Brief History of Artificial Intelligence

Nikesh Muthukrishnan, MEng[a], Farhad Maleki, PhD[a], Katie Ovens, PhD[a,b], Caroline Reinhold, MD, MSc[a,c], Behzad Forghani, MEng[a,d], Reza Forghani, MD, PhD[a,c,d,e,f,*]

KEYWORDS

• Artificial intelligence • History • AI winter • Neural network • Convolutional neural network • Machine learning • Deep learning

KEY POINTS

- There have been several cycles in the past 70 years where artificial intelligence has showed great promise and eventually failed to meet its high expectations.
- Failure to meet expectations owing to limitations of artificial intelligence causes a loss of interest by public and stakeholders.
- These failures have forced the field of artificial intelligence into "artificial intelligence winters," where growth is severely reduced.
- Although the most recent advances in artificial intelligence and successes in deep learning are promising, artificial intelligence has limitations.
- There should realistic expectations set to avoid a third artificial intelligence winter.

INTRODUCTION

Artificial intelligence (AI) is revolutionizing many industries by performing tasks that typically require human intelligence to solve. AI contributes to complex scientific and engineering workflows through simulating, supplementing, or augmenting human intelligence in an efficient and precise manner. Examples of such tasks include fraud detection in banking, conversational bots used in customer service, and precision diagnostics in health care. AI aims at programming intelligence into machines by learning from experiences and adapting to changes in the environment to simulate human decision making and reasoning processes.

Machine learning (ML), a subfield of AI (**Fig. 1**), is concerned with algorithms that are capable of learning complex tasks and developing predictive models through sample data. Through a procedure referred to as feature engineering, often a set of informative features are selected or generated by an expert for building predictive models. The availability of large amounts of data and computational power has led to a surge in successful

Funding Information: R. Forghani is a clinical research scholar (chercheur-boursier clinicien) supported by the Fonds de recherche en santé du Québec (FRQS) and has an operating grant jointly funded by the FRQS and the Fondation de l'Association des radiologistes du Québec (FARQ).

[a] Augmented Intelligence & Precision Health Laboratory (AIPHL), Department of Radiology & Research Institute of the McGill University Health Centre, 5252 Boulevard de Maisonneuve Ouest, Montreal, Quebec H4A 3S5, Canada; [b] University of Saskatchewan, 116 - 110 Science Place, Saskatoon, SK S7N 5C9, Canada; [c] Department of Radiology, McGill University, 1650 Cedar Avenue, Montreal, Quebec H3G 1A4, Canada; [d] Gerald Bronfman Department of Oncology, McGill University, Suite 720, 5100 Maisonneuve Boulevard West, Montreal, Quebec H4A3T2, Canada; [e] Segal Cancer Centre, Lady Davis Institute for Medical Research, Jewish General Hospital, 3755 Cote Ste-Catherine Road, Montreal, Quebec H3T 1E2, Canada; [f] Department of Otolaryngology - Head and Neck Surgery, Royal Victoria Hospital, McGill University Health Centre, 1001 boul. Decarie Boulevard, Montreal, Quebec H3A 3J1, Canada

* Corresponding author. Room C02.5821, 1001 Decarie Boulevard, Montreal, Quebec H4A 3J1, Canada.
E-mail address: reza.forghani@mcgill.ca

Neuroimag Clin N Am 30 (2020) 393–399
https://doi.org/10.1016/j.nic.2020.07.004

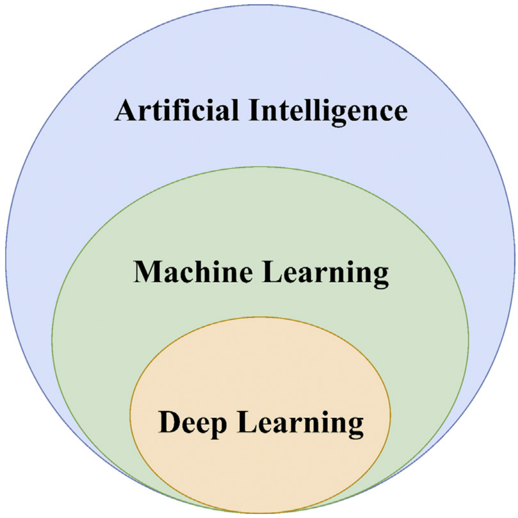

Fig. 1. AI aims at simulating human intelligence. ML, a subfield of AI, is concerned with learning complex associations using data. ML methodologies focus on developing predictive models using sample data. Through a process referred to as feature engineering, human experts select informative features to build predictive models. DL, which is a subfield of ML, tries to eliminate the feature engineering step by learning the optimal set of features from sample data.

