert system that becomes optimal in suggesting "decisions" to man risks reducing human attention with the possible consequence of reducing human skills (or deskilling)[10]. In this sense, it is important to reflect on the synergy between man and machine, and on the search for ways of intelligent "support" that allows man to have "significant human control" in terms of supervision and attention.

2. Developments in AI in the field of healthcare

2.1 Since the '70s, AI has been considered as an "emerging area" (called the area of expert systems) capable of carrying out reasoning on limited knowledge to imitate medical reasoning. Today in medicine there are already many AI applications aimed at improving healthcare practices.

Thus, AI can assist the professional in prevention and in classifying and stratifying the patient's conditions (reducing diagnostic uncertainty[11]); in understanding why and how patients develop diseases (reducing pathophysiological uncertainty); in considering which treatment will be most appropriate for them (reducing therapeutic uncertainty); in predicting whether they will recover with or without specific treatment (reducing prognostic uncertainty and increasing the prediction of the onset or evolution of pathologies), and so on.

In addition, efforts are currently being made to develop support for the doctor and health care workers[12] so that the most updated and appropriate guidelines

[9] Bear in mind that these concerns are found in several international declarations. For example, "the specific characteristics of many AI technologies, including opacity ('black box-effect'), complexity, unpredictability and 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" (European Commission, *White Paper On Artificial Intelligence - A European approach to excellence and trust*, 19.02.2020, p. 12). The UNESCO document on Robotics Ethics of 14, September, 2017 suggests a distinction between traditional deterministic computers and stochastic or probabilistic computers. The *Ethics Guidelines for Trustworthy AI*, developed by the High-Level Expert Group on Artificial Intelligence in April 2019 under the aegis of the European Union suggests the need to preserve the "*human-centric*" dimension of the new technologies. The European Group on Ethics in Science and New Technologies in the document of 20 March 2018 on *Artificial Intelligence, Robotics and 'Autonomous' Systems* also underlines the importance "that humans - and not computers and their algorithms - should ultimately remain in control, and thus be morally responsible".

[10] This is what happened in air transport: “intelligent” assistance to the pilot risks decreasing human attention and, to overcome this problem, the pilot is frequently trained on a simulator.

[11] The Dermosafe system, currently used in multiple hospitals, has a good chance of identifying critical situations and intervening promptly to stem the development of the tumour. And "*embodied*" artificial integration can give positive effects in the interaction with the patient in the prevention of degenerative diseases.

[12] In the text when referring to the doctor, where relevant, the indication also includes the reference to healthcare workers.

can be consulted, even while working on the ward[13]. Proper use of AI could also improve and make flows within healthcare facilities more efficient and smooth, from triage to emergency management or medical device selection, both in the pharmaceutical sector in relation, among other things, to the possible use of intelligent packaging for the packaging and sale of drugs. Furthermore, AI can be usefully employed in clinical trials and in the perspective of precision medicine.

2.2 As was the case with regard to the spread of Ebola, also in the Covid-19 pandemic which has hit the world since the end of 2019, AI is used to make a decisive contribution to the fight against SARS Co-V-2 virus. Since this is an epidemic and with a rapid spread on a global scale, it is essential to have equally fast tools that can be applied simultaneously in different areas. There are many uses made of AI and among these we indicate as prevalent: observation and prediction of the evolution of pandemic trends; diagnostic purposes of the pathology; search for a vaccine or a cure (AI can be of support to know the activity of drugs on metabolic chains and the structure of the receptors on which one wants to act); assistance to healthcare professionals and patients, by providing medicines and food and measuring vital signs; disinfection and decontamination; the sharing of knowledge and detection of misinformation; control and traceability of population behaviour[14].

