

Ethics in AI for medicine

Mariagrazia Fugini, Piercarlo Maggiolini
DEIB - Politecnico di Milano Milano, Italy
[mariagrazia.fugini,piercarlo.maggiolini]@polimi.it

Abstract— The area in which probably, from a social point of view, the greatest benefits for humanity are expected from the applications of Artificial Intelligence (AI) is medicine. If it is true that there are still many people, especially the elderly, not to mention those who live in poor countries, who live, more or less happily, without computers and the , and who use mobile phones – if they own one – practically only to make calls, today, at least in the most economically developed countries, but not only, one cannot think of being cured without the aid of medical technologies that are now widely used. So, it is highly likely that this will happen soon with AI.

If with AI it is possible to have faster and more precise diagnoses, to identify more effective therapies for each specific patient, it will become increasingly difficult to perform health services without AI in hospitals, clinics, pharmacies (and we are not talking about medical schools and research¹ laboratories). But, while in most of the application areas, computers, the Internet and mobile phones, more or less intelligent, and similar devices and applications (e.g., home automation, navigators, etc.) people's lives (and deaths) are not at risk, when you enter the field of medicine we touch human flesh to the core, and every problem can be literally vital, fatal. The risks of medical technologies can be dramatic for people's lives.

Therefore, the ethical challenges posed by the application of AI in medicine are also particularly important and crucial. We discuss AI in medicine and ethical implications in explainable AI (XAI).

Keywords. ethics, health systems, machine learning, Explainable AI

1. Introduction

The applications of AI in medicine, although relatively recent and not fully tested, have had a great impact on clinical practice, healthcare and medical research. The WHO (World Health Organization) recognizes the great benefits that AI could bring to health systems, including improving public health and achieving universal health coverage. However, as described in the WHO guideline "Ethics and governance of artificial intelligence for health: guidance on large multi-modal models", it carries significant risks that could undermine public health and jeopardize the dignity of the individual, privacy and human rights. Although AI applications are relatively new, the speed of their adoption and diffusion has led the WHO to provide these guidelines to ensure that they are used in the best and most sustainable manner around the world. The WHO acknowledges that these guidelines are being issued at a time when there are many conflicting views on the potential benefits and risks of AI, the ethical principles that should be applied to its construction and use, and approaches to governance and

¹ It is no coincidence that the 2024 Nobel Prize in Chemistry was awarded to biochemists who literally produced new proteins with Artificial Intelligence.

regulation. As these guidelines were published shortly after the first applications of AI in healthcare and before more powerful models are released, the WHO will update the guidelines to keep pace with the rapid evolution of the technology, with the ways in which society manages its use, and the health consequences of using AI, beyond healthcare and medicine.

Broadly speaking, AI is revolutionizing healthcare and medical research with significant interventions in the areas of personalized medicine, continuous monitoring of vital parameters, robotic surgery and in general in the management of information regarding patients and the economic/financial management of healthcare.

Technologies in medicine are mainly attributable to the areas classified as Internet of Things (IoT), Medical software, E-healthcare tools (electronic medical record, apps and wearable devices), Remote monitoring (e-health, m-health and telemedicine), Health robotics, and in general, Smart health applications. Smart Health refers to the use of: wearable devices, wearable devices with vital signs monitoring and notification functions; implants, injectable devices or substances that operate in "active ink" mode when illuminated by a special light (for explorations, analyses, etc.); 3D printers, capable of producing tissues and organs modeled in detail, useful in preoperative planning.

In this paper we survey recent applications of AI in medicine and discuss ethical aspects related to explainability

2. AI technologies in medicine

Here we illustrate some of the main AI technologies, with reference to their usability in medical applications.

1) *Machine Learning*

Machine Learning (ML) is a sub-area of AI that trains the machine on how to learn. Part of the market hype around ML includes its subset, *deep learning*, which uses learning algorithms that can classify complex patterns such as images.

