NEURONS SPIKE BACK

The Invention of Inductive Machines and the Artificial Intelligence Controversy

Dominique CARDON
Jean-Philippe COINTET
Antoine MAZIÈRES

La Découverte | « Réseaux »
2018/5 n° 211 | pp. 173-220
ISSN 0751-7971
ISBN 9782348040689

To cite this article :
Dominique Cardon, Jean-Philippe Cointet et Antoine Mazières, «Neurons spike back. The invention of inductive machines and the artificial intelligence controversy », Réseaux 2018/5 (n° 211), pp. 173-220.
DOI 10.3917/res.211.0173

Abstract
Since 2010, machine learning based predictive techniques, and more specifically deep learning neural networks, have achieved spectacular performances in the fields of image recognition or automatic translation, under the umbrella term of "Artificial Intelligence". But their filiation to this field of research is not straightforward. In the tumultuous history of AI, learning techniques using so-called "connectionist" neural networks have long been mocked and ostracized by the "symbolic" movement. This article retraces the history of artificial intelligence through the lens of the tension between symbolic and connectionist approaches. From a social history of science and technology perspective, it seeks to highlight how researchers, relying on the availability of massive data and the multiplication of computing power have undertaken to reformulate the symbolic AI project by reviving the spirit of adaptive and inductive machines dating back from the era of cybernetics.

Keywords
Neural networks, Artificial intelligence, Connectionism, Expert systems, Deep learning

Main figures are available in English here: https://neurovenge.antonomase.fr/

Résumé
Depuis 2010, les techniques prédictives basées sur l'apprentissage artificiel (*machine learning*), et plus spécifiquement des réseaux de neurones (*deep learning*), réalisent des prouesses spectaculaires dans les domaines de la reconnaissance d'image ou de la traduction automatique, sous l'égide du terme d'"Intelligence artificielle". Or l'appartenance de ces techniques à ce domaine de recherche n'a pas toujours été de soi. Dans l'histoire tumultueuse de l'IA, les techniques d'apprentissage utilisant des réseaux de neurones - que l'on qualifie de "connexionnistes" - ont même longtemps été moquées et ostracisées par le courant dit "symbolique". Cet article propose de retracer l'histoire de l'Intelligence artificielle au prisme de la tension entre ces deux approches, symbolique et connexionniste. Dans une perspective d'histoire sociale des sciences et des techniques, il s'attache à mettre en évidence la manière dont les chercheurs, s'appuyant sur l'arrivée de données massives et la démultiplication des capacités de calcul, ont entrepris de reformuler le projet de l'IA symbolique en renouant avec l'esprit des machines adaptatives et inductives de l'époque de la cybernétique.

Mots-clés
Réseaux de neurones, Intelligence artificielle, Connexionnisme, Système expert, Deep learning

The episode has become legendary in computer science history. In October 2012 the ECCV conference brought together researchers specialized in computer vision[1].

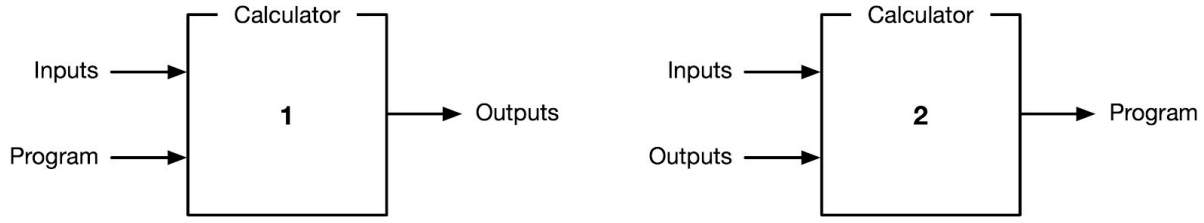
> « So guess who turned up at the 2012 contest? Hinton [*the "father" of neural networks revival*] and that really shook things up. He didn't know anything about the field of computer vision, so he took two young guys to change it all! One of them [*Alex Krizhevsky*] he locked up in a room, telling him: "You can't come out until it works!" He got huge machines to work, machines that had GPUs which at the time weren't great, but he got them to communicate with one another to boost them. It was totally crazy computer stuff. Otherwise, it wouldn't have worked; totally incredible geek knowledge, programming. At the time, computer vision people had been excited about ImageNet for three years [*a database of 1.2 million images tagged with 1,000 categories used as a benchmark to compare the classification results of different competitors*]. Number 1 had an error rate of 27.03%, number 2 had 27.18%, and number 3 had 27.68%. Hinton sent in this guy from nowhere: "we got a really big deep one to work, we got 17%!" He won over everyone by 10 points! So that young geek did it, and he announced the result in front of the jam-packed room. He didn't understand anything at all, like he was 17! He didn't know why those things were there. He'd been locked up in his office and didn't know anything about the field. And then all of a sudden, there he was in front of Fei-Fei, with LeCun sitting at the back of the room and getting up to answer questions [*Li Fei-Fei, professor of computer science and director of SAIL, the Stanford Artificial Intelligence Laboratory; Yann LeCun, today the director of FAIR, Facebook AI Research, and one of the central players in the renewal of neural networks*]. And all the big-wigs of computer vision were trying to react: "But that's not possible. That won't work for recognizing an object when you need…" They were all floored, seeing that basically ten years of intelligence, fine-tuning, and sophistication had been more or less thrown out the window.
>
> They weren't exactly formal logic people, but they were people who had the idea that you have to understand, that you have to know how to explain why you put the branches like that, why you follow that reasoning, why you are

[1] This study was carried out under the ALGODIV project (ANR-15-CE38-0001). The authors would like to thank Telmo Menezes for his advice. As a part of this inquiry, we use three interviews held with French computer science researchers who participated in the revival of neural networks. In order to retain the rawness of their statements, they have been anonymized.

> progressing like that; and that you need to have all of this understanding of the features that go with it and that help you to be able to say that you perfectly understand what you're doing and why it's like that. And then this guy arrives, with a big black box of deep learning with 100 million parameters in it that he'd trained, and he totally disrupted the whole field. "Are your models invariable if the image moves?" The guy didn't even understand the question! Then LeCun responded: "Well, these models are invariable because…" He was really pleased, because Fei-Fei asked him: "but Yann, are these models really fundamentally that different from the models you invented in the 1980s?" To which Yann answered, "Nope, they're exactly the same and we won all the competitions with them!"[2]. »

This colourful account of the announcement of the image classification performance of a deep learning technique (Krizhevsky, Sutskever and Hinton, 2012) bears witness to the effects that the sudden success of a long-marginalized heterodox paradigm has on a scientific community[3]: surprise, faced with the result; questioning of the epistemic validity of the new approach; concern around the future of the orthodox paradigm; mockery faced with the new entrants' ignorance of the theoretical concerns of the field, vertigo faced with the imminent overturning of the paradigm. Starting in 2010, in field after field, deep neural networks have been causing the same disruption in computer science communities dealing with signals, voice, speech, or text. A machine learning method proposing the "rawest" possible processing of inputs, eliminating any explicit modelling of data features and optimizing prediction based on enormous sets of examples has produced spectacular results. A simple way of thinking about this upheaval is to describe it as the transition from hypothetical-deductive machines to inductive machines (Figure 1).

Figure 1. Hypothetical-deductive machines (1) and inductive machines (2)



What was previously thought of as the "human" component in the creation of calculators, program, the rules, or the model was no longer the input into the system but rather its result. The social science perspective on this inductive shift often consists in deconstructing the naturalist illusion of "raw" data and the naiveness of calculation without theory (Gitelman, 2013). While such a precaution is certainly necessary to put into perspective certain heedless discourses stating that "the data speaks for itself", it does not, however, do justice to the determined and intensely artificial work undertaken by the proponents of deep learning techniques to impose the second type of calculation architecture. In this article we will call these *inductive machines* and, more specifically, *connectionist machines*, in order to shine light

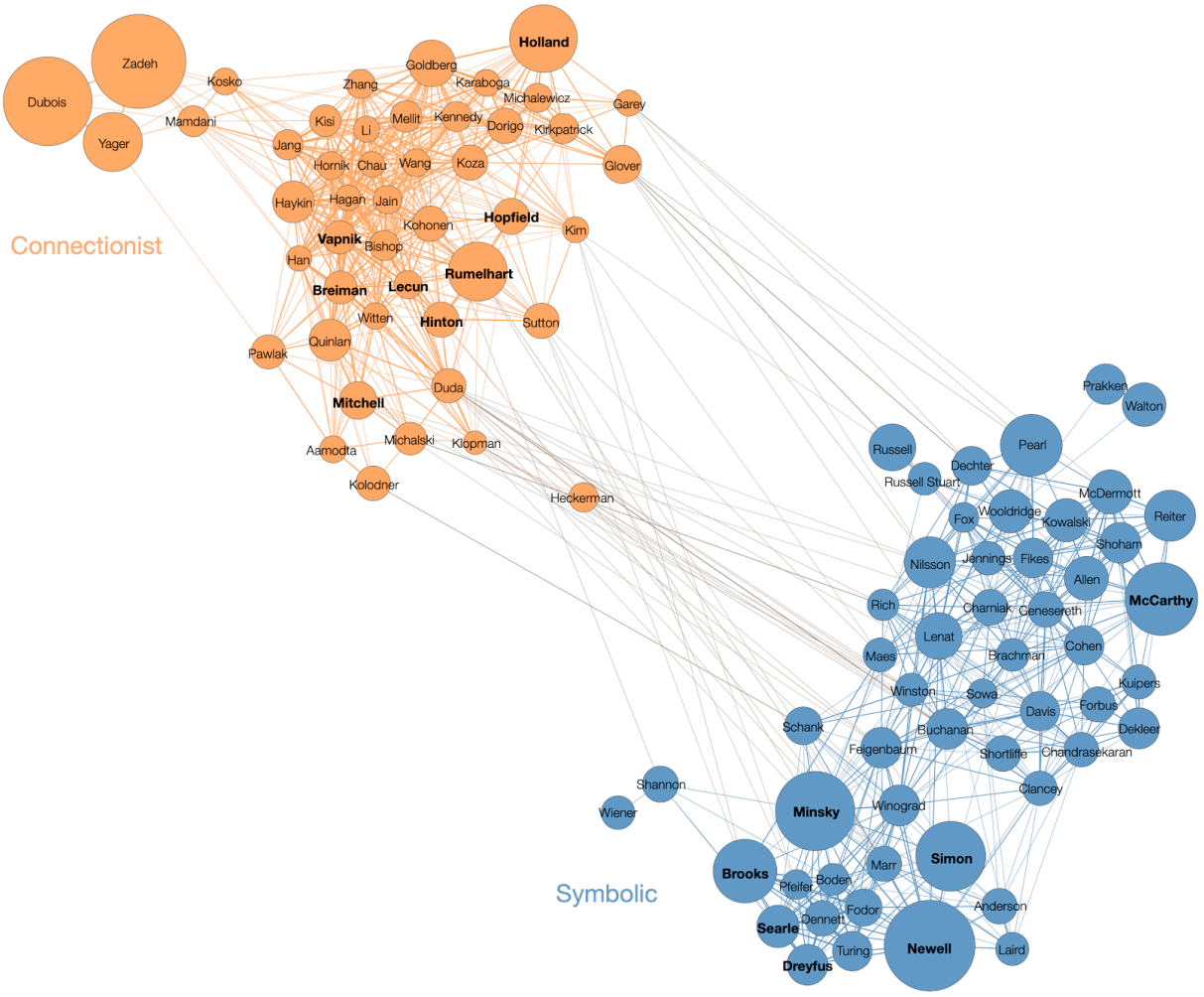
[2] Interview V, computer vision researcher, 12 March 2018.
[3] Y. LeCun gave his version of the same event in a video (starting at minute 20): "Heroes of Deep Learning: Andrew Ng interviews Yann LeCun", YouTube, 7 April 2018.

on the specific type of induction they claim to use. The creation of artefacts able to produce an inductive calculation over large datasets is the result of a conflictive history and a series of highly ingenious constructs. Induction was something that machines had to be constantly steered towards, and which had to be defended against opponents, produced through specific calculations, deployed in specific architectures, and calibrated with suitable data. The designers of these types of machines were not the ingenuous naturalists that constructivist social science often likes to describe them as. The idea of entrusting the production of relevant predictions to machines, by allowing them to learn from the data – *i.e.* inductive calculation – was an undertaking, a theory, and above all, a device with a turbulent history. In order to be implemented and to produce its effects, it required patient work to reconfigure the architecture of "intelligent" machines, which will be the subject of this article.

