

# ARTIFICIAL INTELLIGENCE AND MEDICINE-TEN LESSONS LEARNED (AND FORGOTTEN): 1970-2026

## INTELIGENCIA ARTIFICIAL Y MEDICINA: DIEZ LECCIONES APRENDIDAS (Y OLVIDADAS): 1970-2026

Víctor Maojo<sup>1,2</sup>; Casimir A. Kulikowski<sup>3,4</sup>

1. Professor of Artificial Intelligence, Universidad Politécnica de Madrid.
2. Member, Royal National Academy of Medicine of Spain.
3. Board of Governors Professor of Computer Science, Rutgers University, EE.UU.
4. Miembro de la National Academy of Medicine, EE.UU.

### Keywords:

Artificial Intelligence;  
Medicine;  
Expert systems;  
Machine Learning;  
Deep Learning;  
Generative AI.

### Palabras clave:

Inteligencia Artificial;  
Medicina;  
Sistemas expertos;  
Aprendizaje automático;  
Aprendizaje profundo.  
IA generativa.

### Abstract

The rise of Artificial Intelligence (AI) in recent years, driven by the emergence of generative AI systems, has had an enormous scientific, technological, and social impact, yielding substantial results and promising outcomes across all areas of medicine. This article analyzes the current state of AI in medicine, comparing various topics and areas with past experiences of AI researchers since 1970. We suggest ten lessons learned from the successes and failures of these years, and how some of these shortcomings are being repeated now in similar ways, potentially delaying of the promises of change in future medicine.

### Resumen

El auge de la Inteligencia Artificial (IA) en los últimos años, debido a la aparición de sistemas de la llamada IA generativa, ha causado un enorme impacto científico, tecnológico y social, con sustanciales resultados y promesas en todas las áreas de la medicina. En este artículo se analiza la situación actual de la IA en medicina, comparando diversos temas y áreas con experiencias pasadas vividas por los investigadores en IA desde 1970. Sugerimos diez lecciones aprendidas de los éxitos y fracasos de estos años, y cómo algunas de estas deficiencias se repiten ahora de manera similar, lo que podría retrasar las promesas de cambio en la medicina del futuro.

## INTRODUCTION

Artificial Intelligence (AI) is a scientific and technological discipline, traditionally divided into two conceptual subfields: (1) symbolic and (2) connectionist. From a practical perspective, we can consider two fundamental stages in the history of AI in medicine. The first stage, dominated by knowledge-based systems, spanned from 1970 to approximately the mid-1990s. The second stage, dominated by data-based systems, began in the mid-1990s and continues to the present day with systems developed under the umbrella of terms such as “machine learning,” “deep learning,” and generative AI.

The AI applications in medicine that emerged from 1970 onward were a direct result of prior analyses of cognitive processes underlying medical reasoning and decision-making. Researchers such as Ledley and Lusted (1), Pauker and Gorry (2), Gorry (3), Warner (4), Tversky and Kahneman (5), Feinstein (6), Simon (7), and

others investigated topics such as: the heuristics used by physicians in their reasoning; decision-making under conditions of risk and uncertainty; the use of various types of logical reasoning for medical diagnosis and therapy; the management of diagnostic tests; the use of Bayes' theorem for probability estimation; and the statistical analysis of patient databases, among others.

The first AI systems in medicine had relatively limited clinical successes and have even been disparagingly called a “junkyard” (8). However, these early results did lead to successes such as the development of medical decision support systems, methods and techniques underlying ontologies, biomedical terminologies, electronic health records, and information retrieval techniques in systems like PubMed, among others. These pioneering systems had shortcomings and errors, in an era when there was no prior experience and research was starting from scratch. Ironically and somewhat surprisingly, many of these same shortcomings and misconceptions seem to be

### Autor para la correspondencia

Víctor Maojo  
Universidad Politécnica de Madrid  
Escuela Técnica Superior de Ingenieros Informáticos. Campus de Montegancedo UPM  
Tlf.: +34 910 672 898 | E-Mail: vmaajo@gmail.com

repeating themselves today in deep learning and generative AI applications. In this article, inspired especially by our own experience in advanced AI research in medicine since the 1960s (CK) and 1980s (VM), we highlight a number of lessons learned (and sometimes forgotten) from AI in medicine.

