

JOURNAL > SITE 1: LOGIC GATE, THE POLITICS OF THE ARTIFACTUAL MIND | 2017

Machines that Morph Logic: Neural Networks and the Distorted Automation of Intelligence as Statistical Inference

Matteo Pasquinelli

Perceptrons [artificial neural networks] are not intended to serve as detailed copies of any actual nervous system. They are simplified networks, designed to permit the study of lawful relationships between the organization of a nerve net, the organization of its environment, and the “psychological” performances of which the network is capable. Perceptrons might actually correspond to parts of more extended networks in biological systems... More likely, they represent extreme simplifications of the central nervous system, in which some properties are exaggerated, others suppressed.
—Frank Rosenblatt¹

No algorithm exists for the metaphor, nor can a metaphor be produced by means of a computer’s precise instructions, no matter what the volume of organized information to be fed in.
—Umberto Eco²

The term Artificial Intelligence is often cited in popular press as well as in art and philosophy circles as an alchemic talisman whose functioning is rarely explained. The hegemonic paradigm to date (also crucial to the automation of labor) is not based on GOF AI (Good Old-Fashioned Artificial Intelligence that never succeeded at automating *symbolic deduction*), but on the neural networks designed by Frank Rosenblatt back in 1958 to automate *statistical induction*. The text highlights the role of logic gates in the

distributed architecture of neural networks, in which a generalized control loop affects each node of computation to perform pattern recognition. In this distributed and adaptive architecture of logic gates, rather than applying logic to information top-down, *information turns into logic*, that is, a representation of the world becomes a new function in the same world description. This basic formulation is suggested as a more accurate definition of *learning* to challenge the idealistic definition of (artificial) intelligence. If pattern recognition via statistical induction is the most accurate descriptor of what is popularly termed Artificial Intelligence, the distorting effects of statistical induction on collective perception, intelligence and governance (over-fitting, apophenia, algorithmic bias, “deep dreaming,” etc.) are yet to be fully understood.

More in general, this text advances the hypothesis that new machines enrich and destabilize the mathematical and logical categories that helped to design them. Any machine is always a machine of cognition, a product of the human intellect and unruly component of the gears of extended cognition. Thanks to machines, the human intellect crosses new landscapes of logic in a materialistic way—that is, under the influence of historical artifacts rather than Idealism. As, for instance, the thermal engine prompted the science of thermodynamics (rather than the other way around), computing machines can be expected to cast a new light on the philosophy of the mind and logic itself. When Alan Turing came up with the idea of a universal computing machine, he aimed at the simplest machination to calculate all possible functions. The efficiency of the universal computer catalyzed in Turing the alchemic project for the automation of human intelligence. However, it would be a sweet paradox to see the Turing machine that was born as *Gedankenexperiment* to demonstrate the incompleteness of mathematics aspiring to describe an exhaustive paradigm of intelligence (as the Turing test is often understood).

A Unit of Information Is a Logic Unit of Decision

Rather than reiterating GOFAI—that is, the top-down application of logic to information retrieved from the world—this text tries to frame the transmutation of *external information into internal logic* in the machination of neural networks. Within neural networks (as according also to the classical cybernetic framework), information becomes control; that is, a numerical input retrieved from the world turns into a control function of the same world. More philosophically, it means that a representation of the world (information) becomes a new rule in the same world (function), yet under a good degree

of statistical approximation. *Information becoming logic* is a very crude formulation of intelligence, which however aims to stress openness to the world as a continuous process of learning.

The transformation of information into higher functions can probably be detected at different stages in the history of intelligent machines: this text highlights only the early definition of information and feedback loops before analyzing their ramification into neural networks. The metamorphosis of an information loop into higher forms of knowledge of the world was the concern of Second-Order Cybernetics across the 1970s, but it was already exemplified by Rosenblatt's neural networks at the end of the 1950s.³ In order to understand how neural networks transform information into logic, it might be helpful then to deconstruct the traditional reception of both the concepts of information and information feedback. Usually Claude Shannon is castigated for the reduction of information to a mathematical measure according to channel noise.⁴ In the same period, more interestingly, Norbert Wiener defined information as *decision*.

*What is this in information, and how is it measured? One of the simplest, most unitary forms of information is the recording of a choice between two equally probable simple alternatives, one or the other of which is bound to happen—a choice, for example, between heads and tails in the tossing of a coin. We shall call a single choice of this sort a decision.*⁵

If each unit of information is a unit of decision, an atomic doctrine of control is found within information. If information is decision, any bit of information is a little piece of control logic. Bateson will famously add that “information is a difference that makes a difference,” preparing cybernetics for higher orders of organization.⁶ In fact, Second-Order Cybernetics came to break the spell of the negative feedback loop and the obsession of early cybernetics with keeping biological, technical and social systems constantly in equilibrium. A negative feedback loop is defined as an information loop that is gradually adjusted to adapt a system to its environment (regulating its temperature, energy consumption, etc.). A positive feedback loop, on the contrary, is a loop that grows out of control and brings a system far from equilibrium. Second-Order Cybernetics remarked that only far-from-equilibrium systems make possible the generation of new structures, habits and ideas (it was Nobel Prize Ilya Prigogine who showed that forms of self-organization nevertheless occur also in turbulent and chaotic states).⁷ If already in the basic formulation of early cybernetics, the feedback loop could

be understood as a model of information that turns into logic, that morphs logic itself to invent new rules and habits, only Second-Order Cybernetics seems to suggest that it is the excessive 'pressure' of the external world that forces machinic logic to mutate.

