COMITÉ CONSULTATIF NATIONAL D'ÉTHIQUE
POUR LES SCIENCES DE LA VIE ET DE LA SANTÉ

COMITÉ NATIONAL PILOTE
D'ÉTHIQUE DU NUMÉRIQUE

sous l'égide du
COMITÉ CONSULTATIF NATIONAL D'ÉTHIQUE
POUR LES SCIENCES DE LA VIE ET DE LA SANTÉ

JOINT OPINION
OPINION No.141 CCNE
OPINION NO.4 CNPEN

Medical Diagnosis and Artificial Intelligence: Ethical Issues

Opinion adopted on 23 November 2022 by unanimous vote of the members present at the CNPEN plenary session

Opinion adopted unanimously by the members present at the CCNE plenary meeting on 24 November 2022

How to quote this Opinion:
Medical Diagnosis and Artificial Intelligence: Ethical Issues. Joint opinion of the CCNE and CNPEN, CCNE Opinion 141, CNPEN Opinion 4. November 2022.

TABLE OF CONTENTS

SUMMARY

This Opinion follows a referral from the Prime Minister to the National Digital Ethics Steering Council (CNPEN) on the ethical issues of the use of artificial intelligence (AI) in the field of medical diagnoses, expressed in his mission letter of 15 July 2019. Given the cross-cutting nature of this issue, which is relevant to both bioethics and digital ethics, the National Advisory Ethics Council for Health and Life Sciences (CCNE) and the CNPEN have jointly conducted the reflection that led to this joint Opinion.

The technologies developed around AI concern many fields of application (medicine, transport, cybersecurity, commerce, industry, etc.) and their irruption into our daily lives is accelerating at a steady pace. This context encourages many States and institutions to consider the ethical issues that go hand in hand with this transformation.

The healthcare sector appears to be particularly concerned with the development of artificial intelligence systems. Their application to the medical field is transforming the relationship between doctors and patients and raising many questions about the future of health systems. In accordance with the ministerial referral, this Opinion focuses on the ethical issues of artificial intelligence systems applied to medical diagnosis (AISMD) aimed at improving the technical performance of practitioners.

The CCNE and the CNPEN have built their reflection by determining, firstly, what the use of artificial intelligence systems applied to medical diagnoses refers to by identifying the main foundations of the application of these new techniques (part 1) and their operational scope (part 2). The Opinion then questions the regulatory process for AISMD and its current characteristics (part 3). Finally, an analysis of the impact of AISMD on the diagnostic approach describes the current issues, and the changes they imply (part 4).

In the course of the sixteen recommendations and seven warning points identified by the CCNE and the CNPEN, several areas of ethical tension emerged. First of all, we felt it was important to draw up an overview of what AISMDS are actually capable of doing today. There are many promises, and it is sometimes difficult to distinguish them from the facts. This distinction seemed to us to be a first ethical task. Secondly, we recalled that ADMIS produce results based on probabilistic approaches on the one hand and on the other hand that they can be subject to errors. We emphasised that **health care teams and patients should not deprive themselves of the benefits of these tools, while constantly giving themselves the means to distance themselves from the results provided**. We presented the main resources that allow this perspective to be taken. Any AISMD must be subject to human control. Its results must be explainable. The control of the conformity of an AISMD, which ensures that it is not harmful and thus authorises its marketing, must be improved and, above all, must in the future be accompanied by an evaluation of its clinical effectiveness, showing not only that it is not harmful but also that it contributes effectively to the principle of beneficence.

Artificial intelligence systems applied to medical diagnosis must therefore always be used **first and foremost with a view to improving care**, before organisational, economic or managerial interests.

INTRODUCTION

In his letter of 15 July 2019, the Prime Minister asked the National Digital Ethics Steering Council (CNPEN) to issue an Opinion on the use of artificial intelligence systems in medical diagnosis. Given the cross-cutting nature of this issue, which is relevant to both bioethics and digital ethics, the National Advisory Ethics Council for Health and Life Sciences (CCNE) and the CNPEN have jointly formulated this Opinion.

The notion of artificial intelligence (AI) covers a plurality of techniques based on distinct approaches and pursuing different goals. The debates surrounding the very definition of AI reveal the issues it raises, particularly in legal and industrial terms[1] . The definition of an artificial intelligence system (AIS) used in this Opinion corresponds to the framework proposed by the European Commission, as shall be seen in the first part.

The technologies developed around AI concern many fields of application (medicine, transport, cybersecurity, commerce, industry, etc.) and their irruption into our daily lives is accelerating at a steady pace. Entire areas of human activity are being transformed by the rise of artificial intelligence technologies, opening up horizons that seemed unattainable until now. These upheavals are prompting states to consider AI as a new strategic tool, even though there is no appropriate legal framework to address the complex governance issues raised by AI technologies. This context encourages many States and institutions to propose better regulation in this area and to consider the ethical issues that accompany this transformation.

In 2018, during his presentation of the *AI for Humanity* programme[2] setting out France's strategy for artificial intelligence, the President of the Republic, Emmanuel Macron, announced that he wanted to create "a regulatory and financial environment favourable to the emergence of AI champions"[3] . On this occasion, he designated AI as a strategic tool for national and European economic and industrial development that is called upon to stand out against the great powers, in particular the United States and China[4] . Among the commitments of the French government is the implementation of a "reflection on the ethical issues related to the development of AI technologies", indicating in particular the explicability of AI technologies as one of the factors of their social acceptability.

The desire to develop the conditions for an ethical reflection on artificial intelligence takes place in an international context of increasing debate on the subject. Thus, to date, nearly one hundred institutions have formulated recommendations and guidelines to feed the public debate and guide political decision-makers in terms of AI governance. Among the most prominent positions are the "23 Asilomar Principles", signed in 2017 by hundreds of

[1] There is no commonly accepted definition of AI today, and many institutions have developed their own vision. See: Samoili S. et *al*, (2021), *AI watch defining artificial intelligence 2.0: towards an operational definition and taxonomy for the AI landscape*, JRC technical report, 125 p.
[2] See the dedicated website: https://www.aiforhumanity.fr/.
[3] "Artificial intelligence: making France a champion", 29 March 2018, see the French government website: https://www.gouvernement.fr/argumentaire/intelligence-artificielle-faire-de-la-france-un-champion.
[4] *Ibid*, "This strategy is also part of Europe, as the European Union has the vocation to become a global champion of AI and to carry a singular voice in this field, distinct from that of the United States or China."

scientists specialised in AI and robotics, and echoing the Asilomar Conference organised in 1975 by Paul Berg calling for a moratorium on genetic manipulation[5] .

The health and medical sectors are particularly concerned by the development of artificial intelligence systems. Their application to the medical field suggests a transformation of the relationship between doctors and patients and raises many questions about the future of health systems. AIS in the health field are indeed accompanied by heterogeneous and sometimes contradictory issues, between economic and industrial interests (design of AIS tools), therapeutic promises for patients (expected health benefits, screening, prevention and treatment), changes in the professional practice of the health personnel concerned (activity of the carers who may have to use them, diagnostic assistance, possibly even job cuts or changes) and regulatory objectives for the public authorities. This field requires particular attention because of the values it mobilises, which are, first and foremost, the respect for human life, and the guarantee of fair conditions of access to care.

One of the main tensions raised by the implementation of AI technologies in health is the imbalance between knowledge, public debate and the potential speed of practices transformation brought about by innovative AI-based techniques. There is often a large gap between the promises of AI technologies, the public's perception of artificial intelligence in health, and the reality of what AIS applied to medicine *can* and *cannot* do. The failure of the Watson Health AI programme, which IBM announced at the time of its creation in 2015 as promising unprecedented diagnostic assistance thanks to artificial intelligence, thus invites us to be cautious about announcements raising the spectre of the replacement of humans, and in this case medical professionals, by the machine[6] . On the other hand, certain successes, particularly in medical imaging, which we discuss below, show that these techniques can be useful and explain the enthusiasm they may inspire.

The specific temporality of AIS has two important consequences that are part of a particular context and state of science. Given the rapidity of the transformations observed in the field we are dealing with here, this contextualisation, which usually goes without saying, deserves to be recalled. On the other hand, the very large number of more or less realistic promises that mark the field of AI (to which we will return) forces us, here even more than elsewhere, to make a precise, documented and cold assessment of science. This is an integral part of an ethical reflection insofar as what AI really is and does today is far from clear to many actors. We have therefore decided to give a particularly important place to fact-finding, which is in itself an ethical clarification.

On the other hand, the use of AI in health is such a vast field that it would be perilous to attempt to cover it all. In accordance with the ministerial referral, we have restricted our analysis to the case of artificial intelligence systems applied to medical diagnosis (AISMD) aimed at improving the technical performance of practitioners. In other words, we have not

[5] See for example: Orsini A., *The "23 principles of Asilomar" want to frame the development of artificial intelligence*, Numerama.com, 01/02/2017. We can also mention the UNESCO recommendation on the ethics of AI https://unesdoc.unesco.org/ark:/48223/pf0000381137_fre and the text of the EU *High-Level Expert Group on AI* https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

[6] This division of the computer company IBM, which grew out of the Watson artificial intelligence system that won the American TV game show *Jeopardy* in 2011, promised to be able to diagnose certain cancers more reliably than doctors by analysing large amounts of health data. In the end, the programme did not achieve its ambitious goals and was finally sold in 2022 to a private equity company.

specifically addressed the issue of prognosis, prevention or treatment, except where they are part of the construction of a diagnosis.

AISMD alone raise many ethical issues. From the very first stages of its development (data collection, learning from these data, construction and then evaluation of the algorithm), the design of an AISMD implies that vigilance is exercised with regard to the purposes of the system and the use to which it will guide the physician. The development of the notion of ***ethics by design***, which echoes the ethical tensions involved in the design of a system based on digital technologies, is perfectly suited to AISMD, whose medical field of application involves particularly sensitive subjects. By combining an ethics of intention centred on the subject, an ethics by design centred on the object, and an ethics of mediation centred on the subject-object relationship[7] , the ethical approach to the design of an AISMD raises a number of questions about the chain of responsibilities (of the designer, the evaluator, the certifier, the end user and, more generally, the political decision-maker). *Ethics by design*, through the idea of accountability of system designers that it develops, extends to the purpose of a system involving, for example, the favouring of algorithms that favour the interests of the greatest number over minority interests, or the strict **protection of personal data.**

The effects of an AISMD, both for the practitioner who will use it and for the patient who will benefit from it, and the repercussions of the use of these new technologies on society as a whole, must be examined in the light of the methods aimed at guaranteeing the relationship of **trust** between patients and doctors[8] and **equitable access** to quality care for all. In addition, methods must be developed to guarantee the explicability, which often remains to be clarified, of the results provided by an AISMD at each stage of its dissemination.

An approach to the ethical issues introduced by the development of AISMDs must therefore take into account the plurality of actors involved, the different stages in the process of disseminating these systems and the social implications they entail. Furthermore, our reflection is informed by the work of the European Union on this issue, which may themselves be influenced by national ethics committees.

The CCNE and the CNPEN have built their reflection by determining, firstly, what the use of artificial intelligence systems applied to medical diagnosis refers to by identifying the main foundations of the application of these new techniques (part 1) and their operational scope (part 2). The Opinion then questions the regulatory process for AISMDs and its current characteristics (part 3). Finally, an analysis of the impact of AISMDs on the diagnostic approach describes the current issues and the changes they imply (part 4).

This Opinion sets out the **ethical implications that** motivate the **recommendations** (16) and recommendations for the use of AISMDs that respect patients, the relationship between doctors and users of the healthcare system and our democratic society. These reflections are enriched by **warning points** (7) intended to alert to the main ethical tensions that may arise during the dissemination of an AISMD.

---

[7] See: Fischer F., (2019), 'The ethics by design of digital: genealogy of a concept', *Design Sciences*, 2019/2 (No. 10), pp. 61-67.

[8] B. Mittelstadt, The impact of artificial intelligence on patient-physician relations, Report submitted to the Council of Europe. 2022. https://rm.coe.int/inf-2022-5-report-impact-of-ai-on-doctor-patient-relations-f/1680a6885a

I. ARTIFICIAL INTELLIGENCE SYSTEMS AND DIAGNOSTIC ASSISTANCE: WHAT ARE WE TALKING ABOUT?

1.1 Medical diagnosis assisted by artificial intelligence systems

Medical diagnosis as defined by the National Academy of Medicine "consists in recognising diseases by their symptoms and signs, and in distinguishing them from each other. The primary objective of diagnosis is the appropriate management of the patient. It is an essential element of the medical decision"[9] .

Making a diagnosis is a process that requires a series of steps, each of which provides information (clinical, biological, radiological, etc.), leading to an assessment of a patient's health condition. The accuracy of the identification and characterisation of a given pathology or condition is a decisive element of the diagnostic approach as it determines the therapeutic approach and therefore implies clinical consequences.

As the National Academy of Medicine points out, while new technologies and methods are regularly proposed to refine the final diagnosis and make it safer and more accurate, "none of the new procedures should compromise the primacy of the clinic. The dissemination of a new diagnostic technique should never be recommended without proper verification of its reproducibility, its diagnostic value compared to existing means, its safety for the patient and its economic consequences"[10] .

According to Bonnet and Atlan[11], in the context of a medical examination, two modes of thought are combined to develop a diagnosis. On the one hand, a way of thinking that can be described as **"global"**, which is based on the medical art, scientific expertise (in particular *Evidence-Based Medicine*), experience, intuition, prudence, collegial discussion and observation of the specificities of a situation. On the other hand, a way of thinking that we will describe here as **algorithmic**, based on a "step by step" or probabilistic method. However, this type of thinking ignores the contextual specificities of the case under consideration, which can only be identified through prudence, in the Aristotelian sense of the word.