applications of ML in fields such as natural language processing, machine vision, robotics, and diagnostics.[1–4] Most of the recent successes in ML applications can be attributed to the advances and innovations in deep learning (DL), which is considered a subfield of ML. DL refers to the methodologies that rely on deep neural networks. DL methods eliminate the need for feature engineering by trying to learn the optimal set of features from data. Neural networks were initially designed to simulate neural activities in a human brain. As the development and interest in the field continues to expand, it is important to appreciate its history as well as understand the potential pitfalls.

EARLIEST ATTEMPTS AT ARTIFICIAL INTELLIGENCE

The earliest models of AI tried to simulate the function of a single neuron (**Fig. 2**). The simplest models started as feedforward, simple input–output functions. However, over the following decades, these became more sophisticated with the addition of more sophisticated functions, added layers, and bidirectional feedback, eventually becoming the building blocks of the modern day deep neural networks or DL. One of the first publications alluding to AI was work published by McCulloch and Pitts in 1943.[5] Their publication describes a computer model used to learn based on a process comparable with neurons in the human brain. The model described in their publication was referred to as the MCP neuron, and it functioned by taking in Boolean inputs (Boolean logic refers to a branch of algebra concerning true/false statements), processing them in a preset manner, and if the processed value exceeded a certain threshold, the MCP neuron would output a value. By exceeding the threshold, the MCP neuron is considered to have been fired or activated. Although this simple model was effective for simple processing tasks, it had many limitations. It generated a binary output and required a fixed set of weight and threshold values. Also, they did not provide a methodology for learning those values.

A more sophisticated version of the MCP was published in 1958 by Rosenblatt, called the perceptron.[6] The perceptron, illustrated in (**Fig. 3**) processed non-Boolean inputs $(x_1, x_2, \cdots, x_n)$ and included weights into the model $(w_1, w_2, \cdots, w_n)$ for scaling. In addition, a nonlinear function f processes the sum of the products of the input values and their corresponding weights. This provided more flexibility to the model and later became one of the building blocks for modern neural networks.

With increased development of early AI models and the inevitable progress in the field, a systematic methodology for evaluating the intelligence of a model was necessary. One of the first works introducing the topic of a model's intelligence was published by Alan Turing, entitled "Computing Machinery and Intelligence," in October 1950.[7] Turing raised the question of whether machines can imitate human intelligence and introduced a test for model intelligence. This test, referred to as the Turing Test, involves a blinded human interrogator questioning a human respondent and a machine respondent. The task of the interrogator is to identify which respondent is the machine. If the interrogator is not capable of discerning the machine answers from the human answers more often that what would be expected by chance, the machine is considered to have passed the Turing Test. The Turing Test has typically been considered the goal of AI; however, in modern days, there is a debate on whether the Turing Test distracts AI researchers in addition to unnecessarily raising the public expectation of the field.[8] The 1956 Dartmouth conference is generally considered the moment that AI formally recognized and obtained its name and its mission.[9,10] This conference was organized by Marvin Minsky, John McCarthy, Claude Shannon, and Nathan