Symbolic versus Connectionist

The neural network method that we recently saw triumph at the ECCV'12 is nothing new. By taking advantage of the increase in computer calculation capacity and the accessibility of giant databases, today, it is fulfilling the promise that it made at the beginning of cybernetics. Surprisingly, the term recently adopted to describe these remarkable feats of calculation is *artificial intelligence* (AI). The return of this term – coined by John McCarthy in 1956 – to the front stage is an interesting enigma in the history of science and technology. Specifically, the majority of close observers state that it is only in the field of machine learning methods, and in particular in deep learning, that observable progress in calculated prediction is currently taking place. Yet these techniques have not always been considered to fall within AI. In the turbulent history of this field of research, machine-learning techniques using neural networks – which we will call "connectionist" techniques – were for a long time mocked and ostracized by the "symbolic" school of thought. This tension between these two approaches arose with the emergence of artificial intelligence, which was clearly distinct from early cybernetics. The symbolic approach that constituted the initial reference framework for AI was identified with orthodox cognitivism, in terms of which thinking consists of calculating symbols that have both a material reality and a semantic representation value. By contrast, the connectionist paradigm considers thinking to be similar to a massive parallel calculation of elementary functions – functions that will be distributed across a neural network – the meaningful behaviour of which only appears on the collective level as an emerging effect of the interactions produced by these elementary operations (Andler, 1992). This distinction between two ways of conceiving of and programming the "intelligent" operation of a machine is the basis of a tension that has consistently and very profoundly structured the orientations of research, scientific careers, and the design of calculation infrastructure. We are therefore now witnessing one of the situational about-turns typical of the history of science and technology: a research strategy marginalized by the people who contributed to establishing the conceptual frameworks for artificial intelligences is once again coming to the fore, and is now in a position to very profoundly redefine the field from which it had been excluded. As Michael Jordan (2018) ironically stated, "in an interesting reversal, it is Wiener's intellectual agenda that has come to dominate in the current era, under the banner of McCarthy's terminology".

Figure 2. Co-citation network of the 100 most cited authors in scientific publications mentioning "Artificial Intelligence"[4]



To tell the story of the back and forth between these two schools, we should first outline the chronology based on scientific publications retrieved from *Web of Science* (WoS). The co-citation network of the most cited authors in the articles mentioning "Artificial Intelligence" clearly shows the divide between researchers following the symbolic or connectionist approaches. For example, Figure 2 shows the names of the main actors discussed in this article, clearly distributed according to their community. At the heart of the "connectionists", Rumelhart, LeCun, and Hinton represent the founding core of deep learning and stand alongside researchers (Holland, Hopfield) who promoted this movement at different times, as well as the main contributors to multiple machine learning methods, such as Breiman, Mitchell, and Vapnik. On the "symbolic side" is the founding core of AI (McCarthy, Minsky, Simon, and Newell), set out in a way that reflects their proximities and divergences, surrounded by the main contributors to the production of cognitive modelling, expert systems, and even critique of symbolic AI (Dreyfus, Searle, Brooks).

[4] The "Artificial Intelligence" corpus contains 27,656 publications retrieved from Web of Science in February 2018 using the query TS=("artificial intelligence"). The size of the nodes depends on the frequency of the author's appearance. Authors who are regularly co-cited in the same publications are linked in the network. A community detection algorithm reveals the bi-partition of the network into two cohesive communities.

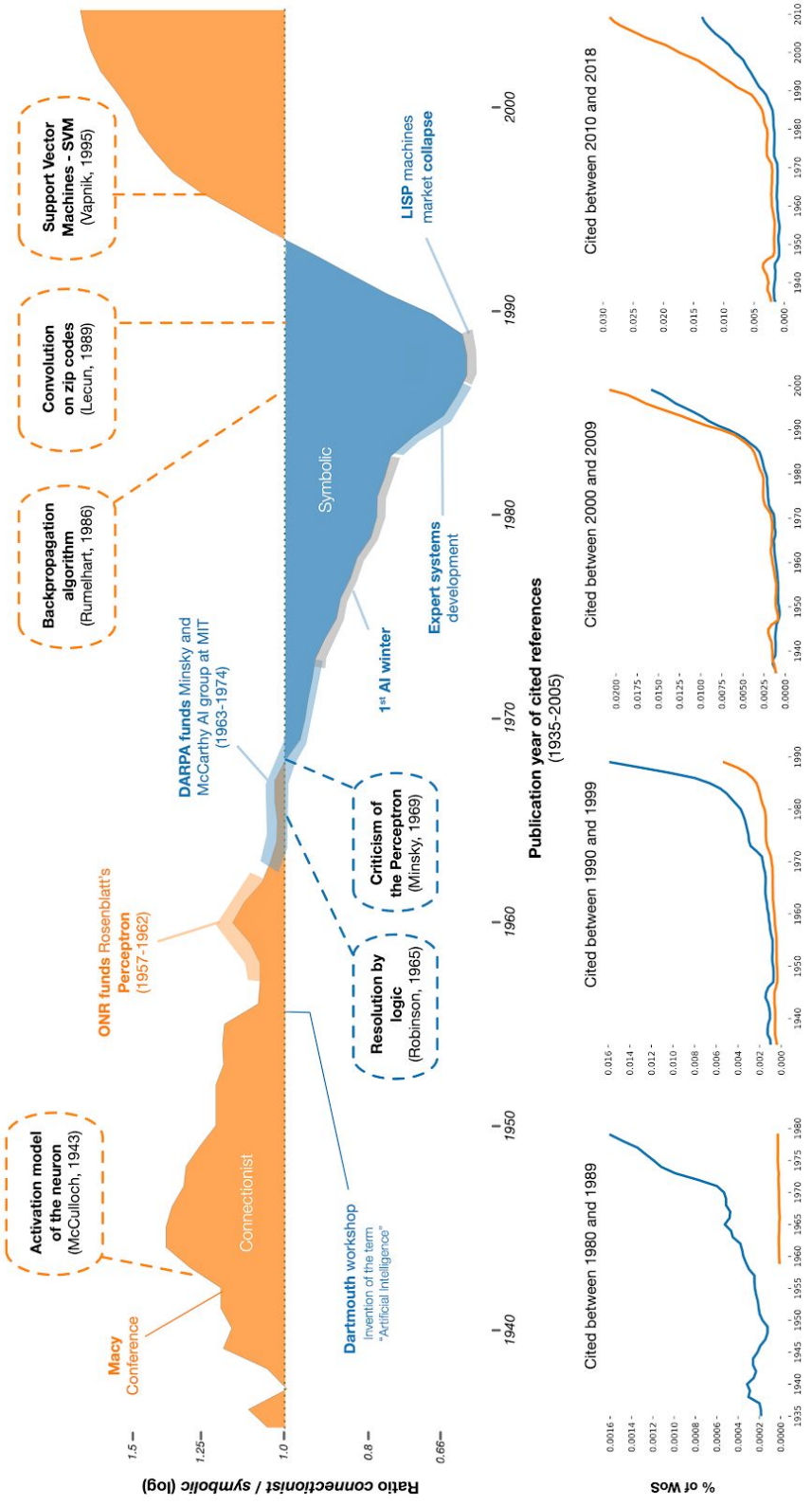
However, the controversy between the two AI communities is even clearer when observing the chronology of the academic impact of the scientific publications in the symbolic[5] and connectionist[6] movements from 1935 to 2005. The chronology in Figure 3 shows the emergence of the connectionist paradigm alongside early cybernetics. Then, starting from the early 1960s, the symbolic paradigm gradually prevailed and defined the main features of AI. It was not until the mid-1990s, following the second AI winter, that the connectionist paradigm once again clearly became dominant in scientific publications, under the banner of deep learning.

To retrace this history, we present a very simple analytical framework which, within a vast array of heterogeneous technologies with a very high level of complexity, isolates a number of reference points allowing us simultaneously to account for the transformation of calculation infrastructures, and different ways of critically analysing their performance. To look at the design of technical systems and their epistemic aim, together, we posit that an "intelligent" machine must articulate a *world*, a *calculator*, and a *target*, based on different configurations. These notions refer to the functional framework within which the design of intelligent artefacts is typically broken down, based on varied terminologies: "environment" /"inputs"/"data"/"knowledge base" (*world*), "calculation"/"program"/"model"/"agent" (*calculator*), and "objectives"/"results"/"outputs" (*target*). Predictive machines can thus be said to establish a calculator in the world by granting it a target. The devices designed throughout the history of AI equip the *world*, the *calculator*, and the *target* with varied and changing entities. They thus propose radically different ways of interrelating the architecture of these sets. The shift in AI research from *symbolic machines* towards *connectionist machines* is therefore the result not of a change in the history of ideas or the validity of one scientific model over another, but of a controversy that led actors to profoundly shift, transform, and redefine the form given to their artefacts. The process that this analytical model allows us to be attentive to is a long historical reconfiguration of alliances and paradigms between competing scientific communities. This affects calculation techniques, but also and above all the form given to these machines, their objectives, the data that they process, and the questions that they address (Latour, 1987). To put it in a way that will become clearer throughout the article: while the designers of symbolic machines sought to insert in the calculator both the world and the target, the current success of connectionist machines is related to the fact that, almost in contrast, their creators empty the *calculator* so that the *world* can adopt its own *target*.

[5] The "Symbolic" corpus contains 65,522 publications retrieved from Web of Science in February 2018 using the query TS=("knowledge representation*" OR "expert system*" OR "knowledge based system*" OR "inference engine*" OR "search tree*" OR "minimax" OR "tree search" OR "Logic programming" OR "theorem prover*" OR ("planning" AND "logic") OR "logic programming" OR "lisp" OR "prolog" OR "deductive database*" OR "nonmonotonic reasoning*").

[6] The "Connectionist" corpus contains 106,278 publications retrieved from Web of Science in February 2018 using the request TS=("artificial neural network*" OR "Deep learning" OR "perceptron*" OR "Backprop*" OR "Deep neural network*" OR "Convolutional neural network*" OR ("CNN" AND "neural network*") OR ("LSTM" AND "neural network*") OR ("recurrent neural network*" OR ("RNN*" AND "neural network*")) OR "Boltzmann machine*" OR "hopfield network*" OR "Autoencoder*" OR "Deep belief network*" OR "recurrent neural network*").

Figure 3. Evolution of the academic influence of the connectionist and symbolic approaches

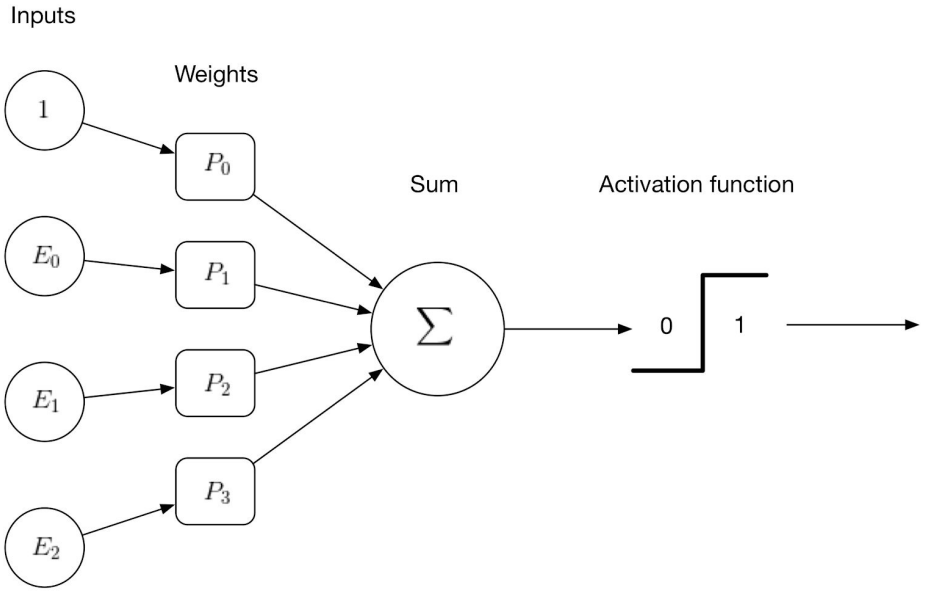


The main graph (top) shows the changes in the ratio between the number of publications cited in the connectionist corpus (orange) and the corresponding number in the symbolic corpus (blue), both adjusted by the total number of publications in WoS. *The additional graphs (bottom) represent the number of publications cited during a given period for each corpus.*

CYBERNETICS AND THE BEGINNINGS OF CONNECTIONISM

The origins of neural networks are found in the pioneering history of computer science and early cybernetics. Even though the term was coined later, cybernetics can effectively be considered "connectionist"[7] and still refers to the goal of mathematically modelling a neural network, set by neurophysiologist Warren McCulloch and logician Walter Pitts in 1943. To this day, that seminal article continues to be quoted as the starting point of the connectionist journey, even in current citations in deep learning articles. The chronology of scientific activity in AI (Figure 3) clearly demonstrates the pre-eminence of the connectionist approach during the early cybernetic period. McCulloch and Pitt's first article proposed a formal model (Figure 4) in which neurons use variables as inputs and weight them to produce a sum that triggers the neuron's activation if it exceeds a certain threshold.

Figure 4. Formal model of an artificial binary threshold neuron



This proposition not formulated as pertaining to artificial intelligence – the term did not exist – but rather as a neurophysiology experimentation tool that was consistent with the biological knowledge of the time regarding the brain's neural processes. It was rapidly associated with the idea of learning through the work of neuropsychologist Donald O. Hebb (1949), which shows that the repeated activation of a neuron by another via a given synapse increases its conductivity and can be considered learning. Biologically inspired, the formal neural model constituted one of the main points of reflection for cyberneticians at the time, and was to become the cornerstone of the calculator of the first "intelligent" machines (Dupuy, 2005).