2.3 In medicine AI, on the basis of data and analysis of data, exerts its effects in the interpretation of complex patterns. For example, in the interpretation of images, AI recognizes signals that the human eye can not distinguish. If AI has great experience of these signals, AI will perform well. There are many applications in the field of personalized medicine, in the analysis of big data in genomics, drug testing[15], in the operation of surgical robots (virtual reality, teleoperation, real-time image analysis). Robotic systems are integrated systems that refer to humans, integrating computational skills (thinking), data storage (memory), sensory systems (sight, hearing, touch) and actuators to generate physical actions (the skeletal muscle system). The synergy of these factors is applied inside the robot to operate according to the needs dictated by the operator (the doctor). The growth of AI in healthcare and in various sectors, in recent

[13] Currently, IBM is trying to make a version of Watson that is useful to the doctor; in particular, a system that provides the doctor, while he is operating, with specific information found in the literature of that area. Each month the Food and Drug Administration finds itself evaluating numerous diagnostic imaging algorithms. Also in the medical field, the FDA has proposed various guidelines and protocols to ensure *best practices* in managing these algorithms that evolve over time.

[14] See the Council of Europe document *AI and control of Covid-19 coronavirus*: https://www.coe.int/en/web/artificial-intelligence/ia-e-lotta-contro-il-coronavirus-covid-19. The initiative of the health facilities of Bergamo for new machinery is particularly significant, with the help of private funding, they have equipped themselves with robots to automate the procedures usually performed by operators, to prepare the saliva and mucus samples taken on swabs before being inserted into scanners that detect the presence of coronavirus. This automatic system is useful not only to lighten part of the lab technicians' work, which can be dedicated to other activities, but also to double the number of swabs analyzed per day.

[15] It is believed that it is possible to shorten the time spent on discovering a drug from 4, 5 years to one, with a cost cut of 80%. For example, Halicina is the first antibiotic discovered by an artificial intelligence algorithm. It is a broad-spectrum antibiotic that acts on difficult to treat bacteria, resistant to antibiotics. An anticancer, BPM31510, obtained with AI systems that sift through thousands of human tissues, has passed a phase 2 trial for patients with advanced pancreatic cancer.

decades, is also due to the growth of robotics in the civil field, used not only for domestic and recreational use (social robots), but also in the health-medical field for both diagnostic functions as well as for clinical practice in a strict sense. Also in this field there are multiple purposes: surgical support activities[16]; in the fields of diagnosis and prevention (with the aid of nanorobots); in the context of care, rehabilitation and personal assistance to the elderly, people with mobility problems, for people with autism[17].

3. Emerging ethical problems

The Committee notes the rapid evolution that is taking place in medicine and appreciates the enormous progress and extraordinary opportunities opened up by AI. The process has started and the transformation seems to be overwhelming, inevitable and irreversible.

In this evolving and transitional context, the Committee intends to recall some elements of ethical reflection, without exalting or hindering the development of technology, but rather to provide "critical" reflection for an understanding and evaluation of new technologies, in an attempt to understand how they really "work" and evaluate what is acquired in terms of potential. The goal is to identify the ethical conditions for a development of AI that does not forsake certain aspects of our humanity, in a new "digital humanism", for medicine "with" machines and not "of" machines. In the awareness that it is man who builds the technology and that technology is not a neutral tool, as it inevitably changes the doctor-patient relationship itself.

3.1 AI in the doctor-patient relationship

The impact on bioethical principles depends as much on the sectors of application of AI in the health field as on the identity of the doctor in light of this new support and on the overall function entrusted to the health service.

The use of intelligent machines and robots in medicine, insofar as they are and will be more efficient, precise, rapid and less expensive, seems desirable if we consider this replacement of man with reference to repetitive, boring, dangerous, demeaning or strenuous activities. If properly used, AI could reduce the time that professionals have to devote to merely routine bureaucratic activities, or activities which expose them to avoidable dangers, allowing them to have fewer risks and more time available for the patient.