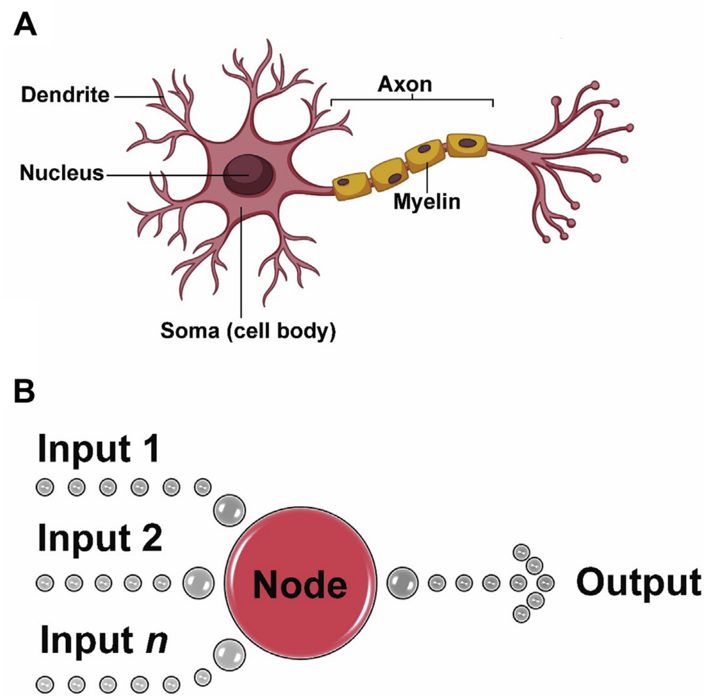

Fig. 2. Neural networks were initially designed to simulate neurons in the human brain (*A*), composed of artificial neurons or nodes (*B*). A node has an activation function, defining the output of that node based on an input or set of inputs. Although in their earliest forms these had feedforward, simple input–output functions, overtime these became more sophisticated, eventually becoming the building blocks of the modern day deep neural networks (see text for additional details). An example of a neural network neuron is depicted in 2B.

Rochester of IBM and may be considered the “birth” of the field of AI, including the assertion that "every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it."

CYCLES OF INTEREST IN ARTIFICIAL INTELLIGENCE AND THE FIRST AND SECOND ARTIFICIAL INTELLIGENCE WINTERS

With media coverage and public expectations being high, AI can often be overhyped, and the slow

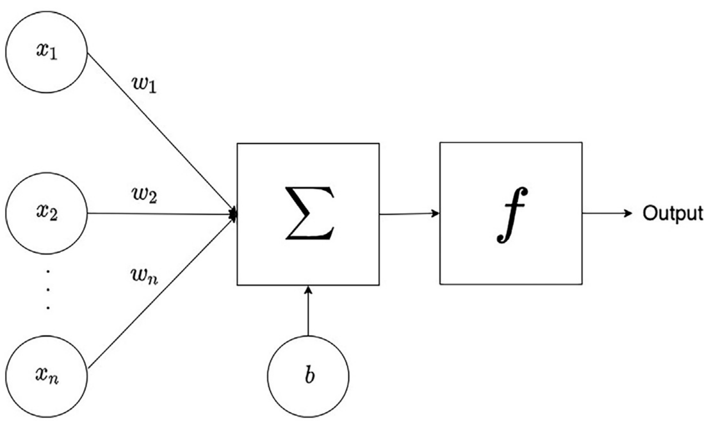

Fig. 3. A visualization of the perceptron, which is one of the building blocks of modern neural networks. A perceptron is a mathematical model for a biological neuron. It takes input values $(x_1, x_2, \cdots, x_n)$, weight values of $(w_1, w_2, \cdots, w_n)$, and a bias value of *b*. First, each input value is multiplied by its corresponding weight; then the summation of the resulting values and the bias value is calculated to generate a single value. This value is then fed to a function *f*, referred to as an activation function. The activation function generates a binary value, that is, 0 or 1, depending on its input. This model can be used as a binary classifier.

but steady progress of the field may be overlooked. This situation leads to unfulfilled overly ambitious expectations and an overall decline in interest in the field by investors, also known as an AI winter. Throughout the history of AI, there have already been 2 AI winters, and some researchers are suggesting that a third AI winter may be arriving soon. The first AI winter was between the years of 1974 and 1980. Leading up to these years, Rosenblatt's perceptron had gathered a lot of hype and demonstrated capacity for modeling simple systems. However, the unrealistic initial expectations of AI at that time also started to become apparent, as discussed in 2 publications that emerged. In 1968, Minsky and Papert[11] demonstrated the limitations of the perceptron by identifying that a Boolean function such as the XOR function that could not be modeled using a perceptron with 2 inputs and without significant user handling. The second significant publication that played a major role in initiating the first AI winter was by Lighthill[12] in 1973, providing an overview of the hype in the field and emphasis on the very small progress in the field. Finally, during these years, many governmental organizations stopped funding for AI research or significantly decreased their AI research funding.[13] After 1980, most AI researchers had given up on AI algorithms learning representations of data and instead shifted toward expert rule-based systems. However, by the mid 1980s, it was recognized that, although expert systems were capable of performing very specific tasks, these models lacked common sense, could not be used to perform more complex tasks, and were not generalizable.[14]