[7] The first use of the term "connectionism" was by D. Hebb in 1949. It was then taken up by F. Rosenblatt in 1958 (Andler, 1992).

The close coupling between the world and the calculator

The characteristic feature of the architecture of these machines is that their coupling with the environment (the *world*) is so organic that it is not necessary to grant the *calculator* its own agentivity. The goal of cybernetics is to create nothing more than a black box of learning and association, the target of which is regulated by measuring the deviation (*i.e.* the error) between the world and the machine's behaviour. This representation of intelligent machines was initially based on a materialistic conception of information that differed from the symbolic conception that prevailed at the time of the emergence of artificial intelligence (Triclot, 2008). As a form of order opposed to entropy, information is a signal rather than a code. With information theory as developed by Shannon (1948), information did not have to be associated with a given meaning; it was conceived of as pure form, independent of all other considerations, limited to "expressing the magnitude of the order or structure in a material agencing" (Triclot, 2008).

Cybernetic machines defined the *target* of their calculation based only on a comparison of inputs from and outputs towards the *world*. Norbert Wiener's (1948) predictive device applied to guiding anti-aircraft missiles was based on continuously updating their trajectory, comparing the real trajectory of the target with prior estimates. The device had to converge towards the best solution on the basis of the available data; this data informed, corrected, and oriented the calculator. Negative feedback – *i.e.* incorporating the measurement of output error as a new input into an adaptive system – would thus constitute the main axiom of cybernetics. It allowed technical systems to be considered in a strictly behaviourist form, echoing the behaviourist psychology of the time (Skinner, 1971). Just like for living organisms, machines inductively adapted to signals from the environment with a coupling that was so tight that it did not require internal representations or intentions; in short, an "intelligence" specific to them. When Arturo Rosenblueth, Norbert Wiener, and Julian Bigelow (1943) formulated the founding principles of cybernetics, they imagined a self-correcting machine capable, through probabilistic operators, of modifying or adopting end goals that were not "internal" but rather produced by adapting its behaviour according to its own mistakes. Rigorously "eliminativist", the design of cybernetic machines could do away with the notions of intention, plans, or reasoning (Galison, 1994). Theorizing the functioning of one of the most famous of these machines, the Homeostat, Ross Ashby (1956: 110) described the calculating portion of the environment/machine system as a "black box"[8]. The configuration of cybernetic prediction machines so tightly coupled the *world* and the *calculator* that their *target* was to optimize the adaptive operation of the system that they formed together. The cybernetic machines of the 1950s (Homeostat, Adaline, etc.) were no more than laboratory artefacts with very limited aims and capacity; by contrast, deep learning calculators would eventually and much more efficiently come to offer a black box around a world of data, turning outputs into inputs.

[8] On the Homeostat, see Pickering (2010) and Rid (2016).

The Perceptron and connectionist machines

Particularly in the field of visual recognition, McCulloch and Pitts' neural networks provided a highly suitable solution for equipping the calculator of the first adaptive machines. At the end of the 1950s, these machines underwent an important development that contributed to the first wave of public interest in brain machines[9]. The connectionist approach inspired the work of Bernard Widrow (Adaline), Charles Rosen at Stanford (Shakey), or even the Pandemonium, Oliver Selfridge's hybrid device (1960). However, it was the Perceptron initiative (1957-1961) of Frank Rosenblatt, a psychologist and computer scientist at Cornell University, that embodied the first true connectionist machine and became the emblem of another way of enabling a calculation artefact with intelligent behaviour. This device, designed for the purpose of image recognition, received much attention and obtained a large amount of financing from the US Navy (the ONR). Frank Rosenblatt's machine was inspired by McCulloch and Pitts' formal neural networks, but added an additional machine learning mechanism. In the Perceptron's superimposed layers, the input neurons stimulated retinal activity and the output neurons classified the "features" recognized by the system; only the hidden, intermediate layers were capable of learning. Contrary to McCulloch and Pitts' logical – and "top-down" – organization, Frank Rosenblatt advocated a "bottom-up" approach that let the learning mechanism statistically organize the network structure. Following an initial software-based implementation, Frank Rosenblatt undertook the construction of the sole hardware version of the Perceptron: the Mark I, which consisted of 400 photoelectric cells connected to neurons. The synaptic weights were encoded in potentiometers, and changes in weight during learning were made by electric engines. However, the concrete implementation of these learning machines remained very rare due to the technical limitations of the time, and above all, was halted by the development of an AI exploring an entirely different direction of research: the "symbolic" school.

SYMBOLIC AI

When the main proponents of the Dartmouth founding meeting, John McCarthy and Marvin Minsky, coined the term "artificial intelligence" (AI) in 1956, their intention was to oppose the connectionism of early cybernetics (Dupuy, 2005)[10]. They very explicitly wanted to give machines a goal other than adaptively adjusting inputs and outputs. The purpose of "symbolic"[11] AI was to implement rules in computers via programs, so that high-level representations could be manipulated. The emergence of AI thus constituted a veritable "anti-inductive" movement in which logic had to counter the "chimera" of the connectionist approach, which was accused of refusing to define data processing independent of physical processes and of proposing a theory of the mind (Minsky, 1986)[12]. As the chronology shows

[9] Note that at the beginning of the 1960s, work on neural networks was considered a potential pathway for AI. It very quickly became a minority field, before being completely marginalized within the emerging field, but the large conferences at the beginning of the 1960s still brought together researchers from both the symbolic and connectionist schools of thought (Anderson and Rosenfeld, 1988).

[10] On the history of the beginnings of AI, see Crevier (1997), McCorduck (1979), and Nilsson (2010).

[11] Also called LGAI for logic-based AI, AGI (artificial general intelligence), "strong" or "full AI", and today, "good old-fashioned AI" (Haugeland, 1985).

[12] The expressions cited are taken from the transcription of the workshop archives: http:// raysolomonoff.com/dartmouth/, retrieved on 05/10/2018. With respect to the desire to break with cybernetics, nobody is more explicit than John McCarthy

(Figure 3), the symbolic approach prevailed in scientific production in the AI field from the mid-1960s up until the early 1990s.

It was initially informed by the work of Herbert Simon, carried out alongside Alan Newell at RAND in the 1950s. In 1956 they wrote the first program intended to simulate machine decision-making, the Logic Theorist (1956), with the announcement – which would become a typical habit among AI researchers – that "over Christmas, Allen Newell and I invented a thinking machine" (McCorduck, 2004: 168). Modelling reasoning was the central feature of this first wave of AI, which spanned the period from 1956 until the early 1970s. This field of research soon consisted of a small group from MIT (Minsky, Papert), Carnegie Mellon (Simon, Newell), and Stanford University (McCarthy). Despite internal differences, this closed circle established a monopoly over defining AI issues and obtained the majority of (large) funds and access to huge computer systems. From 1964 to 1974 they received 75% of the funding for AI research granted by the ARPA and the Air Force (Fleck, 1982: 181), and benefited from the rare calculation capacities needed for their projects. At the ARPA, they enjoyed the unfailing support of Joseph Licklider, who funded symbolic projects while justifying them in terms of their hypothetical military applications.

This seizure of power by the symbolic school over the then fuzzy and very open definition of intelligent machines took on the form of an excommunication, pronounced in the book that Marvin Minsky and Seymour Papert (1969) dedicated to demonstrating the ineffectiveness of neural networks. At the beginning of the 1960s, the connectionist approaches inherited from early cybernetics experienced a certain degree of enthusiasm, driven by the media success of Frank Rosenblatt's Perceptron. Even though, as a student, Marvin Minsky himself developed neural networks (Snarc, 1951), he wished to confirm the mathematical pre-eminence of symbolic AI over the "mystical" nature "surrounded by a romantic atmosphere" of the self-organized and distributed systems of connectionists (Minsky and Papert, 1969, note 13). Targeting a limited and simplified single-layer version of the Perceptron, he and Seymour Papert demonstrated that neural networks were incapable of calculating the XOR (the exclusive OR) function and therefore had no future. As Mikel Olazaran (1996) shows, Minsky and Papert's strategy was to write the pre-eminence of the symbolic school into the definition of artificial intelligence. Even though the book's effects likely went beyond its authors' intentions, its consequences would be definitive. Following the premature death of Frank Rosenblatt in 1971, neural networks were abandoned, their funding was cut, and the work that was to perpetuate their essence would be carried out outside of the AI field.

A space to manipulate symbols

The main feature of the architecture of symbolic machines is that they break the ties with the world and open up an independent space of reasoning within their *calculator*. The so-called "von Neumann" configuration of new computers implemented in the 1950s established this very space. Whereas the ENIAC (1946) was designed to calculate ballistic tables by "programming" the machine into the hardware, the EDVAC project (1952) separated the

(1988): "As for me, one of the reasons why I invented the term "artificial intelligence" was to get away from the association with "cybernetics". This focus on feedback seemed incorrect to me, and I wanted to avoid having to accept Norbert Wiener as a guru or having to talk with him".

logical operations carried out on the symbols (*software*) of the physical structure of machines (*hardware*) (von Neumann, 1945). The program was thus granted its own space independent of the physical operation of the computer. It became a "universal automatic computer with a centralized program" (Goldstine, 1972: 198-199) and the programming, independent of hardware processes, could be freed to be done "on paper", as Alan Turing (2004: 21) put it. Paul Edwards (1996) shows how, with the appearance of sophisticated programming languages similar to human languages, and subsequently compiled into machine language represented by 0s and 1s, the physical machine could be separated from the symbolic machine. Artificial intelligence could thus be considered as the science of the mind in the machine. One of AI's first contributions to computer science was precisely related to the designing of programming languages, the most famous of which was LISP, developed by John McCarthy in 1958, which was fully identified with AI research due to its logical abstraction capabilities[13].

As soon as it was created in the calculator, this programming space was available to manipulate symbols. AI was born in the same year as cognitive science (1956), and together the two fields would shape the efforts to give computers a capacity for reasoning (Gardiner, 1985). Contrary to behaviourist psychology, which inspired the adaptive "black boxes" of cybernetics, cognitive science's aim was to bestow logical and abstract capabilities on machines. And unlike connectionism, these fields showed no interest in human physiology and behaviour, paying attention only to reasoning. The computational theory of the mind established a duality, positing that mental states could be described both in a physical form as a set of physical information-processing instances, and in a symbolic form as mechanically executable operations of comparing, ranking, or inferring meaning (Andler, 2016). This "physical symbol systems" hypothesis states that the mind does not directly access the world but rather consists of internal representations of the world that can be described and organized in the form of symbols inserted in programs.

A "toy" world

The founders of AI did their utmost to separate data from the sensory world and human behaviour[14]. The world of symbolic machines was a theatre backdrop created by the machine in order to project the syntax of its logical rules onto it: chess or checkers games (Arthur Samuel), geometry theorems (with Herbert Gelertner's Geometry Theorem Prover), video game backgrounds. The emblematic projects of this first wave of AI were characterized by the invention of simplified spaces of forms that must be recognized and moved, such as Marvin Minsky's MicroWorlds (MAC) or Terry Winograd's famous SHLURDU language. Just like the limited space with a few rooms and objects in which the Shakey robot is supposed to move around, it is a fictitious, "toy"[15] space in which objects can easily be

[13] Another contribution by J. McCarthy to the development of AI was the invention of time sharing, which allowed programmers to interact directly with the machine and its results, to communicate with it, to test it, and to make it "intelligent" by doing so (Edwards, 1996).

[14] As J. Markoff (2015) emphasizes, the entire history of computer science is underpinned by opposition between people promoting intelligence in machines (artificial intelligence – AI), incarnated by the SAIL, John McCarthy's laboratory at Stanford, and epitomized by the obsession with robotics; and people looking to distribute intelligence between humans and machine interfaces (intelligence amplification – IA), of which D. Engelbard's neighbouring laboratory would be a very productive stronghold, and which would give rise to the human-computer interaction (HCI) school. See also Grudin (2009).

[15] Minsky and Papert described MicroWorlds as "a fairyland in which things are so simplified that every statement about them would be literally false if asserted in the real world" (Minski and Papert, 1970: 36). The hypothesis underpinning this reduction

associated with the syntax of the rules, which are calculated to produce relevant system behaviour.

If the *calculator* projects its own *world*, this is also because its goal is to contain its own *target*. This is how this AI has been able to claim that it is "strong", because the goals given to the system are specific to it and can be deduced from a sort of reasoning incorporated into the logical inferences made by the models. The highly ingenious languages invented to shape the syntax of these systems are all inferential. They organize into stages the elementary processing operations transforming entities, each of which is an inference of a correct calculation (Andler, 1990: 100): a decision tree, intermediate chain of reasoning, breakdown of goals and sub-goals, and means-ends analysis. The rational target of the calculation is enclosed within the program's syntax. The machine must solve the problem, find the true or correct solution, and make the right decision[16]. Therefore, it was not necessary to give it the correct response (as the *examples* of learning techniques would do), because the rules have to lead it to this, following the inferences of the calculator. Because the syntax of the reasoning and the semantics of the manipulated objects were both constructed within the calculator, it was possible to confuse them with each other in correct and more or less deterministic reasonings – but at the expense of an artificial design in which the "intelligent" world was that implemented by the designer; a regulated, precise, and explicit world, so that reasoning could be its target. While these machines were capable of achieving certain performances in a closed environment, they quickly proved to be blind and stupid as soon as they were faced with an external world.