The idea of ML is related to autonomous sensing and decision-making in a dynamic environment. It has rapidly developed in the area of software agents and autonomous vehicles and is growing rapidly in healthcare, thanks to the advent of wearable devices and sensors that use data to verify a patient's health status in real time, which we have previously mentioned. In addition, this technology helps medical advisors with analysis, identifying trends or warning signs that could lead to better drug diagnoses and treatments. The renewed interest in ML is due to the same *factors* that are making *Bayesian*² data mining and analytics³ more popular than ever: the growth in the volume and variety of data, cheaper and more powerful processing processes, and increasingly cheap data storage spaces. This means that you can now automatically build models for the analysis of larger, more complex data, and quickly process more accurate results even at scale. Accurate modeling allows organizations to identify new opportunities or efficiency and effectiveness, or avoid unforeseen risks. By using algorithms to build ML

² *Data mining* is the set of techniques and methodologies that have as their object the extraction of useful information from large amounts of data, through automatic or semi-automatic methods. See the Wikipedia entry: https://it.wikipedia.org/wiki/Data_mining.

³ https://it.wikipedia.org/wiki/Statistica_bayesiana.

models that can unravel correlations, organizations can make better decisions without, or with limited need, for human intervention.

In medicine and healthcare, we talk about *data-driven health care*⁴ which, thanks to ML, benefits from interoperability⁵ to improve outcomes and efficiency.

On the *outcomes* side, it is about having ways of integrating data from multiple sources into a unified and shared platform to gain actionable insights to improve health outcomes, quality of care, processes, and user satisfaction.

Another outcome is the ability to make informed and transparent decisions about patient care, *efficiently* integrating health and non-health data from multiple sources into a unified and shared platform to improve clinical, financial, and operational efficiency.

By considering clinical risk factors and demand forecasting, ML allows you to understand risk factors in individuals and populations through unified and shared platforms to make informed decisions about care pathways. You can forecast demand and services needed, and make transparent decisions about patient care and regulatory compliance of measures.

Integrated bias monitoring and repeatable explanations of data, models, and forecasts give you peace of mind to make equitable and ethical healthcare decisions.

When measuring return on investment (ROI) in healthcare analytics, AI algorithms now enable the creation of personalized visualizations that help quantify the impact and ROI on workflows, care outcomes, user satisfaction, and programs to convince healthcare management to continue investing. Creating a comprehensive *digital patient profile* enables patients, providers, and healthcare providers to use data to make healthcare decisions to improve satisfaction and trust in the healthcare system.

Finally, AI can help understand risk factors in individuals and populations through unified and shared platforms to make informed decisions about care pathways.

2) *Deep Learning*

Deep learning is a type of ML that trains the computer to perform tasks similar to humans, such as recognizing spoken language, identifying images, or making predictions.

Instead of organizing data to run predefined equations, *deep learning* sets some basic parameters on the data and instructs the computer to learn on its own, recognizing patterns through successive layers of processing.

Deep learning may still seem to be in development, as researchers and data scientists are still testing its capabilities. In fact, this branch of ML offers numerous practical applications, which are already in use in various industries. In addition, thanks to continuous advances in research, many more will soon be available.

In the medical/healthcare field, *deep learning* applications ensure better patient care and greater operational efficiency. In fact, healthcare professionals can analyze data faster and more accurately and create electronic health records faster and with fewer errors, using speech recognition with natural language processing tools. In addition, neural networks, combined with image recognition, are able to analyze medical images, rather than just read them, thus helping healthcare personnel, for example, to identify tumors and monitor their *progression*.

Deep learning, known as *Analytics for Healthcare*, is useful in the following areas:

a) *Processes, capacity management and costs*

⁴ See some examples in <https://www.forumipa.it/sanita/bianca-frigerio-microsoft-italia-data-driven-health-un-nuovo-approccio-per-cure-mirate-e-sanita-di-valore/>

⁵ The ability of a computer system or product or service to cooperate with other systems or products or services.