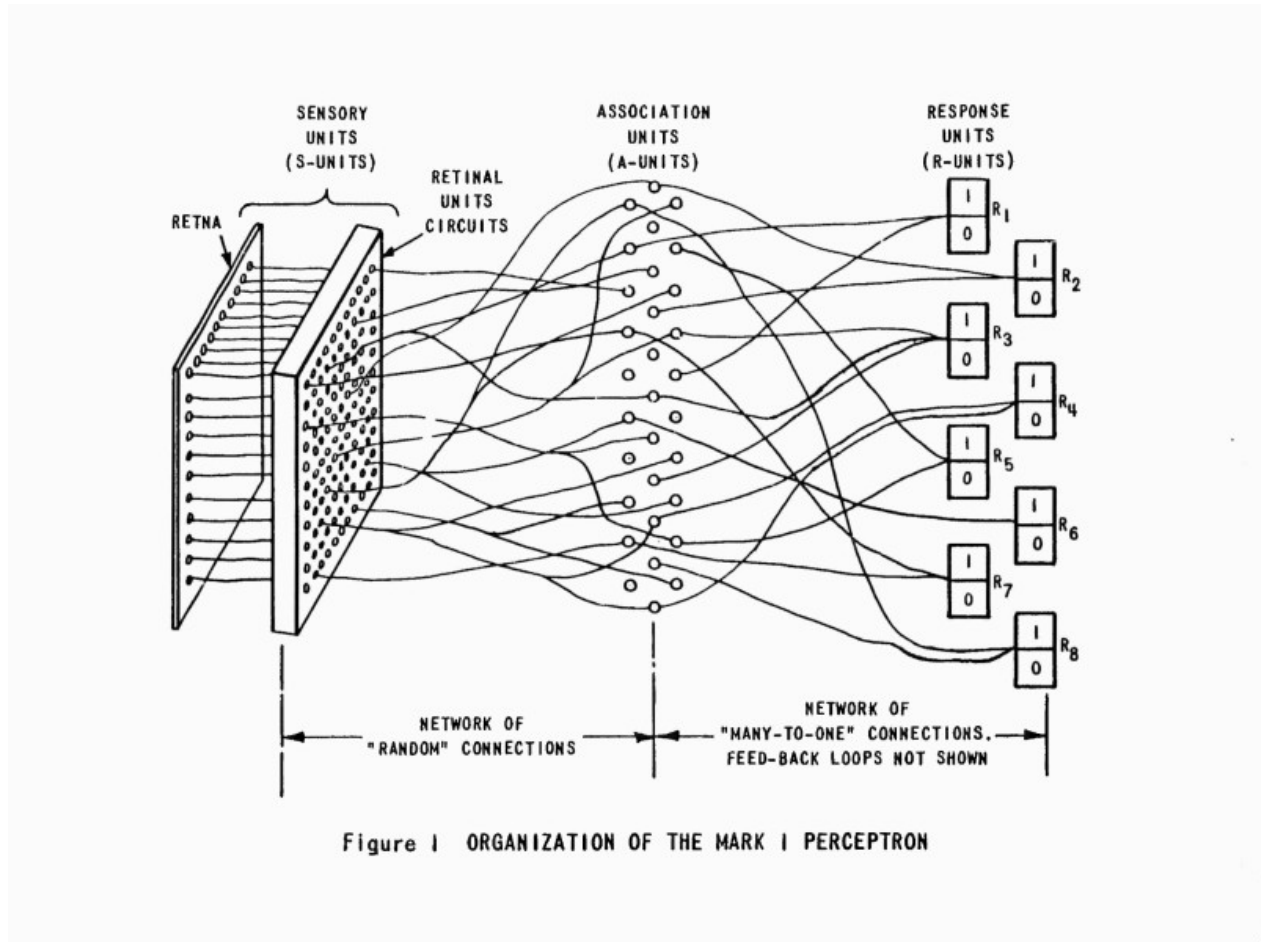


Figure 1 ORGANIZATION OF THE MARK I PERCEPTRON

Diagram of the organisation of the Mark 1 Perceptron. Source with feedback loop not shown. Source: Frank Rosenblatt, Mark I Perceptron Operators' Manual. Buffalo, NY: Cornell Aeronautical Laboratory, 1960.