The first AI winter

At the beginning of the 1970s AI entered its first winter, which froze both the symbolic and connectionist projects. The two streams had both made many promises, and the results were far from meeting expectations. On the connectionist side, Frank Rosenblatt's Perceptron had been harmed by the media exposure in which its proponent – with the complicity of the US Navy – had liberally participated. Among a plethora of media headlines enthusiastic about the imminent arrival of intelligent machines, the *New York Times* announced: "The Navy last week demonstrated the embryo of an electronic computer named the Perceptron which, when completed in about a year, is expected to be the first non-living mechanism able to 'perceive, recognize and identify its surroundings without human training or control'"[17]. However, it was especially within symbolic AI, with Herbert Simon and Marvin Minsky leading it, that the exaggerated prophecies and announcements were quickly disappointing. Giddy with the researchers' promises, the army and the DARPA had thought that they would soon have machines to translate Russian texts, robots for infiltrating enemy lines, or voice command systems for tank and plane pilots, but discovered that the "intelligent" systems announced are only artificial games played in synthetic environments. In 1966 the National Research Council cut the funding for automated translation – a foreboding decision that would trigger a cascade of divestments by the financial and academic supporters of AI.

was that a network representation of abstract concepts within MicroWorlds could then be generalized to a more complete and more detailed world. Connectionists were to use the opposite reasoning: it is the description of information on the most elementary level that subsequently allows the network to generalize.

[16] For example, this is the viewpoint implemented with the ends/means analysis of Newell and Simon's (1963) General Problem Solver.

[17] "Electronic 'Brain' Teaches Itself", New York Times, 13 July 1958.

At the beginning of the 1970s, Minsky and Papert's MicroWorlds project at MIT experienced difficulties and lost its support. At Stanford, the Shakey robot no longer received military financing, and the DARPA SUR speech recognition program benefiting Carnegie Mellon was abruptly shut down. In England, the highly critical *Lighthill report* in 1973 would also play a role in stopping public funding for AI (Crevier, 1997: 133-143).

With the funding crisis, increasingly visible criticism began to be levelled at the very undertaking to logically model reasoning. In 1965, the RAND ordered Hubert Dreyfus to write a report on AI, which he entitled "Alchemy and Artificial Intelligence", and which used a vigorous argument that he later elaborated on in the first edition of his successful book *What Computers Can't Do* (Dreyfus, 1972). Bitter and intense, the controversy between the AI establishment and Hubert Dreyfus considerably undermined the idea that rational rules could make machines "intelligent". The explicit definition of logical rules was completely devoid of the corporeal, situated, implicit, embodied, collective, and contextual forms of the perception, orientation, and decisions of human behaviours[18]. Criticism was also put forward by the first generation of "renegades", who became significant opponents of the hopes that they themselves had expressed; for example Joseph Weizenbaum (1976), the founder of ELIZA, and Terry Winograd, the disappointed designer of SHRDLU (Winograd and Flores, 1986). "Intelligent" machines reasoned according to elegant rules of logic, a deterministic syntax, and rational objectives, but their world did not exist.

THE SECOND WAVE OF AI: A WORLD OF EXPERTS

AI nevertheless experienced a second spring during the 1980s, when it proposed a significant modification to the architecture of symbolic machines under the name of "expert systems"[19]. This renaissance was made possible by access to more powerful calculators allowing far bigger volumes of data to be input into computer memory. The "toy" worlds could thus be replaced with a repertoire of "specialized knowledge" taken from expert knowledge[20]. The artefacts of second-generation AI interacted with an external world that had not been designed and shaped by programmers. It was now composed of knowledge that had to be obtained from specialists in different fields, transformed into a set of declarative propositions, and formulated in a language that was as natural as possible (Winograd, 1972) so that users could interact with it by asking questions (Goldstein and Papert, 1977). This externality of the *world* to calculate led to a modification in the structure of symbolic machines, separating the "inference engine" into what would subsequently constitute the calculator and a series of possible *worlds* called "production systems", according to the terminology proposed by Edward Feigenbaum for DENDRAL, the first expert system that could identify the chemical components of materials. The data that supplied these knowledge bases consisted of long, easily modifiable and revisable lists of rules of the type "IF … THEN" (for example: "IF FEVER, THEN [SEARCH FOR

[18] Following H Dreyfus' book, and often in contact with social science and the humanities, a very productive school of AI criticism developed around the Wittgensteinian critique of rules. It resulted in work on the distribution of intelligence within space (Collins), the collective form of cognition (Brooks), or the embodied mind (Varela).

[19] The other names for intelligent machines during the second wave of AI are: "intelligent knowledge-based systems", "knowledge engineering", "office automation", or "multiagent systems".

[20] In 1967, during a lecture at Carnegie given before A. Newell and H. Simon, E. Feigenbaum challenged his former professors: "You people are working on toy problems. Chess and logic are toy problems. If you solve them, you'll have solved a toy problem. And that's all you'll have done. Get out into the real world and solve real-world problems" (Feigenbaum and McCorduck, 1983: 63).

INFECTION]"), which were dissociated from the mechanism allowing one to decide when and how to apply the rule (inference engine). MYCIN, the first implementation of a knowledge base of 600 rules aimed at diagnosing infectious blood diseases, was the starting point, in the 1980s, of the development of knowledge engineering that would essentially be applied to scientific and industrial contexts: XCON (1980) helped the clients of DEC computers configure them; DELTA (1984) identified locomotive breakdowns; PROSPECTOR detected geological deposits, etc. (Crevier, 1997, starting at p. 233). Large-scale industries developed AI teams as a part of their organization; researchers got started on the industrial adventure; investors rushed towards this new market; companies grew at an exceptional rate (Teknowledge, Intellicorp, Inference) – always with the faithful support of ARPA (Roland and Shiman, 2002) −; and the media seized the phenomenon, once again announcing the imminent arrival of "intelligent machines" (Waldrop, 1987).

The sanctuaries of the rules

Faced with criticism of the rigid computationalism of the first era that invented an abstract universe without realistic ties to the world, AI research undertook a top-down process to complete, intellectualize, and abstract the conceptual systems intended to manipulate the entities of these new knowledge bases. The symbolic movement thus strengthened its goal of rationalization by putting excessive emphasis on modelling in order to encompass a variety of contexts, imperfections in reasoning, and the multiplicity of heuristics, thus moving closer to the user's world through the intermediary of experts. This dedication to programming the calculator was characterized by more flexibility of logical operators (syntax) and the densification of the conceptual networks used to represent knowledge (semantics). The movement observed in AI research sought to de-unify the central, generic, and deterministic mechanism of computational reasoning in order to multiply, decentralize, and probabilize the operations carried out on knowledge. Borrowing from discussions around the modularity of the mind in particular (Fodor, 1983), the systems implemented in calculators broke down the reasoning process into elementary blocks of interacting "agents" which independently could have different ways of mobilizing knowledge and inferring consequences from it[21]. It was thus within the semantic organization of the meanings of heuristics taken from knowledge bases that the main innovations of the second wave of symbolic AI were designed. They used languages (PROLOG, MICROPLANNER, CYCL) and intellectual constructions with a rare degree of sophistication, for example the principle of lists; the notion of "conceptual dependency" as detailed by Robert Schank; Ross Quillian's semantic networks, and so on. The unfinished masterpiece of these multiple initiatives was Douglas Lenat's Cyc, a general common-sense knowledge ontology based on an architecture of "fundamental predicates", "truth functions" and "micro-theories", which everyone in the AI community admired but no one used.

The growing volume of incoming knowledge and the complexification of the networks of concepts intended to manipulate it were the cause of another large-scale shift: logical rules became conditional and could be "probabilitized". With regard to the rational and logical approach represented by John McCarthy, from the 1970s Marvin Minsky and Seymour

[21] M. Minsky's theory of "frames" (1975) was to be highly influential in this process and led to an all-encompassing theory in The Society of Mind (1986).

Papert defended the idea that "the dichotomy right/wrong is too rigid. In dealing with heuristics rather than logic the category true/false is less important than fruitful/sterile. Naturally, the final goal must be to find a true conclusion. But, whether logicians and purists like it or not, the path to truth passes mainly through approximations, simplifications, and plausible hunches which are actually false when taken literally" (Minsky and Papert, 1970: 41). Among the thousands of rules formulated by the experts, it is possible, based on a fixed premise (IF…), to establish a probability of whether the second proposition (THEN…) has a possibility of being true. The probabilization of knowledge rules meant that the deterministic form of the inferential reasoning that had experienced its moment of glory during the first age of AI could be relaxed. By becoming more realistic, diverse, and contradictory, the knowledge entering prediction machines also introduced probability into them (Nilsson, 2010: 475). When the "fruitful/sterile" pair replaced the "true/false" pair, the *target* providing the goal for the calculator appeared to be less of a logical truth than an estimate of the correctness, relevance or verisimilitude of the responses provided by the system. However, this estimate could no longer be taken care of essentially by the rules of the calculator; it had to be externalized towards a world composed of experts, who were mobilized to provide examples and counterexamples for machine learning mechanisms[22].

With the probabilization of inferences, these techniques penetrated deeper into the AI field in order to complete tasks that had become impossible for programmers to complete "by hand" (Carbonnell *et al.*, 1983). Following the work of Tom Mitchell (1977), learning methods could be described as a static solution for finding the best model within a *space of hypotheses* – or "versions" – automatically generated by the calculator. With expert systems, this space of hypotheses was highly structured by the nature of the input data, *i.e.*, the "knowledge". The learning mechanism "explores" the multiple versions of models produced by the calculator to search for a consistent hypothesis, making use of logical inferences to build reasonings (concept generalization, subsumption, inverse deduction). The statistical methods to eliminate potential hypotheses also matured and developed, producing inference-based reasoning such as decision trees (which subsequently gave rise to random forests, "divide and conquer" techniques, or Bayesian networks that served to order dependencies between variables with causalist formalism (Domingos, 2015)). Even when automated, the automatic discovery of a target function conserved the idea that models are hypotheses and that even though machines no longer applied a certain type of deductive reasoning, they chose the best possible reasoning from among a set of potential reasonings. However, starting in the early 1990s, a change in the nature of the data constituting the calculator's input world led to a shift in the field of machine learning. There was more data, it was no longer organized in the form of labelled variables or interdependent concepts, and it soon lost its intelligibility as it became numerical vectors (*infra*). No longer possessing a structure, data could only be collected in the form of statistical proximity. There was consequently a shift in the machine learning field from "exploration-based" methods to "optimization-based" methods (Cornuéjols *et al.*, 2018, p. 22), which would tear down the sanctuaries of the rules to the benefit of mass statistical calculations.

[22] For disciples of logic like A. Newell, such a position was heresy: "you have all these experts working for you and when you have a problem, you decide which expert to call in to solve the problem" (McCorduck, 1979: 267).

By increasingly expanding the volume and realism of the data to calculate, the inductive mechanism changed direction within the calculator. If the data no longer provided information on the relationships between one another (categories, dependencies between variables, conceptual networks), then in order to identify the target function, the inductive mechanism had to rely on the final optimization criteria in order to carry out the correct distribution (Cornuéjols *et al.*, 2018: 22). The transformation in the composition of the world to learn led researchers to modify the inductive method implemented, and by doing so, to propose an entirely different architecture for predictive machines. This shift accelerated with neural networks (*infra*), but the turn had already been prepared within the world of machine learning. Because data were increasingly less "symbolic", the inductive mechanism no longer searched for the model in the structure of initial data, but rather in the optimization factor (Mazières, 2016). The calculation target was no longer internal to the calculator but rather a value that the world assigned to it from outside – and which was very often "human", as demonstrated by all of the manual work to label data: does this image contain a rhinoceros (or not)? Did this user click on this link (or not)? The answer (the optimization criteria) must be input into the calculator along with the data so that the former can discover an adequate "model". The new machine learning methods (SVM, neural networks) thus proved to be more effective at the same time that they became unintelligible, as the inventor of decision trees, Léo Breiman (2001), emphasized in a provocative article on the two cultures of statistical modelling.

The magnificent sanctuaries erected by the builders of expert systems did not fulfil their promises. They soon proved to be extremely complex and very limited in their performance. The highly dynamic market that had developed in the mid-1980s suddenly collapsed and promising AI companies went bankrupt, in particular because to sell expert systems, they also had to sell specialized workstations called "LISP machines" at exorbitant prices, at a time when the PC market was on the rise (Markoff, 2015: 138 onwards). The decrease in cost and increase in calculation capacity during the 1980s made powerful calculators accessible to the heterodox and deviant schools of thought that had been excluded from the funding of large computer science projects as a result of the symbolic school's monopoly (Fleck, 1987: 153). The control of the small circle of influential universities over the "symbolic" definition of AI became weaker, given that expert systems produced only very limited results in the fields of voice synthesis, shape recognition, and other sectors. Symbolic AI was so weak at the beginning of the 1990s that the term almost disappeared from the research lexicon. Creating infinite repositories of explicit rules to convey the thousands of subtleties of perception, language, and human reasoning was increasingly seen as an impossible, unreasonable, and inefficient task (Collins, 1992; Dreyfus, 2007).