The first artificial intelligence systems applied to medical diagnosis were developed in the 1980s. Their functioning by systematic modelling and reaching a conclusion after a series of logical inference steps, or by learning using artificial neural networks, allowed them to reproduce some of the human algorithmic reasoning.

The new AI technologies are profoundly renewing the potential of this "step-by-step" thinking. Applying this thinking to final diagnosis therefore implies finding new, positive and

[9] Académie nationale de médecine, "Le diagnostic en médecine: histoire, mise en œuvre présente, perspectives", Séance du 20 juin 2006 [https://www.academie-medecine.fr/06-12-le-diagnostic-en-medecine-histoire-mise-en-uvre-presente-perspectives/]

[10] National Academy of Medicine, *Ibid*.

[11] Bonnet J.-L., Atlan G., (1980), "L'informatique médicale", In *Médecine Pathologie générale, 1: Présentation des sciences de base*, edited by Pierre de Graciansky and Henri Péquignot, 578-600. Encyclopédie de la Pléiade 45. Paris: Gallimard.

productive, and not destructive, combinations between these two modes of reasoning (so-called "global" thinking and algorithmic thinking) which are articulated throughout the diagnostic process. **A large number of ethical tensions arise not from the limits of these two modes of reasoning but from the difficulties encountered in articulating them appropriately via an AI system**.

This attempt to model algorithmic thinking can in turn lead to a better understanding of the specificities and power of global thinking. Over-reliance on AI in the final decision runs the risk of missing the specificities of each medical case; conversely, to do without AI is to refuse to exploit an additional, perhaps decisive, source of information offered by the new technologies. But, however this is articulated, the use of AI technologies can and should enrich, not impoverish, the way in which carers will use their professional skills in the diagnostic process.

1.2 Artificial intelligence systems: what are the current definitions?

The term "artificial intelligence", which is omnipresent in public debate, carries with it representations, promises and concerns, particularly in the field of health. It is commonly used to designate machine learning methods, i.e. approaches based on algorithms which, based on data, build models (decision trees, neural networks, etc.) which are then applied to new data to calculate a decision based on them. *Strictly speaking*, machine learning is a subset of the vast field of study covered by AI and more generally by digital technology[12] . These systems, which have developed very rapidly in recent years, thanks in particular to an increased capacity to collect, store and analyse huge amounts of data, can be designed to evolve over time, depending on the data provided to them during their use.

The polysemy of the notion of artificial intelligence, due in part to the constant evolution of algorithmic techniques, complicates attempts to define it, particularly in legal matters.

In France, the reference text in terms of AISMDS is Article 17 of **Law No. 2021-1017 of 2 August 2021 on bioethics**[13] . The latter introduces a legal framework for the use of medical devices incorporating a particular artificial intelligence system by inserting into the fourth part of the public health code (devoted to the "health professions") Article L. 4001-3 concerning the use, by a health professional and in the context of "a preventive, diagnostic or therapeutic act" of a "medical device comprising algorithmic data processing learned from massive data". The article also mentions crucial points such as **informing** health professionals about the use of data processing, the **accessibility of** the data for patients and the **results** obtained from it, and **ensuring that the** developers **can explain**[14] how the algorithmic processing works to users.

At the EU level, the draft Common Regulation on AI (or *AI Act*) published on 21 April 2021 by the European Commission[15] , mentioned in the introduction, refers to the need for a "clear definition of high-risk AI applications" in the face of "the new opportunities and challenges offered by artificial intelligence" and proposes a more precise definition of what is an "artificial intelligence system" applied to medical diagnosis (AISMD). In its Article 3,

[12] CNPEN, *Manifesto for a Digital Ethic*, Paris, April 2021, p.5.
[13] Available at the following link: https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043884384/
[14] This term will be explained in 3.1.2.
[15] Available at the following link: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0020.02/DOC_1&format=PDF

the Commission defines an AISMD as a **software "that can, for a given set of human-defined objectives, generate results such as content, predictions, recommendations or decisions influencing the environments with which it interacts"**[16] and includes the following techniques and approaches:

(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods, including deep learning.
(b) Logic and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deduction engines, (symbolic) reasoning and expert systems.
(c) Statistical approaches, Bayesian estimation, search and optimisation methods.

The term "artificial intelligence systems" applied to medical diagnosis, or AISMD, will be used here in the sense determined by this version of the EU Regulation.

1.3 Contextualisation and approach of the Opinion

Previous contributions, on which the CCNE[17] and the CNPEN[18] are relying in this work, have in particular contributed to the reflection on the construction of a framework relating to the use of AI systems in the field of care. The concept of "human guarantee" developed therein was recently taken up by the World Health Organisation (WHO)[19] . The terminology of "human guarantee" can also be found in the notion of **"human control"**, in the sense of human vigilance. This concept has been favoured in the drafting of this Opinion because it corresponds to advances in European law. The European Commission's draft regulation on artificial intelligence (*AI Act* mentioned above), in its Article 14, recognises the principle of *human oversight* that must be ensured over AI systems used to establish a medical diagnosis, a principle that the French translation by the Commission refers to as "contrôle humain".

[16]European Commission, "Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence legislation) and amending certain Union legislation", Brussels, 21 April 2021 [available at: https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:52021PC0206&from=EN].
[17] (1) Chapter "Digital and health" of CCNE Opinion No. 129 *Contribution of the National Advisory Ethics Council to the review of the 2018-2019 bioethics law,* 18 September 2018, pp. 94-106.
(2) CERNA/CCNE report, *Numérique et santé: quels enjeux éthiques pour quelles régulations*, 19 November 2018, 102 p.
(3) CCNE Opinion No. 130 of 29 May 2019, *Massive data and health: state of play, prospects and new ethical issues*, 94 p.
[18] CNPEN, Watch bulletin n° 3 of 21 July 2020, *Ethical issues related to digital tools in telemedicine and telecare in the context of Covid-19*, 21 p.
[19] World Health Organization (WHO), *Ethics and Governance of Artificial Intelligence in Health*, 28 June 2021, p. 13.

English version of Article 14 of the draft IA Act

Article 14
Human oversight

1. High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use.

2. Human oversight shall aim at preventing or minimising the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter.

3. Human oversight shall be ensured through either one or all of the following measures:

(a) identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service;

(b)identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the user.

4. The measures referred to in paragraph 3 shall enable the individuals to whom human oversight is assigned to do the following, as appropriate to the circumstances:

(a)fully understand the capacities and limitations of the high-risk AI system and be able to duly monitor its operation, so that signs of anomalies, dysfunctions and unexpected performance can be detected and addressed as soon as possible;

(b)remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system ('automation bias'), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;

(c)be able to correctly interpret the high-risk AI system's output, taking into account in particular the characteristics of the system and the interpretation tools and methods available;

(d)be able to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system;

(e)be able to intervene on the operation of the high-risk AI system or interrupt the system through a "stop" button or a similar procedure.

5. For high-risk AI systems referred to in point 1(a) of Annex III, the measures referred to in paragraph 3 shall be such as to ensure that, in addition, no action or decision is taken by the user on the basis of the identification resulting from the system unless this has been verified and confirmed by at least two natural persons.

Furthermore, it should be noted that the CNPEN has pointed out that the so-called "human guarantee college" methodologies involving health professionals and patient representatives would ensure this human control at the AI application stage[20].

This Opinion is based more generally on the founding texts in the field of biomedical ethics, such as the Nuremberg Code, the Helsinki Declaration, the Belmont Report and the Universal Declaration on Bioethics and Human Rights. This long tradition of biomedical ethical reflection has inspired and sometimes even influenced the way in which ethical reflection centred on the digital world has been constructed much more recently. These two fields of ethics share a common concern, which is at the root of the ethical approach, to preserve human dignity. This Opinion on the ethical issues linked to the emergence and dissemination of artificial intelligence systems applied to medical diagnosis invites us to examine in greater depth the links between these two areas of ethical reflection. This text ensures that the ethical frameworks used in bioethics and digital ethics are applicable to artificial intelligence.

AISMDs can be used at several stages of a patient's care:

- during the initial referral of patients, **prior to the medical consultation**, with a view to prioritisation, in particular by making it possible to identify situations requiring urgent care; these may be devices proposed to the patients themselves, or used in the emergency services by health professionals[21] ;
- in the **diagnostic stages**, especially those involving imaging;
- during **medical monitoring**, if monitoring devices are used to adjust treatment or detect a change in the patient's condition.

It should be noted that artificial intelligence is now also used in the context of screening campaigns in the general population, for example for breast cancer. In two years, publications on the subject have multiplied with an exponential dynamic and a broadening of the offer available to radiologists[22] . The same applies to the development of these techniques in the context of medical biology[23] . Nevertheless, we shall set aside the ethical reflection on these specific uses of AI for later work, choosing to focus, in this Opinion, on the clinical relationship between the patient and their doctor.

Furthermore, a distinction must be made between AISMDs used in the context of a care programme, and which have been subjected to medical certification, and their many other applications. The latter, which are on the borderline between well-being, recreational use

[20] Opinion issued in July 2020 by the CNPEN in response to the European Commission's White Paper on AI (European Commission, (2020), "White Paper. Artificial Intelligence. A European approach based on excellence and trust", 31 p.).
[21] Other issues in the care pathway could mobilise the CCNE's recommendations, but this Opinion focuses on the issues of AISMDs in the diagnostic framework.
[22] Thomassin-Naggara I., Ceugnart, L., Tardivon L., Verzaux L., Balleyguier C. et al, "Artificial Intelligence: Place in Breast Cancer Screening in France," *Cancer Bulletin* 109, nº 7 (July 1, 2022): 78085-. https://doi.org/10.1016/j.bulcan.2022.04.008.
[23] https://lapbm.org/intelligence-artificielle-et-biologie-medicale/.

or monitoring of health parameters, benefit from a CE mark ("Conformité Européenne") but do not follow the certification process for medical devices.

Most of the approaches presented in the text are still at the experimental stage. However, although their transfer to clinical practice is not yet a reality, it could become one in the near future, and it is therefore appropriate to consider the ethical questions that the use of the AISMDs presented may raise[24] . In particular, we will discuss the main obstacles to their implementation, whether of a technical or of a regulatory nature, with a particular emphasis on the tensions and ethical issues encountered.

[24] Keane P.A., Topol E., (2018), "With an eye to AI and autonomous diagnosis", *npj Digital Medicine*,1, 40.; He J., et *al*., (2019), "The practical implementation of artificial intelligence technologies in medicine", *Nature Medicine*, 25:3.0-36.

II. EXAMPLES OF ARTIFICIAL INTELLIGENCE SYSTEMS FOR MEDICAL DIAGNOSIS: BENEFITS AND ETHICAL ISSUES

Image recognition is particularly well suited to artificial intelligence systems based on machine learning. These systems represent the logical evolution, in the field of imaging, of the computer-aided detection software used by doctors and particularly radiologists for the last ten years. The role of these alert devices - whose detection rules are predetermined by humans based on their knowledge and experience - is to improve diagnostic performance and to draw the radiologist's attention to possible abnormalities. However, there is no consensus on their clinical usefulness, as shall be seen.

An important improvement is made by machine learning on a considerable amount of digitised radiological data. Algorithms can be trained in two ways:

- In the case of supervised learning, the **image recognition** algorithm **uses data annotated by a person**, i.e. the raw data is associated with *a priori* information on the image content;

- In contrast, some unsupervised approaches do not require prior labelling of image features, and the relevant criteria for reaching an outcome **are determined by the algorithm itself without human intervention after its design**.

However, this increases the **risk of opacity in the process of obtaining the result,** which shall be discussed below.

2.1 Learning systems for image analysis

The use of computer-aided detection softwares in the field of imaging has grown over the past decade with the advent of deep learning approaches[25]. The latter are particularly suited to the classification and segmentation of images from radiological examinations such as CT scans, MRI (Magnetic Resonance Imaging), conventional X-rays, ultrasounds, retinal examinations, skin examinations, and the analysis of histological images.

The performance of these learning systems for detection and diagnosis is superior to that of previously available tools. At the moment, only systems with algorithms that cannot be continuously learned are authorised by French regulations, but the emergence of softwares that automatically "learn" offers the prospect of significant developments in this sector, while raising specific ethical questions[26] .

[25] A deep learning system uses hidden layers of artificial neural networks. Deep learning can be either supervised or unsupervised.
[26] Bitterman D.S. et *al*, (2020), "Approaching autonomy in medical artificial intelligence", *The Lancet digital health*, 2: e447-9.

2.1.1 Oncology

Lung, breast and prostate cancer are major public health problems. The key to reducing mortality from these cancers is the early detection of small lesions, before any symptomatic translation, in groups at risk of developing lung or prostate cancer, or in breast cancer screening in the general population.

However, mass screening in the general population - or in a numerically important risk group - requires a large number of practitioners to interpret the data; moreover, this approach requires additional time during an examination or necessitates a specific prevention consultation, an approach which is known to be particularly dependent on the resources of individuals (in terms of information, financial means, etc.) who do not have the same easy access to targeted prevention consultations.

Furthermore, micro lesions may escape the human eye, which makes image interpretation more complex and requires a great deal of targeted experience on the part of radiologists, whose training, experience and availability are very heterogeneous. As a result, some lesions remain undiagnosed, despite a double reading of mammography images[27]. The effectiveness of conventional expert decision support systems has not been demonstrated.

In the last five years, AISMDs based on supervised deep learning techniques have been developed and tested in different cohorts and types of imaging[28]. In addition to reproducibility and time savings, these systems are characterised by a detection accuracy close to 95%[29,30]. This gives them a high sensitivity, particularly in terms of detecting abnormal images, or even markers imperceptible by the eye and by traditional systems, allowing the practitioner to orientate the benign or malignant nature, and to refine the diagnosis and management of the patient (even if, as shall be seen below, this sensitivity may be paid for by a loss of specificity).