The lack of representation learning, during the late 1970s and early 1980s, was one of the drawbacks that had led to the first AI winter. In 1985 a seminal work by Rumelhart, Hinton and Williams ended this long winter.[15] They addressed the concerns presented by Minsky and Papert and introduced gradient descent optimization for minimizing the error of a network. By iteratively updating bias and weight values through a gradient decent optimization method, they minimized the error of the network and were able to systematically learn the bias and weight values. Furthermore, in 1986, Rumelhart and colleagues[16] expanded on this work and introduced the concept of back-propagation in multilayer neural networks, which consists of several layers of neurons stacked together where neurons in each layer were connected to the neurons in the next layer. The back-propagation algorithm revolutionized the learning capabilities of neural networks. Although Rumelhart and colleagues were credited for popularizing back-propagation, it is important to note that there had been several earlier publications of introducing backpropagation by Werbos and John[17] in 1974, Fukushima[18] in 1980, and Parker in 1985.[19] These concepts and publications began to garner more interest and funding for the field, ending the first AI winter. However, in the early 1990s, it was realized that these networks were not scalable. This was in large part related to the lack of computational power. It became apparent that highly complex networks were not feasible, leading to the second AI winter.

The second AI winter was due to the increased hype in the capabilities of neural networks without sufficient advancement in computing power. Researchers began to shift their focus to more practical and simpler algorithms. The support vector machine algorithm (**Figs. 4** and **5**), which was originally introduced in 1963 by Vapnik and Chervonenkis, became popular again with the implementation of nonlinear kernels in 1992.[20] Before Boser and colleagues' publication in 1992,[20] the support vector machine algorithm attempted to solve for a hyperplane by maximizing the marginal distances between 2 separate classes and the hyperplane. The maximized distance between the hyperplane and the classes allows for more robustness because data are always subject to noise. The publication by Boser allowed for a simple modification of the optimization algorithm, now known as the "kernel trick," which enabled the algorithm to solve for nonlinear hyperplanes without significantly increasing the computational requirements of the algorithm. This trick allowed for support vector machine algorithms to capitalize on the low computational power that was available in the early 1990s.

RESURGENCE OF INTEREST IN ARTIFICIAL INTELLIGENCE: THE THIRD WAVE

The interest in AI resurged toward the mid 1990s as computational power increased and could support the development of neural networks. The microcomputer revolution and Moore's law both describe the advancements computers had in this decade that allowed the replacement of the traditional Lisp machines that were throttling AI development. The capabilities of AI paired with sufficient computational power was demonstrated in 1997 when IBM developed the chess playing supercomputer Deep Blue. Deep Blue defeated the chess champion Kasparov, which led to many publications and documentary films that attracted the public's attention to the field once again.[21,22]

Neural networks began to resurface in the late 1990's with the introduction of convolutional

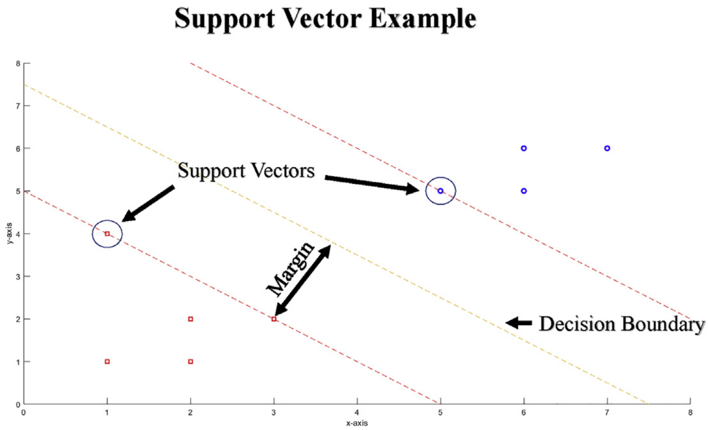

Fig. 4. Example of a support vector machine. (*From* Forghani R, Savadjiev P, Chatterjee A, et al. Radiomics and Artificial Intelligence for Biomarker and Prediction Model Development in Oncology. Comput Struct Biotechnol J. 2019;17:995-1008. Published 2019 Jul 12. https://doi.org/10.1016/j.csbj.2019.07.001; with permission.)

neural networks. LeCun and colleagues[23] published the LeNet-5 network (a 7-level convolutional neural network) in 1998 for document recognition, which used convolutional and subsampling layers processing the input before a fully connected layer for output prediction (**Fig. 6**). Although they were able to achieve the lowest error rates on digit recognition at that time, the algorithm had difficulties scaling to larger problems because they were limited by hardware and data constraints. This issue continued to plague the neural network algorithms for almost a decade. During the last decade, there were 2 main advancements that enabled neural networks to progress: data storage and graphical processing units (GPU). Data became more readily accessible with electronic storage and lower costs. The performance of ML algorithms relies heavily on the available data. Thus, with a high volume of available data, the performance of these algorithms improved. This development enabled the Hinton and colleagues'[24] laboratory to formally introduce DL to the public in 2006 and achieve exceptional performance in speech recognition, which was previously determined to be a challenging AI problem.