THE DISTRIBUTED REPRESENTATIONS OF DEEP LEARNING

It was in this context and the end of the depression phase which had begun in the late 1960s, that the connectionist approaches experienced a comeback in the 1980s and 1990s, with an immense amount of theoretical and algorithmic creativity. Following a meeting in June 1979 in La Jolla (California), organized by Geoff Hinton and James Anderson, an interdisciplinary research group composed of biologists, physicists, and computer scientists once again proposed to turn its attention back to the massively distributed and parallel nature of mental

processes in order to find an alternative to classic cognitivism. This group acquired real visibility in 1986 with the publication of two volumes of research under the name *Parallel Distributed Processing* (PDP), the term chosen to avoid the negative reputation of "connectionism" (Rumelhart *et al.*, 1986b). As opposed to the sequential approaches of computer and symbolic reasoning, PDP explored the micro-structures of cognition, once again using the metaphor of neurons to design a counter-model with original properties: elementary units were linked together via a vast network of connections; knowledge was not statically stored but resided in the strength of the connections between units; these units communicated with one another via a binary activation mechanism ("the currency of our system is not symbols but excitation and inhibition", p. 132); these activations took place all the time, in parallel, and not following the stages of a process; there was no central control over flows; one sub-routine did not trigger the behaviour of another one but instead sub-systems modulated the behaviour of other sub-systems by producing constraints that were factored into the calculations; and the operations carried out by the machine were similar to a relaxation system in which the calculation iteratively proceeded to carry out approximations to satisfy a large number of weak constraints ("the system should be thought of more *as settling into a solution* than *calculating* a solution", p. 135). The connectionists' device did create internal representations, and these representations could be high-level, but they were "sub-symbolic", statistical, and distributed (Smolensky, 1988). As this brief summary conveys, the connectionist approach was not a simple method but rather a highly ambitious intellectual construction intended to totally overturn computational cognitivism:

> « I think in the early days, back in the 50s, people like von Neumann and Turing didn't believe in symbolic AI. They were far more inspired by the brain. Unfortunately, they both died much too young and their voice wasn't heard. In the early days of AI, people were completely convinced that the representations you needed for intelligence were symbolic expressions of some kind, sort of cleaned-up logic where you can do non-monotonic things, and not quite logic, but like logic, and that the essence of intelligence was reasoning. What has happened now is that there's a completely different view, which is that what a thought is, is just a great big vector of neural activity. So, contrast that with a thought being a symbolic expression. I think that the people who thought that thoughts were symbolic expressions just made a huge mistake. What comes in is a string of words and what comes out is a string of words, and because of that, strings of words are the obvious way to represent things. So, they thought what must be in between was a string of words, or something like a string of words. And I think what's in between is nothing like a string of words. […] Thoughts are just these great big vectors and these big vectors have causal powers; they cause other big vectors, and that's utterly unlike the standard AI view[23]. »

While these epistemic references have lost their edge for the new pragmatic users of neural networks of today, who never experienced the exclusion and mockery to which their predecessors were subjected, they were a constant driver of the unrelenting pursuit of the

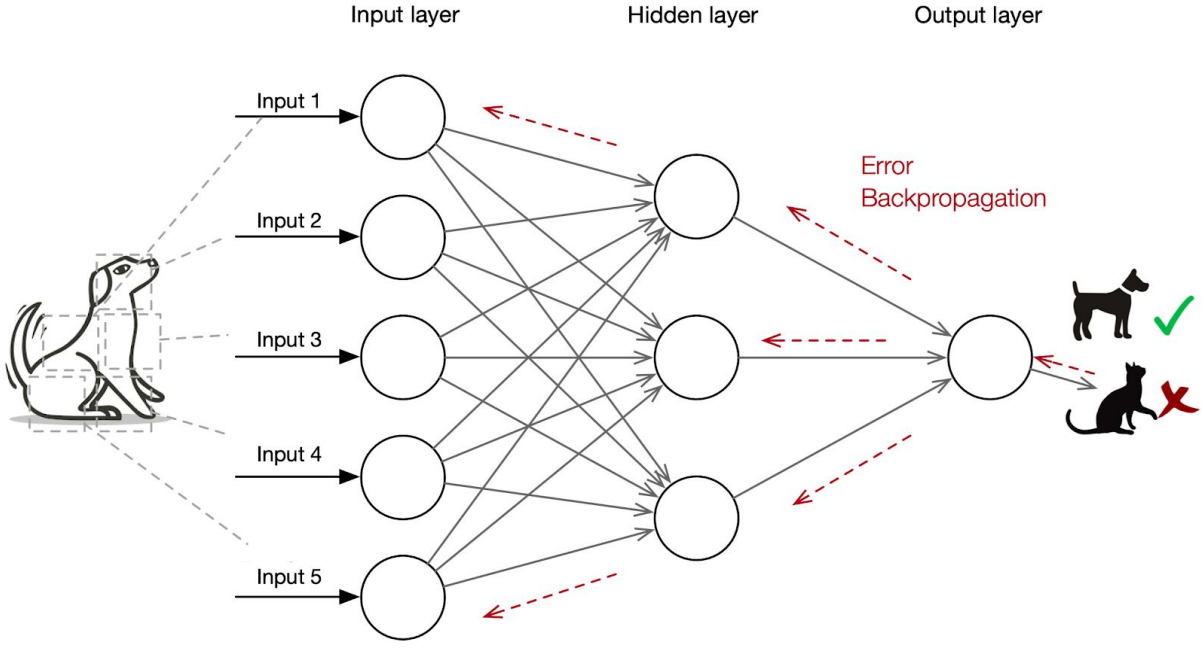
[23] Hinton G., "Heroes of Deep Learning: Andrew Ng interviews Geoffrey Hinton", YouTube, 8 August 2017 (starting at 37'20).

connectionist project. What had to be inserted between the strings of words coming in and those going out was not a model programmed by a logician's mind but a network of elementary entities that adapted its coefficients to inputs and outputs. To the extent possible, it was necessary for it to "do this on its own", and that required many artefacts.

Reconfiguring connectionism based on algorithms

In the early 1980s, in line with the work of John Hopfield, who proposed a revised version of the Perceptron model that gave each neuron the possibility of updating its values independently, the physicist Terry Sejnowski and the English psychologist Geoff Hinton developed new multi-layered architectures for neural networks (called Boltzmann machines). They also designed Nettalk, a system with three layers of neurons and 18,000 synapses that was successful in transforming texts into spoken phrases. However, the true turning point in this re-emergence was the creation of an algorithm called stochastic gradient back-propagation ("backprop" for short), which allowed the weights of coefficients to be calculated (Rumelhart *et al.*, 1986a). Contradicting the criticism of Minsky and Papert (1969), the authors showed that when networks are given multiple layers, they can easily be trained, as the additional layers of neurons make it possible for them to learn non-linear functions. The algorithm works by taking the derivative of the network loss function and "propagates" its error to correct the coefficients in the lower levels of the network[24]. Similarly to cybernetic machines, the output error is "propagated" towards the inputs (Figure 5).

Figure 5. Operation of a simple neural network



With the existence of a general-purpose algorithm that served to optimize any type of neural network, the 1980s and 1990s were a remarkable period of inventiveness that strongly

[24] A debate exists around the anteriority of the "backprop" algorithm. This method had been formulated and used on multiple occasions prior to the publication of this article, in particular by Linnainmaa in 1970, Werbos in 1974, and LeCun in 1985.

influenced the re-emergence of connectionism. One of the first successes was their application by Yann LeCun to zip code recognition carried out at AT&T Bell Labs (Lecun *et al.*, 1989), which "invented" the convolution technique. Using the US Postal Service database, he was successful in training a multi-layer network to recognize the zip code numbers written on packages. His successful approach became one of the first widespread business applications of neural networks, first in the banking (verification of check amounts) and postal sectors. This was followed by a series of proposals to integrate a greater number of hidden layers, to complexify the map of connections (encoders), to diversify optimization functions (ReLU), to integrate memory into network layers (recurrent networks and LSTM), to make unsupervised and supervised learning dependent on the part of the network (beliefs network), and so on (Kurenkov, 2015). In a highly creative way, numerous architectures wiring the relationships between neurons differently were put to the test to explore their properties.

"They might not be convex but they're more effective!"

Even though these algorithms laid the foundations of the majority of the approaches now referred to as deep learning, their invention was not immediately crowned with success. From 1995 to 2007, institutional support became very rare, papers were refused at conferences, and the results obtained remained limited. "*They went through a colossal winter*", a computer vision researcher says. "*The truth is that, at the time, nobody could get those machines to work. There were five laboratories in the world that knew how, but we couldn't manage to train them*"[25]. The researchers maintaining these techniques around Geoff Hinton, Yann LeCun, and Yoshua Bengio were a small, isolated – but cohesive – group, whose exclusive support came from the Canadian Institute for Advanced Research (CIFAR). Their situation became even more difficult in 1992 faced with the emergence of an original learning technique: support-vector machines – also called "kernel methods" –, which proved to be very effective on small datasets. Already exiled from the artificial intelligence community, connectionists once again found themselves on the fringes of the machine learning community.

> « At the time, if you said that you were making a neural network, you couldn't publish a paper. It was like that up until 2010, a has-been field. I remember that one time, LeCun was at our lab as a guest professor, and we had to make the effort of eating with him. Nobody wanted to go. It was bad luck, I swear. He would cry, his publications were refused at the CVPR, his methods weren't trendy, it wasn't sexy. So people gravitated towards what was popular. They gravitated towards kernels, SVM machines. And LeCun would say: "I have a 10-layer neural network and it does the same thing". Then we would say, "Are you sure? What's new?" Because once you have a neural network, even though it might have 10 layers this time, it doesn't work any better than the last one. It sucked! Then he would say, "Yeah, but there isn't as much data!"[26]. »

[25] Interview V, computer vision researcher, 12 March 2018.
[26] *Ibid.*

One argument constantly appears in the criticism levelled at the rare proponents of neural networks:

> « They [*SVM proponents*] would always say, "they [*neural networks*] aren't convex, they're just a shortcut". That's all that came out of their mouths. We would submit papers and they'd say, "they're not convex!" Maths wizards, obsessed with optimization, who'd never seen anything else in their life! It was like that for years. But we didn't give a damn.[27] »

Due to their non-linear nature[28], neural networks could not guarantee that the overall minimum had been found during the loss function optimization phase; it could just as well converge towards a local minimum or plateau[29]. From 2005 to 2008, a veritable policy of reconquest was initiated by the small group of "neural conspirators" (Markoff, 2015: 150) who set out to convince the machine learning community that it had been the victim of an epidemic of "convexitis" (LeCun, 2007). When their papers were refused at the NIPS in 2007, they organized an offshoot conference, transporting participants to the Hyatt Hotel in Vancouver by vehicle to defend an approach that the proponents of the dominant SVMs at the time considered archaic and alchemistic. Yann LeCun led the way with the title of his paper: "Who Is Afraid of Non-convex Loss Functions?" After presenting multiple results showing that neural networks were more effective than SVMs, he argued that an excessive attachment to the theoretical requisites resulting from linearized models was hindering the creation of innovative calculation architectures and the ability to consider other optimization methods. The very simple technique of stochastic gradient descent could not guarantee convergence towards a global minimum, yet "when empirical evidence suggests a fact for which you don't have theoretical guarantees, that precisely means that the theory is maladapted […], if that means that you have to throw convexity out the window, then that's okay!" (LeCun, 2017, 11'19).

> « Creative people are always crazy. At the beginning, that group, the creative people, were very tumultuous. After that, people from fields other than AI arrived, coming from maths and dismissing gradient descent to tell you about their methods: "my theorem is more elegant than yours". In optimization, people spent something like ten years searching for a more effective convex method and doing highly sophisticated but very costly things [*in terms of calculation capacity*]. That does have its advantages, but it had been bled dry, with thousands of papers, and when the big wave of data arrived, all of a sudden, none of their machines worked[30]! »

Transforming the world into vectors

In this way, connectionists shifted the scientific controversy around convexity, requiring the new data flows knocking at the doors of laboratories to contain the choice of the best

[27] Interview F., computer science researcher, one of the pioneers of deep learning in France, 20 July 2018.
[28] The uniqueness of neural networks lies in the fact that the neuron activation function creates discontinuities that produce non-linear transformations; an output cannot be reproduced by a linear combination of inputs.
[29] The property that ensured the reputation of SVMs was that they offered a linear system that could be standardized to guarantee convexity (Boser *et al.*, 1992).
[30] Interview F., one of the pioneers of deep learning in France, 20 July 2018.

calculation method. The architecture of predictive machines was transformed to cater for big data. It bared no resemblance to the small, calibrated, and highly-artificial datasets of the traditional competitions between researchers. This is because during this debate, the computerization of society and the development of web services triggered the emergence of new engineering problems based on large data volumes, such as spam detection, collaborative filtering techniques for making recommendations, inventory prediction, information searches, or the analysis of social networks. In the industrial context, the statistical methods of the new field of data science borrowed from and developed machine learning techniques (Bayesian methods, decision trees, random forests, etc.) without worrying about positioning themselves with respect to AI concerns (Dagiral and Parasie, 2017). On the other hand, it was clear that faced with the volume and heterogeneity of data features, as opposed to "confirmatory" techniques, it was necessary to use more "exploratory" and inductive methods (Tuckey, 1962). It was also in contact with industry players (AT&T originally, followed by Google, Facebook, and Baidu) that the neural network conspirators addressed problems, calculation capacities, and datasets allowing them to demonstrate the potential of their machines and to assert their viewpoint in the scientific controversy. They brought in a new referee: the effectiveness of predictions, in this case when applied to the "real" world.