AI-powered automation supports operational processes, workflows, and asset management across the entire healthcare ecosystem. Simplify operations by applying AI-powered predictions and optimization algorithms.

b) *Workforce demand and planning*

It predicts which patients will use services to optimize workforce planning. It enables bed managers and hospital managers to use organizational resources efficiently using insights from an intuitive, *low-code/no-code interface*⁶.

c) *Patient outcomes and experience*

Efficiently integrate data from different health and non-health data sources into shared platforms to better understand patient needs and improve outcomes and user satisfaction.

d) *Analysis of health documentation*

It enables text analysis to structure information for medical staff, improving accuracy and efficiency.

e) *Contact Center Analytics*

It provides insights into the most common topics and issues to improve user satisfaction and call resolution times.

3) *Natural Language Processing*

Natural Language Processing (NLP) is a branch of AI that supports computers in understanding, interpreting, and using human language. NLP helps computers communicate with humans in their own language, enabling them to read text, listen to voices, interpret them, measure sentiment, and determine what content is most important. NLP draws from many disciplines, including computer science and computational linguistics, and aims to bridge the gap between human communication and computer understanding.

It uses various techniques: from statistical and ML methods to rule-based and algorithm-based approaches. The multiplicity of approaches is essential, as textual and vocal data vary greatly, as do practical applications.

The basic tasks of NLP include tokenization (inserting tokens or labels that classify a text or parts of it)⁷ and parsification⁸ (or syntactic analysis), lemmatization⁹ (or root detection), detection of parts of speech, language recognition and identification of semantic relationships, activities similar to logical and grammatical analysis. These tasks are used for more complex NLP capabilities¹⁰. In all cases, the goal is to have raw linguistic data as input and transform or enrich the text, using linguistics and algorithms, in order to give them greater value.

NLP has an enormous, yet untapped potential for unstructured word processing. As examples of the application of NLP in medical diagnosis, we mention the systems of

⁶ Main article: *Low-Code and No-Code Technologies: Differences and Benefits – Digital Future* <https://futurodigitale.infocert.it/agenda-digitale/tecnologie-low-code-e-no-code/>.

⁷ See "tokenization" – Accademia della Crusca (<https://accademiadellacrusca.it/parole-nuove/tokenizzazione/23537>).

⁸ See "parsificare" – De Mauro Italian Dictionary (<https://dizionario.internazionale.it/parola/parsificare>)

⁹ Vedi Wikipedia: <https://en.wikipedia.org/wiki/Lemmatization>

¹⁰ such as: *Content categorization*: Provides a language-based document summary, including search and indexing, content warnings, and duplicate detection. *Automatic topic modeling and discovery*: Accurately identify meaning and themes in a collection of texts and apply advanced analytics, such as optimization and forecasting, to content. *Contextual extraction*: Automatically extracts structured information from textual sources. *Sentiment analysis*: Identify mood or subjective opinions within large volumes of text, including average sentiment and opinion mining. *Speech-to-text and text-to-speech conversion*: Turn voice commands into written text and vice versa. *Document summary*: Automatically generate synopsis from large text sources. *Automatic translation*: Automatically translates content from one language to another, starting with written text or voice input.

assistance to doctors in the diagnosis of various diseases. A recent study by the non-profit *Kaiser Permanente*¹¹ has demonstrated the validity of NLP technology in hospitals for doctors who, thanks to it, have been able to identify more than fifty thousand patients with aortic stenosis. Researchers at *Permanente Medical Group* trained NLP to sort and — through more than one million electronic health records and echocardiograms — identify certain abbreviations, words, and phrases associated with aortic stenosis. In a few minutes, the software developed by the Californian hospital group identified the approximately fifty thousand patients suffering from aortic stenosis, a process that, manually, would have required years of analysis.