Frank Rosenblatt and the Invention of the Perceptron

Whereas the evolution of artificial intelligence is made of multiple lineages, this text recalls only the crucial confrontation between two classmates of The Bronx High School of Science, namely Marvin Minsky, founder of the MIT Artificial Intelligence Lab, and Frank Rosenblatt, the inventor of the first operative neural network, the Perceptron. The clash between Minsky and Rosenblatt is often simplified as the dispute between a top-down rule-based paradigm (symbolic AI) and distributed parallel computation (connectionism). Rather than incarnating a fully intelligent algorithm from start, in the

latter model a machine *learns* from the environment and gradually becomes partially ‘intelligent.’ In logic terms, here runs the tension between *symbolic deduction* and *statistical induction*.⁸

In 1951, Minsky developed the first artificial neural network SNARC (a maze solver), but then he abandoned the project convinced that neural networks would require excessive computing power.⁹ In 1957, Rosenblatt described the first successful neural network in a report for Cornell Aeronautical Laboratory titled “The Perceptron: A Perceiving and Recognizing Automaton.” Similar to Minsky, Rosenblatt sketched his neural network, giving a bottom-up and distributed structure to the artificial neuron idea of Warren McCulloch and Walter Pitts that was itself inspired by the eye’s neurons.¹⁰ The first neural machine, the Mark I Perceptron, was born in fact as a vision machine.¹¹

*A primary requirement of such a system is that it must be able to recognize complex patterns of information which are phenomenally similar [...] a process which corresponds to the psychological phenomena of “association” and “stimulus generalization.” The system must recognize the “same” object in different orientations, sizes, colors, or transformations, and against a variety of different backgrounds. [It] should be feasible to construct an electronic or electromechanical system which will learn to recognize similarities or identities between patterns of optical, electrical, or tonal information, in a manner which may be closely analogous to the perceptual processes of a biological brain. The proposed system depends on probabilistic rather than deterministic principles for its operation, and gains its reliability from the properties of statistical measurements obtained from large populations of elements.*¹²

It must be clarified that the Perceptron was not a machine to recognize simple shapes like letters (optical character recognition already existed at the time), but a machine that could *learn* how to recognize shapes by calculating one single statistical file rather than saving multiple ones in its memory. Speculating beyond image recognition, Rosenblatt prophetically added: “Devices of this sort are expected ultimately to be capable of concept formation, language translation, collation of military intelligence, and the solution of problems through inductive Logic.”¹³

In 1961, Rosenblatt published *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanism*, which would influence neural computation until today (the term Multi-Layer Perceptron, for example, is already here). The book moves from psychological and neurological findings on neuroplasticity and applies them to the design of neural networks. The Perceptron was an artifactual model of the brain that was intended to explain some of its mechanisms without being taken for the brain itself. (In

fact, neural networks were conceived by imitating the eye's rather than the brain's neurons, and without knowing how the visual cortex actually elaborates visual inputs). Rosenblatt stressed that artificial neural networks are both a simplification and exaggeration of nervous systems and this approximation (that is the recognition of limits in model-based thinking) should be a guideline for any philosophy of the (artefactual) mind. Ultimately Rosenblatt proposed *neurodynamics* as a discipline against the hype of artificial intelligence.

The perceptron program is not primarily concerned with the invention of devices for "artificial intelligence", but rather with investigating the physical structures and neurodynamic principles which underlie "natural intelligence." A perceptron is first and foremost a brain model, not an invention for pattern recognition. As a brain model, its utility is in enabling us to determine the physical conditions for the emergence of various psychological properties. It is by no means a "complete" model, and we are fully aware of the simplifications which have been made from biological systems; but it is, at least, an analyzable model.¹⁴

In 1969 Marvin Minsky and Seymour Papert's book, titled *Perceptrons*, attacked Rosenblatt's neural network model by wrongly claiming that a Perceptron (although a simple single-layer one) could not learn the XOR function and solve classifications in higher dimensions. This recalcitrant book had a devastating impact, also because of Rosenblatt's premature death in 1971, and blocked funds to neural network research for decades. What is termed as the first 'winter of Artificial Intelligence' would be better described as the 'winter of neural networks,' which lasted until 1986 when the two volumes *Parallel Distributed Processing* clarified that (multilayer) Perceptrons can actually learn complex logic functions.¹⁵ Half a century and many more neurons later, pace Minsky, Papert and the fundamentalists of symbolic AI, multilayer Perceptrons are capable of better-than-human image recognition, and they constitute the core of Deep Learning systems such as automatic translation and self-driving cars.¹⁶