Some AISMDs are approved by the competent authorities in Europe and the United States, but comparison of the results obtained by these AISMDs and of an analysis made only by radiologists is not systematic, and the variability in the level of validation hinders a generalisation of the clinical implementation as discussed below. The French Society of Radiology has recently published its recommendations on this subject[31]. In the long term, we can hope to achieve individualisation of risk analysis by combining clinical,

---

[27] Hickman S.E., et al, (2021), "Adoption of artificial intelligence in breast imaging: evaluation, ethical constraints and limitations", British Journal of Cancer, 125(1):15-22; Lotter W., et al, (2021), "Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach", *Nature Medicine*, 27, 244-249; Kim H.E., et al, (2020), "Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study", The Lancet digital health, 2, 138-48; Yu K.H., et *al*, (2018), "Artificial intelligence in healthcare", *Nature Biomedical Engineering*, 2: 719-31; Thomassin-Nagara I., et *al*, (2019), "Artificial intelligence and breast screening: French Radiology Community position paper", *Diagnostic and Interventional Imaging*, 100, 553-66.

[28] See for example the solution developed by the French start-up Therapixel: https://www.therapixel.com/product-technology/ .

[29] E. Svoboda, (2020), "Deep learning delivers early detection", *Nature,* 587, S20-2.

[30] Ardila, D. et *al*, (2019), "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography", *Nature Medicine,* 25, 954-61; Blanc D., et *al*, (2020), "Artificial intelligence solution to classify pulmonary nodules on CT", *Diagnostic and Interventional Imaging*, 101: 803-810.

[31] Thomassin-Nagara I., et *al*, (2019), 'Artificial intelligence and breast screening: French Radiology Community position paper', *Diagnostic and Interventional Imaging*, 100, 553-66.

epidemiological and radiological data for each patient in a "multitasking" algorithm, so as to adapt the screening modalities.

Many other applications based on image processing are becoming widespread in the field of oncology (e.g. interpretation of colonoscopy images), as well as in other medical specialties.

Because of the performance and time savings offered by AISMDs, their development seems particularly well suited to the early detection of certain cancers, especially for large-scale screening campaigns with a public health objective.

2.1.2 Retinal diseases

Large-scale screening for three retinal diseases that cause early blindness is certainly one of the most advanced and promising applications of AISMDs. These are diabetic retinopathy (a serious complication affecting 50% of type 2 diabetics), age-related macular degeneration (AMD, affecting 20% of people over the age of 80) and glaucoma[32]. These conditions can be effectively treated if detected early.

AISMDs developed for the interpretation of images for these diseases are used in the context of routine screenings[33]. These screenings by digital retinography can already be carried out by health professionals who are not doctors or ophthalmology specialists and can benefit from off-line and remote interpretation of images by the ophthalmologist in the event of difficulties in accessing care.

The performance and reliability of some algorithms for detecting diabetic retinopathy from retinal images have been demonstrated[34]. In addition, a prospective study[35] in real-life conditions tested and demonstrated the effectiveness of an AISMD that performs the examination and interprets it locally and simultaneously, without the intervention of a human specialist, thus facilitating a rapid medical decision. The Food and Drug Administration (FDA) validated and cleared such an AISMD (IDx-DR) for marketing in 2017. It is also CE marked (conforms to the requirements set by European Community regulations), as is the case with a French AISMD, OphtAI, which obtained CE marking in 2019. However, the certification of such an AISMD is, for the moment, an exception.

2.1.3 Dermatology

The importance of early recognition of skin lesions, particularly three types of cancer, including melanoma, a fearsome malignant tumour, justifies the current screening campaigns for these diseases by dermatologists. For several years, it has been demonstrated that an AISMD can distinguish between malignant and benign skin lesions

[32] It is a progressive degeneration of the optic nerve which, if left untreated, leads to a narrowing of the visual field and blindness. Systematic screening is recommended from the age of 40.
[33] Deep learning is applied to fundus images, optical coherence tomography of the retina in 3D, or visual field analysis.
[34] Quellec G., et *al*, (2017), "Deep image mining for diabetic retinopathy screening", *Medical Image Analysis*, 39:178-193.
[35] Abramoff M.D., et *al*, (2018), "Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices", *npj Digital Medicine*, 1:1-8.

under experimental conditions[36]. However, the applications proposed for ambulatory use, particularly those that interpret images of skin lesions collected by a *smartphone*, are not as advanced or as rigorous in their evaluation as those used in ophthalmology and are still at the experimental stage. The US Food and Drug Administration (FDA) has not yet cleared any detection systems, and warnings have been issued[37].

2.1.4 Cardiology

The electrocardiogram (ECG, a tracing obtained by recording and transcribing the electrical currents of the heart) and echocardiography (ultrasound of the structures of the heart) are common practices for the diagnosis, monitoring and detection of risk factors for cardiovascular diseases. While the ECG is not *strictly* an image, the interpretation of the highly stereotyped characteristics of the pattern lends itself particularly well to the application of algorithms and, in particular, to the detection of cardiac rhythm disorders[38] (such as atrial fibrillation) that expose patients to serious complications[39]. Repeated prolonged ECG recordings with wearable devices allow their detection. Coupled with deep learning algorithms that automate interpretation, they can help detect these arrhythmias and prevent disabling complications. Several of these devices have been approved[40].

2.1.5 Image analysis in histopathology

In practice, histopathological analysis[41] is essential to consolidating diagnostic certainty, particularly in oncology. Histological diagnosis is based on tissue marking techniques combined with molecular or genomic techniques offering a very fine, but also very complex, classification of cancers. The interpretation of histopathological diagnosis can, as in all fields, be subjective when performed by humans[42]. However, it is essential, since it is the basis not only of the therapeutic decision but also of the prediction of the response to a treatment. Several studies have shown that AISMDs applied to digitised microscopic images from histopathology slides allows the optimisation of histological classification, and can distinguish a benign lesion from a malignant one. By combining the analysis of clinical, microscopic and molecular data, artificial intelligence technologies can also predict the

[36] Esteva A., et *al*, (2017), "Dermatologist-level classification of skin cancer with deep neural networks", *Nature,* 542**:** 115-18.

[37] Matin R.N., Dinnes J., (2021), "AI-based smartphone apps for risk assessment of skin cancer need more evaluation and better regulation", *British Journal of Cancer*, 124, 1749-1750; Walter F.M., et *al*., (2020), "Further evaluation is required for smartphone-aided diagnosis of skin cancer", *The Lancet Digital Health*, 2; e104**.**

[38] For example, atrial fibrillation (AF), which is often asymptomatic and transient, can lead to thrombotic strokes (30% of strokes) or cardiac decompensation, and can be treated if detected.

[39] Attia Z.I. et *al*, (2019), 'An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction', *The Lancet*, 394: 861-7; Hannun A.Y., et *al*, (2019), 'Cardiologist level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network', *Nature Medicine*, 25: 65-69.

[40] The French start-up Cardiologs has developed a web-based ECG analysis platform, using artificial intelligence techniques, which has recently been approved by the FDA and CE marked, and can be viewed at the following link: https://cardiologs.com/.

[41] It is established from tissue or cell samples taken during surgery, endoscopy or any other sampling, fixed on a glass slide. The diagnosis is based on the distinction of the different components of the tissue and cell architecture under the optical microscope after staining and marking. These slides are nowadays digitised.

[42] Laurent C., et *al*, (2017), "Impact of expert pathologic review of lymphoma diagnosis: study of patients from the French lymphopath Network", *Journal of Clinical Oncology*, 35, 2008-2017; Elmore, J. G. et *al*, (2015), "Diagnostic concordance among pathologists interpreting breast biopsy specimens", *Journal of the American Medical Association,* 313, 1122-32.

behaviour of a tumour[43,44]. In addition, as in radiology and all the fields of application analysed above, softwares optimise the management of professionals' time by automating reading, and by selecting only samples whose interpretation is open to discussion for human examination.

2.1.6 Learning artificial intelligence systems, uncertainties and performance

The functioning of the artificial intelligence systems discussed in this Opinion is based on statistical processing of data, which are usually very large in number. Many of these AISMDs use artificial neural networks organised in layers, which implement mathematical functions, such as classifiers.

The input data can be of different kinds: images, text, sounds, etc. In so-called "supervised learning", the most commonly used, each piece of training data is outlined and then labelled. A fracture on an X-ray will thus be labelled "fracture", a malignant tumour on a CT scan will be labelled "tumour", etc.

In a first phase, called learning or training, the numerous parameters of the network (the number of synaptic weights can be several thousands, millions or billions depending on the size of the neural network used) are calculated by means of an optimisation algorithm in order to obtain a satisfactory classification of the training data, hence the name "supervised" learning. As an example, let us consider a classification of the parts of images labelled "fracture" or labelled "malignant tumour". The objective of this phase is for the system to adjust its parameters to identify common elements, in terms of parameters characterising the components of the images (intensity, pixel alignments, etc.), frequently found in this or that characteristic of the images. The result is a statistical "model" of the data represented by the optimised values of the synaptic weights.

Some models require the setting of a number of parameters external to the training, called hyperparameters (typically the number of neurons or the batch size, etc.); these hyperparameters are determined on a separate (validation) data set.

The model is then ready to be used to classify new data according to the learned categories. This is called prediction: it involves predicting to which category the new data belongs. The performance of this prediction depends on probabilistic parameters characterising the system. These predictions are therefore always accompanied by uncertainties. To evaluate them, the following notions are used in the case of a classification with two categories:

- True positive (tp): the number of cases where the prediction finds a feature actually present in the data;
- False negative (fn): the number of cases where the prediction does not find a feature although it is present in the data;
- False positive (fp): the number of cases where the prediction finds a feature when it is not present in the data;

[43] Syryk C., et *al*, (2020), "Accurate diagnosis of lymphoma on whole-slide histopathology images using deep learning", *npj Digital Medicine*, 3, 63; Coudray N., et al, (2018), "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning", *Nature Medicine*, 24, 1559-67.
[44] Colling R., et *al*, (2019), 'Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice', *The Journal of Pathology*, 249: 143-50; Van der Laak J., et *al*, (2021), 'Deep learning in histopathology: the path to the clinic', *Nature Medicine*, 27, 775-84.

- True negative (tn): the number of cases where the prediction does not find a feature that is not actually not present in the data.

The performance of the system is then evaluated by several metrics, the most common of which are the following:

- *Accuracy* A = (tp+vn)/(tp+tn+fp+fn), i.e. the rate of correct predictions out of all predictions;
- *Precision* P = tp/(tp+fp) is the rate of correct predictions of the features of interest among all the features found. Precision is a measure of relevance;
- *Recall* R = tp/(tp+fn) (or sensitivity) is the rate of correct predictions of the features of interest among all predictions of those features. Recall is a measure of comprehensiveness;
- F_1 (called F1 measure) = 2 P.R/(P+R) which is the harmonic means of precision and recall that allows them to be combined with equal importance.

These systems can be useful in that they achieve a good level of reproducibility of the decision[45].

The assistance of AISMDs is also likely to be decisive for moderately specialised centres. Indeed, for highly specialised and competent centres, it is certainly an interesting but complementary aid. On the other hand, we must insist on the need to democratise the use of these AISMDs, which can sometimes help to compensate for the relative lack of specialised skills.

At the same time, however, they raise an ethical tension because **the performance of AISMDs always carries a probability of error**. It is therefore essential that the medical team using it is aware of these possibilities of error and assert control over the results proposed by the AISMD. This tension can be explained by means of the notions of **incompleteness** and **incidentaloma**.

Of course, humans also make mistakes, which raises ethical questions. However, this suggests that it is not the risk of error in itself that is problematic, since humans make them. But intellectual and mental procedures in human beings put their decisions into perspective. The problem with decision devices using AISMDs is that they can be binary when not used with a critical eye. This raises an ethical problem in that there is currently no ethically satisfactory way of ensuring that the diagnosis made using these technologies can be assessed fairly, minimising the risk, to provide a medical decision that often has a considerable impact on the patient. Furthermore, the issue of discharging responsibility to the machine raises the moral question of the origin and meaning of human action.

Firstly, any AISMD can produce **false negatives,** i.e. the recall (or sensitivity) can be too low. This is called **incompleteness**. For example, a lesion or abnormality escapes the AISMD, which indicates that the case is "negative" in the sense that it is not problematic, although the case is in fact problematic. Secondly, the algorithm may make **false positives**, i.e. identify lesions that are not actually lesions.

[45] Morris, Alan H., Brian Stagg, Michael Lanspa, James Orme, Terry P. Clemmer, Lindell K. Weaver, Frank Thomas, et al. "Enabling a Learning Healthcare System with Automated Computer Protocols That Produce Replicable and Personalized Clinician Actions. *Journal of the American Medical Informatics Association: JAMIA* 28, no.º 6 (June 12, 2021): 133044-. https://doi.org/10.1093/jamia/ocaa294.

In addition, imaging of the human body, whatever the technique used, exposes the discovery of **certain incidental elements, normal or abnormal, which they contain and which were not sought after**. These **unexpected discoveries**, with no relation to the context in which the imaging was performed, are classically called ***incidentalomas*** in the medical community. This is the case, for example, when a trivial bone cyst is discovered during the exploration of a liver mass on CT. These discoveries are becoming more and more widespread, which implies that their occurrence should be taken into account in the case of image analysis by an AISMD.

The rates of incompleteness and false positives go in the opposite direction. The more we try to ensure that the AISMD captures all the images of lesions, the more it will be likely to capture false ones; conversely, the more we want it to capture only true lesions, the more it will be likely to leave out real ones. The AISMD is therefore adjusted, in most cases by the developers, rarely by the doctors, to find a happy medium between the two.