In addition, there is a large amount of important clinical data in patient records that remain unusable due to their unstructured nature. NLP technologies can improve the efficiency and feasibility of research by allowing large groups of patients to be analyzed, without the need for manual reading of medical records.

NLP can also help doctors overcome the limitations of procedure and diagnosis codes, codes that are not currently designed to include detailed data on a specific medical condition. For example, the medical status of a patient with moderate or severe aortic stenosis is completely different from a patient with mild disease affecting the aortic valve. These variables are not included in the diagnosis or procedure codes. In addition, some codes may simply label the pathology as "aortic valve disease," which could be applied to a clinical problem completely different from aortic stenosis.

Population health management and research into chronic conditions are only possible with accurate and systematic identification of clinical cases. Unstructured data in medical records is the most suitable for NLP-based solutions to identify conditions in patients¹².

4) *Computer Vision*

Computer Vision is a field of AI that trains computers to interpret and understand the visual world. Using digital images from cameras and video, and through *deep learning* models, machines can accurately identify and classify objects and then react to what they "see." From facial recognition to analyzing a live video, *computer vision* rivals and surpasses human visual abilities in many areas.

In the medical field, *computer vision* systems can meticulously examine MRI, CT and X-ray images to detect abnormalities with the same precision as a doctor's eye. In addition, healthcare providers use neural networks on three-dimensional images such as ultrasound, to detect visual differences in heart rate and more.

Current vision systems are able to go a step further and take actions based on image understanding using different techniques:

- Image segmentation breaks down an image into multiple regions or parts, which are examined separately.
- Object detection identifies a specific object in an image. Advanced object detection recognizes multiple objects within a single image.
- Facial recognition is an evolution of object detection that allows you to not only recognize the human face in an image, but also identify a specific individual.

¹¹ <https://about.kaiserpermanente.org>

¹² What Are the Benefits of Natural Language Processing Technology? - *HealthITAnalytics.com* by *TechTarget* (<https://healthitanalytics.com/features/what-are-the-benefits-of-natural-language-processing-technology>).

- Edge detection is a technique used to identify the outer edge of an object or the lines of a landscape, in order to understand exactly what elements are present within the image.
- Pattern detection is the process of recognizing shapes, colors, and other visual indicators that occur in images.

5) *Generative AI*

The generative AI model is a very promising example of unsupervised ML. Previous generative¹³ models have shown various limitations as they lacked the fundamental capacity for generalization¹⁴.

The need to adopt generative AI in healthcare has been evident since the Covid-19 pandemic: it has become even more evident how much there is a need for medical and nursing staff. 81% of physicians reported levels of burnout.

Generative AI in healthcare is an opportunity that has long been used in various fields of health. The adoption of generative AI models in medicine, on the other hand, is recent and still limited, but it is very promising. By using machine learning algorithms to *generate new* textual, sound and image data, it allows you to analyze large data sets and generate new (and valuable) information, allowing you to improve diagnosis, care and treatment.

Generative AI models are useful for creating new drugs: there are already those who are implementing it. A team of researchers from Stanford Medicine and McMaster University have adopted generative AI to create a model, called *SyntheMol*¹⁵ (to synthesize molecules), capable of creating indications for chemists, useful for synthesizing drugs in the laboratory. The team created structures and chemical recipes for six new drugs, aimed at annihilating one of the main pathogens responsible for many deaths due to antibacterial resistance. Generative models are useful for improving diagnoses. When trained on large datasets of medical records and images, they allow you to identify and abstract disease-related patterns. In this regard, GANs were used to reconstruct, synthesize, segment, record and classify images. In particular, they made it possible to generate synthetic medical images that can be used to train machine learning models for image-based diagnoses or to improve medical datasets. In addition, generative AI can be applied for personalized medicine. Using generative AI, a patient's genetic makeup, lifestyle, and medical history can be analyzed to predict how they might respond to different treatments.