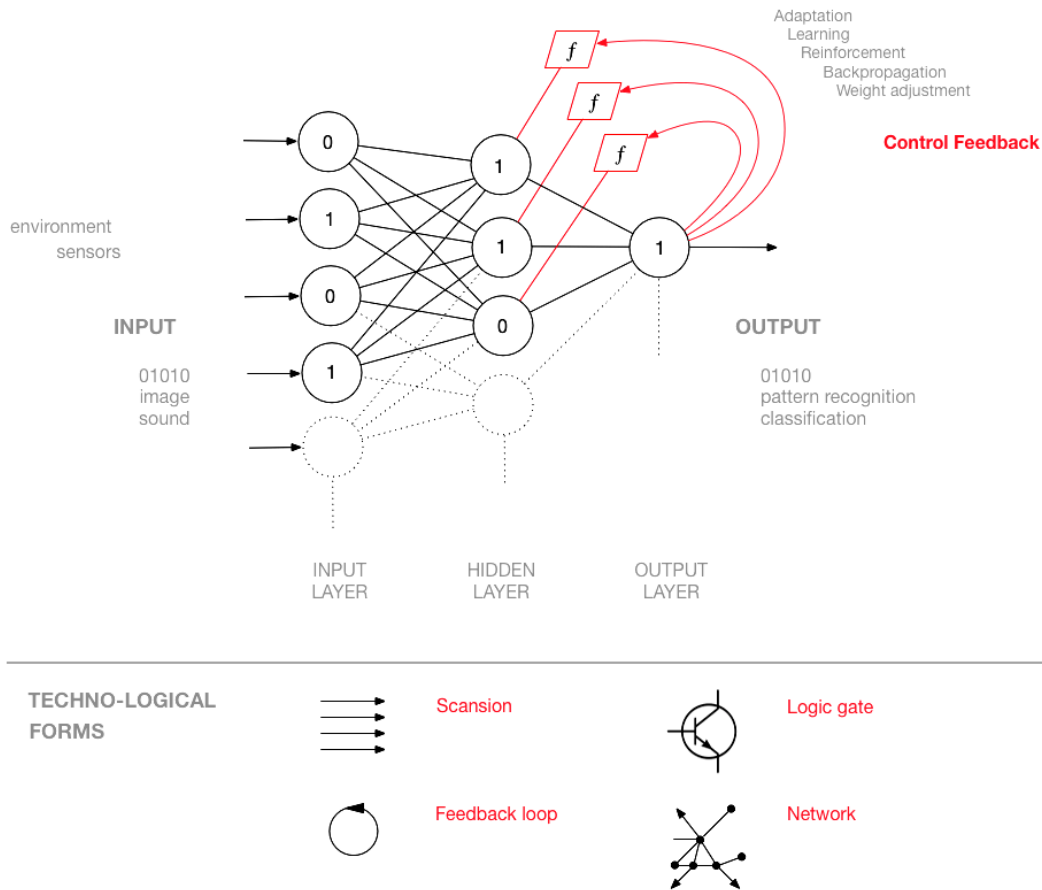


Diagram of a simple neural network showing feedback loops. Matteo Pasquinelli, HfG Karlsruhe. See: www.academia.edu/33205589

Anatomy of a Neural Network

In terms of media archaeology, the neural network invention can be described as the composition of four techno-logical forms: scansion (discretization or digitization of analog inputs), logic gate (that can be realized as potentiometer, valve, transistor, etc.), feedback loop (the basic idea of cybernetics), and network (inspired here by the arrangement of neurons and synapses). Nonetheless, the purpose of a neural network is to calculate a statistico-topological construct that is more complex than the disposition of such forms. The function of a neural network is to record similar input patterns (training dataset) as an *inner state* of its nodes. Once an inner state has been calculated (i.e., the neural network has been ‘trained’ for the recognition of a specific pattern), this statistical construct can be installed in neural networks with identical structure and used to recognize patterns in new data.

The logic gates that are usually part of linear structures of computation acquire, in the parallel computing of neural networks, new properties. In this sense, Rosenblatt gave probably one of the first descriptions of machine intelligence as emergent property: “It is significant that the individual elements, or cells, of a nerve network have never been demonstrated to possess any specifically psychological functions, such as ‘memory,’ ‘awareness,’ or ‘intelligence.’ Such properties, therefore, presumably reside in the organization and functioning of the network as a whole, rather than in its elementary parts.”¹⁷ Yet neural networks are not horizontal but hierarchical (layered) networks. The neural network is composed of three types of neuron layers: input layer, hidden layers (that can be many, from which the term ‘deep learning’), and output layer. Since the first Perceptron (and revealing the influence of the visual paradigm) the input layer is often called retina, even if it does not compute visual data. The neurons of the first layer are connected to the neurons of the next one, following a flow of information in which a complex input is encoded to match a given output. The structure that emerges is not really a network (or a rhizome) but an *arborescent network*, that grows as a *hierarchical cone* in which information is pipelined and distilled into higher forms of abstraction.¹⁸ Each neuron of the network is a transmission node, but also a computational node; it is information gate and logic gate. Each node has then two roles: to transmit information and to apply logic. The neural network ‘learns’ as the wrong output is redirected to adjust the error of each node of computation until the desired output is reached. Neural networks are much more complex than traditional cybernetic systems, since they instantiate a *generalized feedback loop* that affects a multitude of nodes of computation. In this sense, the neural network is the most adaptive architecture of computation designed for machine learning.

The generalized feedback affects the *function* of each node or neuron; that is, the way a node computes (its ‘weight’). The feedback that controls the computation of each node (what is variously termed weight adjustment, error backpropagation, etc.) can be an equation, an algorithm, or even a human operator. In one specific instance of neural network, by modifying a node threshold, the control feedback can change an OR gate into an AND gate, for example—which means that the control feedback changes the way a node ‘thinks.’¹⁹ The logic gates of neural networks compute information in order to affect the way they will compute future information. In this way, *information affects logic*. The business core of the main IT companies today is about finding the most effective formula of the neural control feedback.

More specifically, the neural network learns how to recognize a picture by recording the dependencies or relations between pixels and statistically composing an internal representation. In a photo of an apple, for instance, a red pixel may be surrounded by other red pixels 80% of the time, and so on. In this way also, unusual relations can be combined in more complex graphical features (edges, lines, curves, etc.). Just as an apple has to be recognized from different angles, an actual picture is never memorized, only its statistical dependencies. The statistical graph of dependencies is recorded as a multidimensional internal representation that is then associated to a human-readable output (the word 'apple'). This model of training is called supervised learning, as a human decides if each output is correct. Unsupervised learning is when the neural network has to discover the most common patterns of dependencies in a training dataset without following a previous classification (given a dataset of cat pictures, it will extract the features of a generic cat).