Neo-connectionists first imposed their own terms in the debate. According to them, it was necessary to distinguish the "width" of the "shallow" architecture of SVMs from the "depth" (the term "deep learning" was coined by Geoff Hinton in 2006) of architectures based on layers of neurons. By doing so, they were able to demonstrate that depth is preferable to width: only the former is calculable when the data and dimensions increase, and is capable of capturing the diversity of data features. However convex SVMs may be, they do not give good results on large datasets: the dimensions increase too quickly and become incalculable; poor examples trigger considerable disturbances in predictions; and the solution consisting of linearizing a non-linear method deprives the system of its capacity to learn complex representations (Bengio and LeCun, 2007). The crusaders of connectionism thus managed to convince people that it was preferable to sacrifice the intelligibility of the calculator and rigorously controlled optimization for better perception of the complexity of dimensions present in this new form of data. When the volume of training data increases considerably, many local minimums exist, but there are enough redundancies and symmetries for the representations learned by the network to be robust and tolerant to errors in learning data. At the heart of the debate with the machine learning community, one thing went without saying: only laboratories used linear models; the world, the "real world" where data are produced by the digitization of images, sounds, speech, and text, is non-linear. It is noisy; the information contained in it is redundant; data flows are not categorized according to the attributes of homogeneous, clear, and intelligibly-constructed variables; examples are sometimes false. As Yoshua Bengio *et al.* wrote, "an AI must fundamentally understand the world around us, and we argue that this can only be achieved if it can learn to identify and disentangle the underlying explanatory factors hidden in the observed milieu of low-level sensory data" (2014, p. 1). This is why a "deep" architecture has more calculation power and is more "expressive" than a "shallow" architecture (LeCun and Benigo, 2007). Decreasing the intelligibility of the calculator to increase its ability to capture the complexity of the world, this controversy around convexity, clearly demonstrates that as opposed to being an

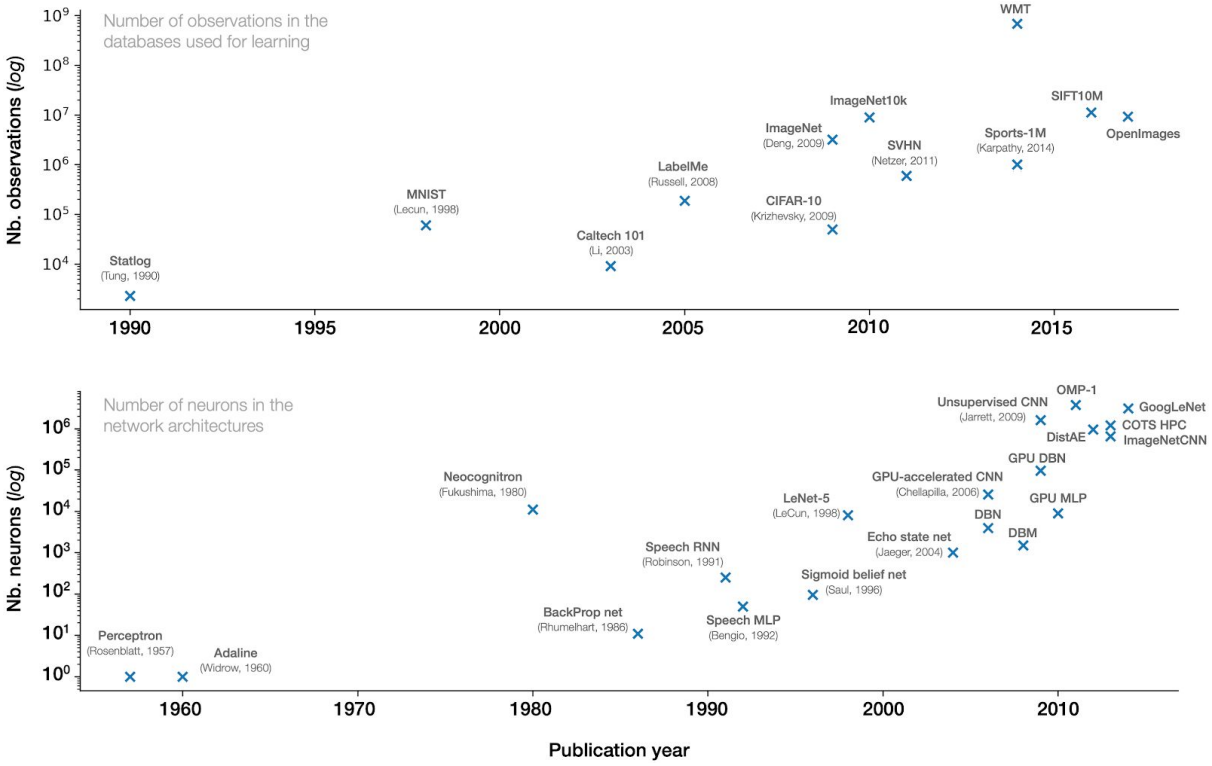
example of naive empiricism, the production of inductive machines was the result of intense work to convince people of the need to fundamentally reformulate the relationship between the calculator and the world.

In order for data to shift the scientific debate, it was therefore necessary to radically increase the volume of research datasets. In a 1988 article on character recognition, Yann LeCun used a database of 9,298 handwritten zip code numbers. The database used for character recognition since 2012 (MNIST) contained 60,000 labelled pieces of data on 28-pixel wide black and white images. It served to demonstrate the effectiveness of neural networks, but did not overcome the support for other techniques such as SVMs. In addition, scientific communities took advantage of the Internet to produce much more voluminous datasets and explicitly to build them for machine learning tasks by creating input/output pairs. This systematic collection of the broadest and most elementary digital data possible allowed gave more meaning to Hubert Dreyfus' statement that the "the best model of the world is the world itself" (Dreyfus, 2007: 1140). As the heterodox approaches critical of representational AI had long argued, representations are found in data from the world, as opposed to being internal to the calculator (Brooks, 1988). The creation of ImageNet, the dataset used during the challenge presented at the beginning of this article, which was initiated by Li Fei-Fei (Deng *et al.*, 2009), is exemplary of this. Today, this database contains 14 million images, the elements of which were manually annotated into 21,841 categories by using the hierarchical structure of another classic database in natural language processing, Wordnet (Miller, 1995). To be successful in this immense task of qualifying elements identified by hand-drawn squares in images, it was necessary to crowdsource the tasks to thousands of annotators via Mechanical Turk (Su *et al.*, 2012; Jaton, 2017). From 9,298 pieces of data to 14 million, such a massive change in the volume of datasets – and therefore in the dimensions of the data –became meaningful only when accompanied by an exponential growth in the power of calculators, offered by parallel computing and the development of GPUs (Figure 6). In 2009, "backprop" was implemented on graphics cards that enabled a neural network to be trained up to 70 times faster (Raina *et al.*, 2009). Today, it is considered good practice to learn a category in a classification task with 5,000 examples per category, which quickly leads datasets to contain several million examples. The exponential growth in datasets accompanied a parallel change in calculator architectures: the number of neurons in a network doubles every 2.4 years (Goodfellow *et al.*, 2016: 27).

However, another transformation in data was also initiated by connectionists, this time to granularize data and transform it into a calculable format through "embedding" operations. A neural network requires the inputs of the calculator to take on the form of a vector. Therefore, the world must be coded in advance in the form of a purely digital vectorial representation. While certain objects such as images are naturally broken down into vectors, other objects need to be "embedded" within a vectorial space before it is possible to calculate or classify them with neural networks. This is the case of text, which is the prototypical example. To input a word into a neural network, the *Word2vec* technique "embeds" it into a vectorial space that measures its distance from the other words in the corpus (Mikolov *et al.*, 2013). Words thus inherit a position within a space with several hundreds dimensions. The advantage of such a representation resides in the numerous operations offered by such a transformation. Two terms whose inferred positions are near one another in this space are

equally similar semantically; these representations are said to be distributed: the vector of the concept "apartment" [-0.2, 0.3, -4.2, 5.1...] will be similar to that of "house" [-0.2, 0.3, -4.0, 5.1...]. Semantic proximity is not deduced from a symbolic categorization but rather induced from the statistical proximity between all of the terms in the corpus. Vectors can thus advantageously replace the words that they represent to resolve complex tasks, such as automated document classification, translation, or automatic summarization. The designers of connectionist machines thus carried out highly artificial operations to transform data into another representation system and to "rawificate" them (Denis and Goëta, 2017). While natural language processing was pioneering for "embedding" words in a vectorial space, today we are witnessing a generalization of the embedding process which is progressively extending to all applications fields: networks are becoming simple points in a vectorial space with *graph2vec*, texts with *paragraph2vec*, films with *movie2vec*, meanings of words with *sens2vec*, molecular structures with *mol2vec*, etc. According to Yann LeCun, the goal of the designers of connectionist machines is to put the world in a vector (*world2vec*). Instead of transforming inputs into symbols interrelated via a fabric of interdependent concepts, this vectorization creates neighbourhood proximities between the internal properties of the elements in the learning corpus[31].

Figure 6. Growth in the number of observations in research datasets from 1990 to 2015 (above) and in the number of neurons in calculation architectures implemented from 1960 to 2015



[31] True to the cognitive model of connectionism, the three main proponents of deep learning, Y. LeCun, G. Hinton, and Y. Bengio, translate it into calculatory terms: "The issue of representation lies at the heart of the debate between the logic-inspired and the neural-network-inspired paradigms for cognition. In the logic-inspired paradigm, an instance of a symbol is something for which the only property is that it is either identical or non-identical to other symbol instances. It has no internal structure that is relevant to its use; and to reason with symbols, they must be bound to the variables in judiciously chosen rules of inference. By contrast, neural networks just use big activity vectors, big weight matrices and scalar non-linearities to perform the type of fast "intuitive" inference that underpins effortless commonsense reasoning" (LeCun *et al.*, 2015: 436).

These data were partially taken from Goodfellow et al. *(2016: 21 and 24) and were completed drawing on the Wikipedia article "List of datasets for machine learning research".*

From modelling to architecture

Through a real shift, that which was offered by the variety and the volume of data had to be removed from the calculator. The designers of neuron-based architectures therefore proceeded to systematically and strictly eliminate all the explicit rules "intentionally" integrated into calculators for the purpose of identifying, describing, or aggregating data in advance. A researcher in the field explained:

> « There was a force behind it. There was a wave, the wave of data, a sort of giant background wave that washed everything away. And that completely threw out the schools of thought that had been based on human modelling, explicit modelling. I worked in multiple application fields, including for speech, writing, text, social data, and I saw the same thing every time. For a time, people thought about putting knowledge into their system, but that was swept away. Systematically! And it has been crumbling for thirty years since, in field after field. That's how things are. It's a funny thing, you know. It's like when people spend their whole life believing in a socialist regime and then it collapses right in front of them… It's the same sort of thing[32]. »

From the end of the 2000s, the destabilizing feeling that arose from watching a technique without theory replace years of efforts to patiently model behaviour, spread throughout the signal, voice, image, and machine translation communities one by one. In field after field, neural network calculations became more efficient, transferring the operations that had previously been the main focus of the attention of scientific activity – feature engineering and pattern recognition – to the distribution of weight in the network. These techniques consisted of "handcrafting" algorithms to identify the features of initial data – an extraction process that facilitated learning by simplifying the relationship between the features and the objective of the problem. The increasingly effective automation of feature recognition allowed statistical machine learning techniques to become more powerful than modellers within calculators (*supra*)[33]. However, neural networks took this shift to a radical degree, this time eliminating any feature extraction process to the benefit of "end-to-end" processing: going from the "raw" piece of digital data to the "labelled" example without explicitly aiming at producing intermediate representations of data guiding calculations towards the objective.

An example of this shift is the principle of convolution used in the opening illustration of this article. The computer vision community developed extremely subtle extraction methods to identify the borders, corners, transitions in contrast, and specific points of interest in images in order to associate them with bags of words used as features for the task entrusted to the calculator. These operations became the implicit responsibility of the specific structure given to convolutional networks: breaking the image down into little tiles of pixels entrusted

[32] Interview F., one of the pioneers of deep learning in France, 20 July 2018.

