Special article

The long journey of artificial intelligence in medicine: an overview

E. Grossi

Villa Santa Maria Foundation, Tavernerio, Italy.

E. Grossi, MD

Please address correspondence to:
Enzo Grossi
Via IV Novembre,
22038 Tavernerio (CO), Italy.
E-mail: enzo.grossi@bracco.com

Received on March 11, 2025; accepted in revised form on April 10, 2025.

Clin Exp Rheumatol 2025; 43: 815-821.

Key words: artificial intelligence, medicine

Competing interests: none declared.

ABSTRACT

Artificial intelligence (AI) has its roots in the history of philosophy and of applied mathematics of the 17th, 18th and 19th centuries. Throughout the 20th century, significant advancements in mathematics and computer science laid the groundwork for modern AI, culminating in the establishment of the field as a formal discipline during the Dartmouth Conference in 1956.

This pivotal event brought together leading researchers who envisioned creating machines capable of simulating human intelligence, setting the stage for decades of research and innovation in the field. The development of early AI systems focused on problem-solving and symbolic reasoning, leading to the creation of programmes that could play games like chess and solve mathematical equations, which showcased the potential of machines to perform tasks previously thought to require human intellect.

As these foundational systems evolved, researchers began to explore more complex algorithms and learning models, paving the way for advancements in machine learning and neural networks that would eventually revolutionise AI applications across various fields among which medicine. The growth of big data and increased computational power further accelerated these advancements, enabling machines to analyse vast amounts of health information and learn from patterns at unprecedented speeds.

The revolution of deep learning and soon after large language models has enabled machines to achieve remarkable feats, such as image and speech recognition, natural language processing, and even creative tasks like art generation, pushing the boundaries of what was once thought possible. As organisations grapple with these challenges, there is growing emphasis on developing frameworks that ensure responsible AI deployment while maximising its potential benefits for human health.

Introduction: the long journey toward AI through applied mathematics

The attention to artificial intelligence (AI) has reached its peak in recent years, whether we are dealing with articles, the lay press or the man on the street. Doctors are faced with the task of deciding where, when and how to employ AI and of understanding its risks, problems and possibilities.

Just fifty years ago, the idea that a computer could learn and understand was still the stuff of science fiction. Today it is an integral part of our lives, helping us do everything from finding photos to driving cars. The explosion of AI in all areas of our lives, including medicine, is the culmination of a long journey of mathematical and philosophical thinking that laid the theoretical foundations, and of computer science that incorporated these advances.

Already in the 17th and 18th centuries, there were crucial philosophical developments that would later influence AI. René Descartes proposed that animals and the human body were essentially complex machines, laying the groundwork for considering intelligence as potentially replicable through mechanical means. Gottfried Wilhelm Leibniz envisioned a universal language of human thought that could be manipulated logically, foreshadowing modern computational approaches to reasoning and natural language processing. Galileo Galilei for his part declared the existence of a language of the universe, written in mathematical terms:

"*questo grandissimo libro che continuamente ci sta aperto innanzi a gli occhi* *(io dico l'universo), non si può intendere se prima non s'impara a intender la lingua, e conoscer i caratteri, ne' quali è scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi, ed altre figure geometriche, senza i quali mezi è impossibile a intenderne umanamente parola; senza questi è un aggirarsi vanamente per un oscuro labirinto*" [this great book which ever lies open before our eyes (I mean the universe), cannot be understood unless we first learn to understand the language and recognise the characters in which it is written. It is is written in mathematical language and its characters are triangles, circles and other geometrical figures, without whose help it is impossible to comprehend a single word of it; without this, we wander in vain through a dark labyrinth] (1).

We must wait till the 19th century when mathematics formally enters medicine, and some outstanding pioneers, came onto the scene. The French military physician Pierre-Charles-Alexandre Louis with the discovery of the 'numerical method' in 1837 (2) contributed enormously to the development of modern medical statistics. His research on chest diseases, typhoid and leeches, make him one of the greatest French clinicians of the 19th century.

Another legendary figure in the history of public health is Dr John Snow (1813-1858), considered to be one of the fathers of epidemiology, demonstrated the mechanisms by which cholera spread thanks to the statistical mapping methods he invented (3).

At the same time, other non-medical scientists were laying the foundations for the development of computer science. In 1834 the Englishman Charles Babbage conceived of a general-purpose device that could be programmed with punch cards (4). His analytical machine was never built, but almost all modern computers are based on its logical structure.