Dependencies and patterns can be traced across the most diverse types of data: visual datasets are the most intuitive to understand but the same procedures are applied, for instance, to social, medical, and economic data. Current techniques of Artificial Intelligence are clearly a sophisticated form of pattern recognition rather than intelligence, if intelligence is understood as the *discovery and invention of new rules*. To be precise in terms of logic, what neural networks calculate is a form of *statistical induction*. Of course, such an extraordinary form of automated inference can be a precious ally for human creativity and science (and it is the closest approximation to what is known as Peirce's weak abduction), but it does not represent per se the automation of intelligence qua invention, precisely as it remains within 'too human' categories.²⁰

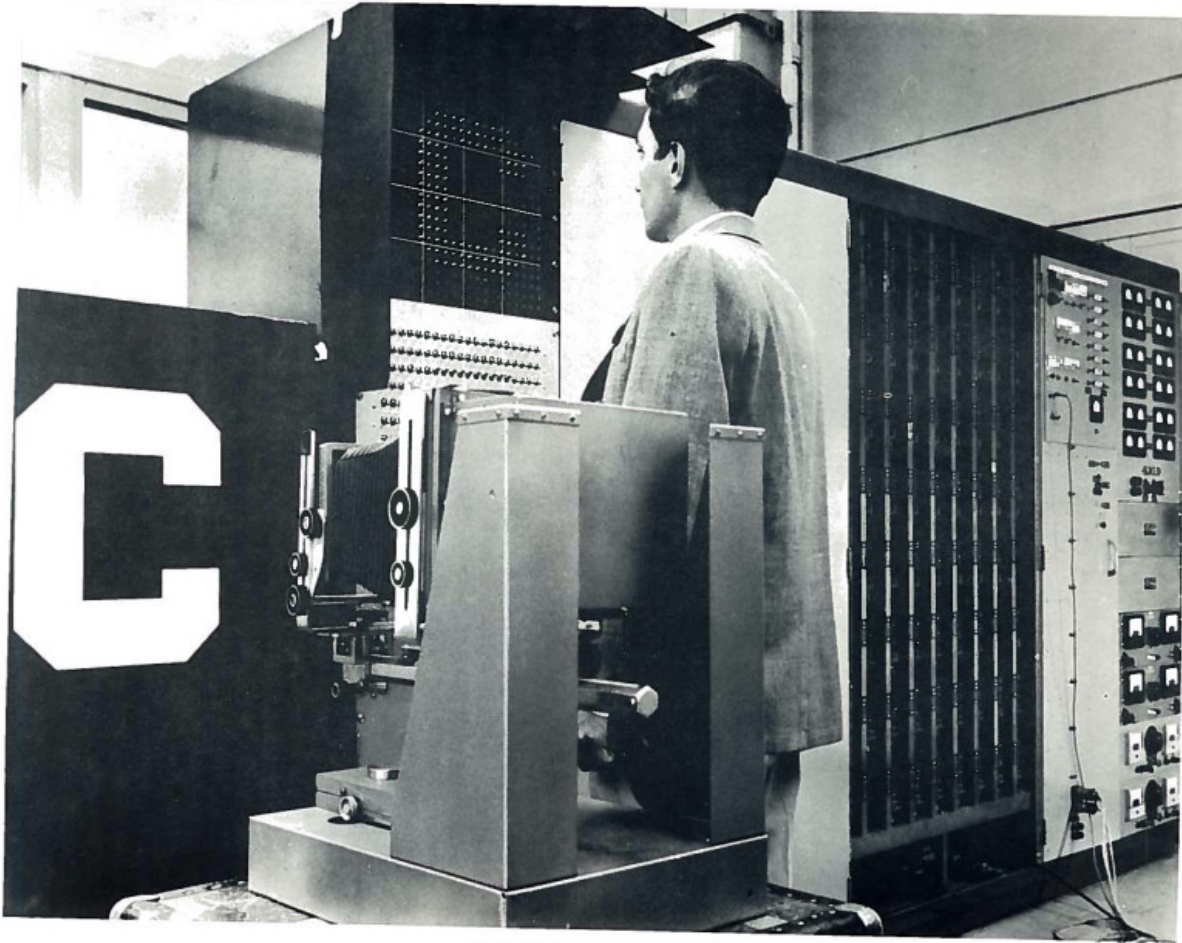
Human, Too Human Computation

Peirce said that "man is an external sign."²¹ If this intuition encouraged philosophers to stress that the human mind is an artifactual project that extends into technology, however, the human mind's actual imbrication with the external machines of cognition happened to be rarely *empirically* illustrated. This has produced simplistic poses in which ideas such as Artificial General Intelligence and Superintelligence are evoked as alchemic talismans of posthumanism with little explanation of the inner workings and postulates of computation. A fascinating aspect of neural computation is actually the way it amplifies the categories of human knowledge rather than supersedes them in autonomous forms. Contrary to the naïve conception of the autonomy of artificial intelligence, in the architecture of neural networks many elements are still deeply

affected by human intervention. If one wants to understand how much neural computation extends into the ‘inhuman,’ one should discern how much it is still ‘too human.’ The role of the human (and also the locus of power) is clearly visible in (1) the design of the training dataset and its categories, (2) the error correction technique and (3) the classification of the desired output. For reasons of space, only the first point is discussed here.