[33] "Many developers of AI systems now recognize that, for many applications, it can be far easier to train a system by showing it examples of desired input-output behavior than to program it manually by anticipating the desired response for all possible input" (Jordan and Mitchell, 2015: 255).

to separate segments of neurons in order to reassemble them in another network layer. Rather than modelling a rhinoceros, or the features of the blocks of pixels that govern the rhinoceros shape, several thousand photos of rhinoceroses moving within the image, a portion of the body of which is truncated, viewed from varied angles and positions, will do a much better job of imprinting the "rhinoceros" shape-concept in the weight of neurons than a feature preprocessing procedure that does not know how to deal with problems related to invariance of the scale, transformation, or rotation. The relationship between the piece of data and its feature is not sought out but rather obtained. Neural networks do extract features – the edges are often "seen" by the first layer of neurons, the corners by another, the more complex shape elements by another –, but these operations, without having been explicitly implemented, are the emergent effects of the network under architectural constraints.

The pre-processing of calculation "parameters" was thus transformed into defining the "hyper-parameters" of the calculator. The more the human component of modelling decreased, the more complex specifying the architecture of inductive machines became. A fully connected neural network does not produce anything; it is necessary to sculpt it to adapt its architecture to the machine learning task entrusted to it: number of hidden layers, number of neurons per layer, map of connections, choice of activation function, type of optimization, coefficients at the beginning of learning, choice of objective function, number of times that the learning dataset will be shown to the model, etc. These configurations are adjusted via trial/error. The technique of pruning, for example, consists in removing neurons to see whether this changes the performance of the network. The dropout technique suggests, during the learning phase, not sending signals towards certain neurons on the input layer or hidden layers randomly, in order to avoid over-fitting when the network has to generalize towards fresh data. These recipes, good practices, and industry standards are to a large extent the subject of the discussions in the community and are still do-it-yourself in nature (Domingos, 2012). Faced with mathematical refining of feature extraction, the creation of neural networks may thus appear to be the job of a hacker, an activity for gifted programmers endowed with a sort of black magic.

> « The thing that they did to remove all feature extraction to adopt the raw image, the guys that did that with Hinton, they were crazy, because it's one thing to reproduce something, but to go about it like that by exploring! They created systems of unimaginable complexity and were able to get them to work. If you take a paper from one of those people and look at it, you'll say, it's scary, I'm too old for this! Those guys almost even talk to you like they're programming. They don't create a description with three equations that make sense for me. But in 5 lines, they can describe something that's hyper-complex. So, in other words, he created an architecture in which he placed 100 elements linked to one another, and to link them, for each one, you have ten possible choices. He played with that and managed to get it to work. That's a hacker; that's the job of a hacker[34] ! »

[34] Interview F., one of the pioneers of deep learning in France, 20 July 2018.

Hyper-parameters were therefore the place to which the new explicability requirements for neural networks were moved. The data only "speaks for itself" when it is submitted to an architecture which cannot be learned from the data, and which and from that point on was the focus of a large portion of AI research. At the NIPS conference, one noteworthy paper was an article proposing a new architecture, to which, like the planets, researchers systematically provided names, thus establishing a strange kind of bestiary (Figure 7). By shifting from modelling to architecture, which was the place where researchers' inventiveness could be expressed, the skills and qualities required by their design were also transformed. This enabled a new population of data scientists, do-it-yourselfers, and programmers to enter into the previously very closed field of AI producers, particularly due to the availability of open and easy-to-use tools. By transforming the architecture of predictive machines, connectionists contributed to shifting the social worlds of AI: first, because "real" data, and in particular data from digital industries, (partially) replaced the "toy" datasets of academic laboratories; and second, because the know-how required to create connectionist machines required computer development skills other than those of the previous AI generations.

THE WORK OF INDUCTION

The path of intelligent machines, the history of which we have just summarized in four successive configurations, shows the profound transformation in their architecture (Table 1 below). The *world*, the *calculator*, and the *target* of these devices have been profoundly reorganized, and the interrelations between these components shape devices that offer markedly different definitions of intelligence, reasoning, and prediction.

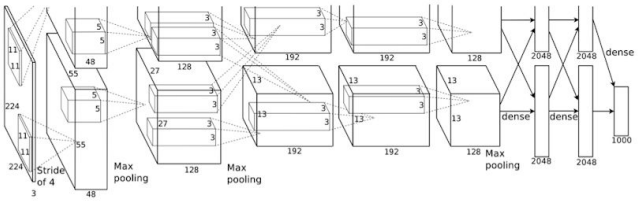
Table 1. The four ages of predictive machines

Machine	World	Calculator	Target
Cybernetics (connectionist)	*Environment*	*"Black box"*	*Negative feedback*
Symbolic AI (symbolic)	*"Toy" world*	*Logical reasoning*	*Problem-solving*
Expert systems (symbolic)	*World of expert knowledge*	*Selection of hypothesis*	*Examples / counter-examples*
Deep Learning (connectionist)	*The world as a vector of big data*	*Deep neural network*	*Objective-based error optimization*

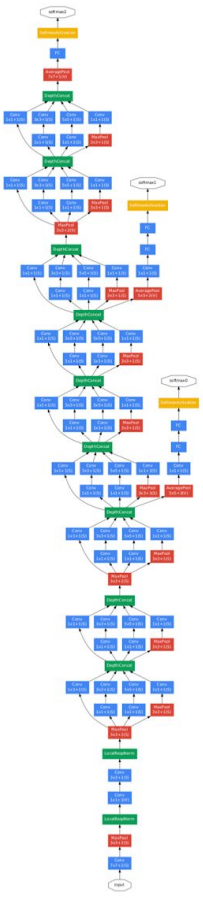
A global dynamic nevertheless appears in this shifting history. The materialistic undertaking to represent the mind computationally has adopted a resolutely connectionist approach today. But the current success of inductive machines certainly does not mean that a final point or a "solution" has been found. Despite their prowess, deep learning techniques are very far from satisfying the needs of the general artificial intelligence program, a source of constant criticism from "symbolists" who, clinging to the cliff, claim that the two approaches

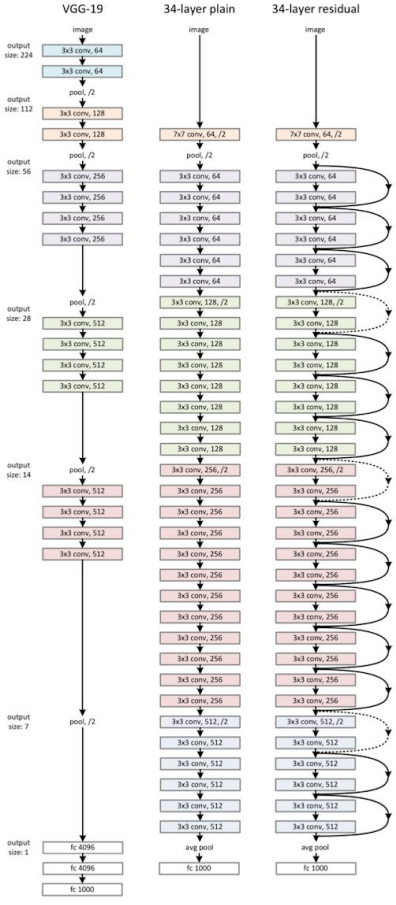
need to be hybridized[35]. However, what is clear from the history we have followed in this article is that this inductive reorganization of predictive calculation could not have been done without considerable and ambitious efforts to modify the balance between the world of data and the form of calculation.

Figure 7. Examples of three victorious neural network architectures at the ILSVRC challenge from 2012 to 2015



(a) 9 layers in AlexNet network architecture (ImageNet 2012)





(b) 40 layers in GoogLeNet (ImageNet 2014)

(c) 152 layers by Microsoft (ImageNet 2015)

[35] See the debate between Y. LeCun and G. Markus (2017). Markus called for a hybridization of the symbolic and connectionist approaches because the latter had numerous weaknesses that were creating new research concerns in the field: it allowed one to interpolate between two known examples, but was bad at extrapolating in situations that had not been the subject of learning; its models consumed a considerable amount of labelled data which was often not always accessible; it was not capable of establishing a hierarchy of reasonings by isolating rules and abstractions; it was not capable of integrating pre-existing knowledge relative to the data calculated; it lacked transparency and explicability; it predicted in a stable and static world without being prepared for unexpected elements; it was probabilistic and incapable of predicting with certainty and precision (Markus, 2018).

First of all, as an input to the calculator, the composition of the world has undergone a profound shift in terms of atomization and granularization. While the "toy" and expert knowledge worlds of symbolic machines consisted of small, limited worlds that had been cleaned up and domesticated via a framework of intelligible and interdependent features, connectionist machines operate in a world in which the data must not only come in huge volumes, but must also be as atomized as possible in order to deprive it of any explicit structure. Even if the data contains regularities, compositional relationships, global styles, etc., these elements must be highlighted by the calculator and not by the programmer. The first component of achieving induction therefore had to consist in inputting data into the system in the most elementary way possible: pixels rather than shapes, frequencies rather than phonemes, letters rather than words, clicks rather than statements by Internet users, behaviours rather than categories, and so on (Cardon, 2017). The fact that data may be heterogeneous, redundant, and often incorrect is no longer a problem; each signal can be added in the form of a new column in the input matrix that forms the world of connectionist machines. Therefore, data is not made available to the perception of calculators in a "raw" and "immediate" form, but rather is subject to atomization and dissociation in order to transform it into the most elementary possible standardized digital signs. To create these inputs, a new metrology of sensors, recordings, and databases constitutes an essential infrastructure for transforming images, sounds, movements, clicks, or variables of all types into these giant vectors required by connectionist machines (Mackenzie, 2017).

The second feature of this shift as a whole is the disappearance of the *a priori* mobilization of the activities of the calculator (a phenomenon often referred to as the "end of theory" (Anderson, 2008)) to the benefit of the probabilitization of models within an increasingly broader space of hypotheses, followed by a more radical dispersion in models when the various dimensions of data are taken into account throughout multiple layers of the neural networks. The immense intellectual undertaking to model reasoning, typical of the early ages of AI, has crumbled, paving the way for important contributions to computer science research. Connectionist machines have shifted AI concerns from resolving the abstract problems that were the focus of orthodox cognitive science, to the perception of features within enormous volumes of sensory signals. The second feature of the undertaking to achieve induction was likely successfully attaining the conditions that would overturn the calculation device of AI in order to make programs outputs and not inputs. Yet neural networks by no means eliminate "theory"; instead, they shift it towards the hyper-parameters of the calculator's architecture, giving the word "theory" a less "symbolizable" meaning. This topic makes issues relating to the understanding and interoperability of the processes that they implement in making their predictions particularly delicate (Burrel, 2016; Cardon, 2015). As the PDP of the 1980s and much research on complex systems urge, we may very well have to learn how to turn modelling forms that no longer have the properties that we have been accustomed to (linearity, readability, completeness, parsimony, etc.) by the – very "symbolic" – idea of the intelligibility of models in social science, into elements that are perceptible, appropriable, and discussable.

The third shift is related to the target given to the calculator. While the intelligent machines devised by symbolic AI assigned themselves the goal of the rational expectations of logic – a rationality internal to calculations that allowed AI proponents to claim that these machines

were "autonomous" –, in the connectionist model the calculation target belongs not to the calculator but rather to the world that has given it "labelled" examples. The outputs – produced, symbolized, and biased by humans – today constitute one of the most important inputs of connectionist machines. The third feature of the undertaking to achieve induction consisted in basing the performance of prediction on the world itself, renewing the adaptive promises of the *reflection machines* of cybernetics: to form a system with the environment to calculate in order to implement a new type of feedback loop. It is therefore somewhat paradoxical that by perpetuating a "symbolic" conception of machine intelligence, much of the critical debate around the biases of new calculation forms was directed at the strategic intentions of programmers, whereas the latter were constantly seeking to eliminate all traces of prior "human" intervention (knowledge free) in the calculator's operations. Admittedly, it is wise to be very vigilant with respect to the strategic objectives that digital economy companies seek to slip into their calculations. However, to be more relevant and efficient, criticism should also adapt to the "inductive" revolution of predictive machines, because while calculated predictions are not the "natural" reflection of the data, the supervision of learning to which it is necessary to pay attention has been becoming increasingly focused on the composition of input data, the architecture retained by different systems, and the objectives. Whether apologetic or critical, the representations of artificial intelligence informed by a science-fiction genre drawing its worldview from symbolic AI – Marvin Minsky was the scientific adviser for *2001: A Space Odyssey* – appear to be highly inappropriate, obsolete, and above all unimaginative faced with the much more intriguing and unique reality of these new machines.

REFERENCES

ANDERSON C. (2008), *The end of theory: Will the data deluge makes the scientific method obsolete?*, June 23, http://www.wired.com/science/discoveries/magazine/16-07/pb_theory.

ANDERSON J. A., ROSENFELD E. (eds.) (1988), *Neurocomputing: Foundations of Research*, Cambridge, The MIT Press.

ANDLER D. (1990), « Connexionnisme et cognition. À la recherche des bonnes questions », *Revue de synthèse*, n° 1-2, pp. 95-127.

ANDLER D. (1992), « From paleo to neo-connectionism », in G. VAN DER VIJVER (ed.), *Perspectives on Cybernetics*, Dordrecht, Kluwer, pp. 125-146.

ANDLER D. (2016), *La silhouette de l'humain. Quelle place pour le naturalisme dans le monde d'aujourd'hui ?*, Paris, Gallimard.