The young 27-year-old mathematician Ada Lovelace described a sequence of operations to solve mathematical problems using Charles Babbage's theoretical punch-card machine in 1842 (5). In the 1970s, the US Department of Defense paid tribute to her by naming a new software language ADA.

Table I. Development periods of artificial intelligence..

Period	Phases	Main focus
1940-1970	Heroic attempts with early computers	Brain simulation
1970-1985	Expert systems	Medical problem solving
1986-2000	ANN and other MLS	Diagnosis and prognosis prediction at individual level
2000-2012	Deep learning	Image recognition and understanding
2012-now	Large language models	Language recognition and understanding

Just five years later, philosopher and mystic George Boole created a form of algebra in which all values can be reduced to 'true' or 'false'. Essential to modern computing, Boolean logic helps a CPU decide how to process new input (6).

In the 1930s, inspired by how we follow specific processes to perform tasks, English logician and cryptanalyst Alan Turing, one of the fathers of computer science and known for having decrypted the Enigma code used for communications in Nazi Germany during World War II, theorised how a machine could decipher and execute a series of instructions. His published proof is considered the basis of computer science (7).

In the long journey that has led up to the present scenario of artificial intelligence we can recognise five periods, as described in Table I.

The first period: simulation of human brain

In a famous article written in 1943, Warren McCulloch, a neurophysiologist and mathematician, explained how human neurons might work (8). To illustrate the theory, he modelled a network with electrical circuits with the help of Walter Pitts. The goal of the network was to solve a problem that had been posed by John von Neumann and others: how could computers be made to communicate with each other?

This early model showed that it was possible for two computers to communicate without any human interaction. This event is important because it paved the way for the development of machine learning. However, these early artificial neural networks were not capable of learning and the synaptic values of their connections had to be predetermined by the experimenter. McCulloch and Pitts' neurons, simple as they are, represent the model on which almost all of today's neural models are based.

In the 1950s, some visionary computer scientists and mathematicians set themselves the goal with their first computers to simulate the functioning of the human brain. Among these, in 1955, John McCarthy proposed and organised the first famous seminar which took place that summer at Dartmouth College in Hanover, New Hampshire: "*An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.*" The following year he created the term "*artificial intelligence*".

This conference at Dartmouth College was attended by researchers who were either already established (such as Claude Shannon, the 'father' of information theory) or destined to become famous in their field of expertise (such as Marvin Minsky and John McCarthy). Subsequent developments were promising, *e.g.* programmes were created that could prove certain mathematical theorems or play chess.

The reference language was LISP, which influenced numerous programming languages over the decades.

In 1957, Frank Rosenblatt – affiliated with the Cornell Aeronautical Laboratory – created the perceptron by fusing Donald Hebb's algorithm on brain cell interaction with Arthur Samuel's theories of machine learning (9). The perceptron was a real machine that had thousands of wires rather than software. Called the perceptron Mark 1, it was built for image recognition, anticipating what would, 40 years later, become the idea underlying the deep learning. It was the first neurocomputer, which unfortunately, due to a series of vicissitudes, was never successful. Those who were responsible for this rejection were Marvin Minsky and Seymour Papert, both mathematicians. In their famous book published in 1969 (12), they uncovered the intrinsic weaknesses of the perceptron which lacked the ability to handle non-linear relations and which therefore, in their opinion, could not be considered a model of the brain at all. However. they realised that building a multi-level perceptron network could solve more complex problems, but at the time, the increasing computational complexity required to train networks using algorithms had not yet found an infrastructural answer (there were no hardware systems capable of 'supporting' such operations).

The second period: expert systems

With the 1970s came expert systems, *i.e.* (intelligent) programmes capable of providing a solution to complex problems falling within a specific field or domain without the intervention of a 'human' expert.

But this decade only saw a first generation of expert systems, linked to Boolean logic and logical reasoning under conditions of certainty by means of a deterministic model, which would undergo developments in the following years with the acquisition of knowledge through the contributions of domain experts (10). These systems achieved a notable degree of success by enabling the encoding and preservation of human expert knowledge, thereby rendering it accessible to a diverse array of users.

Towards the end of the 1970s, there was an exponential increase in the use of minicomputers which, being smaller and cheaper, were widely purchased, especially by companies. Consequently, the amount of documentation produced also began to grow, necessitating tools that could organise and consult it quickly.