### AI IS NOT NEW —AND WHAT THIS IMPLIES

The emergence of ChatGPT in 2022 led to a huge increase in the number of AI users. Most people new to the field are unaware of the history of AI, which publicly emerged as a discipline in 1956 at a meeting held at Dartmouth College by ten now-almost--legendary pioneers (9).

What happened at this meeting cannot be understood without considering the development of 20th-century research in mathematics, logic, and computing such as Wiener's earlier development of cybernetics, and the Macy Conference on Cognitive Science, held that same summer of 1956. However, few professionals (even academics) seem concerned with the history of AI and its relevance to current challenges, ignoring the fact that without this background, it is impossible to fully grasp the complexities of AI.

Machine Learning for AI is based on statistics, and for this reason, it cannot be claimed to be error-free (10). As Jelinek, a pioneer in natural language processing (NLP), has described, statistical approaches have dominated this field for practical reasons (11). In essence, generative AI systems are statistically combinatorial efficient systems that often produce spectacular results, but they lack deep understandable reasoning of how those results have been computationally generated.

Can current generative AI systems, based on natural language processing or digitized images, without human control over their interaction with the external world, achieve the same level of reasoning as humans? For now, it doesn't appear to be the case, but AI is not merely a "stochastic parrot," as some practitioners claim. AI scientists cannot explain all the complexities and emergent properties of the unexpected calculations that arise within complex artificial neural networks (12), which creates a substantial new field of research. AI is a highly complex scientific and technological discipline and not just—as is frequently stated—a tool capable of performing complex intelligent tasks. In medicine, these complexities are even significantly greater.

### CONSCIOUS AND UNCONSCIOUS PROCESSES AND AI

The first, "practical" AI systems, or expert systems, had the goal to acquire the knowledge and

reasoning methods of experts in a specific domain and represent them in a computer. To achieve this, a person, called a knowledge engineer, had to extract this knowledge from the experts using methods such as interviews, protocol analysis, and other techniques (13) and produce a computer representation. These systems achieved academic success, with a few applications becoming commercially available during those first twenty years, but few were ever routinely used in clinical practice—something that still occurs today with many machine learning-based AI systems.

Why were these expert systems not widely successful for clinical practice? A powerful underlying cognitive reason is that it is impossible for experts to fully verbalize the knowledge and reasoning methods they use, and therefore, to transfer them to a computer. The acquisition of cognitive skills occurs in several stages (14,15), and after a period of approximately ten years, an expert's reasoning—and cognitive processes in general—in a specific domain becomes automated and partially unconscious. At this later stage, it is not possible to fully verbalize a person's reasoning. This limitation has been a central problem often described as the "knowledge acquisition bottleneck", becoming a barrier to the development of generalizable expert systems (16).

The authors recall a colleague, an experienced physician, who commented that he sometimes felt a difficult-to-define sensation, which he called a "gut feeling," when he thought something was missing in the diagnosis of a specific patient, but he couldn't understand or articulate what it was until he had evaluated them in depth. These types of intuitions, often referred to under the broad concept of "clinical eye," correspond to cognitive reasoning about the patient's clinical situation, often unconscious, that doctors use when examining a patient, and involve a series of issues—ethical, social, psychological, empathic, personal understanding and experiences, among others—that, as of now, are very difficult either people, or past and present AI systems to truly grasp or understand.

### AI WINTERS

The neural networks of McCulloch and Pitts (17), along with Shannon's information theory and Wiener's cybernetics, spurred an emerging field of automatic pattern recognition for modeling perception during the 1950s. In 1958, pattern recognition reached a milestone with the retina-inspired Perceptron (18). The parallels between this heuristic model and statistical inference methods were rapidly recognized, although in 1968 Minsky published the book "Perceptrons" (19), in which he pointed out the serious limitations of the simple linear mathematical models they used. This, along with the shortcomings of automatic language translation, contributed to triggering the first "AI Winter" in the 1970's.