ASHBY R. (1956), *Introduction to Cybernetics*, London, Chapman & Hall.

BENGIO Y., COURVILLE A., VINCENT P. (2013), « Representation Learning: A Review and New Perspectives », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, n° 8.

BENGIO Y., LECUN Y. (2007), « Scaling Learning Algorithms towards AI », in L. BOTTOU, O. CHAPELLE, D. DECOSTE, J. WESTON, *Large-Scale Kernel Machines*, Cambridge, MIT Press.

BOSER B. E., GUYON I. M., VAPNIK V. N. (1992), « A Training Algorithm for Optimal Margin Classifiers », *Fifth Annual Workshop on Computational Learning Theory*, Pittsburgh, pp. 144-152.

BREIMAN L. (2001), « Statistical Modeling: The Two Cultures », *Statistical Science*, vol. 16, n° 3, pp. 199-215.

BROOKS R. A. (1988), « Intelligence without Representation », *Mind Design*, in J. HAUGELAND (ed.), *Mind Design*, Cambridge MA, The MIT Press.

BURRELL J. (2016), « How the machine 'thinks': Understanding opacity in machine learning algorithms », *Big Data & Society*, January-June, pp. 1-12.

CARBONELL J. G., MICHALSKI R. S., MITCHELL T. (1983), « Machine Learning: A Historical and Methodological Analysis », *AI Magazine*, vol. 4, n° 3, pp. 69-79.

CARDON D. (2015), *À quoi rêvent les algorithmes. Promesses et limites*, Paris, Seuil, coll. « République des idées ».

CARDON D. (2017), « Infrastructures numériques et production d'environnements personnalisés », in K. CHATZISI, G. JEANNOT, V. NOVEMBER, P. UGHETTO (dir.), *Les métamorphoses des infrastructures, entre béton et numérique*, Bruxelles, Peter Lang, pp. 351-368.

COLLINS H. M. (1992), *Experts artificiels. Machines intelligentes et savoir social*, Paris, Seuil.

CORNUÉJOLS A., MICLET L., BARRA V. (2018), *Apprentissage artificiel. Concept et algorithmes*, Paris, Eyrolles (3e éd.).

CREVIER D. (1997), *À la recherche de l'intelligence artificielle*, Paris, Champs/ Flammarion [1re éd. américaine 1993].

DAGIRAL É., PARASIE S. (2017), « La "science des données" à la conquête des mondes sociaux. Ce que le "Big Data" doit aux épistémologies locales », in P.-M. MENGER, S. PAYE (dir.), *Big data et traçabilité numérique. Les sciences sociales face à la quantification massive des individus*, Paris, Collège de France.

DENG J., DONG W., SOCHER R., LI L. J., LI K., FEI-FEI L. (2009). « Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition », *CVPR 2009*, pp. 248-255.

DENIS J., GOËTA S. (2017), « Les facettes de l'Open Data : émergence, fondements et travail en coulisses », in P.-M. MENGER, S. PAYE (dir.), *Big data et traçabilité numérique. Les sciences sociales face à la quantification massive des individus*, Paris, Collège de France.

DOMINGOS P. (2012), « A Few Useful Things to Know about Machine Learning », *Communication of the ACM*, vol. 55, n° 10, pp. 78-87.

DOMINGOS P. (2015), *The Master Algorithm. How the question for the ultimate machine will remake our world*, London, Penguin Random House UK.

DREYFUS H. (1972), *What Computers Can't Do: The Limits of Artificial Intelligence*, New York, Harper and Row.

DREYFUS H. (2007), « Why Heideggerian AI failed and how fixing it would require making it more Heideggerian », *Artificial Intelligence*, n° 171, pp. 1137-1160.

DUPUY J.-P. (2005), *Aux origines des sciences cognitives*, Paris, La Découverte.

EDWARDS P. N. (1996), *The Closed World. Computers and the Politics of Discourses in Cold War America*, Cambridge MA, The MIT Press.

FEIGENBAUM E. A., McCORDUCK P. (1983), *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World*, Reading, Addison Wesley.

FLECK J. (1982), « Development and Establishment in Artificial Intelligence », in N.ELIAS, H.MARTINS, R.WHITLEY (eds.), *Scientific Establishments and Hiérarchies, Sociology of the Sciences Yearbook*, vol. 6, Dordrecht, Reidel, pp. 169-217.

FLECK J. (1987), « Postscript: The Commercialisation of Artificial Intelligence », in B. BLOMFIELD (ed.), *The Question of AI*, London, Croom-Helm, pp. 149-64.

FODOR J. A. (1983). *Modularity of Mind: An Essay on Faculty Psychology*, Cambridge MA, MIT Press.

GALISON P. (1994), « The ontology of the enemy: Norbert Wiener and the cybernetic vision », *Critical Inquiry*, vol. 21, n° 1, pp. 228-266.

GARDNER H. (1985), *The Mind's New Science. A History of Cognitive Revolution*, New York, Basic Books.

GITELMAN L. (ed.) (2013), *Raw data is an oxymoron*, Cambridge MA, MIT Press.

GOLDSTEIN I., PAPERT S. (1977), « Artificial Intelligence. Language and the Study of Knowledge », *Cognitive Science*, vol. 1, n° 1.

GOLDSTINE H. (1972), *The Computer From Pascal to Von Neumann*, Princeton, Princeton University Press.

GOODFELLOW I., BENGIO Y., COURVILLE A. (2016), *Deep Learning*, Cambridge MA, MIT Press.

GRUDIN J. (2009), « AI and HCI: Two fields divided by a common focus », *AI Magazine*, vol. 30, n° 4, pp. 48-57.

HAUGELAND J. (1985), *Artificial Intelligence: The Very Idea*, Cambridge MA, MIT Press.

HEBB D. O. (1949), *The Organization of Behavior*, New York, Wiley.

HOPFIELD J. J. (1982), « Neural Networks and Physical Systems with Emergent Collective Computational Abilities », *Proc. Natl. Acad. Sc. USA*, vol. 79.

JATON F. (2017), « We get the algorithms of our ground truths: Designing referential databases in Digital Image Processing », *Social Studies of Science*, vol. 47, n° 6, pp. 811-840.

JORDAN M. (2018), « Artificial Intelligence: The Revolution hasn't happened yet », *Medium*, April 19.

JORDAN M. I, MITCHELL T. M. (2015), « Machine learning: Trends, perspectives, and prospects », *Science*, vol. 349, n° 6245, pp. 255-260.

KRIZHEVSKY A., SUTSKEVER I., HINTON G. (2012), « ImageNet Classification with Deep Convolutional Neural Networks », *NIPS 2012*, Lake Tahoe, December 3-6.

KURENKOV A. (2015), « A 'Brief' History of Neural Nets and Deep Learning », *andreykurenkov.com*, December 24.

LATOUR B. (1987), *Science in Action: How to Follow Scientists and Engineers through Society*, Cambridge MA, Harvard University Press.

LECUN Y. (2007), « Who is Afraid of Non-Convex Loss Functions? », *2007 NIPS workshop on Efficient Learning*, Vancouver, December 7.

LECUN Y., BENGIO Y., HINTON G. (2015), « Deep learning », *Nature*, vol. 521, n° 7553.

LECUN Y., BOSER B., DENKER J., HENDERSON D., HOWARD R., HUBBARD W. JACKEL L. (1989), « Backpropagation Applied to Handwritten Zip Code Recognition », *Neural Computation*, vol. 1, n° 4, pp. 541-551.

LECUN Y., MARKUS G. (2017), «Debate: "Does AI Need More Innate Machinery?" », *YouTube*, October 20.

MARKOFF J. (2015), *Machines of loving grace. Between human and robots*, HarperCollins Publishers, 2015.

MACKENZIE A. (2017), *Machine Learners. Archaeology of a Data Practice*, Cambridge MA, The MIT Press.

MARKUS G. (2018), « Deep Learning: A Critical Appraisal », *arXiv :1801.00631*, January 2.

MAZIÈRES, A. (2016). *Cartographie de l'apprentissage artificiel et de ses algorithmes*. Manuscrit de thèse, Université Paris Diderot.

McCARTHY J. (1988), « [Review of] Bloomfield Brian ed. The Question of Artificial Intelligence... », *Annals of the History of Computing*, vol. 10, n° 3, pp. 221-233.

McCORDUCK P. (1979), *Machines Who Think. A Personal Inquiry into the History and Prospects of Artificial Intelligence*, Natick, AK Peters.

MIKOLOV T., SUTSKEVER I., CHEN K., CORRADO G. S., DEAN J. (2013), « Distributed representations of words and phrases and their compositionality », *Proceedings of the 26th International Conference on Neural Information Processing Systems*, pp. 3111-3119.

MILLER G. A. (1995), « WordNet: A Lexical Database for English », *Communications of the ACM*, vol. 38, n° 11, pp. 39-41.

MINSKY M. (1975), « A Framework for Representing Knowledge », in P. WINSTON (ed.), *The Psychology of Computer Vision*, New York, McGraw-Hill.

MINSKY M. (1986), *The Society of Mind*, New York, Simon & Schuster.

MINSKY M., PAPERT S. (1969), *Perceptrons: An Introduction to Computational Geometry*, Cambridge MA, The MIT Press.

MINSKY M., PAPERT S. (1970), « Draft of a Proposal to ARPA for Research on Artificial Intelligence at MIT, 1970-1971 », *Artificial Intelligence Lab Publication*, MIT.

MITCHELL T. (1977), « Version Spaces: A Candidate Elimination Approach to Rule Learning », *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, Cambridge, August, pp. 305-310.

NEWELL A., SIMON H., SHAW J. C. (1956), « The Logic Theory Machine », *IRE Transactions on Information Theory*, vol. IT-2, n° 3.

NEWELL A., SIMON H. A. (1963), « GPS: A Program That Simulates Human Thought », in E. A. FEIGENBAUM, J. FELDMAN (eds.), *Computers and Thought*, New York, McGraw-Hill, pp. 279-283.

NILSSON N. J. (2010), *The Quest for Artificial Intelligence. A history of ideas and achievements*, Cambridge, Cambridge University Press.

OLAZARAN M. (1996), « A Sociological Study of the Official History of the Perceptron Controversy », *Social Studies of Science*, vol. 26, n° 3, pp. 611-659.

PICKERING A. (2010), *The Cybernetic Brain. Sketches of another Future*, Chicago IL, The Chicago University Press.

RAINA R., MADHAVAN A., NG A. Y. (2009). Large-scale deep unsupervised learning using graphics processors. In *Proceedings of the 26th annual international conference on machine learning*, ACM, pp. 873-880.

RID T. (2016), *Rises of the Machines. The lost history of cybernetics*, London, Scribe Publications.

ROLAND A., SHIMAN P. (2002), *Strategic Computing. DARPA and the Quest for Machine Intelligence, 1893-1993*, London, The MIT Press.

ROSENBLUETH A., WIENER N., BIGELOW J., (1943), « Behavior, Purpose and Teleology », *Philosophy of Science*, vol. 10, n° 1, pp. 18-24.

RUMELHART D. E., HINTON G., WILLIAMS R. J. (1986a), « Learning representations by back-propagating errors », *Nature*, n° 323, pp. 533-536.

RUMELHART D. E., McCLELLAND J. L. (1986b), « PDP Models and General Issues in Cognitive Science », in PDP RESEARCH GROUP (1986), *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*, Cambridge MA, MIT Press.

SHANNON C. (1948), « A mathematical theory of communication », *Bell System Technical Journal*, n° 27, pp. 379-423.

SKINNER B. F. (1971), *Beyond Freedom and Dignity*, New York, Bantam.

SMOLENSKY P. (1988), « The proper treatment of connectionism », *The Behavioral and Brain Sciences*, vol. 11, pp. 1-74.

SU H., DENG J., FEI-FEI L. (2012), « Crowdsourcing Annotation for Visual Object Detection », *AAAI Workshops*, Toronto.

TRICLOT M. (2008), *Le moment cybernétique. La constitution de la notion d'information*, Paris, Champ Vallon.

TUKEY J. W. (1962), « The future of data analysis », *The Annals of Mathematical Statistics*, vol. 33, n° 1, pp. 1-67.

TURING A. (2004), « Proposal for Development in the Mathematics of an Automatic Computing Engine (ACE) », in J. COPELAND (ed.), *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and And Artificial Life plus The Secret of Enigma*, New York, Oxford University Press.

VON NEUMANN J. (1945), « First Draft of a Report on EDVAC », *Contract n° W-670-ORD-4926 Between the United States Army Ordnance Department and the University of Pennsylvania*, Moore School of Electrical Engineering.

WALDROP M. (1987), *Man-made Minds: The Promise of Artificial Intelligence*, New York, Walker.

WEIZENBAUM J. (1976), *Computer and Human Reason*, San Francisco, Freeman.

WIENER N. (1948), *Cybernetics, or control and communication in the animal and the machine*, Cambridge, Cambridge University Press.

WINOGRAD T. (1972), *Understanding Natural Language*, Edinburgh, Edinburgh University Press.

WINOGRAD T., FLORES F. (1986), *Understanding Computers and Cognition: A New Foundation for Design*, Norwood, Ablex Publishing Corporation.