It was in this climate that expert systems acquired a vital role, as they differed significantly from the procedural method of programming (which was widespread at the time) in their use of 'a natural application of the concept of symbolic systems' (11).

Those of the 1980s are second-generation expert systems in that they introduced the probabilistic model which, unlike the deterministic model, reasons on the causes and the possible effects. However, this model, like that of the 1970s, also has limitations, such as the fact that the most probable answer may not always be the most useful one.

Following a period of optimism regarding their potential during the 1980s, it became apparent that these systems possessed significant limitations, primarily stemming from their inability to adequately consider the intricate, multifaceted, and often implicit contextual factors that are relevant to issues pertaining to health and disease, which are predominantly resolved by human agents through the application of common sense, a fact that poses the greatest challenge when pursued through the lens of inductive and deductive reasoning. It is not coincidental that the most compelling outcomes of expert systems within the medical domain were observed in applications that exhibited considerable requirements for pattern recognition, such as the formulation of diagnoses.

The culmination of these events led to a diminished set of expectations that manifested in the first winter of artificial intelligence.

The third period: artificial neural networks and other machine learning systems

Subsequent significant breakthroughs in the ensuing years reignited interest and fervour for neural networks. I am alluding to the formulation of the backpropagation algorithm by Rumelhart and Hinton (13). Backpropagation constitutes a computational procedure that facilitates the network's capacity to learn from its errors by recalibrating its internal parameters to minimise inaccuracies, thereby enhancing the operational efficacy of artificial neural networks (ANNs). This signifies that the network progressively improves its aptitude for identifying patterns, generating predictions, and executing tasks as it continually engages with data. Hence, backpropagation can be regarded as the fundamental mechanism that rendered contemporary AI practical. It was later acknowledged that, while neural networks could not accurately mimic the human brain, they possessed formidable capabilities in addressing intricate problems.

Towards the end of the 1980s there was an unequivocal demonstration of the ability of neural networks to interpolate any function problem given enough hidden units (14). Neural networks became universal approximators. AI was experiencing its second boom. With artificial neural networks the assumptions of classical statistics were overturned.

If we could ask an artificial neural network "which assumptions underly your reasoning?" we would probably receive this answer: the variables under study are all dependent; their relationships are almost always non-linear; data must be processed dynamically; learning arises from errors; the focus is on the subject and not the variables; apparent learning must be confirmed blindly. These statements are clearly orthogonal to classical statistics assumptions.

The most powerful and well-established statistical methods were developed in the first half of the past century when the size and the quality coming from clinical observations was rather limited and certainly negligible in comparison with the big data explosion.

It is noteworthy that all these methods rely on the basic assumption that in order to apply statistical tests the variables must be independent of each other, normally distributed and, more importantly, having linear relationships between them. But how one can imagine that the variables related to a single subject are independent? Also, the normal distribution and the linearity turned out to be more an exception rather than a rule.

Complex chronic diseases have a heterogeneous origin where various mechanisms participate to a different extent in different patients. Consequently, techniques belonging to classical statistics, such as discriminant analysis, which assume linear functions underlying key pathogenetic factors, normal or quasi normal values distribution, and reduce the contribution of outliers through average computation, might incorrectly represent the complex dynamics of socio-demographic, clinical, genetic and environmental features which may interact with each other in these patients. Artificial neural networks on the contrary take advantage of modern mathematical theories inspired by life sciences and seem to be able to extract information from the data that are not considered useful for traditional approaches.

The progressive pervasive entrance of artificial neural networks and of the other machine learning systems in the medical scenario can be explained by the growing awareness of their properties: the ability to adapt to complex problems through progressive approximations and to model building by the data themselves, to analyse data more and more in depth, allowing new discoveries.

Artificial neural networks are systems of interconnected mathematical equations based on a principle inspired by the highly interactive processes of the human brain, generally composed of three processing layers with a finite number of non-linear units (*i.e.* artificial neurons). Like the brain, neural networks can recognise patterns, use data and learn by examples, just as a doctor does in the initial phase of his work.