The discovery of incidentalomas has intensified in both clinical medicine and biomedical research, correlating with the development of imaging examinations[46] and the improvement of image resolution. A fortuitous discovery is often a source of complex situations for the patient and the medical team. This question, which is particularly relevant for radiology and genetics, was previously addressed by the CCNE in its Opinion n° 133 on "Ethical issues of targeted genome modifications: between hope and vigilance".

It is likely that the use of AISMDs will alter the frequency of incidentalomas, either because it will minimise their detection (e.g. AISMDs are not trained to detect a bone lesion during a liver scan), or because it will detect them inappropriately (e.g. detection of a normal growth plate in paediatrics mistaken for a fracture).

Warning points:

1) To avoid the risks of incompleteness and incidentaloma, it is essential that the meaning of a measure that assesses the trade-off between precision and sensitivity is fully clarified by the industry.

2.1.7 Interpretation of genomic data

Artificial intelligence technologies have the unique ability to exploit the vast amounts of data generated by Next Generation Sequencing *(*NGS) techniques. This convergence between digital and genomic[47] has led to a considerable advance in knowledge, not only for the prediction of gene function, but also for the description of associations between genotype and phenotype in multiple diseases, biomarkers for certain patients, and the individualised therapeutic management of patients.

[46] O'Sullivan J.W. et *al*, (2018), 'Prevalence and outcomes of incidental imaging findings: umbrella review', *British Medical Journal*, 361: k2387.
[47] Genomics is discussed in section 4.2. of this document.

2.1.8 Ethical implications: incompleteness and incidentaloma

Although, technically, several existing learning systems are able to detect a lesion and sometimes suggest its malignant or benign nature without direct human intervention, the complete transfer of the diagnosis to the algorithm in clinical practice raises many ethical and legal questions and is not yet authorised, especially in France, with the exception of the ophthalmology sector. Similarly, algorithms that determine in an unsupervised manner (without human annotation of the data) the criteria enabling them to arrive at a result (unsupervised learning) are not currently authorised.

The medical team must be able to use what we have called above its "global" thinking, which gives it a point of support for evaluating the diagnostic proposals made by an AISMD. It is essential that the development of these diagnostic assistance systems does not replace the professional skills of practitioners and does not run the risk of a loss of contextualised diagnostic elements for patients, and a loss of expertise over time for doctors.

Warning point:

2) An AISMD, while it may be reassuring because of its rigorous and automatic operation, nonetheless plunges both the patient and the care team into a certain degree of uncertainty. Maintaining human control during the use of an AISMD appears to be essential but will not necessarily remove uncertainties.

Recommendations:

1) It is essential that **already established diagnostic methods,** not involving a priori AISMDs, **continue to be taught and researched,** in order to push them forward.

2) The **growing importance of AISMDs** in the field of medical skills **requires in-depth studies** on the interaction between humans and artificial intelligence technologies in order to assess the impact of AISMDs in the practice of medicine.

3) In the interests of **transparency and traceability**, the use of an AISMD should be indicated in the medical report of a consultation.

4) These elements converge in favour of **human control** at all stages of care, from the indication of examinations to the results of analyses and the contextual interpretation of these results.

2.2 AISMDs in improving the care pathway

In addition to AISMDs dedicated to healthcare professionals, applications or devices are offered to people themselves as part of their care pathway. In addition to making it easier and safer for patients to stay at home, these new tools could also help improve compliance with certain treatments and involve patients in the day-to-day management of their illnesses. These medical benefits are, however, closely linked to the resulting economic

benefits and this intertwining must constitute an important point of ethical vigilance, to which we shall return.

2.2.1 Pre-diagnostic orientation

Simultaneously with the democratisation of the Internet, patients have appropriated online medical information and have adopted, prior to their medical consultations, active research behaviours concerning their state of health.

This change in the relationship between patients and doctors has led to the development by companies of a pre-diagnostic guidance offer (diagnostic support system) made available through applications which, on the basis of questionnaires or conversational agents (or chatbots) collecting their symptoms, formulate a diagnostic guidance and a strategy to be adopted, based on a knowledge base combining data from the literature and disease models. The primary objective is to correctly direct patients to the appropriate professional and, in particular, to identify situations requiring emergency intervention. Examples in the United Kingdom include Ada[48] and Babylon[49] , platforms that present themselves as universal and are preparing to expand internationally, including in France.

These devices represent a very important economic field that raises many ethical questions (see section 2.2.4).

2.2.2 Support for the management of life-threatening emergencies

At the emergency department level, the challenge is to identify patients requiring immediate care, as the time to initiate life-saving treatment may be very short in the case of a stroke or a septic shock, for example. This prioritisation is a complex exercise and can be slowed down by the overload of emergency departments[50] or by the lack of information on the patient's previous comorbidities.

Several prioritisation algorithms have been tested, trained on data integrating demographic parameters, vital signs, symptoms, medical history and comorbidities, and more recently textual analysis of shared medical records[51] and/or observations of health professionals, or even brain imaging data[52]. They have been compared to traditional decision trees with the efficiency criterion of efficient prioritisation of patients arriving at the emergency department, their correct referral to intensive care or conventional hospitalisation, or the time taken for therapeutic management. This last criterion is

[48] Available at the following link: https://ada.com/

[49] Available at the following link: https://www.babylonhealth.com/en-us

[50] According to several studies, "40% of admission diagnoses on first presentation to the emergency room do not agree with the patient's final diagnosis", see this summary: Faqar-Uz-Zaman S.F. et *al*., (2021), "Study protocol for a prospective, double-blinded, observational study investigating the diagnostic accuracy of an app-based diagnostic health care application in an emergency room setting: the eRadaR trial", *British Medical Journal Open,* 11:e041396.

[51] Klang E., et *al*, (2021), "Predicting adult neuroscience intensive care unit admission from emergency department triage using a retrospective, tabular-free text machine learning approach", *Scientific Reports*, 11, 1381; Liang H., et *al*, (2019), "Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence", *Nature Medicine*, 25: 433-438

[52] Lynch C.L., et *al*, (2018), "New machine-learning technologies for computer-aided diagnosis", *Nature Medicine*, 24, 1304-12.

essential in the emergency room, particularly in the case of a stroke or cardiovascular accident, where there are only three or four hours[53] before treatment can be initiated. Similarly, these systems could help to better detect patients at risk of rapid life-threatening deterioration in intensive care units.

At least two French experiments of AISMD application are in progress within the Emergency Medical Service (SAMU) regulation services. In Besançon, the Franco-Swiss "SIA REMU" (AI system for the regulation of medical emergencies) project aims to optimise the regulation of medical emergency calls by helping doctors make decisions based on the evaluation of the seriousness of calls, using a medical decision support software developed using AI. Another study conducted jointly by the Bordeaux University Hospital and Inserm plans to deploy a system for monitoring public health events (car accidents, respiratory symptoms, etc.) aimed at lightening the workload of regulators. The AI systems in question have "learned" based on the analysis of call reports.

However, the research mentioned above is still at an experimental level. It is mainly based on retrospective (previously collected for another purpose) and monocentric (from a single facility or group of patients) data. Their validation in prospective studies (for which new data are collected) in real life, multicentric (data from several centres or groups of patients), is essential to judge their effectiveness and their benefit for the relevance of the medical decision. Their performance and their capacity to accelerate the decision while improving the flow must be compared to those of the methods in use. These strategies could then be integrated into clinical practice for the benefit of the patient.

2.2.3 Helping to monitor patients at home

Independently of an emergency context, monitoring devices allow doctors to continuously monitor their patients' biological indicators (blood pressure, blood sugar, etc.), measured in a real-life context that can range from home to the pharmacy and even the supermarket, where self-diagnosis booths are beginning to be deployed. This essential approach, given the prevalence of chronic diseases and medical desertification, could be particularly interesting in cardiology or diabetology. In general, they are relevant in areas where environmental or behavioural factors are decisive for the evolution of the disease (diabetes, cardiovascular diseases, cancer, or Parkinson's disease, for example, via the interpretation of movement or voice tone) or at home after a surgical operation.

Beyond simple monitoring, AISMDs could make it possible to automate and adapt treatment according to the activity of the patient. Systems used for detecting unusual symptoms or signals (decreased activity, increased heart rate, decreased mobility) can also alert caregivers and thus contribute to faster intervention. This is the case, for example, with the self-learning algorithm produced by Diabeloop for the automated management of type 1 diabetes, which was approved by the French National Authority for Health (HAS) in 2020 to be reimbursed by the national health insurance scheme[54]. Another example is the

[53] Raita Y., et *al*, (2019), "Emergency department triage prediction of clinical outcomes using machine learning models", *Critical Care*, 23:64; Fernandes M., et *al*, (2020), "Predicting Intensive Care Unit admission among patients presenting to the emergency department using machine learning and natural language processing", *PLoS One*, 15: e0229331.

[54] "The sensor measures interstitial glucose every 5 minutes, which is transmitted to the mobile terminal. This mobile terminal aims to automatically control the insulin pump, or propose recommendations based on the sensor data, the patient's historical data, the insulin pump and the data entered by the patient", see:

management of patient monitoring at home after surgery, notably through remote monitoring by text messages[55] and chatbots[56] .

AISMDs have contributed to the deployment of automated home monitoring of patients hospitalised on an outpatient basis, both before and after their hospital stay, whether for postoperative monitoring or cancer treatment, for example. In the latter case, assessment of the patient's quality of life and consideration of their behaviour and tolerance of the treatment in their daily environment are important factors.

Applications in the field of prediction can also be considered. Whether it is the early detection or evolution of a disease, the response of a patient to a personalised treatment or the storage and integration of multiple measurements taken continuously, AISMD make it possible to constitute an alert system. It can then be used in the intensive care unit to more accurately predict unfavourable events or the deterioration of the patient's condition. It thus facilitates an appropriate and timely decision.

As mentioned in the introduction of this section, economic considerations in a context of scarce hospital resources play a crucial role in this "outpatient" transfer of care previously provided in hospital. The influence of this economic pressure in the transformation of the health system through digital technology represents a major ethical issue.

2.2.4 Ethical implications: economic versus health benefits

AISMDs are subject to particular attention in the medical field, notably because of the time saving they promise in a context of scarce hospital resources. The time freed up for doctors could be used to invest in complex and/or singular medical situations which sensitivities escape the approach proposed by AI. The time freed up by AISMDs would thus make it possible to rediscover and develop a major source of information that arises from the relationship that the doctor - and any health professional - forms with the patient. This freed-up time would also make it possible to improve the time spent informing and explaining to the patient in the light of the notion of informed consent. Finally, it could be used to develop the interdisciplinary and interprofessional exchanges that form the basis of correct decisions in highly ethical situations.

Some questions remain regarding the real performance of AISMDs. This has not been definitively established and, on the contrary, numerous studies have shown their technical limitations. We shall return to these evaluation problems in part 3. On the other hand, AISMD may raise ethical questions about the situation and the role of the patient. Finally, they will inevitably raise questions about supplementary health insurance - two points that will be addressed below.

By being considered an "actor of their health" through digital tools in the framework of tele-medicine and tele-care using AI, the patient can feel invested with a responsibility. This can lead to anxiety and, due to a lack of medical and/or digital skills, the patient can feel

HAS, Opinion by the National Commission for the Evaluation of Medical Devices and of Health Technologies (CNEDiMTS) of 28 January 2020, p.45.

[55] Calmedica's MemoQuest tool, adopted by several APHP hospitals, is a chatbot capable of exchanging text messages with patients before and after their outpatient hospitalisation.

[56] See : CNPEN, Opinion No.3 of 15 September 2021, *Conversational agents: ethical issues*, 38 p.

helpless, without a contact person in the healthcare team. Vulnerable people could be particularly affected by this situation[57] .

This is the reason why AISMDs raise a dual ethical issue of inequalities in the face of digital technology, but also in the face of increasingly difficult access to the healthcare system, including the issue of medical deserts. Barriers to accessing healthcare cannot be removed by digital tools alone, as patients' appropriation of them is uneven and depends on a number of economic, geographical, social and generational factors. It is well known that these difficulties tend to accumulate in the same populations. People who have difficulty accessing hospitals will also have difficulty using digital health devices or self-diagnosis booths on their own, to take just one example. In addition, the remote collection of consent from these patients also raises new points of tension. Finally, the risk of misuse and drift towards coercive measures (concerning non-adherence to a treatment or a prescribed measure, risky behaviour, etc.) must be taken into account.

Another specific tension raised by these systems is that of the balance between the benefit in terms of medical care promised to patients and the economic benefit targeted (in terms of medical staff resources). Pre-diagnostic referral, which proposes an optimised intervention in a context of hospital crisis and scarcity of practitioners, could become an obstacle to the admission of patients to the hospital if the criteria established were too demanding, or in cases where the specific characteristics of certain patients were not taken into account in the established algorithms.

Possible errors at the pre-diagnostic referral stage can generate a potentially serious risk for the patient. This risk is all the more important as hospital facilities are experiencing an exponential increase in the number of users whose pathologies could be treated by town doctors, who cannot meet all the demand. Some patients who do not have access to a general practitioner are also more likely to experience difficulties in accessing hospitals and therefore find themselves without alternative care. Artificial intelligence technologies applied to health would then have the effect of transferring the burden of the disorganisation of the healthcare system to these most vulnerable patients.

The increased possibility of outpatient monitoring by AISMDs may encourage the use of this form of care, which is much cheaper than hospitalisation, by muddying the specificities of certain categories of patients or specific cases requiring a hospitalisation. Thus, in general, flow management systems, mainly promoted by hospitals for economic reasons, are at risk of adding to the multiple barriers to access healthcare for the most vulnerable users. It is true, however, that a patient should not be kept in a critical care bed if they do not need it. We therefore find ourselves in a situation where technology is seen as a means of compensating for the deficient organisation of a healthcare system that would evolve with digital innovations.