Automation in the acquisition and interpretation of data, in the elaboration of diagnoses and in the identification of therapies or in the performing of the intervention itself cannot be completely independent of man, but requires constant verification, therefore it does not exclude the specificity of the relationship between doctor and patient. It is impossible to forget that each patient is sick "in his/her own way" and that personal contact is the essential element of every diagnosis and therapy. In this sense, the machine cannot replace the

[16] The robot-surgeon, equipped with forms of AI, carries out innovative support activities also allowing to perform the surgical operation by means of "virtual reality", with the aid of a pen connected to a tiny robot that directs the laser beam in the direction desired by the surgeon. Therefore, the doctor is immersed within the operating field and can carry out these activities remotely from the operating room.

[17] See a previous opinion of the ICB and ICBBSL, *Developments of Robotics and Roboethics,* (2017).

human being in a relationship that is built on the meeting of complementary areas of autonomy, competence and responsibility.

AI should be considered exclusively as an aid to the doctor's decisions, which remain controlled and supervised by man. It is for the doctor, 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 and therapeutic 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 doctor.

A problem arises: what happens when AI proves to perform better than the doctor? In some circumstances this is technically possible and this should be taken into account. It is in this specific space that the dreaded 'replacement' of man by machine could happen in the future.

But a further, more immediate consequence may be the delegating of decisions to technology. Delegating complex tasks to intelligent systems can lead to the loss of human and professional qualities. If the relationship of care is configured as a relationship of trust, as well as of care (Law 219/2017), the substantial role of the "human doctor" must be preserved as only the doctor possesses the skills of empathy and true understanding, which cannot be expressed by AI and it is precisely these skills alone which make such a relationship real. As suggested by some of the foremost experts on these issues four main components must be guaranteed: *Deep Phenotyping*, *Deep Learning*, *Deep Empathy* and *Connection*.

The predetermination of canons of behaviour and codes of conduct, such as protocols and guidelines, constitute support for the knowledge and experience of professional activity, but the requirements of diagnosis and care often oblige to go beyond predetermined models. It would be extremely serious if the space left to the supposed neutrality of machines led to the "neutralization" of the patient. The enormous potential offered by AI should be considered as a precious opportunity in which technique broadens the horizons of ethics, allowing to increase the patient's listening spaces and contact with the course of his/her illness. In this sense, AI would be a very useful tool that saves the time employed by the doctor in routine operations in order to gain more time for the relationship with the patient.

3.2 The reliability of AI and the opacity of algorithms

As mentioned, AI is made up of a series of algorithms: precise instructions and mathematical expressions to find associations, identify trends, extract dynamics from the data collected and entered. When the algorithms operate, they are considered ‘trustworthy and neutral' in themselves, only for the fact that their methods are represented through measurable, mathematical systems[18].

But it must be remembered that it is man (with the help of the machine) that collects and selects the data, and who builds the algorithms. In this sense, the AI system can be "opaque". "Opacity" refers to: the steps through which data are interpreted not being always explainable (transparent) and that they can also give

[18] As already highlighted, it is not always possible to reduce AI to synthetic logic such as that of an equation, precisely because AI is closer to a dynamic concept in which individual observations are compared to the atoms that make up the knowledge base.

discriminatory results. The discrimination does not come from the machine but from man who selects the data and develops the algorithms. This step implies a reflection on the "data ethics" supporting AI (which require both quality and interoperability) and on "algorithm ethics" (also called "algor-ethics"), which should be based on data that are not selected, alternatively on inclusive and non-discriminatory selections.

AI, even if it can reach a high degree of accuracy, it is not and cannot always be explainable. It is impossible for programmers and technicians themselves to explain how the system has achieved certain results (black box problem). Automation can lead to a lack of transparency on the logic followed by the machine: the machine does not provide, nor is it possible to trace, information on the correlations indicated or on the logic adopted to reach a conclusion or propose a decision (addressed to the doctor and/or patient).