ANNs gained popularity in problems where the relationships between the variables of interest are very complex (15-17). Based on a set of simple rules, the system attempts to learn using some of the data and apply its 'knowledge' to the rest of the available information. Their main feature is the ability to modify their internal structure in response to the data presented (18). Artificial neural networks (ANNs) are sophisticated systems comprised of interconnected mathematical formulations, drawing inspiration from the intricate and dynamic functionalities of the human brain. Typically, ANN architectures consist of three processing layers, each encompassing a finite number of non-linear units, commonly referred to as artificial neurons. Analogous to the cognitive processes of the brain, these neural networks possess the capacity to identify patterns, utilise data, and, most importantly, learn through experiential examples.

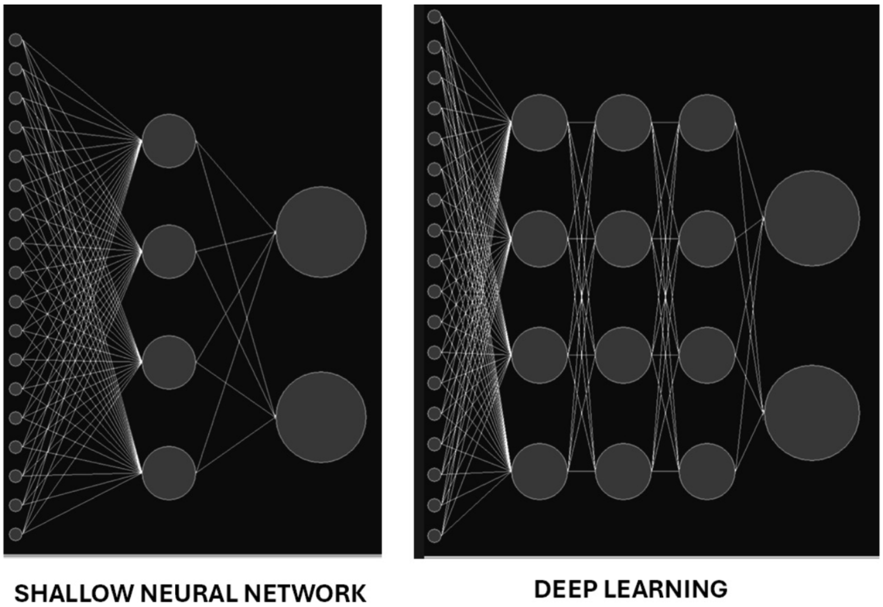

Fig. 1. Schematic of shallow neural network and deep neural network handling 18 inputs and 2 outputs.

In this framework, the initial layer and the terminal layer of the network are classified as the input and output layers, respectively; whereas, the intermediary layer, which accommodates a variable quantity of artificial neurons termed hidden units, is proficient in executing the computations related to the non-linearity inherent in a specific problem.

The utilisation of ANNs has surged in domains where the interrelationships between the variables of interest exhibit significant complexity. By employing a set of fundamental principles, the system endeavours to assimilate knowledge from a subset of the data and subsequently apply its acquired 'understanding' to the remaining available information. A salient characteristic of these networks is their inherent ability to adapt their internal architecture in response to the data encountered. In contrast to conventional statistical methodologies employed in the field of epidemiology, these models exhibit the capability to concurrently analyse all signals and accommodate the non-linear interrelations between all variables under consideration (19).

There are many examples of predictive applications of ANNs in the medical field. Referring to the author's experience, interesting results have been obtained in diseases such as dyspeptic syndrome (20), atrophic gastritis (21), venous thrombosis (22), gastro-oesophageal reflux syndrome (23), irritable bowel syndrome (24), Alzheimer's disease (25) and mild cognitive impairment (26), cardiovascular diseases (27), gastrointestinal bleeding (28), gastric cancer (29), hypercortisolism (30), anticoagulant dose prediction (31), energy expenditure in children (32), Covid diagnosis (33), among others.

The fourth period: deep learning

Around the turn of the new millennium, a significant paradigm shift emerged with the advent of deep learning (34). the feasibility of which has been facilitated by the proliferation of neuronal layers, systematically arranged in hierarchical constructs organised into 'cascades' that progressively enhance generalisation capacity. This collectively endows these architectures with remarkably refined pattern recognition capabilities, as well as the neurons and their interconnections, thus creating opportunities that were previously believed to be implausible. The corpus of literature pertaining to deep learning has experienced an unprecedented surge in recent years. A notable and exhaustive