Having personally experienced these so-called AI winters, we can refute the notion that AI actually disappeared or was completely eclipsed during these periods, as is often claimed. When Minsky's book "Perceptrons" was published, funding plummeted for about 20 years across the entire field of artificial neural networks; but during that time, knowledge-based systems, flourished with substantial funding. A reversal took place starting in the 1990s, when knowledge-based systems encountered scalability difficulties, leading to their "AI Winter", while artificial neural networks made a successful comeback, driven by advances such as the backpropagation algorithm (20) for heuristically updating network weights based on empirical performance criteria.

Will there be a third AI inter, as some suggest? It is quite likely that, among the numerous AI companies being created, only a few will survive; however, the current economic bubble does not necessarily mean that AI will disappear or face a massive crisis. Certain companies and fields of biomedical research and other application will continue to thrive, while others will disappear.

#### DATA, INFORMATION, KNOWLEDGE... AND THEORIES

In the 1920s and 1930s, the pioneers of quantum physics were obsessed with accumulating data to advance the discipline's theories. This idea stemmed from a suggestion by Einstein. Yet, when many years later, during Heisenberg's visiting him in Princeton, he reminded Einstein that many theoretical physicists had focused on collecting and analyzing data, following Einstein's earlier suggestion, during their conversation, Einstein admitted his mistake, having subsequently understood that his initial idea that theories must guide scientific research was more correct, since data alone cannot contribute to the advancement of science (21).

In the first half of the 20th century, scientists often analyzed data, making corrections that frequently favored existing theories. A significant example is that of Eddington and his team, who went to Príncipe Island in 1919 to verify, during an eclipse, Einstein's predictions in his theory of relativity, and made errors that brought them closer to Einstein's theory. Eddington was convinced of the validity of the theory of relativity, and this possible bias led to the empirical confirmation, perhaps prematurely, of the theory of relativity (22).

The authors themselves experimented with a machine learning project (1994-97) to extract clinical prediction rules from a thousand medical records of patients with rheumatoid arthritis, and found that this dataset presented numerous problems. After eliminating cases with numerous errors, only 340 cases were selected (23). Among the most significant problems with current

electronic health records (EHRs) are inaccurate data collection; misinterpretations by physicians and patients; changes in treatments over time; and transcription errors. Inadequate assumptions and analytical models, and especially different types of biases—selection, classification, measurement, demographic, temporal, data availability, algorithmic, publication, etc.—are common problems. Some current AI systems have been developed using millions of medical records (24), but many of the aforementioned problems may be hidden in these retrospective databases, whose data have been recorded at different times, in different contexts by different authors, and under different circumstances. Both the quantity and quality of medical data are key to ensuring that machine learning projects produce clinically valid results. All developers and users of these systems should remember this fundamental characteristic, which is unfortunately often ignored.

In 2024, the Nobel Prizes in Physics and Chemistry were awarded, both related to AI. The Nobel Prize in Chemistry was awarded to Hassabis and Jumper (along with Baker) for the development of the AlphaFold algorithms (25), which have made it possible to predict the three-dimensional structure of hundreds of thousands of proteins. This achievement was primarily statistical, through massive AI-based combinatorial analyses. However, the theory underlying protein folding remains unknown and has not yet been discovered using machine learning algorithms like AlphaFold or others. Discovering scientific theories remains more complex than analyzing data and extracting statistical patterns, -even such complex ones.

#### AI EVALUATIONS

When IBM's AI-based Watson system won the American television quiz show Jeopardy! in 2011, IBM decided to extend Watson to address medical applications, particularly in oncology. Following considerable media attention, the system did not achieve the expected results. In numerous public presentations given by IBM professionals, the evaluation was limited to a few cases in specific hospitals—and the present authors pointed this out to the developers in one of these presentations. This represented a significant drawback, similar to that experienced with many early AI applications in medicine. These applications tended to underperform in hospitals and universities outside their primary development environments due to the different characteristics of the knowledge and data available at each center (data diversity, inherent biases, procedures and protocols used by physicians, different technologies, etc.).