Compared to fears in the face of the application potential, the merits of generative AI in healthcare are different, but already a priori, doubts and critical points persist to date regarding its concrete application. Despite the potential benefits, the utility and impact of generative AI in healthcare remains poorly understood due to ethical and medico-legal concerns. It is a shame because, as evidenced by an article published in *Nature*, by a team from the *Duke-Margolis Institute for Health Policy*, Duke University in the United States, it was highlighted that:

¹³ Sakirin T., Siddhartha Kusuma (2023) - A survey of generative artificial intelligence techniques - *Babylonian Journal of Artificial Intelligence*. pp. 10-14 (<https://mesopotamian.press/journals/index.php/BJAI>).

¹⁴ Alter S. (2024) - Validating generalizations about AI and its uses - Proceedings of the 57th Hawaii International Conference on System Sciences | 2024 (<https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/7808da71-2301-46cb-8731-4ab471ada144/content>).

¹⁵ See: <https://www.labworld.it/synthemol-il-nuovo-modello-di-intelligenza-artificiale-generativa-che-sintetizza-nuovi-farmaci/>

AI algorithms will only work based on the accuracy and completeness of the underlying data, and the quality of the data depends on actions and workflows that encourage trust

In any case, the use of generative AI in the medical-health field has already taken hold. According to the results of a survey conducted by the Deloitte Center for Health Solutions, 75% of major healthcare companies are already experimenting with generative AI or attempting to expand it at the enterprise level. In addition, 82% have or plan to implement a governance and oversight structure based on generative AI; 92% of managers believe it is promising for improving efficiency, and 65% of them see the possibility of faster decision-making.

Models such as GPT-4 are already assisting physicians in the analysis and interpretation of complex clinical data and, in the context of medical education, simplifying the phases of searching, selecting and processing sources, offering quick answers to clinical questions during medical practice.

Another generative AI tool in medicine is *MedQuestio*¹⁶ which, compared to *ChatGPT*, while using the latest AI language model platforms, differs in that it is focused on providing answers based on validated scientific sources (*MedQuestio* extracts answers directly from scientific articles). This ensures that the information you provide is always based on authoritative research and studies. Basically, it is structured as a Virtual Assistant (not a simple search engine) that helps in the process of searching for reliable scientific information on medical libraries such as *PubMed* allowing a dynamic scientific update, typical of medical practice in the field.

A professional *chatbot*¹⁷ provides immediate answers to queries (while searching on *PubMed* may take longer to filter and analyze the results).

What illustrated so far is summarized in Table 1.

¹⁶ see: <https://www.medquestio.it>

¹⁷ software that simulates and processes human conversations allowing users to interact with digital devices

Table 1. AI technologies and their applications

Areas of AI Technology	Experience				
Machine Learning	Understand risk factors in individuals and populations	Wearable devices	data-driven health care Digital patient profile	Demand and services forecasts	Development of unified and shared platforms to make informed decisions about care pathways.
Deep Learning	processes, capacity management and costs	Healthcare Workforce Demand and Planning	Patient outcomes and experience	analysis of health documentation	Contact Center Analytics
Natural Language Processing	communication with humans in their own language, reading a text, listening to voices, measuring sentiment	statistical, machine learning, rule-based and algorithm-based methods.	A variety of approaches is essential	systems of assistance to doctors in the diagnosis of various pathologies	Benefit in interpreting unstructured data in folders
Computer Vision	Training to interpret/understand the visual world Using deep learning models	Using digital images from cameras and videos	facial recognition, analysis of a live video,	Exceeds human visual abilities:	Machines can accurately identify and classify objects and react to what they see
Generative AI	Unsupervised machine learning overcoming previous generative models thanks to the fundamental capacity of <i>generalization</i>	Allows you to generate realistic and anonymous synthetic (artificial) patient data for research and training	Wide range of versatile applications thanks to the generation of synthetic data (e.g., creating new drugs)	improve patient care because it generates new data, learning from the input data	Revolutionizes The diagnostic process of diseases and treatment options

3. Ethical aspects in AI

Considering the ethical aspects of what has been illustrated so far, the question arises:

does medical technology dequalify doctors or discriminate?