The design of the training dataset is the most critical and vulnerable component of the architecture of neural networks. The neural network is trained to recognize patterns in past data with the hope of extending this capability on future data. But, as has already occurred many times, if training data show a racial, gender and class bias, neural networks will reflect, amplify and distort such a bias. Facial recognition systems that were trained on databases of white people’s faces failed miserably at recognizing black people as humans. This is a problem called ‘over-fitting’: given abundant computing power, a neural network will show the tendency to learn too much, that is to fixate on a super-specific pattern: it is therefore necessary to drop out some of its results to make its recognition impetus more relaxed. Similar to over-fitting can be considered the case of ‘apophenia,’ such as Google DeepDream psychedelic landscapes, in which neural neural networks ‘see’ patterns that are not there or, better, generate patterns against a noisy background. Over-fitting and apophenia are an example of intrinsic limits in neural computation: they show how neural networks can paranoically spiral around embedded patterns rather than helping to reveal new correlations.

The issue of over-fitting points to a more fundamental issue in the constitution of the training dataset: the boundary of the categories within which the neural network operates. The way a training dataset represents a sample of the world marks, at the same time, a closed universe. What is the relation of such a closed data universe with the outside? A neural network is considered ‘trained’ when it is able to generalize its results to unknown data with a very low margin of error, yet such a generalization is possible due to the homogeneity between training and test dataset. A neural network is never asked to perform across categories that do not belong to its ‘education.’ The question is then: How much is a neural network (and AI in general) capable of escaping the categorical ontology in which it operates?²²



THE MARK I PERCEPTON

Mark 1 Perceptron. Source: Rosenblatt, Frank (1961) Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Buffalo, NY: Cornell Aeronautical Laboratory.

Abduction of the Unknown

Usually a neural network calculates statistical induction out of a homogenous dataset; that is, it extrapolates patterns that are consistent with the dataset nature (a visual pattern out of visual data, for example), but if the dataset is not homogenous and contains multidimensional features (for a very basic example, social data describing age, gender, income, education, health conditions of the population, etc.), neural networks can discover patterns among data that human cognition does not tend to correlate. Even if neural networks show correlations unforeseen to the human mind, they operate within the implicit grid of (human) postulates and categories that are in the training dataset and, in this sense, they cannot make the necessary leap for the invention of radically new categories.

Charles S. Peirce's distinction between deduction, induction and abduction (hypothesis) is the best way to frame the limits and potentialities of machine intelligence. Peirce remarkably noticed that the classic logical forms of inference—deduction and induction—never invent new ideas but just repeat quantitative facts. Only abduction (hypothesis) is capable of breaking into new worldviews and inventing new rules.

*The only thing that induction accomplishes is to determine the value of a quantity. It sets out with a theory and it measures the degree of concordance of that theory with fact. It never can originate any idea whatever. No more can deduction. All the ideas of science come to it by the way of Abduction. Abduction consists in studying facts and devising a theory to explain them.*²³

Specifically, Peirce's distinction between abduction and induction can illuminate the logic form of neural networks, as since their invention by Rosemblaat they were designed to automate complex forms of induction.

*By induction, we conclude that facts, similar to observed facts, are true in cases not examined. By hypothesis, we conclude the existence of a fact quite different from anything observed, from which, according to known laws, something observed would necessary result. The former, is reasoning from particulars to the general law; the latter, from effect to cause. The former classifies, the latter explains.*²⁴

The distinction between induction as classifier and abduction as explainer frames very well also the nature of the results of neural networks (and the core problem of Artificial Intelligence). The complex statistical induction that is performed by neural networks gets close to a form of weak abduction, where new categories and ideas loom on the horizon, but it appears invention and creativity are far from being fully automated. The invention of new rules (an acceptable definition of intelligence) is not just a matter of generalization of a specific rule (as in the case of induction and weak abduction) but of breaking through semiotic planes that were not connected or conceivable beforehand, as in scientific discoveries or the creation of metaphors (strong abduction).

In his critique of artificial intelligence, Umberto Eco remarked: "No algorithm exists for the metaphor, nor can a metaphor be produced by means of a computer's precise instructions, no matter what the volume of organized information to be fed in."²⁵ Eco stressed that algorithms are not able to escape the straitjacket of the categories that are implicitly or explicitly embodied by the "organized information" of the dataset. Inventing a new metaphor is about making a leap and connecting categories that never happened to be logically related. Breaking a linguistic rule is the invention of a new rule, only when

it encompasses the creation of a more complex order in which the old rule appears as a simplified and primitive case. Neural networks can *a posteriori* compute metaphors²⁶ but cannot *a priori* automate the invention of new metaphors (without falling into comic results such as random text generation). The automation of (strong) abduction remains the philosopher's stone of Artificial Intelligence.