The opacity surrounding the essential elements and the decision-making process by which an AI system can draw a conclusion, involve**s** the risk that health workers cannot validate and confirm, or reasonably discard, the proposal made by the system in an attempt to make their own decision. It is practically impossible for a human being to analyze the huge amount of calculations made by the algorithm and find out exactly how the machine managed to decide. This raises problems for the doctor in relation to the machine (whether or not to rely on the algorithms) and in relation to the patient, to whom the doctor cannot provide an explanation and transparent information.

Furthermore, algorithms, with the classification of people into groups or subgroups with profiles similar to those associated with certain schemes (clustering), may not take into account the variations that a particular patient may present. A care decision based exclusively on profiles elaborated on patients and in an automated way (through algorithms) can lead to the exclusion of treatment without offering in exchange an alternative, albeit presumed less effective, but nevertheless an indicated alternative. The risk arises, by classifying or stratifying patients into groups or subgroups on the basis of personal profiles obtained by them on the basis of various criteria or purposes, that discriminatory, stigmatizing or arbitrary decisions are made exclusively on the basis of these profiles or on the basis of considerations not related to healthcare (also indirectly linked, for example, to ethnic origin or gender). “Algorithmic discrimination" is possible, even in the medical field, with an impact on equity and inclusiveness. Inequities already exist in the health sector, but AI could accentuate and worsen them by creating and/or increasing the "gap" and inequalities. It is possible to avoid this drift with a broad and representative approach of useful data, continuously updated, for the development of algorithms.

Furthermore, it should not be forgotten that medical care also involves major economic interests[19], therefore AI can be oriented, through the construction of algorithms, to influence the doctor's decisions in various ways, for example by facilitating prescriptions through an increase or a decrease in normal values for a series of functional or biochemical parameters. Therefore, AI can bring to favour one class of drugs over another that has the same indications for a particular symptom or pathology. It can give preference to a diagnostic path which favours the use of certain reagents rather than others. It may suggest the use of certain more expensive equipment and technologies as an alternative to other cheaper

[19] One has only to recall that the drug market alone is worth at least 30 billion euros, with the addition of the market for diagnostic instruments, medical and rehabilitation devices.

ones. It can influence the doctor to prescribe treatments rather than stimulate the patient to improve good lifestyles[20].

These, among other risk reasons, push the Committee to believe that accurate controls must be made, also through the validation of the algorithms, in order to obtain the most probable certainty that the introduction of various forms of AI are beneficial to improve the quality of the services of the National Health Service. In other words, all the "products" of AI must be compared, through studies conducted with the rules of controlled clinical trials, with decisions that are made independently of AI by groups of competent doctors[21]. Without prejudice to the fact that controlled clinical studies remain the "gold standard" for the demonstration of the efficacy and safety of treatments, it must be borne in mind that when we talk about the application of AI in medicine, it refers to software[22]. With the problem that the mechanism changes over time and validation requires monitoring and further checks.

Only if it emerges from these studies that AI has a better performance than that of doctors, should it be accepted and used. However, the meaning of "performance" must be considered: for example, the segmentation of diagnostic images can be done with high level of quality by a doctor, but an AI system takes a fraction of time to perform the same operation and never tires.

This is particularly important for the improvement of the quality of the services of the National Health Service in the interests of citizens. It is also essential to set up facilities suitable for public interest research to take charge of the development of AI through public funds. It will therefore be necessary to demonstrate AI safety on the basis of control starting from the data base, the advantage in terms of benefits and risks, in a clinical sense, and cost-effectiveness, as well as the diffusion and sustainability of these technologies throughout the territory and over time. Only in this way will it be possible to demonstrate the reliability of these systems through certifications 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.

3.3 AI and data: between privacy and data sharing

In medicine, AI "feeds" on data: data is indispensable for the "training" of the machine and are the basic elements of the construction of algorithms, mathematical models that interpret the data. The availability of data (clinical data, images, genetic data, etc.), the accuracy and quality of the data, the interoperability of the data (through standardization and classification criteria) are the necessary conditions for the developments and applications of AI. Since every AI system is based on data, the problem of verifying, selecting, preparing and supervising data from human beings emerges, avoiding the errors of data

[20] See R. SPARROW, J. HATHERLEY, *High Hopes for "Deep Medicine"? AI, Economics and the Future of Care,* in "The Hastings Centre Report", 18 February 2020.