Furthermore, with the introduction of GPUs and the steady improvement to computers as per Moore's law, the hardware limitations that constrained the performance of neural networks were overcome. The additional computational power enabled researchers to run larger networks with more complex layers. Cireşan and colleagues[25] were the first to implement GPUs with DL using a GTX 280 graphics card. Finally, in 2012, Krizhevsky and colleagues[26] presented AlexNet, which used the massively available dataset from ImageNet (1.2 million images with 1000 classes at the time) with GPUs to win the ImageNet Large Scale Visual Recognition Challenge. Krizhevsky and colleagues used the rectified linear units to introduce nonlinearities to convolutional neural networks as well as the dropout technique to avoid overfitting. AlexNet was able to achieve an error rate of 15.3% on the dataset, which was

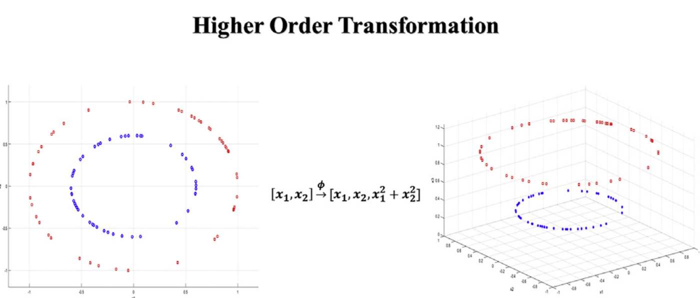

Fig. 5. Example of support vector machines: nonlinearly separable data transformed to a higher dimension. (*From* Forghani R, Savadjiev P, Chatterjee A, et al. Radiomics and Artificial Intelligence for Biomarker and Prediction Model Development in Oncology. Comput Struct Biotechnol J. 2019;17:995-1008. Published 2019 Jul 12. https://doi.org/10.1016/j.csbj.2019.07.001; with permission.)

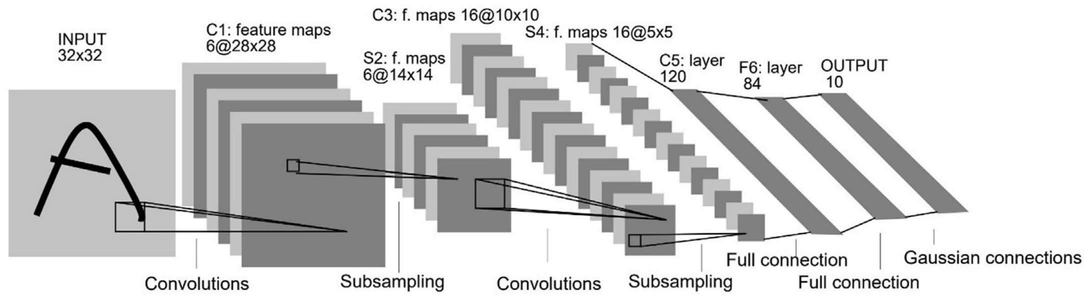

Fig. 6. Architecture of LeNet-5, a CNN, here for digits recognition. Each plane is a feature map, that is, a set of units whose weights are constrained to be identical. (*Adapted from* Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE.* 1998;86(11):2278-2324; with permission.)

10.9% lower than the runner up. The drastic difference in performance caused a revolution within the field of AI and interest shifted back to DL. This led to many modified state-of-the-art DL networks such as the VGG network, Inception network, ResNet and others that form the cornerstones of many medical and neuroimaging applications of deep neural networks that are described in various articles elsewhere in this issue.