(Quasi) Explainable Artificial Intelligence

The current debate on Artificial Intelligence is basically still elaborating the epistemic traumas provoked by the rise of neural computation. It is claimed that machine intelligence opens up new perspectives of knowledge that have to be recognized as posthuman patrimony (see Lyotard's notion of the inhuman), but there is little attention to the symbolic forms of pattern recognition, statistical inference, and weak abduction that constitute such a posthuman shift. Besides, it is claimed that such new scales of computation constitute a black box that is beyond human (and political) control, without realizing that the architecture of such black box can be reverse-engineered. The following passages remark that the human can still break into the 'inhuman' abyss of deep computation and that human influence is still recognizable in a good part of the 'inhuman' results of computation.

It is true that layers and layers of artificial neurons intricate so much computation that it is hard to *look back* into such structure and find out where and how a specific 'decision' was computed. Artificial neural networks are regarded as *black boxes* because they have little to no ability to explain causation, or which features are important in generating an inference like classification. The programmer has often no control over which features are extracted, as they are deduced by the neural network on its own.²⁷

The problem is once again clearly perceived by the military. DARPA (the research agency of the US Defense) is studying a solution to the black box effect under the program Explainable Artificial Intelligence (XAI).²⁸ The scenario to address is, for example, a self-driving tank that turns into an unusual direction, or the unexpected detection of enemy weapons out of a neutral landscape. The idea of XAI is that neural networks have to provide not just an unambiguous output but also a rationale (part of the computational context for that output). If, for example, the figure of an enemy is recognized ("this image is a soldier with a gun"), the system will say why it thinks so—that is, according to which features. Similar systems can be applied also to email monitoring to spot potential terrorists, traitors, and double agents. The system will try not just to detect anomalies of behavior against a normal social pattern, but also to give

explanations of which context of elements describe a person as suspect. As the automation of anomaly detection has already bred its casualties (see the Skynet affair in Pakistan),²⁹ it is clear that XAI is supposed to preempt also further algorithmic disasters in the context of predictive policing.

Explainable Artificial Intelligence (to be termed more correctly, Explainable Deep Learning) adds a further control loop on top of the architecture of neural networks, and it is preparing a new generation of epistemic mediators. This is already part of a multi-billion business interest as insurance companies, for instance, will cover only those self-driving cars that will provide “computational black box” featuring not just video and audio recordings but also the rationale for their driving decisions (imagine the case of the first accident between two self-driving vehicles). Inhuman scales of computation and the new dark age aesthetics have already found their legal representatives.

Conclusion

In order to understand the historical impact of Artificial Intelligence, this text stresses that its hegemonic and dominant paradigm to date is not symbolic (GOFAI) but connectionist, namely the neural networks that constitute also Deep Learning systems. What mainstream media call Artificial Intelligence is a folkloristic way to refer to neural networks for pattern recognition (a specific task within the broader definition of intelligence and, for sure, not an exhaustive one). Pattern recognition is possible thanks to the calculus of the inner state of a neural network that embodies the logical form of statistical induction. The ‘intelligence’ of neural networks is, therefore, just a statistical inference of the correlations of a training dataset. The intrinsic limits of statistical induction are found in between over-fitting and apophenia, whose effects are gradually emerging in collective perception and governance. The extrinsic limits of statistical induction can be illustrated thanks to Peirce’s distinction of induction, deduction, and abduction (hypothesis). It is suggested that statistical induction gets closer to forms of weak abduction (e.g., medical diagnosis), but it is unable to automate strong abduction, as it happens in the discovery of scientific laws or the invention of linguistic metaphors. This is because neural networks cannot escape the boundary of the categories that are implicitly embedded in the training dataset. Neural networks display a relative degree of autonomy in their computation: they are still directed by human factors and they are components in a system of human power. For sure, they do not show signs of

‘autonomous intelligence’ or consciousness. Super-human scales of knowledge are acquired only in collaboration with the human observer, suggesting that Augmented Intelligence would be a more precise term than Artificial Intelligence.

Statistical inference via neural networks has enabled computational capitalism to imitate and automate both low and hi-skill labor.³⁰ Nobody expected that even a bus driver could become a source of cognitive labor to be automated by neural networks in self-driving vehicles. Automation of intelligence via statistical inference is the new eye that capital casts on the data ocean of global labor, logistics, and markets with novel effects of *abnormalization*—that is, distortion of collective perception and social representations, as it happens in the algorithmic magnification of class, race and gender bias.³¹ Statistical inference is the distorted, new eye of the capital’s Master.³²

The author wishes to thank Anil Bawa-Cavia, Nina Franz, and Nikos Patelis for their comments.

Footnotes

1. Frank Rosenblatt. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Buffalo, NY: Cornell Aeronautical Laboratory, 1961. 28. Print.
2. Umberto Eco. *Semiotics and the Philosophy of Language*. Bloomington: Indiana University Press, 1986. 127. Print.
3. See Francis Heylighen and Cliff Joslyn. “Cybernetics and Second-Order Cybernetics” in *Encyclopedia of Physical Science and Technology* 19. Ed. R.A. Meyers. New York: Academic Press, 2001. Print.
4. Claude Shannon. “A Mathematical Theory of Communication.” *Bell System Technical Journal* 27/3 (1948). Print.
5. Norbert Wiener. *Cybernetics: Or Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press, 1948. 61. Print. Wiener’s formulation happened to influence also Jacques Lacan’s 1955 lecture on cybernetics and psychoanalysis, in which logic gates are literally understood as “doors” that open onto or close off new destinies within the Symbolic order. Jacques Lacan. “Psychoanalysis and cybernetics, or on the nature of language.” *The Seminar of Jacques Lacan* 2. New York: Norton, 1988. Print.
6. Gregory Bateson. *Steps to an Ecology of Mind*. Chicago: University of Chicago Press, 1972. Print.
7. Gregoire Nicolis and Ilya Prigogine. *Self-Organization in Nonequilibrium Systems*. New York: Wiley, 1977. Print.
8. Artificial General Intelligence (AGI) is often the attempt to meet top-down (symbolic) with bottom-up (connectionist) approaches halfway, that is to combine symbolic deduction with statistical induction. To date, however, only the connectionist paradigm of neural networks happened to be successfully automated, casting doubts on some metaphysical and centralizing premises of AGI.