review was published in 2023 by Haug *et al*. (35).
Figure 1 shows the transition from shallow neural networks and deep learning. Deep learning models are distinguished by an extensive array of parameters, necessitating the acquisition of knowledge from substantial datasets. Owing to the substantial number of parameters employed, these models facilitate the identification of intricate configurations within the input data and the corresponding complex relationships with the output variables. Furthermore, they enable the utilisation of all forms of data in digital format as inputs to the system, encompassing images, signals, and textual data. An intriguing characteristic of these deep models is the potential for reusing the parameters from the initial layers, which serve to 'transform' the input variables by accentuating their salient features, even in contexts that differ from those utilised during the model training phase.
These models possess noteworthy characteristics: specifically, they facilitate the mathematical transformation of input variables, thereby accentuating certain attributes and subsequently correlating these attributes with the resultant outputs.
For example, when the input data consists of the pixels from a digitised image, these algorithms permit the extraction of intermediary representations, such as the delineation of edges or the textural features inherent to the image, which can later be utilised in the ultimate prediction phase.
The augmented accessibility of digitised data, coupled with the concomitant enhancement in computational capabilities (which is further bolstered by processors engineered in the realm of digital gaming, such as Graphics Processing Units, GPUs) and the progressive refinement of deep learning methodologies have culminated in a substantial enhancement of the performance of machine learning systems, particularly within the domains of image and text analysis. In the clinical arena, this progression has engendered a marked proliferation of AI systems that, subsequent to rigorous clinical trials, have attained the status of software medical devices, thereby receiving recognition from certification entities and approval by the Food and Drug Administration (FDA) or having been CE-marked. A recent update to the catalogue of FDA-approved AI products indicates that the field of radiology is the most populous domain (679), closely followed by cardiology (90) and neurology (32). Therefore, the advancement of imaging technology epitomises the domain in which AI has now evolved into a practical reality. Within the realm of radiology, it has become feasible to conduct clinical studies adhering to high quality standards, facilitated by both the availability of sophisticated algorithms and the quality and reproducibility of the underlying data.
Other significant advancements that have emerged alongside the evolution of deep learning encompass:
(a) the development of generative adversarial networks, distinguished by the collaborative interplay between two distinct types of networks, one assigned the responsibility of identifying valid solutions, while the other is designated to enhance the former's capacity for discrimination (36).
(b) the formulation of attention mechanisms – integrated within transformer-type neural networks – which are able to deploy a selective allocation of computational resources, thereby enhancing the proficiency in learning contextual relationships (37).
(c) the introduction of Energy-based generative neural networks, a category of generative models designed to acquire explicit probability distributions of data through the utilisation of energy-based frameworks (38).
On the one hand, strategies have been developed to implement so-called 'generative' learning models. These models have been widely used, for example, to increase the quality of images by providing both the input and output of the supervised learning model with the same examples. In other words, the model is able to reproduce the input data. With a few tricks in terms of the underlying mathematical tools, the reproduction of the output data can be changed, providing a new 'version' of the input data, *e.g.* less noisy and, in the case of images, more defined. With further appropriate computational strategies, it is then possible to use these models to generate new data, *i.e.* data that deviates more from the original data. Generative models can then be used to create so-called 'synthetic data', which are data with the same statistical characteristics as the original data, but which can have a larger numerosity and have far fewer personal and clinical data protection issues.
This technological revolution has significantly transformed several aspects of our everyday life. In contemporary society, it has become exceedingly easy to peruse thousands of images on a mobile device without needing to manually annotate each one with its content. An individual is even capable of recognising objects in images with which they have no prior familiarity, such as specific varieties of flora. The utilisation of voice recognition technology has become widespread.
Individuals are enabled to translate among more than one hundred languages, whether through the input of text or by directing their camera towards written words in an unfamiliar language.
Furthermore, advancements in deep learning have rendered novel applications feasible within the healthcare sector. A seminal article published by JAMA, recognised as one of the decade's most impactful contributions, illustrated the capability of ophthalmologist-equivalent identification of diabetic retinopathy through the analysis of retinal photographs (39). Research endeavours have also yielded significant advancements in the screening of breast (40) and lung cancers (41), pathology (42), identification of dermatological conditions (43), and predictive analytics derived from electronic health record data (44), among a plethora of other domains.

The fifth period: large language models

The integration of methodologies employed in text analysis alongside generative models has facilitated the development of computational systems adept at synthesising text based on specified input descriptions while concurrently addressing user inquiries. Such methodologies are distinguished by their models exhibiting exceedingly high complexity, often encompassing multiples of billions of parameters, which are acquired from hundreds or thousands of terabytes of data. These systems are designated as Large Language Models (LLM), and their implementation signifies one of the contemporary frontiers of artificial intelligence, thereby garnering significant interest within the clinical domain (45, 46).