[21] For example, if a *learning machine* is programmed to diagnose and treat lung disease, it must be evaluated against the decisions of a group of doctors with specializations in pneumology.

[22] It is the same theme of *digital therapeutics* that is prompting towards some reflection (real world evidence) which also have regulatory implications. Beyond the methodological question, there is still uncertainty as to which regulatory body will have to deal with the issue. A discussion is underway both at European Medicines Agency (EMA) but also at Italian Medicines Agency (AIFA).

collection and classification, also providing AI mechanisms for checking and verifying correctness.

The insistence on the protection of privacy and confidentiality is often pointed out as being an obstacle to the development of AI. Those who intend to apply AI insist on the need to dispose of data in a broad field of action, on a global level (therefore also with transfer of data to other countries) and storage of data, but also storage of samples and associated images, over time. Data not fully anonymous but pseudonymised that allow traceability, identification in cases of importance of communication of the results, with appropriate conditions to prevent improper disclosures.

The huge collection of data, necessary for AI, also highlights the risk, related to the use of data and the crossing of data, of both intentional and accidental re-identification, raising the problem of privacy, which in this context tends to "vanish"[23]. To the point that it is believed that technologies are becoming increasingly "opaque" and the users "transparent".

In the AI era and the need for the use of data for medical research, questions arise with regard to the possibility of "sharing" data (*data sharing*) as a "social good" for the advancement of scientific knowledge.

There are methods and technologies for performing data transactions while preserving data security (one of the technologies is the family of *block-chain* applications).

This sharing is, in any case, guaranteed by the exclusive use for research purposes, which enables a return of information and sharing of clinically relevant results (*benefit sharing*).

There is wide debate, even on a regulatory level, of the applicability of the General Data Protection Regulation (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.

3.4 Consent and autonomy

The favourable efficacy/risk ratio and the requirement of informed consent and autonomy are fundamental rules in the doctor-patient/healthcare worker relationship, as they protect the right to life, health, dignity of the person, self-determination.

It follows that this must be central even if the doctor wishes to make use of the data collected by AI and the development of robotics in the healthcare treatment. But the informative process is far from easy to implement and autonomy/consent is complicated by AI which arouses a sense of disorientation given the speed with which technologies are radically changing the known world. It is not easy for the patient-person to imagine the consequences that could arise from these new technological advances: It is the doctor who must act as mediator in this communication. Complex terminologies, words that may sound

[23] On the subject of privacy, the ICB intervened in the document *Information and communication technologies and big data: bioethical issues* (2016) underlining that as part of the "data processing 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 apps: 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".

mysterious, are found in the new healthcare procedures (*machine learning, deep learning, neural networks, big data, algorithms, cloud, etc.*) making the consent to new healthcare treatments increasingly complex and given perhaps more through trusting the doctor than on actual understanding. Informed consent to AI-based healthcare treatments may impact patient autonomy. Certainly the patient sees the traditional relationship of alliance with the doctor change: still having very confused ideas about the applications of AI, the patient appreciates its advantages, but does not fully understand its risks.

It is therefore an ethical and legal obligation that those who undergo such innovative health treatments, through AI, are informed in the most appropriate and comprehensible way for the patient as to what is happening, to be (if necessary) subject to experimentation and validation; to be aware that what is applied to them (on a diagnostic and therapeutic level) implies advantages, but also risks. It must be explicitly specified in the informed consent if the applied treatments (be they diagnostic or therapeutic) are only from a machine (AI, robot) or if and what the areas and limits are in human supervision or control over the machine. These difficulties in the providing of understandable and exhaustive information, given by the doctor to the patient (difficulties regarding both the doctor's communication and the patient's reception) when employing treatments that make use of AI are augmented by the opacity of the algorithms.