9. See Marvin Minsky. *Theory of Neural-Analog Reinforcement Systems and Its Application to the Brain Model Problem*. Dissertation. Princeton University, 1954. Print.
10. Warren McCulloch and Walter Pitts. "A Logical Calculus of the Ideas Immanent in Nervous Activity." *Bulletin of Mathematical Biophysics* 5/4 (1943). Print. And Warren McCulloch and Walter Pitts. "How We Know Universals the Perception of Auditory and Visual Forms." *Bulletin of Mathematical Biophysics* 9/3 (1947). Print.
11. Remarkably Paul Virilio's book on machine vision was inspired also by the Perceptron (yet Virilio could not foresee that the Perceptron would become the hegemonic paradigm of machine intelligence in the early 21st century). See Ch. 5 in Paul Virilio. *La Machine de vision: essai sur les nouvelles techniques de représentation*. Paris: Galilée, 1988. Print. English translation: *The Vision Machine*. Bloomington: Indiana University Press, 1994. Print.
12. Frank Rosenblatt. "The Perceptron a Perceiving and Recognizing Automaton." *Technical Report* 85/460/1 (1957). 1-2. Print.
13. *Ibid.* 30.
14. Frank Rosenblatt. *Op. cit.* 1961. vii. Print.
15. David Rumelhart and PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* 1-2. Cambridge, MA: MIT Press, 1986. Print.
16. Neural networks keep on growing towards even more complex topologies and have inaugurated a true computational ars combinatoria (see the diagrams of auto-encoder, Boltzmann machine, recurrent and Long Short-Term Memory neural network, Adversarial Generative Networks, etc.). Neural networks are the most articulated and sophisticated machines along the tradition of computable knowledge, as in the ancient Arab device *zairja* or Ramon Llull's book *Ars Magna* (1305). See David Link. "Scrambling T-R-U-T-H: Rotating Letters as a Material Form of Thought." *Variantology* 4. *On Deep Time Relations of Arts, Sciences and Technologies in the Arabic-Islamic World*. Eds. Siegfried Zielinski and Eckhard Fülus. Cologne: König, 2010. Print.
17. Frank Rosenblatt. *Op. cit.* 1961. 9.
18. See Ethem Alpaydin. *Introduction to Machine Learning*. 2nd ed. Cambridge, MA: MIT Press, 2014. 260. Print.
19. This is one specific case for illustration purposes. Activation functions operate also in different ways.
20. On the attempts to automate weak abduction, see "Automatic Abductive Scientists" in: Lorenzo Magnani. *Abductive Cognition*. Springer Science & Business Media, 2009. 112. Print.
21. Charles S. Peirce. "Some Consequences of Four Incapacities" (1868). *The Essential Peirce* 1 (1867-1893). Eds. Nathan Houser and Christian Kloesel. Bloomington: Indiana University Press, 1992. 54. Print.
22. Ontology is here used in the sense of information science.
23. Charles S. Peirce. *Collected Papers*. Cambridge, MA: Belknap, 1965. 5, 145. Print.
24. Charles S. Peirce. "Deduction, Induction, and Hypothesis" (1878). *Op. cit.* 1992. 194.
25. Umberto Eco. *Op. cit.* 127.

26. See *Word2vec*, a framework for the mapping of word embedding in vectorial space.
27. This is true both in supervised and unsupervised learning. Thanks to Anil Bawa-Cavia for clarifying it.
28. See www.darpa.mil/program/explainable-artificial-intelligence
29. Matteo Pasquinelli. "Arcana Mathematica Imperii: The Evolution of Western Computational Norms." *Former West*. Eds. Maria Hlavajova, et al. Cambridge, MA: MIT Press, 2017. Print.
30. There are different approaches to machine intelligence, yet the hegemony of connectionism in automation is manifest. For an accessible introduction to the different families of machine learning, see Pedro Domingos. *The Master Algorithm*. New York: Basic Books, 2015. Print.
31. Matteo Pasquinelli. "Abnormal Encephalization in the Age of Machine Learning." *e-flux* 75 (September 2016). Web.
32. "The special skill of each individual machine-operator, who has now been deprived of all significance, vanishes as an infinitesimal quantity in the face of the science, the gigantic natural forces, and the mass of the social labor embodied in the system of machinery, which, together with these three forces, constitutes the power of the master". Karl Marx. *Capital 1* (1867). London: Penguin, 1982. 549. Print.

Matteo Pasquinelli is Professor in Media Theory at the University of Arts and Design, Karlsruhe.