For example, does the ethics of the machine learning model ensure that individuals are not treated based on race, gender, and disabilities? Does model transparency provide the end user with information about how the model works? An important, essential factor – as we have widely said – is the explainability of machine learning, i.e. providing the user with information on why the model works in a certain way. A reliable model provides security to the user, allowing them to operate the system with no worry.

Another area of focus in machine learning is privacy, which must ensure the authorized use of services.

Further reflections concern the skills of doctors. Studies have shown that healthcare personnel are progressively exposed to technology for different purposes, such as

collecting medical records or diagnosing them. In addition, the discussion expands with considerations of how excessive use of technology could hinder the expansion of physicians' expertise and clinical procedures. Among the main problems that emerge in the literature is the possible dequalification of health personnel due to the reduced autonomy in the decision-making process concerning patients.

Therefore, the challenges are also broadened by considering the ethical implications of technology and the role of skills.

B. Legal framework

Among the key regulatory aspects is the GDPR, which regulates the use of personal data. Its application varies at national level, especially in the area of sensitive data used for the public interest. The EU AI Regulation introduces a risk-based approach to classify AI systems according to four levels of risk (low, medium, high and unacceptable), indicating the specific rules for each category. High-risk systems, such as some AI-based medical devices, must comply with stringent requirements for transparency, safety and reliability.

The European regulatory framework also includes other regulations, such as the Open Data Directive (Directive (EU) 2019/1024) and the Data Governance Act, which promote secure health data sharing and interoperability. The latter is crucial to ensure that data from different sources, such as medical devices and health information systems, can be exchanged and used effectively.

The study suggests the need for more detailed regulation to address issues related to legal liability in the event of AI malfunctions. In addition, it is important to set clear standards to ensure that patients have control over their data, especially in the context of machine learning and the use of algorithms that can influence clinical decisions.

Current EU policies reflect a balance between innovation and the protection of patients' rights, but the rapid development of AI technologies requires constant updating of regulations. The European institutions, together with the member states, are working to create a regulatory ecosystem that fosters the responsible adoption of AI, promoting research and the ethical use of new technologies in the health sector.

C. Principles and challenges

As seen above, AI is rapidly emerging as a fundamental tool in medicine, capable of revolutionizing the way patient care is provided and clinical activities are managed. A survey by Univadis Medscape Italia – an information portal for health professionals with news, tools, updates and continuing education – investigated the perception of Italian doctors with respect to the use of AI, interviewing a sample of 1133 doctors, which highlighted how AI is on the one hand considered an opportunity to improve patient care but on the other hand generates concerns about its implementation and ethical and professional impacts.

The results show that a significant portion (40%) has a neutral position, while 18% express concerns especially with regard to ethics and professional autonomy. Currently, the use of AI in medical practice is still in its early stages: only 21% of Italian doctors use it for research on pathologies and 11% use it for administrative tasks. 6% of doctors said they have already used it for the electronic health record before a visit, while 66% of doctors rate the use of AI for the interpretation of imaging examinations positively

because of AI's ability to improve the quality and accuracy of diagnoses. Doctors' opinions are more divergent in relation to communication with patients, a particularly sensitive aspect of the medical profession that many professionals fear could be compromised by the excessive use of automated technologies: 21% remain concerned about the possible negative impact on the independence of clinical judgment. One fact that respondents agree on is the ability of AI to reduce errors. There is therefore a strong demand for regulation and supervision. 85% of doctors believe that the use of AI should be closely monitored by governments or medical institutions to ensure its application in a safe way and in accordance with professional and ethical standards. In addition, 88% support the need for specific legislation regulating the use of AI in healthcare, especially to ensure the protection of sensitive data.