One example is PERFEX, an expert system for analyzing cardiac SPECT images developed in the

1990s at Emory University and Georgia Tech, and successfully commercialized by General Electric. It required two years of design and implementation and approximately five years of multicenter evaluation before its approval for clinical use (26). These multicenter evaluations have proven essential to ensure the validity of AI projects in medicine. In recent years, thousands of AI systems have been developed for clinical use, but many still need to undergo systematic evaluation in routine clinical practice and at centers other than those that developed them, including compliance with requirements imposed by public agencies. It may well take many years to determine their true impact on clinical practice.

### AI AS THE ORACLE OF DELPHI: A SUBSTITUTE FOR DOCTORS

One of the first and best-known AI systems in medicine was INTERNIST-I, an expert system designed to manage knowledge of numerous diseases, similar to the broad scope of practice of an internist. Its creators claimed that the diagnostic consultation style of the INTERNIST-I program resembled a "Greek oracle" (27), but it ultimately evolved into a reference system called Quick Medical Reference (QMR), rather than a true medical oracle.

In recent years, generative AI systems have emerged, capable of answering numerous medical questions (diagnoses, treatments, prevention, etc.). The oracle concept was not useful in the pioneering AI systems, and we find ourselves facing a parallel today with generative AI systems. Some current AI systems, whose authors claim to be able to predict outcomes for hundreds of diseases—in one of them, called DELPHI 2-M (28), the name chosen by its creators seems no coincidence—still need to demonstrate these results through systematic evaluations.

Proponents of generative AI systems argue that, with their analysis of large amounts of data, they contain within them the kind of reasoning and logical approaches used by human experts; but these human processes include processes such as intuition, common sense, emotions, empathy, ethical considerations, generalizations to new cases, understanding of past psychosocial problems and the environment, etc., which are still very different from AI.

### EXPLANABILITY AND INTERPRETABILITY

A key component of AI, still under development and unresolved, is so-called explainable AI, with the objective to understand and clearly describe how and why an AI system makes a specific decision or prediction.

A fundamental problem in medical AI in recent years is how to get an AI system to explain the reasons behind a decision or outcome—for example, a specific diagnosis or therapeutic recommendation—and how to ensure that the system's conclusions are easily interpretable by the user. The authors recently published a review on explainable AI in medicine (29).

Some—or many—people working in AI, without many years of experience in the field, think that the explainability of AI systems is a recent topic, forgetting that it has been studied for more than fifty years. The concept of a "black box"—a system that does not allow for an explanation of the details of its internal processes—is common in neural network-based systems, but explainability was already essential in the design of knowledge-based systems. In fact, two of the first medical expert systems, MYCIN (30) and CASNET (31,32), included explanations of their results. MYCIN had an explanation module that showed the user the list of rules activated and used to reach the system's final result. CASNET, developed by one of the authors, provided the user with an explanation of the causal or associative basis of its conclusions.

The following graph shows the different types of explainability/interpretability. We can observe that systems based on traditional IF...THEN... rules (typical of expert systems) exhibit lower precision, although with greater explanatory and interpretive capabilities, while the opposite is true in deep learning systems.

The conclusions reached by expert systems could be explained because their reasoning and knowledge were explicit and known in detail, whereas generative models, based on their predictive capabilities, still behave like black boxes as artificial neural networks. The explainability of large language models is limited, although techniques such as the so-called "chain-of-thought," models, the use of other LLMs, or hybrid, neurosymbolic models have been proposed.

A recent study showed that a large percentage of physicians who had used a generative AI system considered the system's correct explanations not to be a particularly relevant problem, as long as the system provided good decisions (33). We could say that, while pioneering AI systems encountered overwhelming distrust from clinicians, in many cases we now find the opposite situation.

### MEDICAL REASONING AND AI

The problem-solving reasoning strategies used by physicians in clinical practice include various approaches, such as the use of heuristics, the application of different types of logic (deductive, inductive, and abductive), the estimation of disease



Figure 1. Relationship between accuracy and interpretability for different AI techniques. Originally published by the authors in (29).

probabilities based on comparisons between them rather than statistics, rapid recognition of textual and visual patterns, causal reasoning, experience with previous cases, and common sense, among others.