3.5 Responsibility

Automation in medicine can contribute to the reduction of accidents and mortality (increase attention and the accuracy of the doctor's actions, enhancing its use even in routine procedures, etc.), but, as mentioned above, it is not without its risks. Machines can be poorly planned and poorly employed. Therefore, the issue of liability is one of the most delicate and complex problems that arise with the use and development of new AI systems. In particular, the problem receives increasing attention in terms of policy and legislative strategy. The attempt is above all to clarify whether accountability for certain decisions made through an intelligent system should be attributed to the designer, the software vendor, the owner, the user (the doctor) or third parties. The possible occurrence of accidents should be traced and analyzed as is the case for any medical error.

Any evolution that in the medical field ends up changing the doctor-patient relationship, should provide for an intervention of the law that governs innovations in accordance with the existing "system", thus creating guarantees both for the patient and the covering of new risks, and at the same time for the work of the doctor. The factor that poses new requests for legal mediation is not so much the presence of an intelligence with the possibility of self-learning, but the fact that AI has an "author" who creates it and who may not coincide with the "producer" of the good that incorporates it, the "seller" and the user and for whom the problem arises of outlining rights, limits and responsibilities. In many cases, command of action remains with the doctor, who is however not directly the agent of the action, being in fact sometimes in a different distance place. The fact that there is a "chain of command" to which the responsibility of the action can be traced may suggest that the action is less subject to chance and improvisation, but each link in the chain has its fragile point and given the complexity of the gestational structure and of the action it is not obvious to say who in the end is responsible for what eventually happens to the detriment of the patient or if this responsibility is unique. These autonomous and distinct responsibilities should be able to be

directly asserted by the end user of that product, the patient, not only through the traditional contractual plan that binds the doctor and the healthcare facility. In addition to these responsibilities there are also those differentiated in the context of the relationships between professionals (designer, validator, software vendor, programmer, etc.), who have contributed to the formation of the doctor-patient chain, without there being a prior proposal, a prior formal act, so that it is extremely difficult to recognise a contract equally. We are in a logic that brings us closer to the category, developed by jurisprudence, of "social contact responsibility", which hypothesises an obligation that is linked to the duty of diligence in observing the rules of art that it professes.

These are all aspects in which the analysis of new responsibilities and conceivable new rules, as well as new evolutionary interpretations of existing rules, make collaboration between legal and medical sciences indispensable, since the former have to deal with the latter and vice versa. An interdisciplinary continuous reflection that sees the two competencies "talking to one another" is both opportune and indeed necessary, also in order to outline the future structure of possible multiple medical responsibilities connected with AI.

3.6 Medical, technological and social training

Today, the medical world and healthcare professionals are not fully trained, with few exceptions, to use the results of AI research responsibly. It is therefore very important to act on two fronts: on the one hand to insert the problems deriving from AI in the activities of Continuing Medical Education (ECM) carried out independently and on the other to undertake a reform of Medical Schools, as well as the schools of healthcare workers.

The inclusion of AI in the education of doctors and health professionals[24] falls under the so-called *reskilling* of employees, i.e. reconverting workers (in this case healthcare workers) in the face of developments in emerging technologies. However, training salaried healthcare workers to occupy the same positions, but which imply new needs and professionalism, will be more complicated and expensive than creating new jobs for people already trained in understanding AI in the medical field[25]. 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 "*skills polarization*" that can hide new forms of discrimination, excluding those who are unable to learn the new required "skills". The problem of new professions, even in the medical field, remains therefore that high-level skills will be required. This discussion falls within the field of new diagnoses and therapies; continuous updating is essential for doctors and healthcare professionals.