It is also essential that the quality of care prevails over the economic interests associated with the development of AISMDs. This concerns both the sector of technological and industrial innovation, which raises ethical tensions discussed above, and that of supplementary health insurance, where other problems arise.

[57] See : CNPEN, Watch bulletin No.3 of 21 July 2020, *Ethical issues related to digital tools in telemedicine and telecare in the context of Covid-19*, p.21.

Indeed, the dissemination of AI for medical diagnosis leads to highlighting two types of ethical questions regarding the role that supplementary health insurance can or cannot play. Firstly, there is the risk of discrimination due to the fact that supplementary health insurance providers would have access to data about the state of health of users. This data can theoretically allow supplementary health insurance providers to offer individualised and evolving products. The current French legal framework prohibits, in principle, such individualisation[58].

However, an ethical reflection on AISMDs also implies integrating these issues into their medium-term perspectives. However, the possible evolution of supplementary health insurance confronts us with new problems, and in particular with a risk of discrimination if the possibility of this modification is not thought out in advance. The current framework aims, in particular, to avoid the risk of adverse selection of profiles based on health status. Moreover, such a development could lead to an amplification of existing biases and would also be potentially detrimental to the protection of users' personal data.

Secondly, the role of supplementary health insurance actors should not be considered only in terms of ethical risks. The absence of intervention on their part would also be highly problematic. Supplementary health insurance companies also have a major role to play in making health innovation accessible to all. However, if they do not take part in the financing of innovation, witness a breach of equality could be witnessed between users regarding access to innovation in medical diagnosis AI or the development of a form of alternative care system materialised by a multiplication of unlicensed or comfort AISMD.

Better access to innovation for all is essential. If the principle of equal access for users is not taken into account, we could be faced with a situation where some people could benefit from cutting-edge algorithmic medicine and others would be sidelined for economic reasons. Health insurance as a whole has an essential role to play in these matters, even if it is natural that compulsory health insurance should cover at least part of the cost of access to innovation in AI in medical diagnosis, as a *sine qua non* condition for preserving the principle of solidarity governing our health system.

Warning points:

3) AISMDs must be regarded as **complementary tools** for responding to the inadequacies of the healthcare system, in particular medical desertification, but **must not be regarded as substitutes** for medical teams.
4) The decision-making process leading to the use of AISMDs must imperatively take into account **concern for care before economic considerations**. It must be considered as a means of fighting social and territorial inequalities in health.
5) Vigilance must be maintained against the risks of adverse selection that could be operated by supplementary insurance companies, in particular by means of actuarial diagnostic tools based on AI techniques.

[58] CCNE, Opinion No.124 of 21 January 2016, *Ethical reflection on the evolution of genetic testing linked to very high throughput human DNA sequencing*, p.84.

Recommendation:

5) Initiate a **national consultation** on the affordability of AISMDs for users of the healthcare system. In this respect, the identification of a **section dedicated to the financing of AI in the Social Security financing laws** could be examined. In this debate, the role of complementary protection actors in identifying and stimulating the accessibility of AI innovation should be carefully examined.

2.3 Wearable devices using AISMDs outside the care pathway

There is now a proliferation of sensors, mobile applications, social networks, geolocation, or simple questionnaires, which regularly collect information in real time and measure multiple body parameters[59] which they integrate to provide a result that can be described as health or well-being parameters. They may be used in the context of monitoring or diagnosing diseases (see above) or they may be "recreational" and outside of any medical supervision. Among the range of devices available, some can estimate the degree of anaemia by analysing photos of the nail bed, others offer smartphone-based ultrasound systems that can potentially be manipulated by non-physicians and can generate and interpret ultrasound scans or electrocardiograms remotely, providing the patient with initial diagnostic guidance. Finally, other applications can be used in dermatology (see 2.1.3. of this document). These wearable devices (digital wearables such as connected watches, recording bracelets, smartphone applications, etc.) are sometimes referred to as "medical selfies".

The data measured by these devices can be collected by non-medical professionals. They can therefore help in the screening of chronic diseases (or not, such as neuro-evolutionary diseases) that are often neglected or diagnosed late (e.g. hypertension, diabetes) but also in the monitoring of these conditions when access to care is difficult and/or expensive. In this respect, these tools can have a positive impact on our health system by avoiding long, costly and tiring journeys and by reducing appointment times. On the other hand, the public development of these devices is accompanied by a risk of an increase in medical consultations due to the anxiety that data from portable devices using AISMDs may generate in some people.

In order to be truly useful to health professionals, in addition to regular and rigorous validation, it would be necessary to be able to cross-reference these mobile data with those of the patient's shared medical file, which raises considerable logistical, legal and ethical questions concerning health data, which are beyond the scope of this Opinion.

More subjective data can also be collected through questionnaires implemented in smartphones and connected object applications or through the collection of behavioural indicators that are supposed to detect signs of depression, for example. Analysis of a person's gait or voice, for example, can be used to diagnose Parkinson's disease or a

[59] Heart rate, breathing rate, blood pressure, temperature, sleep patterns, voice, movement, etc. For example, a connected watch can detect heart rhythm anomalies (atrial fibrillation). This is the case of the Apple watch: Campion E.W., Jarcho J. A., (2019), "Watched by Apple", *The New England Journal of Medicine*, 381; 1964-5.; Perez M.V., et *al*., (2019), "Large-scale assessment of a smartwatch to identify atrial fibrillation", *The New England Journal of Medicine*, 381: 1909-17. See also: Rodin J.M., et *al*, (2021), "The hopes and hazards of using personal health technologies in the diagnosis and prognosis of infections", *The Lancet Digital Health*, 3, 7, e455-e461.

depressive syndrome. Many studies have also focused on the area of personal sensing, where biomarkers detecting sleep, social context, mood or stress are most commonly used. These data could help identify people in need of care more easily and quickly, but the processing or dissemination of this information may also expose them to risks in terms of personal data protection and therefore raise important ethical questions. On the other hand, the quality of the information, which will often not have been checked as part of a care device, is very uneven.

As mentioned above, some of these devices are already used for remote monitoring of patients with various chronic diseases. The clinical or personal interest (in the case of wellness applications) of these tools must be critically analysed and be part of a health policy. Their development, in a rapidly expanding market, should not be an end in itself and it is important to ensure that their development is aimed at making them a medical device and not solely for recreational or commercial purposes.

2.3.1 Ethical implications: risks of unprofessional advice

There is a market of objects and sensors at the edge of the established health system, or even outside it, which is spreading and which could possibly benefit the health of the population. Nevertheless, some studies point to serious limitations[60] and risks of inappropriate use. Let us emphasise three of the ethical issues that follow the development of this market.

Firstly, the **protection of personal data** produced and collected in this context must be ensured. However, most of the time, there are no educational tools that make it easy to understand what these devices do with the data they have: how they store it, with whom they share it, etc. It is therefore necessary to ensure that the data is treated, in the same way as data produced within the health care system, as sensitive health data which treatment is more restrictive in terms of protection than the simple data that we produce in other recreational settings (e.g. social networks). Since data can become medical by destination, it is important to implement the means to identify it as such when this is the case. Users must therefore be informed of the security procedures that govern the production and storage of such data, and they must be able to control the use of such data by expressing consent, or at least the possibility of opposition. One of the questions that arises is to what extent this data could eventually be associated with the digital health space "My Health Space" currently being deployed[61]. If a rapprochement between the two were to be an objective, the rules for production, storage and use would have to be made to converge in the direction that is most protective of users' personal data and users would have to be informed of this.

Secondly and more generally, beyond the sole problem of data protection, the eventual dissemination of **these portable AISMDs could contribute to distancing the most sophisticated patients (in terms of digital use) from the strictly regulated conventional healthcare system**. Their performance as health tools is rarely certified and, on the other hand, they are likely to widen the gap between the healthcare offer subject to state regulation and other unregulated care practices. There is therefore a risk of medical disintermediation or disruption of the patient-doctor relationship. Clearly, however,

[60] For example: Y. Demeure, "According to a study, self-diagnosis on the Web works in only one third of cases", *SciencePost*, May 2020.
[61] Available at the following link: https://esante.gouv.fr/strategie-nationale/mon-espace-sante.

portable medical information systems will not replace the relationship that a patient may have with a doctor. On the contrary, the way forward is to support the intervention of health professionals in order to free up more human time with the patient through the use of technology.

Thirdly, **these AISMDs can generate inequalities of treatment between** the most favoured **people** - in economic capital, but also in relational and cultural capital (including digital culture) - and the others. The former would be able to afford the services provided by private companies to meet their specific needs, while the latter would be excluded - private companies only developing services if they are profitable.

Warning point:

6) Take into account **the risk of drifting into non-professional advice, unsecured data processing, and the illegal practice of medicine,** which accentuates the vulnerability of individuals. AISMDs must fall within the scope of strict medical practices.

III. A FRAMEWORK FOR AISMDs THAT REMAINS TO BE DEVELOPED

In the medical disciplines mentioned above, the literature dealing with AISMDs in the diagnostic approach remains above all research literature, the quality of which is very heterogeneous. The majority of the applications studied have no official validation[62]. There is still a wide gap between the promises announced[63]- often at the initiative of industrialists, given the very significant growth of this market - when designing a scientifically robust algorithm, and the reality of its transfer to clinical practice[64]. It is possible to identify two interrelated stages through which an AISMD must pass to be adopted in practice. Firstly, the evaluation of the clinical efficacy of the AISMD, which paves the way for its eventual reimbursement by the social security system, and secondly, the certification of conformity, which authorises its marketing.

3.1 For an evaluation of AISMDs that protects the care relationship

The use of AISMDs lies at the crossroads of two fields, artificial intelligence technologies and medicine, which have different scientific and technical evaluation procedures. There is a specific difficulty, still under discussion at international level, in defining an evaluation of AISMDs that is scientifically and socially acceptable. In particular, clinical development cannot be carried out without the AISMD having been evaluated with regard to its real effectiveness in terms of care. With regard to this evaluation, a twofold question arises. Firstly, how to produce a sincere and useful evaluation of its performance, and secondly, how to control the predominant influence of the market in these judgements, with the risk that it will dictate the transformations of the care system.

3.1.1 The role of the French National Authority for Health (HAS) in the evaluation of AISMDs

In France, the body responsible for assessing the clinical benefits of AISMDs is the French National Authority for Health (HAS). Its role is to identify the contribution of a product from the point of view of its usefulness for the patient[65]. It evaluates AI algorithms when they are included in a medical device. However, it only evaluates medical devices that are reimbursed by the health insurance system for patients. **In fact, almost all the algorithms used by healthcare professionals are therefore excluded from the HAS evaluation since most of them are not intended to be sold to patients**[66].

[62] According to Rezazade Mehrizi, M.H., et *al*, (2021), "Applications of artificial intelligence (AI) in diagnostic radiology: a technography study", *European Radiology*, 31, 1805-1811: "More than half of the applications (60%) have not been validated by an official body, and of those that have, 60% are validated by the FDA, 62% have a CE mark, and 32% have both". Thus, as of October 2020, 29 of these systems have been validated by the FDA and 35 are pending (compared to 2 in 2012).

[63] Wilkinson J., et *al*, (2020), "Time to reality check the promises of machine learning-powered precision medicine", *The Lancet Digital Health*, e677-e680.

[64] Keane P.A., Topol E., (2018), "With an eye to AI and autonomous diagnosis", npj Digital Medicine,1, 40.; He J., et al., (2019), "The practical implementation of artificial intelligence technologies in medicine", Nature Medicine, 25:3.0-36.

[65] Hearing of HAS representatives by the CCNE and CNPEN working group; see also: https://www.has-sante.fr/jcms/c_928541/fr/comprendre-l-evaluation-des-dispositifs-medicaux.

[66] See the website of the High Authority for Health: https://www.has-sante.fr/jcms/p_3212876/fr/un-nouvel-outil-pour-l-evaluation-des-dispositifs-medicaux-embarquant-de-l-intelligence-artificielle.

This lack of evaluation - currently optional and at the initiative of the companies - is likely to legitimately hinder the public's confidence in AISMDs. In response to these shortcomings, the HAS has developed a basic grid for evaluation that contains an ethical section in addition to that provided by the European regulation; it aims to evaluate what a genuinely useful innovation for the patient is, and which may be financed by the health insurance system.

Without a widening of the HAS evaluation perimeter to all health establishments or health professionals who want to equip themselves with AI software, manufacturers can make promises that are difficult to measure. Two major problems arise: first, for devices intended for professionals, no control of clinical benefits is imposed; and second, according to the HAS, most of the devices being presented as containing AI have not been submitted to any evaluation.

3.1.2 Ethical implications: for greater explicability and temperance of promises

One of the solutions imagined for evaluating AISMDs would be to be able to guarantee a certain level of explicability of the functioning of the algorithms and their results. Explicability is the possibility of providing a reason understandable by the human user for the result produced by the system which is also representative of its operation. **For example, if an image analysis AISMD identifies a lesion, the doctor must be able to explain this lesion and whether it should be ignored or taken into account.** Explicability is the ability to make clinical sense of computer results. It depends both on what is explained and on the degree of computer literacy of the practitioner. In fact, to indicate whether an identified lesion is an error or not, it is necessary to know the specialty on the one hand and, on the other, to understand what, in the operation of the AISMD, led to the result produced. The level of the doctor's knowledge in computer science may be inexistant or almost inexistant, but it may also be informed, which is what some computer science researchers are trying to do by producing "guides" to help with explicability[67] as well as numerous research projects[68].

Intermediation might also be necessary. It would consist in creating, if not new professions, at least new skills for the doctor to cover the understanding of AISMD devices.

Note that this notion of explicability cannot be extended from short-term diagnosis to longer-term treatment. Indeed, it is conceivable, particularly in the case of rare diseases, that a treatment will be proposed on the basis of the information collected, which will not be very comprehensible and which results will never be seen immediately. This treatment will not be "explicable" in the sense proposed above, but it is of course still necessary. A clear distinction must therefore be made between the two situations, namely short-term explicability, which can be controlled, and medium- and long-term explicability, which is impossible to control today.