Several AI systems developed in the first thirty years of medical applications (knowledge-based systems) could address each of these problems with some success. For example, pioneering AI systems for diagnosis included hypothetical-deductive reasoning capabilities; expert systems such as CASNET or ABEL included causal reasoning that linked the symptoms, signs, and tests used to the underlying pathophysiological processes of a specific disease; Case-based reasoning systems could compare a new patient's case to a pre-stored library of cases, selecting the most similar ones and proposing a treatment strategy. The difference

with the reasoning used by medical professionals is evident, since doctors are able to integrate all the aforementioned tasks (and others), while these systems could only simulate one of them.

Generative AI seeks to unify all these strategies, extracting knowledge and reasoning strategies from vast amounts of data, including scientific articles, medical histories, clinical records, books, websites, and so on. Scientists developing generative AI argue that this knowledge and these reasoning strategies are ultimately integrated into the scientific data processed by the systems, which are capable of internally making complex inferences and associations that lead to results similar to those reached by doctors. Yet, they lack real-world clinical experience (like that of physicians, who can assess the full psychological and social context of their patients), and

they cannot grasp the causal relationships and pathophysiological processes already represented in expert systems like CASNET or ABEL, nor do they have the capacity to understand all the ethical aspects of medicine, which physicians learn over years of experience.

Large language models (LLMs) like ChatGPT, Claude, and others can generally produce highly consistent simulations of deductive reasoning, but they lack intrinsic knowledge of their logical foundations, since they extract this knowledge from linguistic pattern statistics of millions of texts. Therefore, they can make mistakes in their deductions because they cannot internally verify the truth of the premises or do not have sufficient correct information, and they may be prone to proposing inconsistent answers or fallacies such as “hallucinations”—conclusions that appear correct but are logically flawed. Again, although some of these LLMs can clearly outperform human professionals in specific tasks—particularly when complex calculations and large amounts of data or knowledge need to be processed—these LLM reasoning methods cannot replace the multiple capabilities, skills, and judged experience of the best physicians.

#### LIMITATIONS AND RISKS OF AI

Pioneers of expert systems and clinical decision support, such as Feigenbaum, Buchanan, Kulikowski, Shortliffe, Szolovits, Barnett, Greenes, Pople, Myers and Miller, and others, already understood the numerous limitations of AI and the dangers of accepting its decisions without a deep understanding of its operation and without the valid explanations generated by these systems. As early as 2004, the authors warned about these limitations (34). Below, we highlight some critical limitations of current AI.

#### COMMON SENSE

A popular (multicultural) saying is that “common sense is the least common of the senses.” Aristotle considered it the ability of people to form coherent judgments about the world.

Turing stated that mathematical logic cannot be sufficient to support reason without considering common sense (35). He was the first in the scientific world related to computers and AI to explicitly mention common sense. Valiant (10) mentions that we face two problems related to common sense: identifying what logic fails to grasp—a consequence of mathematical logic requiring a solid theoretical framework to function correctly—and determining the necessary scientific path to address the common-sense problem—for which we need, paradoxically, a general theory of the non-theoretical. At the dawn of symbolic AI in the 1950s, researchers

such as John McCarthy (36) and Marvin Minsky (37) identified the “common-sense problem”: the difficulty of endowing machines with basic knowledge about the world.

Humans are particularly adept at generalizing, even from a few examples, but in the era of knowledge-based systems, machines could not reason beyond the set of rules provided by their designers, and thus often made mistakes in novel or implicit situations. On the other hand, generative models do not use explicit rules; instead, they learn statistical patterns from massive amounts of data (texts or images). Generative AI developers argue that these systems capture the common-sense knowledge implicit in the texts they are trained on or in the inferences they can draw; however, while it is proposed that generative AI could improve upon this approach, it cannot yet be considered equivalent to human intelligence, since these models do not truly reason or have their own understanding of the physical world.

#### HALLUCINATIONS

Hallucinations in large language models such as ChatGPT, Claude, and many others, occur when these models generate false or inaccurate information, but do so in a way that may seem plausible and coherent to users. The causes can be diverse (38), such as limitations in the training data, errors, biases, incomplete information or the fabrication of false information, a lack of understanding of the context, inadequate formulation of the question posed to the system, or when the system cannot find an answer and responds incorrectly. It is important to remember that these systems do not truly understand the answers they provide. Therefore, they can provide answers that, from a linguistic point of view, seem correct, but which, in reality, contain false information.