SUMMARY

There is a tremendous amount of interest in the potential of AI in health care. There is reason to be excited about AI and its future impact on health care, including in neuroimaging. At the same time, one must exercise caution and not fall into the trap of unrealistic expectations and hype as it pertains to AI applications. With many aspiring researchers and an expanding number of publications on DL, it is important to recognize the challenges and potential pitfalls. AI over the last few decades has repeatedly undergone moments of tremendous hype followed by the realization that some of the initial goals were out of scope. This realization causes investors to pull away from the field leading to AI winters. Projects such as IBM Watson can serve as an important example of the need for setting appropriate and realistic expectations. In 2011, after Watson gathered a lot of public attention by winning on the game show *Jeopardy!*, IBM had announced that it would begin to focus attention toward the health care field and revolutionize AI in health care in the upcoming years. Although there were several successful collaborations between hospitals and IBM, some collaborations resulted in failures.[27] Because of the high profile of the latter failures and some of the other unrealized expectations promoted both by industry and academia, there is the potential for the public and key stakeholders to lose trust and interest, beginning a third AI winter, and overlooking and undermining the many potential important and realizable AI supported applications in health care. By providing a brief history of ML applications and AI, this article aims to provide a broader perspective on the cycles of progress in AI, in addition to many developments that preceded the current wave of innovations using DL and form the basis for many applications described in the accompanying articles in this issue. By keeping the right perspective, and understanding both the strengths and weaknesses of this important technology, one will be able to ensure the steady growth of the field without researchers and investors abandoning the field again. This could help the field of AI realize its full potential in the transformation of neuroimaging and more broadly health care.

REFERENCES

1. Devlin J, Chang M-W, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv 2018. 1810.04805. Available at: https://ui.adsabs.harvard.edu/abs/2018arXiv181004805D. Accessed October 1, 2018.
2. Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1; Palais des Congres de Montreal, Montreal, Canada, December 7-12, 2015.
3. Pierson HA, Gashler MS. Deep learning in robotics: a review of recent research. Adv Robot 2017;31(16): 821–35.
4. Miotto R, Wang F, Wang S, et al. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinformatics 2018;19(6):1236–46.
5. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics 1943;5:115–33.

6. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 1958;65(6):386–408.
7. Turing AM. I.—Computing machinery and intelligence. Mind 1950;LIX(236):433–60.
8. Pinar Saygin A, Cicekli I, Akman V. Turing test: 50 years later. Minds Mach (Dordr) 2000;10(4):463–518.
9. McCorduck P. Machines who think. 2nd edition. Natick (MA): A K Peters, Ltd.; 2004.
10. Stone P, Brooks R, Brynjolfsson E, et al. Artificial intelligence and life in 2030. In: One hundred year study on artificial intelligence: report of the 2015-2016 study panel, Stanford University. Stanford (CA): Stanford University; 2016. Available at: https://ai100.sites.stanford.edu/sites/g/files/sbiybj9861/f/ai100report10032016fnl_singles.pdf. Accessed September 6, 2016.
11. Minsky M, Papert S. Perceptrons; an introduction to computational geometry. Cambridge (MA): MIT Press; 1969.
12. Lighthill J. Artificial intelligence: a general survey. London: Science Research Council; 1973.
13. Hendler J. Avoiding Another AI Winter. IEEE Intell Syst 2008;23(2):2–4.
14. McCarthy J. Some expert systems need common sense. Ann N Y Acad Sci 1984;426:129–37.
15. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, The PDP Research Group, editors. Parallel distributed processing: explorations in the microstructure of cognition, vol. 1. Cambridge (MA): MIT Press; 1985. p. 318–62.
16. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986;323(6088):533–6.
17. Werbos P, John P. Beyond regression: new tools for prediction and analysis in the behavioral sciences. PhD thesis, Harvard University; 1974.
18. Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern 1980;36(4):193–202.
19. Parker DB. Learning-logic: casting the cortex of the human brain in silicon. Technical report Tr-47. Cambridge (MA): Center for Computational Research in Economics and Management Science. MIT; 1985.
20. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory. Pittsburgh, PA, July 27-29, 1992.
21. Schaeffer J, Plaat A. Kasparov versus deep blue: the rematch. J Int Comput Games Assoc 1997;20:95–101.
22. Campbell M, Hoane AJ, Hsu F-h. Deep Blue. Artif Intelligence 2002;134(1):57–83.
23. Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proc IEEE 1998;86(11):2278–324.
24. Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. Neural Comput 2006;18(7):1527–54.
25. Cireşan DC, Meier U, Gambardella LM, et al. Deep, big, simple neural nets for handwritten digit recognition. Neural Comput 2010;22(12):3207–20.
26. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1. Lake Tahoe (NV), December 3-6, 2012.
27. Strickland E. IBM Watson, heal thyself: how IBM overpromised and underdelivered on AI health care. IEEE Spectr 2019;56(4):24–31.