At the same time, it is worth asking whether the health status of patients and, more broadly, the healthcare system, benefit directly from the use of these artificial intelligence

[67] Bennetot, Adrien, Ivan Donadello, Ayoub El Qadi, Mauro Dragoni, Thomas Frossard, Benedikt Wagner, Anna Saranti, et al. A Practical Guide on Explainable Ai Techniques Applied on Biomedical Use Case Applications". *SSRN Electronic Journal*, 2022.

[68] A. Barredo Arrieta et *al*, (2020), "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI", *Information Fusion*, 58, 82-115.

technologies. Few studies make a rigorous comparison of their usefulness or relevance to the existing arsenal, and few analyse the performance of the algorithm on independent, multi-centre external cohorts. Important non-medical criteria such as economic criteria, time management or human and material resources must be integrated into the evaluation grid for these systems.

The increase in the workload of radiologists, the explosion in the demand for imaging examinations, and the sophistication of imaging techniques play an important role in the multiplication of these systems to help interpret radiological images. This is particularly true for large-scale screening campaigns, which can only be managed via digital tools. Time savings have often been mentioned for the benefit of health professionals, who are more available to study complex cases or to exchange with their patients.

In this sense, the sometimes exorbitant promises made by industrialists to market their products raise profound ethical tensions. Many economists today refer to an "economics of promise" in the field of new technologies to designate the sometimes spectacular financing granted to AISMDs which are still only promises of a solution.

This economy must be accompanied by **an *ethics of promise***. Firstly, the promise should only be expressed when a set of pragmatic conditions have been met for it to be realistic. In particular, the accessibility of the databases used to develop the AISMD must be proven. On the other hand, a broken promise must be sanctioned one way or another. However, as things stand, manufacturers are rarely required to prove the realism of their project. To remedy these shortcomings, it would be necessary to require at least that industrialists be able to account for the use made of the funding and explain why the goal pursued has not been achieved.

Recommendations:

6) Create the **conditions for trust** by encouraging developers to provide a certain level of **explicability** of the AISMD they put on the market.

7) Promote the **development of new professional skills** to make the properties of AISMDs explicit to carers.

8) Create the conditions for trust by promoting the development of **justifications for promises** and, on the other hand, the **publication of negative research results, i**.e. those that do not confirm the announced promises.

3.2 Certification, a crucial step for trusted AISMDs

The medical sector is subject to strict regulatory processes, particularly concerning health products intended for patients and mainly medication. *On the other hand,* in the case of innovative devices based on AI techniques, health authorities are confronted with legal uncertainty associated with their novelty and ill-adapted procedures.

With regard to the certification of AISMDs, prior to marketing, there are two regulations at the European level, the GDPR and the CE mark. Compliance with the GDPR does not fall within the scope of this Opinion. For CE marking, the certification of AISMDs requires that they are categorised as medical devices or *in vitro* diagnostic medical devices. It aims to assess the benefit-risk balance associated with an AISMD. At the national level, the

competent authority in France is the National Drugs and Health Products Security Agency (ANSM). It is not a question here of questioning the characterisation of AISMDs as medical devices, but of questioning the marking processes that are implemented for these very singular systems compared to, for example, a tongue depressor or crutches.

3.2.1 Regulatory developments that are struggling to keep pace with the diffusion of AISMDs

The evolution of regulations concerning AISMDs accelerated from 2007 onwards with the publication of a Council of the European Union directive[69] on medical devices. This introduced a definition of softwares as medical devices in their own right and incorporated the idea that softwares may, in view of its purpose, have a purpose compatible with a medical diagnostic device or an *in vitro* diagnostic medical device.

In 2010, European and international work completed these texts by questioning the definition of medical device softwares and the way to assess the benefit-risk balance in this area. The positioning of AI technologies is still in its infancy: it is then referred to as external softwares that provide slightly more subtle information than a simple display of results, the term used being "diagnostic assistance".

In 2017, a new regulation started to address AI and a number of points were revised to adapt the regulation more specifically to softwares. Among the important steps of these new texts, one can note the introduction of a risk classification (according to 4 classes) concerning the use of the medical devices, the purpose of which is to direct the manufacturer towards conformity assessment procedures.

The same requirements apply to medical devices and *in vitro* diagnostic devices, however, compared to devices that show a low risk, a higher level of evidence is required from the manufacturer. Previous directives included these devices in very low risk classes, including softwares that did for example mammography image processing previously classified at the same level as a simple tongue depressor.

There has therefore been a significant change since the classification conditions developed in the 1990s for physical medical diagnosis (e.g. ultrasound, scalpel or tongue depressor). With the new regulation, the classification is made according to the risk linked to the use of this software by taking into account, for example, in the design aspects, the protection against cyber security flaws.

In this rapidly evolving context, the **term AI does not currently appear in French regulations: it is referred to as diagnostic or therapeutic decision support software.**

3.2.2 A conformity procedure materialised by the CE marking

The marketing authorisation (MA) system applied to medicinal products is not valid for AISMDs. These devices involve a "new approach" regulation that calls for a **conformity**

[69] Directive 2007/47/EC of the European Parliament and of the Council of 5 September 2007 amending Council Directive 90/385/EEC on the approximation of the laws of the Member States relating to active implantable medical devices, Council Directive 93/42/EEC concerning medical devices and Directive 98/8/EC concerning the placing of biocidal products on the market

procedure based on a CE marking, a procedure followed by the manufacturer with more or less important considerations depending on the risk class of the medical device.

In France, and in several European countries, applications are subject to coordinated authorisations during which evaluation processes involve ethics councils that give their opinion on the ethical aspects of AISMDs. The competent authority - in France, the National Drugs and Health Products Security Agency (ANSM) - then assesses the device in terms of its benefits and the questions it must answer. The conformity check requires the intervention of a so-called notified third-party body[70]. The relevance of the demonstration of the clinical evaluation is thus the result of both the manufacturer's own clinical investigations and of bibliographic work which are evaluated by the notified body which decides whether or not the device is compliant. During this process, the manufacturer must, for example, justify the relevance of an image bank on which he could validate his software by justifying that his sample is compliant.

It is thus up to the manufacturer to prove that his device is capable, for example on X-rays, of detecting a certain type of fracture. It is assumed that by holding the know-how of the device, the manufacturer should be able to assess its conformity.

Among the shortcomings of compliance monitoring: database bias

The issue of possible database bias, for example, which is supposed to be addressed in the compliance check, is only imperfectly addressed. **Ethical dilemmas** may arise here. Let us elaborate on this step: any machine learning algorithm is first designed based on an initial set of data, called the learning base. In the case of a breast cancer detection algorithm, for example, ultrasound scans will be submitted to an algorithm after being labelled by a human.

From these images, the algorithm will identify specific patterns in the images designated by humans as having cancer, as opposed to those that do not. A second set of data, known as the "test base", will then be submitted to the algorithm to see if it classifies the images correctly. If the results obtained are satisfactory, the algorithm can then be tested on a validation database and then be used to process real data. The processing parameters are weighted according to the particularities of the databases to be processed, which requires other specific data.

On the other hand, the initial learning base or the nature of the successive algorithmic analyses must also be assessed with regard to the equal treatment of possible cases (i.e. that they do not introduce a representation bias of certain sub-populations) and improved if necessary.

These different datasets may indeed contain biases of different kinds that need to be identified and weighted and that need to be addressed. These biases may relate to different characteristics. For example, the symptoms of heart attacks are on average not quite the same for men and women. If one of the two categories is less represented in the

[70] According to Article 35 of Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC, available at the following link: https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32017R0745&from=FR.

learning base, the AISMD will be likely to produce more false negatives for this category[71]. Differences between social categories can also lead to this type of problem. For example, disadvantaged people are in a situation of under-use of care, which can be interpreted by the AISMD as a state of good health[72]. Finally, biases relating to the origin of the patients can also be observed[73]. To avoid and control this kind of bias, it is necessary to have access not only to training databases, but also to test and validation databases. While these difficulties are not unique to machine learning applied to medical diagnosis, the sensitivity of the data in question, being health data, makes auditing possibilities all the more complex. Furthermore, the comparison of the results obtained by two AI systems is all the more relevant if they are evaluated from the same training or test databases - this point argues for the definition of a number of fixed, shareable, double-blind databases.

Finally, these systems also produce data. In the case of medical imaging, the data labelled by the AISMD is sometimes sent back to the central archiving system - Picture Archiving and Communication System (PACS)- even though the algorithms that produced it have not been certified. When this is the case, the question then arises of the accessibility of these marked images for all the actors in the medical chain via the central archiving system, without prior control of the relevance of the marking. To date, the availability in an archiving system of images marked by an AI system - for example, images labelled as showing a fracture or a breast lesion - is not subject to prior validation by a radiologist, which does not meet the principle of human control (see section 5).

In conclusion, concerning the "AI" regulation which is currently under discussion at the community level, it appears that additional requirements on medical devices will enrich the current regulation. This implies that, for manufacturers of software with functionalities that meet the definition of AI given in this regulation, the device will have to be brought into conformity and meet, with a notified body, the criteria for CE marking for certain risk classes. *Ultimately,* the outcome of the conformity check involving a CE marking for AISMDs is similar to a marketing authorisation.

3.2.3 Ethical implications: a compliance check in the making

The compliance procedure raises several ethical issues:

- Compliance monitoring is a work in progress and requires professional skills involving training and the development of academic and scientific research to perfect its definition.

- Manufacturers are called upon to design the demonstrations of conformity of their devices themselves, which implies that the control that the demonstration is not biased

[71] Cirillo, Davide, Silvina Catuara-Solarz, Czuee Morey, Emre Guney, Laia Subirats, Simona Mellino, Annalisa Gigante, et al. "Sex and Gender Differences and Biases in Artificial Intelligence for Biomedicine and Healthcare." *Npj Digital Medicine* 3, nº 1 (1 June 2020): -111. https://doi.org/10.1038/s41746-020-0288-5.

[72] Gianfrancesco, Milena A., Suzanne Tamang, Jinoos Yazdany, and Gabriela Schmajuk. "Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data". *JAMA internal medicine* 178, nº 11 (November 1, 2018): -154447. https://doi.org/10.1001/jamainternmed.2018.3763.

[73] Sjoding, Michael W., Robert P. Dickson, Theodore J. Iwashyna, Steven E. Gay, and Thomas S. Valley. "Racial Bias in Pulse Oximetry Measurement. *New England Journal of Medicine* 383, nº 25 (December 17, 2020): 247778-. https://doi.org/10.1056/NEJMc2029240.

is weak, especially as this control is carried out in the first place by a private notified body.

- The compliance of an AISMD does not mean that there is any clinical benefit to its use. It may be able to do what it claims to be able to do, but there is no assurance that what it claims to be able to do is useful or beneficial to the patient or to the healthcare system.

Recommendations:

9) Encourage **multi-disciplinary research and professional training** to develop an unbiased definition of compliance monitoring for AISMD.

10) Promote the use of clinical trials to **assess the benefit-risk ratio** of AISMD in the same way as for medication.

3.3 Liability regime

On 28 September 2022, the European Commission published a proposal for a directive on the establishment of a harmonised EU-wide compensation regime for damage that may be caused by AI systems. This directive would significantly strengthen the scope of the human control requirement for high-risk AI systems, which applies to the specific field of health. The absence of an established system of human control would allow the patient to rely on a rebuttable presumption of a causal link between the manufacturer's fault and their damage, making it easier for them to obtain compensation.

Under these conditions, developers and users of AI solutions would be strongly encouraged to deploy supervisory mechanisms such as the "human guarantee colleges" recommended by the CCNE's Opinion No. 129 and by the WHO, and which associate algorithm designers, representatives of professionals and representatives of the final beneficiaries of the AI system.

The provisions of the European regulation draft on AI already provide for a mechanism to sanction the absence of human control within the framework of a system of financial penalties - up to 4% of the group's annual worldwide turnover - decided by the regulatory authorities on the model of the GDPR. The directive would thus complete the enforceability of human control by giving health patients and, more broadly, the final beneficiaries of AI the ability to bring a civil liability action to obtain compensation in the event of damage due to the absence or inadequacy of human control.

3.4 For compliance monitoring and evaluation of AISMDs that places quality of care above all other considerations

As has been seen above, prior to their introduction on the health market, and therefore to their use in a real situation, AISMDs are subject to a control of conformity ("CE" marking) and evaluation (concerning clinical effectiveness) during which ethical questions are called upon to occupy a particular place.

In terms of **compliance monitoring** (see 3.2), the European (EU) and US (Food and drug administration, FDA) systems differ, but no precise comparison of the two systems is available. It would be useful to develop research on this topic.

In the current state of affairs in Europe, it is desirable that a certain degree of explicability be achieved when presenting the AISMD to the supervisory institution (National Drugs and Health Products Security Agency (ANSM) in the French case) so that the latter can ensure that the end-users of the AISMD have access to the best possible understanding of its operation.

In addition, the audit should specify the degree of diagnostic performance achieved by the AISMD. This degree of performance is often derived from a comparison between a diagnosis made by a practitioner or a group of practitioners and the results calculated by the AISMD. This degree of performance can then be expressed in many different ways: sensitivity, specificity, stability, robustness, etc. However, the definition of an opposable and widely accepted reference (or gold standard) of control is delicate for AISMDs, in particular because of the complexity and diversity of the training data. However, it is essential to stabilise the stages and to seek, as far as possible, stable and effective control criteria.

As regards **evaluation**, this is carried out by the French National Authority for Health (HAS) in the French case. However, the HAS is only interested in AISMD that are likely to be reimbursed by social security. The large number of AISMD used by carers - most of those we have presented in this Opinion - are therefore not covered by the HAS since they are not sold to patients. The question arises as to whether the competences of the HAS should be extended, which would of course imply providing it with the means to do so.