Among the immense amount of data used to train these systems—which could one day encompass all documents created by humans, like a universal Borges library—there are frequently numerous errors, biases, and gaps in scientific knowledge on many subjects. LLMs can provide answers that often seem plausible, but are actually flaws in the model's generation or reasoning, due to problems such as limitations in the training data—in quality or quantity, for example, or in the lack of widely accepted knowledge on the subject—among other causes. In this way, hallucinations can lead doctors and patients to poor clinical decisions, which can add to other causes of potential malpractice from the use of AI in clinical practice, with shared responsibility on the part of the healthcare professional (39).

A clear need in clinical practice is understanding how hallucinations occur and their causes, something a physician must know to use an AI system while being aware of its limitations. These errors are inherent in any system based on data and statistical analysis.

## UNCERTAINTY AND RISK

"Uncertainty" is one of the words that best summarizes the numerous difficulties faced by AI researchers in biomedicine. Uncertainty is found in many aspects of medical documentation and reasoning, such as the multiple errors and inconsistencies found in electronic health record data; patients' subjective statements recorded in their charts; the implicit reasoning embedded in the collected data, corresponding to decisions made by physicians but not recorded in these documents; the lack of knowledge, yet to be discovered, in so many areas of medicine, including the causes of diseases; the differences in protocols and patient management among different professionals in different medical practices; the errors and discrepancies between the medical devices used in each hospital or clinic, and so on. All these factors contribute to the risk of medical decisions and proposed therapies, for which current statistical theories, still based on utilitarian economics, are completely inadequate to take into account the particular considerations of the individual patient and their ethical treatment subject to the Hippocratic Oath.

Generative AI models are essentially probabilistic. They have no way of knowing when they are uncertain about an answer, can often mislead users, and implicitly represent uncertainty through probability distributions in the dataset they handle. To overcome this key problem, developers of generative AI models have created different methods, but these still do not include new representations of the common sense and human context needed for the ethical practice of medicine.

## CONCLUSIONS

The accelerated acquisition of new cognitive skills through AI can also lead to the loss of other traditional skills (40). For example, recent articles (41, 42) warn about the loss of skills—e.g., advanced reasoning in complex medical cases. In the few years since ChatGPT's emergence, numerous AI "experts" have appeared, lacking real experience, and, worse still, thousands of scientific articles on the use of AI in medicine are published each year, with study designs of highly questionable validity in many cases.

There is currently likely an excess of expectations regarding AI and its capabilities, especially in medicine, but there is also great hope that AI will become the center of a scientific revolution comparable to previous ones in medicine. Numerous articles have suggested, even dating back to the 1970s (43), that AI systems outperform physicians in their results—a clear example being the diagnosis of mammograms in recent studies (44)—although ultimately, physicians must

still supervise the AI's results to avoid potential negative consequences for patients. In African countries, for example, the use of AI could lead to substantial improvement and change in future healthcare due resulting from the shortage of physicians.

In this article, we have selected ten lessons learned, but there could be more. For example, at the time of writing this article, a large-scale experiment has been published demonstrating the problems with using an AI-based stethoscope (45). While it achieved better results than the physicians with whom it was compared, it is not efficient due to the time required for its use and its lack of clinical integration—a problem already observed with the first medical AI systems in the 1970s (46). Keeping this in mind, healthcare and medical professionals need to understand the basic principles and limitations of these AI systems, as well as the lessons learned from over fifty years of experience.

## ACKNOWLEDGMENTS

Víctor Maojo is supported personal projects at the Polytechnic University of Madrid and by the European project SHIELD (European Union's Horizon Europe research and innovation program under grant agreement NA 101156751).

## CONFLICT OF INTEREST

The authors of this article declare that they have no conflicts of interest regarding the content presented in this work.