The application of generative AI within clinical settings constitutes one of the cutting-edge topics of scholarly inquiry. Following the introduction of ChatGPT, there has been a remarkable increase in the volume of published literature, together with the emergence of various clinical studies that have explored its application for diverse purposes. For example, one investigation contrasted the responses provided to patients' inquiries by an AI-driven system with those offered by clinical practitioners, revealing that the former were deemed not only accurate but also exhibited greater empathy compared to the latter.

The methodical application of generative models and LLM necessitates the utmost prudence, owing to several pivotal considerations that warrant meticulous evaluation.

These considerations include: i) the challenges associated with providing assurances regarding the accuracy of the responses; ii) the substantial adverse ramifications stemming from diagnostic and therapeutic inaccuracies; iii) the imperative to elucidate the rationale behind the provided responses; iv) the safeguarding of intellectual property rights; v) the complexities inherent in modifying clinical workflows, and finally vi) the privacy and security concerns pertaining to patient information.

It is therefore unsurprising that in certain institutions, tools such as GPT are employed to facilitate functions that do not have a direct bearing on diagnostic or therapeutic decisions. For example, these tools may be utilised to synthesise existing documents, to generate administrative paperwork, or to automatically document and transcribe dialogues with patients.

A recent review has characterised the current landscape of evaluation efforts of the performance of LLMs in clinical health care settings, including uniformity, thoroughness, and robustness, to guide their deployment and propose a framework for the testing and evaluation of LLMs across health care applications (47).

The objective of this study was to summarise existing evaluations of LLMs in health care in terms of 5 components: (i) evaluation of type of data, (ii) health care task, (iii) natural language processing (NLP) and natural language understanding (NLU) tasks, (iv) dimension of evaluation, and (v) medical specialty. A systematic search of PubMed and Web of Science was performed for studies published between January 1, 2022, and February 19, 2024.

The results of this study pointed out several unexpected facts: only 5% of the studies used real patient data. Real patient care data encompasses the complexities of clinical practice, providing a thorough evaluation of LLM performance that mirrors real-world performance; only 15.8% of studies evaluated bias. Accurate bias quantification is crucial for policymaking and regulation. No platform exists for reporting LLM failure modes in health care. Reporting failure modes is essential for root cause analysis in health care settings. No consensus exists as to which evaluation dimensions to examine for a given health care or NLP task. Standardisation enables objective comparison, leading to reliable conclusions.

This systematic review highlights the need for considerable caution in appreciating the added value of studies on large language models.

Conclusions

As stated above, AI is a tool available to us, it is at our service to improve our lives and activities, not the other way around. But such a tool must give people all the information they need so that they can make correct decisions and can trust the output provided. Decisions that must be made following the results of a black box (where it is unclear how the output was arrived at) need to be supported by verification and analysis.