On the other hand, the evaluation must take into account the properties of the databases, which must also be adapted, quantitatively and qualitatively, to the prevalence of the disease under study. Indeed, these databases can mask or even generate biases of various kinds: bias in the representativeness of the population, activity bias (e.g. oriented towards the best valued care according to the care structure) or organisational bias (e.g. collection of data that is easy to collect but not necessarily useful for care) (see the insert in 3.2.2). It is therefore necessary for these databases themselves to be evaluated.

Moreover, when they are submitted for evaluation, AISMD are accompanied by promises made by their promoters (see 3.2.2.). Admittedly, these promises meet the financial constraints of investment in the projects, but they may be exaggerated in relation to the benefits actually provided to patients. Moreover, this discrepancy poses a specific challenge of information which should be pointed out by the evaluation authorities.

Finally, AISMDs are inserted into interprofessional relations of the care system, which they transform. It is therefore essential for the evaluating institution to have socio-economic information nature on these transformations so as to anticipate the organisational changes that result from their use. This remark becomes even more important if we consider that AISMDs are products that can be delocalised and used in under-connected countries for which they will not be adapted. Indeed, one of the crucial blind spots of AI is that the databases of AI systems are mainly from Western countries that are largely digitised, effectively excluding health data on people living in regions with poor health data collection infrastructures such as Africa and Latin America, which can lead to bias. It is then very likely that implicit social elements embedded in an AISMD do not correspond to the

characteristics of the society where it is used. For example, a quarter of African hospitals[74] do not have an inventory of their computer equipment, which restricts the capacity to aggregate data (since it is then difficult to know where it can be produced and stored) and drastically limits the use of AISMDs.

Very generally, the evaluation, as it should be practised, will have the function of verifying that health rationality is not only not overshadowed by algorithmic rationality, but also that it is not replaced by economic and organisational rationalities. The evaluation must always be based on the principle of beneficence.

One of the criteria for the evaluation of AISMDs on which France has distinguished itself by its reactivity[75], in particular with the law on bioethics of 2 August 2021, lies in what the European Commission has translated by the expression "**human control**" or *Human Oversight*[76] under Article 14 of its draft common regulation on AI. After much hesitation and discussion within both the CCNE and the CNPEN, which proves that the concept is not yet firmly established, the expression "human control" has been chosen because it is the translation of the concept of “human oversight” that has been chosen by the European Commission in the text quoted above. Paragraph 1 of this article states that artificial intelligence systems must be designed and developed in such a way that they can be controlled by humans. The next paragraph specifies that human control will prevent or minimise risks to health, safety or fundamental rights that may emerge from an AI system that may present a high level of risk. Paragraph 3 provides guidance on the implementation of human control of AISMDs by recommending, where technically feasible, measures that are identified and, where possible, built into the AI system by the vendor prior to the system being placed on the market or put into service and that are suitable for implementation by the user.

The principle of human supervision is an extension of the recommendations made by the CCNE in its Opinion No. 129 concerning what it described as "human supervision"[77]. In its Opinion No. 130, the CCNE also specified the content of this principle with regard to data processing more specifically, stating that "a 'human guarantee' is essential to ensure the methodological rigour of the various stages of the data collection and processing process, namely (i) the quality and suitability of the data selected to train the algorithms; (ii) the suitability of the choice of algorithmic treatments to the question raised; (iii) the verification

[74] Sagbo J-Y., (2007), "Etude sur les contraintes de gestion des équipements biomédicaux en Afrique", 64 p.
[75] In the law on bioethics of 2 August 2021, the provisions concerned appear in Article 17. The words "Human Guarantee" were proposed and then disappeared from the law itself during the parliamentary shuttles but are, on the other hand, very directly included in the explanatory memorandum of the bill [1] and in its impact study [2]. This position corresponds to the distinction that can be made between a general ethical principle and its operational implementation in positive law, which is found directly in the body of the legislative provisions.
[1] The article "aims to ensure that the patient is properly informed when algorithmic processing of massive data ("artificial intelligence") is used in the course of a healthcare procedure. It also declines to guarantee human intervention.
[2] https://www.legifrance.gouv.fr/dossierlegislatif/JORFDOLE000038811571/.
[76] See: European Commission, "Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence legislation) and amending certain Union legislation", Brussels, 21 April 2021 [available at: https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:52021PC0206&from=EN].
[77] CCNE, Opinion n° 129 of 18 September 2018, *Contribution of the National Advisory Ethics Council to the review of the bioethics law 2018-2019*, 165 p.

on an independent data set of the robustness and accuracy of the result given by the algorithm. "

The CNPEN, for its part, in its contribution in response to the European Commission's White Paper on artificial intelligence[78], had taken up these recommendations and specified, with regard to the implementation of the principle, the expectations of the "human guarantee colleges" by stressing the interest of associating AI innovators, representatives of health professionals and patients' representatives in the exercise of this control.

Recommendations:

11) **Ensure that potential biases related to the programming of algorithms themselves and to different databases** used for the development of AISMDs, which training data is specific to their date of conception, are **limited.**

12) Promote the definition, by the industry, of criteria that will allow the evaluation of an AISMD from its conception ("**Ethics by design**"). These criteria may be generic or specific to an AISMD.

13) Establish a certain number of **social criteria for the evaluation** of AISMDs (medical criteria for the patient, organisational criteria for health establishments and professionals, health criteria for the population, economic criteria for all the stakeholders, etc.) enabling an ethical assessment to be made.

[78] European Commission, (2020), *White Paper. Artificial Intelligence. A European approach based on excellence and trust*, 31 p.

VI. ETHICAL ISSUES RAISED BY AISMDs IN THE TRANSFORMATION OF THE DIAGNOSTIC PROCESS

In this section devoted to the changes brought about by AISMDs in the diagnostic approach, we shall describe the issues that appeared to us to be the most salient with regard to the current data both in terms of potential, the state of progress of the technical processes and also in terms of expectations and social needs.

4.1 Issues in radiomics

In addition to their contribution in terms of diagnostic assistance in the detection of lesions, algorithms applied to image analysis can also extract, from medical imaging data (ultrasound, scanner or MRI), a large number of quantitative tissue parameters, inaccessible to the human eye, which significance in terms of prognosis and response to treatment, particularly in tumours, is currently the subject of promising research.

The image of a tumour is a "phenotype", i.e. an "appearance" corresponding to numerous heterogeneous tissue parameters. This parametric computer analysis, known as "radiomics" (a term based on the model of genomics for the genome and proteomics for the proteome), may eventually make it possible to find biomarkers capable of predicting the prognosis of a tumour (risk of metastasis in particular), and therefore to adapt the treatment accordingly and then to monitor the therapeutic response in detail. In parallel, radiomics aims to study parametric correlations with the genetic profile of a tumour.

Thus, radiomics *ultimately* aims to develop a therapeutic response individually and precisely adapted to each tumour thanks to the knowledge of its parametric (radiomic) and genetic (genomic) "identity card".

The majority of studies in this field are still at the experimental stage, most of the potential biomarkers have yet to be validated, but the potential application of radiomics is significant, with the prerequisite of organising and processing large imaging databases in multidisciplinary teams (medical, bioinformatics, biostatistics etc.). Correlatively, the opening of such fields implies challenges for the interpretation of these massive data: on the one hand, over-accumulation can saturate the health professional and, on the other hand, a possible over-interpretation if the field of use of these new data is not perfectly mastered. In this respect, attention must be drawn to the importance of contextualisation of this new information by the referring doctor.

4.2 Issues in the interpretation of genomic data

Researchers have recently been considering the large-scale measurement of many other components of the human body using so-called "omics" techniques, but also environmental techniques (food, pollution, climate, etc.) based on the concept of the exposome[79], which is recognised as a major public health issue. Whereas in the past, when faced with a given phenotype (disease symptomatology), the sequencing of a targeted region of the genome

[79] Vermeulen R., et *al*, (2020), "The exposome and health: Where chemistry meets biology", *Science,* 367: 392-396.

confirmed the existence of mutations, the opposite approach is now applied: it is known that the occurrence of a disease is based on both genetic and environmental factors, in the sense of an exposome. If reliable measurements of these elements are known, it becomes possible not only to make an early diagnosis, but also to propose a "score" correlated with a risk of developing this disease[80].

4.2.1 From symptom diagnosis to risk diagnosis

In the field of genomics, a minority of diseases, known as "monogenic" because they involve only one gene, are causally linked to rare genetic mutations. But most chronic diseases are complex diseases known as polygenic, i.e. involving several variants[81] on several identified genes. Statistical methods that aggregate the cumulative risk due to the presence of these variants give a "polygenic score" for a given disease and for a given individual. For some diseases, a high polygenic score correlates with as significant a risk of developing the disease as the presence of a rare mutation[82]. These variations will only be expressed as "symptomatic disease" under the pressure of other factors, notably environmental. This data is all based on the analysis by AI technologies of different types of massive data (genomic, clinical, environmental) and the determination of correlations.

The development of a pre-clinical diagnosis of a pathology and the anticipation of the onset of diseases by measuring their "risk" of occurrence in order to avoid their development are among the ethical challenges posed by the development of predictive health models in terms of informing the people potentially concerned. This "polygenic score" may represent a major clinical marker and it could be envisaged that genome analysis will be part of every individual's care pathway in a few years' time, or even of preventive screening strategies. This emphasis on early risk assessment is the basis of the establishment of large genome banks, often supported by governments and public institutions[83]. This approach, initially practised in oncology and now applied to other diseases or disabilities, is expanding rapidly. However, it should be borne in mind that algorithmic genome analysis, while allowing for more accurate diagnosis or risk assessment, uses so much information that it is difficult for the geneticist to understand how the prediction was made. This element refers to the question of the explicability of the solutions proposed by artificial intelligence discussed in section 3.1.2. It also raises delicate ethical questions, in particular regarding the management of incidental data (see section 2.1.8. on the incidentaloma), a subject addressed by the CCNE in a previous Opinion (see CCNE Opinion n° 130).

AI algorithms have also made it possible, particularly in oncology, to highlight correlations between the occurrence of diseases and other potential markers (microRNAs, circulating metabolites, proteins, circulating genomic DNA fragments) that can be measured in

[80] See : Steinfeldt J. et *al*, (2022), "Neural network-based integration of polygenic and clinical information: development and validation of a prediction model for 10-year risk of major adverse cardiac events in the UK Biobank cohort", *The Lancet Digital Health*, 4, 2, e84-e94.

[81] A variant is a change in the sequence of one or more nucleic acids in the DNA, relative to a so-called reference sequence.

[82] Khera A.V., *et al*, (2018), "Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations", *Nature Genetics,* 50: 1219-24.

[83] Examples include the UK Biobank (500 000 people), but also in the United States the *100K Wellness Project* and the *All of Us Research Program* (one million people), in Asia the Kadoorie biobank financed by the Wellcome Trust-funded China (515 000 Chinese), and others. In the UK, the National Health Service and the Department of Health also encourage individuals to pay to have their genome analysed (Genomics England) with guarantees of reliability. This data has been available to researchers since 2017.

individuals by a simple blood test. From these correlations, biomarkers can be deduced, which, when integrated with the polygenic score and 'monitoring' by sensors, could increase the accuracy of an early diagnosis of certain diseases, but also predict their evolution. But beyond the technical feat[84], the question remains of what each individual does with this information, which raises serious ethical questions to which we shall return (see section 4.3).

4.2.2 Ethical implications: explicability, explanation and importance of genetic counselling

The spread of AI technologies, the widespread access to genetic data and the acceleration of the identification of risk factors leads to a growing need for medical support and explanation of related diagnoses to patients and their families. This acceleration in the identification of risk factors leads to a potential change in the perception of what a "healthy state" is. This early detection of risk is a major issue from an ethical point of view, both from a societal and an individual point of view, but also particularly from the point of view of patient autonomy and anxiety management.

Furthermore, as discussed earlier in this Opinion (section 3.1.2), the need for the explicability of AISMDs for professionals is significant in genomics given the increasingly rapid spread of algorithmic processing in the discipline. Tasks such as establishing the clinical significance of genomic variants are now devolved to them and involve a large number of assumptions that must be made explicit for the user. For patients, the development of the profession of 'genetic counsellor' may partly address this point, particularly if their skills are extended to understanding the complexity of AISMDs.

The interpretation of genomic data, in particular resulting from high-throughput whole-genome or exome sequencing, raises the question of incidental findings in a particularly salient way. Although this subject is not specific to artificial intelligence and genomic data, the acceleration of genomic data processing thanks to AI is causing a change in scale of these particularly sensitive issues. For example, it is possible that an AISMD carrying out a systematic reading of the genome in the context of clinical research will identify real risk factors on strands of the genome that are *a priori* unrelated to the symptoms for which the patient is consulting.

Warning point:
7) Particular attention should be paid to **the risks of surveillance and adverse selection in the field of insurance and public health** (this is developed in Part 5).

4.3 Issues concerning the emergence of new spaces for health democracy

The digital transformation of the health sector in general, and the spread of artificial intelligence in medical diagnosis in particular, will necessarily modify the relationship between health professionals and their patients. Article 17 of the Bioethics Act introduces a duty to inform patients about the use of algorithmic treatment in their treatment

[84] For example, such analyses have shown that anomalies were detectable more than 10 years before clinical manifestations (cancer, Alzheimer's), or that the date of death could be anticipated.

protocol[85]. On a more collective level, the dissemination of AI systems in health may also be an opportunity to open up spaces for participation by patients and their representatives in the reflection on ethics and regulation of the use of artificial intelligence in medical diagnosis. In addition, new rights could be set out in the context of legislation on health democracy, i.e. the participation of patients in medical decisions concerning them, in line with Law No. 2002-303 of 4 March 2002 on the rights of patients and the quality of the health system.