In this perspective, as already mentioned, in January 2024 the WHO (World Health Organization) published the document "Ethics and governance of artificial intelligence for health: guidance on large multi-modal models", a document of guidelines on generative AI (also known as Large Multimodal Models, large multimodal models, which also includes chatGPT). The Italian edition of these guidelines has been published by Zadig, in collaboration with the Italian Society of AI in Medicine (SIIAM).

The document (WHO) addresses the ethical and governance challenges associated with the use of AI in health, updating the contents of the previous "Ethics and governance of artificial intelligence for health" published in 2021 by the WHO. The document contains over 40 recommendations addressed to governments, technology companies in the artificial intelligence sector, health professionals, with the aim of ensuring a responsible, sustainable and inclusive use of generative AI for health and in healthcare, in compliance with some key ethical principles:

1. protect autonomy: control over health systems and medical decisions must always remain with humans;
2. promote well-being and safety: the effects on use are positive for health without risk;
3. ensuring transparency: reference is made to the explainability and comprehensibility of AI technologies;
4. to promote responsibility and accountability: in order to ensure appropriate use and by adequately trained people;
5. ensuring inclusiveness and fairness by avoiding prejudices and monitoring the discriminatory effects that are already present in these models;
6. promoting sustainable AI for health systems, the environment and workplaces.

AI algorithms today are being implemented faster than our ability to fully understand their potential and fragility. Governance must therefore aim to keep pace with their rapid development and increasing use, without favoring either governments, which seek a technological advantage, or companies, which seek commercial gain. Ethical principles and human rights obligations must be at the heart of appropriate governance, including

both procedures and practices that may be introduced by companies, as well as laws and policies enacted by governments.

4 Concluding Remarks

We have illustrated AI applications in medicine and discussed aspects of ethical relevance related to the use of AI in medicine.

Among the main concerns, we mentioned the fact that the systems used have not been sufficiently tested and supported by scientific evidence. Methodologically more robust clinical trials involving multiple centers, hospitals and institutes (multicenter) should be conducted and evaluating the effects randomly (randomized) on a representative sample of the population examined, from the beginning to the end of the studies.

These are some of the considerations contained in the document Guidelines on the use of artificial intelligence systems in diagnostics, published by the Ministry of Health,¹⁸ which represents an important tool for those who develop AI systems and for those (doctors, patients, family members, caregivers) who will then use them.

AI systems, then, must be *adequately trained* to avoid bias in evaluation. In fact, various cases of AI tools that failed to answer questions (diagnostic, prognostic, predictive) because the patients for whom the answer was sought were not adequately represented in the sample with which the system had been trained are documented.

Then there is the phenomenon of the black box, i.e. the propensity of the artificial intelligence system to provide difficult answers.

Finally, it is necessary that these tools are also regulated at an institutional level in accordance with the new European legislation on medical devices (to which these tools largely belong), through more stringent rules, in terms of requiring safety and efficacy tests, for the purpose of their approval and introduction on the market.

References

- [1] Chauhan, Chhavi, and Rama R. Gullapalli. "Ethics of AI in pathology: current paradigms and emerging issues." *Artificial Intelligence in Pathology* (2025): 159-180. S. Liu, "Wi-Fi Energy Detection Testbed (12MTC)," 2023, gitHub repository. [Online]. Available: <https://github.com/liustone99/Wi-Fi-Energy-Detection-Testbed-12MTC>
- [2] "Savulescu, J., Giubilini, A., Vandersluis, R., & Mishra, A. (2024). Ethics of artificial intelligence in medicine. *Singapore Medical Journal*, 65(3), 150-158.
- [3] Stahl, B. C. (2021). *Artificial intelligence for a better future: an ecosystem perspective on the ethics of AI and emerging digital technologies* (p. 124). Springer Nature.

¹⁸ https://www.salute.gov.it/imgs/C_17_pubblicazioni_3218_allegato.pdf