References

1. GALILEI G: Il Saggiatore. 1623, 171.
2. LOUIS PCA: Recherches sur l'emphysème des poumons. Mem Soc Med Obs 1837; 1: 160-261.
3. SNOW J: Mode of communication of cholera. In: CHURCHILL J (Ed.), 2nd Edition, London, 1855, 162.
4. BABBAGE C: A comparative view of the various institutions for the assurance of lives. London, J. Mawman & J. Booth, 1826.
5. FUEGI J, FRANCIS J: Lovelace & Babbage and the creation of the 1843 'notes'. *IEEE Ann Hist Comput* 2003; 25(4): 16-26. https://doi.org/10.1109/mahc.2003.1253887
6. BOOLE G: The calculus of logic. *The Cambridge and Dublin Mathematical Journa*l 1848; 3: 183-98.
7. TURING AM: On computable numbers, with an application to the Entscheidungsproblem. *Proc Lond Math Soc* 1936; 58: 230-265. https://doi.org/10.1112/plms/s2-42.1.230
8. MCCULLOCH W, PITTS W: A logical calculus of ideas immanent in nervous activity. *Bull Math Biophys* 1943; 5: 115-13.
9. ROSENBLATT F: The Perceptron - a perceiving and recognizing automaton. *Cornell Aeronautical Lab* 1957; 85: 460-61.
10. NEWELL A, SIMON HA: Human problem solving. Prentice-Hall, New York, 1972.
11. NEWELL A: The knowledge level. *J Artif Intell* 1982; 18(1): 87-127. https://doi.org/10.1016/0004-3702(82)90012-1
12. MINSKY M, PAPERT S: Perceptrons, an introduction to computational geometry. MIT Press, Cambridge, Mass., 1969.
13. RUMELHART D, HINTON G, WILLIAMS R: Learning representations by back-propagating errors. *Nature* 1986; 323: 533-36. https://doi.org/10.1038/323533a0
14. HORNIK K, STINCHCOMBE M, WHITE H: Multilayer feedforward networks are universal approximators. *Neural Networks* 1989; 2: 359-66. http://doi.org/10.1016/0893-6080(89)90020-8
15. ALMEIDA JS: Predictive non-linear modeling of complex data by artificial neural networks. *Curr Opin Biotechnol* 2002; 13(1): 72-76. https://doi.org/10.1016/s0958-1669(02)00288-4c
16. BUSCEMA M, GROSSI E (Eds.): Artificial adaptive systems in medicine: New theories and models for new applications in the real world. Bentham Science Publishers; 2009. https://doi.org/10.2174/97816080504201090101
17. TASTLE WJ: Data mining applications using artificial adaptive systems. Springer Science & Business Media New York, 2013.
18. HUNT KJ, SBARBARO D, ŻBIKOWSKI R, GAWTHROP PJ: Neural networks for control systems - a survey. *Automatica* 1992; 28(6): 1083-112. https://doi.org/10.1016/0005-1098(92)90053-I
19. GROSSI E: Artificial adaptive systems andpredictive medicine: a revolutionary paradigm shift. *Immun Ageing* 2010; 7(Suppl 1): S3. https://doi.org/10.1186/1742-4933-7-s1-s3
20. ANDRIULLI A, GROSSI E, BUSCEMA M, PILOTTO A, FESTA V, PERRI F: Artificial neuralnetworks can classify uninvestigated patients