4.3.1 The emergence of a new status: the digital auxiliary

A new status could be promoted by the care system, the "digital auxiliary" or "digital helper", who would help patients better understand the challenges of AISMDs they might face. This would be an auxiliary whose function would be to prevent difficulty of access to digital technology from adding to difficulty of access to care. It would also be advisable to establish a method of certification of the skills associated with this role.

The person carrying out this task could be chosen from the patient's close circle. However, if the patient cannot choose a family member, it should be possible to propose an external help: members of patients' associations could then be involved. This new role could also be attributed to the "trusted person", instituted by the law of 4 March 2002, in order to enable a patient to benefit from their support in expressing consent, which is more complex because of the multiplicity of circumstances. Finally, this function could be assigned to a representative of a patients' association, competent in communicating information on algorithmic data processing. In this respect, the effectiveness of this new role for patient representatives implies the commitment of a broad training programme on the challenges of artificial intelligence in health within the associations. The organisation representing patients and users of the health system, France Assos Santé, has already committed itself to this[86].

The context of the health crisis has shown an increasingly wide diffusion of digital tools while revealing the persistence of situations for which the use of these tools may prove problematic (lack of digital literacy, network infrastructure failure, economic and social conditionality for digital access). Thus, by developing more and more widely, medical diagnostic AI makes it even more imperative to address the digital divide and illiteracy, which INSEE considered in 2019 to affect 17% of the French population[87].

4.3.2 Ethical implications: promoting health democracy initiatives

The low level of acculturation of the population and health professionals to the field of digital health and particularly to artificial intelligence implies that the dissemination of AISMDs carries with it an intrinsic risk of exclusion of patients, their assistants and health professionals. The danger of the development and accessibility of medical devices

[85] The article thus states: "The health professional who decides to use, for an act of prevention, diagnosis or care, a medical device comprising algorithmic data processing whose learning has been carried out on the basis of massive data shall ensure that the person concerned has been informed of this and that they are, where appropriate, informed of the resulting interpretation.

[86] France Assos santé, (2019), *White paper. Contribution of digital tools to the transformation of health organisations*, p.189.

[87] Vie-publique.fr, "Fracture numérique : l'illectronisme touche 17% de la population selon l'INSEE", 13 November 2019 [https://www.vie-publique.fr/en-bref/271657-fracture-numerique-lillectronisme-touche-17-de-la-population].

integrating AI is that of a deepening of the distance between professionals and technicians in AI and health on the one hand, and patients on the other, or between doctors familiar with digital technology and those who remain strangers to it.

However, at this stage, there is a reverse movement of the emergence of new spaces of democracy in health and the consolidation of patient participation at an individual level but also at a collective level. This dynamic is not, of course, a given in principle and must be supported, in particular to consolidate the effort to train patient representatives in the challenges of artificial intelligence in health.

Recommendation:

14) Promote the **status of digital auxiliary**. This person would provide an explanation understandable for the patient, or their legal representative, or trusted person, of the result produced by the AISMD, that would also be representative of its functioning. This could be compared to the "advanced practice nurse" status currently being developed.

4.4 AISMDs should not degrade the quality of the diagnostic framework

AISMDs must remain a **human decision aid**. AISMDs can and should be deliberately ignored by a practitioner if necessary[88]. The importance of prudence in the use of AI is one of the justifications for human control, but it also induces a certain number of corollaries in the positioning of the professional vis-à-vis the AI tool.

During the course of care, it is necessary to ensure that the patient has been informed that the medical team may use an AISMD, for what reason, and what the benefits and risks are. This is the meaning of Article 17 of the Bioethics Act, which introduces a duty for the professional to inform the patient about the use of algorithmic treatment, with this information being provided at the time of obtaining consent for the entire care protocol.

The relationship between the medical teams and the patient must not suffer from the use of an AISMD by the carers or by the patient[89]. If the latter has used an AISMD outside the care pathway, they cannot be satisfied with non-professional information to manage their pathology. They must first be able to understand the limits of the information produced outside the medical framework and then, if they suspect a problem based on this information, the AISMD must direct them to a health professional. This issue is particularly relevant for comfort systems.

The use of an AISMD during the diagnostic process has a number of consequences, which also raise ethical issues. The data and information produced by the AISMD, particularly if

[88] This is included in the European regulation draft for AI: (d) to be able to decide, in a particular situation, not to use the high-risk AI system or to disregard, override or reverse the outcome provided by that system.
[89] B. Mittelstadt, *op. cit*.

not expected or sought out in the initial diagnostic process, may lead the medical team to make decisions that are not always necessary (incidentalomas are an example). The medical team should be careful not to order interventions without checking the reality of the identified risk and seeking informed consent from the patient.

To better understand these structural transformations associated with the dissemination of AI technologies applied to medical diagnosis, a new approach to the training of health professionals is imperative. The impacts of the spread of artificial intelligence on the health professions are obvious and even disruptive[90]. These innovative technologies do not necessarily replace human intervention but can increase the doctor's capacities (analysis, diagnosis, prescription). However, they question the **responsibility of the decision**. Indeed, if the AISMD has very good precision, there is a risk of overconfidence on the part of the practitioner. Conversely, if the practitioner has more confidence in their own judgement, there is a risk of under-use of the technology. From this point of view, training in the collegial debate around AI, through, if necessary, the classic vector of multidisciplinary consultation meetings (RCPs) where different medical disciplines meet, is an essential issue for the appropriation and control of these tools. It will then be necessary to integrate other professions (engineers, computer scientists) and users of the healthcare system into all or part of the RCPs[91] .

The conditions under which medical specialties and care support functions are practised could change permanently, even if the impact of the use of AISMDs differs according to the specialties. The impact may be significant for medical disciplines that are already very demanding in terms of digital technology, such as radiology, histopathology and ophthalmology. Consequently, these specialties are naturally the first to be transformed by AI, but it is not possible at this stage to grasp all the concrete implications of this wider diffusion. The development of AISMDs is taking place in a context of scarcity of these specialists. One likely consequence is that the activity of these professionals will focus on the analysis of complex diagnostic situations or the development of acts with higher added value, with AISMDs being used to carry out simple or routine diagnostic tasks, under human control, the precise modalities of which will have to be specified. In the short and medium term, while AI technologies do not directly threaten to lead to the disappearance of doctors, the profession will undergo changes and their effects will be even more obvious on care support functions, as has been observed in other sectors where several AI systems for managing administrative or logistical tasks already exist.

Recommendations:

15) **Ensure that AISMDs developers clearly state the performance and limitations of their AISMDs and explain them in terms that are understandable to their users.**

16) Adapt the teaching curricula (initial and continuing training) for the medical and paramedical professions so that they train **in AI technologies** by **integrating consideration for ethical issues.**

[90] See in particular the Institut Montaigne study on the HR impacts of AI and robotisation in health: https://www.institutmontaigne.org/publications/ia-et-emploi-en-sante-quoi-de-neuf-docteur

[91] HAS, "Réunion de concertation disciplinaire", November 2017, 3 p.

CONCLUSION

AISMDs are the source of many advances. They make it possible to identify lesions that escape the human eye. They make it possible to treat very quickly cases that are fairly easy to diagnose but repetitive, and thus free up medical time for discussion with the patient. Some AISMDs also make it possible to rethink the organisation of care, particularly in hospitals, and thus improve the patient intake phases.

But a number of ethical issues associated with AI technologies applied to medical diagnosis have been presented in the four parts of this Opinion, in relation to the different situations of possible uses of these technologies. Two main areas of ethical tension have emerged.

On the one hand, there are tensions regarding **respect for the individual**, who must be respected with their particularities and in their unique context and who, in particular, cannot be transformed into a mere "consumer of care" acting within the framework of a market. The patient is of a different nature from the consumer.

On the other hand, there are tensions relating to the **need to treat all the people** concerned **fairly** so as not to create discrimination. There is a fundamental issue in artificial intelligence diagnosis which echoes a constant tension in public health between the individual and the collective: how to treat each person considered in a specific situation while treating all people fairly? It is to this difficult exercise that the introduction of AISMDs into medical diagnosis should make it possible to respond, thus associating singular thinking and algorithmic thinking, as described in the first part of this Opinion.

Finally, the CCNE and the CNPEN draw attention to the fact that many AI developments in the field of health prevention are emerging. It would be of the utmost importance to complement this Opinion on diagnosis with a reflection that would cover this complementary issue.

APPENDIX

Appendix No.1: Members of the working group

Catherine Adamsbaum - Guest on behalf of the CCNE
Gilles Adda (CCNE)
Erik Boucher de Crèvecœur - Guest of the CNPEN
Raja Chatila (CNPEN)
Laure Coulombel (CCNE)
Laurence Devillers (CNPEN)
Emmanuel Didier (CCNE rapporteur)
Karine Dognin-Sauze (CNPEN)
Valeria Faure-Muntian (CNPEN)
Christine Froidevaux (CNPEN)
Jean-Gabriel Ganascia (CNPEN)
David Gruson (rapporteur CNPEN)
Emmanuel Hirsch (CNPEN)
Claude Kirchner (CNPEN)

Camille Darche (CNPEN editor)
Lucie Guimier (CCNE editor)
Marguerite Schweizer (CCNE intern)

Appendix No.2: List of persons heard by the working group

Isabelle Adenot, President of the National Commission for the Evaluation of Medical Devices and of Health Technologies (CNEDiMTS)
Hélène Bruyere, team leader, compliance assessor (ANSM);
Guillaume Chassagnon, Senior Lecturer, hospital practitioner, Thoracic Imaging Unit, APHP, Centre-Université de Paris - Hôpital Cochin;
Corinne Collignon, Head of the Department of the Digital Health Mission at the French National Authority for Health (HAS);
Stéphanie Combes, Director of the *Health Data Hub;*
Frédéric Dittenit, Data Protection Officer (ANSM);
Antoine Flahault, epidemiologist and professor of public health at the University of Geneva;
Virginie Gaiffe, Head of the Diagnostic Products, Radiogenic Systems and Information Systems Team (ANSM);
Dominique Leguludec, President of the HAS College;
Thierry Sirdey, Head of the Market Evaluation and Control Unit at the Medical Devices Evaluation Directorate (ANSM).

Appendix No.3: Recommendations and warning points of the Opinion

Warning points:

1) To avoid the risks of incompleteness and incidentaloma, it is essential that the meaning of a measure that assesses the trade-off between precision and sensitivity is fully clarified by industry.

2) An AISMD, while it may be reassuring because of its rigorous and automatic operation, nonetheless plunges both the patient and the care team into a certain degree of uncertainty. The maintenance of human control during the use of an AISMD appears to be essential but will not necessarily remove the uncertainties.

3) AISMDs must be considered as complementary tools for responding to the inadequacies of the health care system, in particular medical desertification, but must not be considered as a substitute for medical teams.

4) The decision-making process leading to the use of AISMD must imperatively take into account the concern for care before economic considerations. It must be considered as a means of combating social and territorial inequalities in health.
5) Maintain vigilance against the risks of adverse selection that could be operated by supplementary insurance companies, in particular by means of actuarial diagnostic tools based on AI techniques.

6) Take into account the risk of drifting into non-professional advice, unsecured data processing, and the illegal practice of medicine, which accentuates the vulnerability of individuals. AISMDs must be subject to strict medical practices.

7) Particular attention should be paid to the risks of surveillance and adverse selection in insurance and public health.

Recommendations:

1) It is essential that already established diagnostic methods, not involving a priori AISMD, continue to be taught and researched to advance them.

2) The growing importance of AISMD in the field of medical skills requires in-depth studies on the interaction between humans and artificial intelligence technologies to assess the impact of AISMD in the practice of medicine.

3) In the interests of transparency and traceability, the use of an AISMD should be indicated in the medical report of a consultation.

4) These elements converge in favour of human control at all stages of care, from the indication of examinations to the results of analyses and the contextual interpretation of these results.

5) Initiate a national consultation on the affordability of AISMDs for users of the healthcare system. In this respect, the identification of a section dedicated to the financing of AI in the Social Security financing laws could be examined. In this debate, the role of complementary

protection actors in identifying and stimulating the accessibility of AI innovation should be carefully examined.

6) Create the conditions for trust by encouraging developers to provide a certain level of explicability of the AISMD they put on the market.

7) Promote the development of new professional skills to make the properties of AISMDs explicit to carers.

8) Create the conditions for trust by promoting the development of justifications for promises and, on the other hand, the publication of negative research results, i.e. those that do not confirm the announced promises.

9) Encourage multi-disciplinary research and professional training to develop an unbiased definition of compliance monitoring for AISMD.

10) Promote the use of clinical trials to assess the benefit-risk ratio of AISMD in the same way as for medication.

11) Ensure that potential biases related to the programming of the algorithms themselves and to the different databases used for the development of AISMDs, which training data is specific to their date of conception, are limited.

12) Promote the definition, by the industry, of criteria that will allow the evaluation of an AISMD from its conception ("Ethics by design"). These criteria may be generic or specific to an AISMD.

13) Establish a certain number of social criteria for the evaluation of AISMDs (medical criteria for the patient, organisational criteria for health establishments and professionals, health criteria for the population, economic criteria for all the stakeholders, etc.) enabling an ethical assessment to be made.

14) Promote the status of digital auxiliary. This person would provide an explanation understandable for the patient or their legal representative or trusted person, of the result produced by the AISMD, that would also be representative of its functioning. This could be compared to the "advanced practice nurse" status currently being developed.

15) Ensure that AISMD developers clearly state the performance and limitations of their AISMD and explain them in terms that are understandable to their users.

16) Adapt the teaching curricula (initial and continuing training) for the medical and paramedical professions so that they train in AI technologies by integrating consideration of ethical issues.