## BIBLIOGRAPHY

1. Ledley, RS and Lusted, LB. Reasoning foundations of medical diagnosis: symbolic, logic, probability and value theory aid our understanding of how physicians reason. *Science* 1959; 130 (3366): 9-21.
2. Pauker SG, Gorry GA, Kassirer JP et al. Towards the simulation of clinical cognition. Taking a present illness by computer. *Am J Med.* 1976 Jun;60(7):981-96. doi: 10.1016/0002-9343(76)90570-2. PMID: 779466.
3. Gorry A. Strategies for computer-aided diagnosis. *Math Biosci* 1968; 2:293-318.
4. Warner, HR, Toronto, AF, Veasey, LG. Et al. A mathematical approach to medical diagnosis: application to congenital heart disease. *JAMA* 1961; 177 (3): 177-83
5. Tversky, A. and Kahneman, D. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking

- under uncertainty. *Science*. 1974 Vol 185, Issue 4157. pp. 1124-1131
6. Feinstein AR. *Clinical Judgment*. Baltimore: Williams and Wilkins; 1967.
  7. Simon HA, CA Kaplan. C.A. in *Foundations of Cognitive Science*. M. Posner (Ed): MIT Press, Cambridge, MA; 1990.
  8. Wachter, R. *The Digital Doctor: Hope, Hype and Harm at the Dawn of Medicine's Computer Age*. McGraw Hill. 2015.
  9. McCarthy, J., Minsky, M. L., Rochester, N., et al. *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. Unpublished report. 1956
  10. Valiant, L.; *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World*. Basic Books. 2014
  11. Jelinek, F. *Statistical Methods for Speech Recognition*. The MIT Press, 1997
  12. Amodei, D. *The Urgency of Interpretability*. <https://www.darioamodei.com/post/the-urgency-of-interpretability>
  13. Pazos, J. *Inteligencia Artificial*. Paraninfo. 1987
  14. Musen MA, van der Lei J. Knowledge engineering for clinical consultation programs: modeling the application area. *Methods Inf Med*. 1989 Jan;28(1):28-35.
  15. Maojo, V. *Cerebro y Música. Entre la neurociencia, la tecnología y el arte*. EMSE EDAPP. 2018
  16. Gomez, A.; Juristo, N; Montes, C. et al. *Ingeniería del Conocimiento*. Editorial Centro de Estudios Ramón Areces. Madrid, Spain. 1997
  17. McCulloch, W. and Pitts, W. A Logical Calculus of the Ideas Immanent in Nervous Activity". *The Bulletin of Mathematical Biophysics*. 1943 Vol. 5, pp. 115-133
  18. Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*. 1958 Vol. 65, No. 6, pp. 386-408
  19. Minsky, M. and Papert, S.A. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Cambridge. 1969
  20. Rumelhart, D., Hinton, G. & Williams, R. Learning representations by back-propagating errors. *Nature* 1986 323, 533-536 . <https://doi.org/10.1038/323533a0>
  21. Heisenberg *Encuentros y conversaciones con Einstein y otros ensayos* . Alianza Editorial., 1979
  22. Dyson, F.W.; Eddington, A.S.; Davidson, C.R. (1920) "A determination of the deflection of light by the sun's gravitational field, from observations made at the solar eclipse of May 29, 1919," *Philosophical Transactions of the Royal Society A* 220: 571-581.
  23. Sanandres, Ja.; Maojo, V.; Crespo, J et al.. A clustering-based constructive induction method and its application to Rheumatoid arthritis. *Proceedings of AI in Medicine*. 2101. 2001 pp. 59 - 62.
  24. Callaway E. Medical AI trained on whopping 57 million health records. *Nature*. 2025 May 6. doi: 10.1038/d41586-025-01422-3
  25. Jumper J, Evans R, Pritzel et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021 Aug;596(7873):583-589. doi: 10.1038/s41586-021-03819-2.
  26. Garcia EV, Cooke CD, Folks RD et al. Diagnostic performance of an expert system for the interpretation of myocardial perfusion SPECT studies. *J Nucl Med*. 2001 Aug;42(8):1185-91.
  27. Miller RA, Masarie FE Jr. The demise of the "Greek Oracle" model for medical diagnostic systems. *Methods Inf Med*. 1990 Jan;29(1):1-2.
  28. Shmatko A, Jung AW, Gaurav K, et al. Learning the natural history of human disease with generative transformers. *Nature*. 2025 Nov;647(8088):248-256. doi: 10.1038/s41586-025-09529-3.
  29. González-Alday, R., García-Cuesta, E., Kulikowski, C. A., et al. A Scoping Review on the Progress, Applicability, and Future of Explainable Artificial Intelligence in Medicine. *Applied Sciences* 2023, 13(19), 10778.
  30. Shortliffe EH. *Computer Based Medical Consultations: MYCIN* New York: Elsevier; 1976.
  31. Kulikowski CA, Weiss SM. Representation of expert knowledge for consultation: the CASNET and EXPERT projects. In: Szolovits P, editor. *Artificial Intelligence in Medicine*. AAAS Selected Symposia Series. Boulder CO: Westview Press; 1982:21-55.
  32. Kulikowski, C.A. *Strategies for test selection in causal network models*. Tech. Report #TR-11. *Computers in Biomedicine*. Rutgers University. 1972
  33. Sumner J, Wang Y, Tan SY, Chew EHH et al. Perspectives and Experiences With Large Language Models in Health Care: Survey Study. *J Med Internet Res*. 2025 May 1;27:e67383. doi: 10.2196/67383.
  34. Maojo, V. *Domain-Specific Particularities of Data Mining: Lessons Learned*. *Proceedings of ISBMDA (2004)*. *Lecture Notes in Computer Science*. Pp 235-242.
  35. Turing, A. M. *Solvable and Unsolvable Problems*. *Science News*, 31, 7-23. Penguin Books, Melbourne-London-Baltimore. 1954
  36. McCarthy, J. *Circumscription—A form of non-monotonic reasoning*. *Artificial Intelligence* 1980, 13(1-2), 27-39
  37. Minsky, M. *The Emotion Machine: Common-sense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Simon & Schuster. 2006.
  38. Yubin K., Hyewon J., Shan Chen, Sh. et al. *Medical Hallucination in Foundation Models and Their Impact on Healthcare*. medRxiv 2025.02.28.25323115
  39. Wu D, Haredasht FN, Maharaj SK, et al. First, do NOHARM: towards clinically safe large language models. *ArXiv [Preprint]*. 2025 Dec 17:arXiv:2512.01241v2.
  40. Bastani H, Bastani O, Sungu A et al. Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proc Natl Acad Sci U S A*. 2025 Jul;122(26):e2422633122. doi: 10.1073/pnas.2422633122.
  41. Abdunour RE, Gin B, Boscardin CK. Educational Strategies for Clinical Supervision of Artificial Intelligence Use. *N Engl J Med*.