with dyspepsia. *Eur J Gastroenterol Hepatol* 2007; 19(12): 1055-58. https://doi.org/10.1097/meg.0b013e3282f198b2
21. LAHNER E, GROSSI E, INTRALIGI M *et al*.: Possible contribution of artificial neural networks and linear discriminant analysis in recognition of patients with suspected atrophic body gastritis. *World J Gastroenterol* 2005; 11(37): 5867-73. https://doi.org/10.3748/wjg.v11.i37.5867
22. PENCO S, GROSSI E, CHENG S *et al*.: Assessment of the role of genetic polymorphism in venous thrombosis through artificial neural-networks. *Ann Hum Genet* 2005; 69(Pt 6): 693-706. https://doi.org/10.1111/j.1529-8817.2005.00206.x
23. PACE F, BUSCEMA M, DOMINICI P *et al*.: Artificial neural networks are able to recognize gastro-oesophageal reflux disease patients solely on the basis of clinical data. *Eur J Gastroenterol Hepatol* 2005; 17(6): 605-10. https://doi.org/10.1097/00042737-200506000-00003
24. GROSSI E, ASTEGIANO M, DEMARCHI B *et al*.: Dolore addominale e alterazione dell'alvo: è possibile semplificare l'iter diagnostico attraverso l'uso di Sistemi di Intelligenza Artificiale? *Sistemi Artificiali Adattivi in Biomedicina* 2005; 1: 66-102.
25. COPPEDÈ F, GROSSI E, BUSCEMA M, MIGLIORE L: Application of artificial neural networks to investigate one-carbon metabolism in Alzheimer's disease and healthy matched individuals. *PLoS One* 2013; 8(8): e74012. https://doi.org/10.1371/journal.pone.0074012
26. TABATON M, ODETTI P, CAMMARATA S *et al*.: Artificial neural networks identify the predictive values of risk factors on the conversion of amnestic mild cognitive impairment. *J Alzheimers Dis* 2010; 19(3): 1035-40. https://doi.org/10.3233/jad-2010-1300
27. BALDASSARRE D, GROSSI E, BUSCEMA M *et al*.: Recognition of patients with cardiovascular disease by artificial neural networks. *Ann Med* 2004; 36(8): 630-40. https://doi.org/10.1080/07853890410018880
28. ROTONDANO G, CIPOLLETTA L, GROSSI E *et al*.: Artificial neural networks accurately predict mortality in patients with nonvariceal upper GI bleeding. *Gastrointest Endosc* 2011; 73(2): 218-26, 226.e1-2. https://doi.org/10.1016/j.gie.2010.10.006
29. BURI L, ZULLO A, HASSAN C *et al*.: Upper GI endoscopy in elderly patients: predictive factors of relevant endoscopic findings. *Intern Emerg Med* 2013; 8(2): 141-46. https://doi.org/10.1007/s11739-011-0598-3
30. MORELLI V, PALMIERI S, LANIA A *et al*.: Cardiovascular events in patients with mild autonomous cortisol secretion: analysis with artificial neural networks. *Eur J Endocrinol* 2017; 177(1): 73-83. https://doi.org/10.1530/eje-17-0047
31. GROSSI E, PODDA GM, PUGLIANO M *et al*.: Prediction of optimal warfarin maintenance dose using advanced artificial neural networks. *Pharmacogenomics* 2014; 15(1): 29-37. https://doi.org/10.2217/pgs.13.212
32. SPOLIDORO GCI, D'ORIA V, DE COSMI V*et al*.: Artificial neural network algorithms to predict resting energy expenditure in critically ill children. *Nutrients* 2021; 13(11): 3797. https://doi.org/10.3390/nu13113797
33. LANGER T, FAVARATO M, GIUDICI R *et al*.: Development of machine learning models to predict RT-PCR results for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in patients with influenza-like symptoms using only basic clinical data. *Scand J Trauma Resusc Emerg Med* 2020; 28(1): 113. https://doi.org/10.1186/s13049-020-00808-8
34. LECUN Y, BENGIO Y, HINTON G: Deep learning. *Nature* 2015; 521: 436-44. https://doi.org/10.1038/nature14539
35. HAUG CJ, DRAZEN JM: Artificial intelligence and machine learning in clinical medicine. *N Engl J Med* 2023; 388(13): 1201-8. https://doi.org/10.1056/nejmra2302038
36. GOODFELLOW I, POUGET-ABADIE J, MIRZA M *et al*.: Generative adversarial nets. Proceedings of the *International Conference on Neural Information Processing Systems* 2014; 2: 2672-80.
37. VASWANI A, SHAZEER N, PARMAR N *et al*.: Attention is all you need. *arXiv* 2017. https://doi.org/10.48550/arxiv.1706.03762
38. XIE J, LU Y, ZHU S-C, WU YN: A theory of generative ConvNet. *arXiv* 2016. https://doi.org/10.48550/arxiv.1602.03264
39. GULSHAN V, PENG L, CORAM M *et al*.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016; 316(22): 2402-10. https://doi.org/10.1001/jama.2016.17216
40. MCKINNEY SM, SIENIEK M, GODBOLE V *et al*.: International evaluation of an AI system for breast cancer screening. *Nature* 2020; 577(7788): 89-94. https://doi.org/10.1038/s41586-019-1799-6
41. ARDILA D, KIRALY AP, BHARADWAJ S *et al*.: End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019; 25(6): 954-61. https://doi.org/10.1038/s41591-019-0447-x
42. NAGPAL K, FOOTE D, TAN F *et al*.: Development and validation of a deep learning algorithm for Gleason grading of prostate cancer from biopsy specimens. *JAMA Oncol* 2020; 6(9): 1372-80. https://doi.org/10.1001/jamaoncol.2020.248543
43. LIU Y, JAIN A, ENG C *et al*.: A deep learning system for differential diagnosis of skin diseases. *Nat Med* 2020; 26(6): 900-8. https://doi.org/10.1038/s41591-020-0842-3
44. RAJKOMAR A, OREN E, CHEN K *et al*.: Scalable and accurate deep learning with electronic health records. *NPJ Digit Med* 2018; 1: 18. https://doi.org/10.1038/s41746-018-0029-1
45. LEE P, GOLDBERG C, KOHANE I: The AI revolution in medicine: GPT-4 and beyond. Pearson, 2023.
46. SCOTT IA, ZUCCON G: The new paradigm in machine learning - foundation models, large language models and beyond: a primer for physicians. *Intern Med J* 2024; 54(5): 705-15. https://doi.org/10.1111/imj.1639347.
47. BEDI S, LIU Y, ORR-EWING L *et al*.: Testing and evaluation of health care applications of large language models: a systematic review. *JAMA* 2025; 333(4): 319-28. https://doi.org/10.1001/jama.2024.21700