The other path is the reformulation of medical education programs, allocating a significant part of the training of future doctors to the problems deriving from the digitalisation of medicine which is the basis of the AI technologies that future doctors will have to take advantage of, being able to understand its advantages, limits and dangers. The institutionalisation of interdisciplinary courses for the

[24] In recent years Europe seems to have become increasingly aware of the importance of the problem, just as several committees are beginning to take an interest in it, outlining the orientations of ethical and legal reflection in the field of AI.

[25] It follows that despite the emergence of many new professions, we may be witnesses to the formation, as already suggested by several parties, of a class of unemployed and "useless" individuals.

training of health professionals to a constant adaptation to technological change and to the possible "convergence" and transversality of traditional disciplinary sectors (e.g. medicine and computer science or physics or data science with foundations of computer science and AI, components of clinical ethics, bioethics and biolaw) is desirable.

Training must also be renewed in the field of technology, introducing ethics and bioethics training courses for engineers, computer technicians, computer scientists, with particular reference to ethics in the design of technologies (*ethics by design/in design/for designers*) and in the planning, methodology and application of technologies. This is the only way to ensure the ethical awareness of those building the technologies, in order to allow principles and values to be present from the beginning of the technological design.

It is also desirable to promote public debate on the developments and limits of AI in medicine, so that all citizens can acquire the basics of "*AI literacy*", active participation in social discussion. In the long term, it is hoped that the introduction of science as an essential part of culture in schools can lay the foundations for an understanding of the presence of AI within various sectors. These are the prerequisites for a possible overcoming of the "digital divide" in medicine, promoting greater inclusiveness.

4. Recommendations

In the light of the previous analysis, the ICB intends to recall some ethical principles of reference in the context of the use of AI in medicine. In the face of the progress in this area considered as "transformative" and "disruptive", especially in the field of health protection, the Committee intends to promote ethical reflection in balancing the human dimension and the artificial dimension, without mutual exclusion. In the belief that: exclusion of the artificial takes away many opportunities for man; exclusion of the human raises many critical issues given the limits of the artificial. We must avoid excessive hopes, but also excessive fears, adopting an attitude of trust and caution.

Committee recommendations:

- prepare *ex ante* accurate controls for the "training" of machines on the basis of quality data, that are updated and interoperable and conduct adequate experiments in the context of AI to guarantee safety and efficacy in the use of these new technologies as well as encouraging 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; 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;
- in the context of the doctor-patient relationship, informing patients in the correct way, especially during this transition period, regarding the risks and benefits of using AI with reference to specific applications (and also of the limits of explainability of "opaque" technologies), in order to ensure full awareness of the choices and also assuring alternative paths to the extent that resistance to accept the new technologies emerges; guarantee, in the applications of AI for health, a broad and representative (non-selective and discriminating) approach and an area of "significant human control" of human-machine interaction and collaboration, to protect overall correctness and patient-doctor communication as a field of care;

- rethink the training of health professionals in a dynamic way with a flexible review of the study programs by interdisciplinary commissions, for constant adaptation to technological change, also thinking about the possible "convergence" of paths in traditional disciplinary sectors (e.g. in the faculty of medicine, medicine and computer science or physics or data science and symmetrically, in the faculty of law/human sciences with fundamentals of computer science and AI);
- introduce the importance of the ethical principles of autonomy, responsibility, transparency, justice in the codes of conduct and the training courses of engineers, computer scientists, developers, with particular reference to ethics in the design of technologies (*ethics by design/in design/for designers*), ensuring technology that is oriented towards incorporating values and ensuring the centrality of the patient;
- create public awareness in society regarding the opportunities and risks of new technologies, so that citizens can participate critically in the debate on AI, without blind trust and not even an excess of concern, being aware of the choices and implications of digital healthcare: such promotion can also take place through organising conferences for schools and meetings with citizens, which the ICB regularly proposes;
- request, on a regulatory level, an update on the profiles concerning responsibility in the application of new technologies;
- promote research on AI, not only in the private sector, but also and above all in the public sphere of the National Health Service (NHS).