- 2025 Aug 21;393(8):786-797. doi: 10.1056/NEJMra2503232.
42. Fogo AB, Kronbichler A, Bajema IM. AI's Threat to the Medical Profession. *JAMA*. 2024;331(6):471-472. doi:10.1001/jama.2024.0018
  43. Yu VL, Fagan LM, Wraith SM et al. Antimicrobial selection by a computer. A blinded evaluation by infectious diseases experts. *JAMA*. 1979 Sep 21;242(12):1279-82.
  44. Gommers J, Hernström V, Josefsson et al. Interval cancer, sensitivity, and specificity comparing AI-supported mammography screening with standard double reading without AI in the MASAI study: a randomised, controlled, non-inferiority, single-blinded, population-based, screening-accuracy trial. *Lancet*. 2026 Jan 31;407(10527):505-514. doi: 10.1016/S0140-6736(25)02464-X.
  45. Kelshiker MA, Bächtiger P, Petri CF et al. Triple cardiovascular disease detection with an artificial intelligence-enabled stethoscope (TRICORDER) in the UK: a cluster-randomised controlled implementation trial. *Lancet*. 2026 Feb 14;407(10529):704-715. doi: 10.1016/S0140-6736(25)02156-7.
  46. Buchanan, B.G. and Shortliffe, E.H. Rule-based expert systems: The Mycin experiments of the Stanford heuristic programming Project. Addison-Wesley, Reading, MA.1984

---

**Si desea citar nuestro artículo:**

**Majo V, Kulikowski CA. Artificial Intelligence and medicine: ten lessons learned (and forgotten): 1970-2026. An RANM. 2026;143(01): 76-84. DOI: 10.32440/ar.2026.143.01.rev05en**

---