Vivien García teaches philosophy at the Faculty of Medicine of the Université de Lorraine. His research lies at the intersection of technology, ethics, and politics. His publications include *Que faire de l'intelligence artificielle ? Petite histoire critique de la raison artificielle* (Rivages, 2024) and *Le sort de la philosophie : Quatre parcours de jeunes hégéliens : Bakounine, Engels, Marx, Stirner* (Hermann, 2018).
Contact: vivien.garcia@univ.lorraine.fr

Soft Power
Revista euro-americana de teoría
e historia de la política y del derecho

Vol. 11(2). Julio-Diciembre 2024
ISSN (online): 2539/2239 ISSN (print): 2389-8232
https://doi.org/10.14718/SoftPower.2024.11.2.2

ARTIFICIAL INTELLIGENCE: FROM SEMINAL SCIENTIFIC HYPOTHESES TO SOCIO-TECHNICAL REALISATION*

Vivien García
Université de Lorraine

INTELIGENCIA ARTIFICIAL: DE LAS HIPÓTESIS CIENTÍFICAS FUNDAMENTALES A LA REALIZACIÓN SOCIOTÉCNICA

Abstract

This paper, excerpted from *Que faire de l'intelligence artificielle ? Petite histoire critique de la raison artificielle*, offers an account and genealogy of artificial intelligence (AI)'s technical foundations and philosophical implications from its mid-20th-century origins to the present. The narrative underscores how early scientific hypotheses continue to influence current AI paradigms. Part one uncovers foundational hypotheses—predating the 1956 Dartmouth workshop—that informed both symbolic AI (prevalent until the late 1980s) and the surge of connectionist methods in recent decades, demonstrating their shared premise of mind and brain as logical machines. Part two examines

* Reception date: 24th March 2024; acceptance date: 27th March 2024.

the sociotechnical labor of AI development, using the ImageNet database as a case study to reveal the hidden work of crowdsourcing in training modern models. By juxtaposing these theoretical foundations with practical industry practices, the paper highlights enduring continuities and transformations in AI research and invites reflection on the technical and social processes shaping the AI industry today.

Keywords

artificial intelligence; mind; logical machines; machine learning

Resumen

El presente artículo, extraído del libro *Que faire de l'intelligence artificielle ? Petite histoire critique de la raison artificielle*, ofrece una exposición didáctica y una genealogía crítica de la inteligencia artificial (IA) desde mediados del siglo XX hasta la actualidad. En la primera parte, se desvelan las hipótesis científicas fundamentales que, antes de la conferencia de Dartmouth de 1956, sustentaron la evolución de la IA simbólica (dominante hasta finales de los 80) y la IA conexionista (impulsada en las últimas dos décadas), mientras se muestra su convergencia al concebir mente y cerebro como máquinas lógicas. En la segunda parte, se analizan aspectos sociotécnicos recientes con base en el caso de la base de datos ImageNet para ilustrar el trabajo invisible del *crowdsourcing* en el entrenamiento de modelos. Al conjugar fundamentos teóricos y prácticas industriales, el texto subraya las continuidades y transformaciones que definen la industria de la IA hoy.

Palabras clave

inteligencia artificial; mente; máquinas lógicas; aprendizaje automático

The Brain and the Mind as Machines

The artificial intelligence (AI) project[1] aims to create machines capable of performing all sorts of tasks that, at first sight, only humans seem capable of: solving problems, identifying objects, using natural languages, playing games…. In this sense, we could say that it aims to imbue machines with a human dimension. However, it is also possible to consider that, by assimilating all or part of human beings to machines, AI ultimately amounts to creating other thinking machines. Its theoretical inspirations lead more clearly towards this second perspective.

It is first necessary to agree on the kind of machines in question. Warren McCulloch (who, in collaboration with Walter Pitts, provided a mathematical model of neural activity that inspired the artificial neural network idea on which connectionist AI is based) offers some explanations in this regard: "Everything we learn of organisms leads us to conclude not merely that they are analogous to machines but that they are machines. Man-made machines are not brains, but brains are a very ill-understood variety of computing machines" (McCulloch, 1965, p. 159).

As might be expected, the brain, the starting point of connectionism, is at the heart of the discussion; the first purpose of the formal neural network was to explain how it works. It might be more surprising to see it described as a machine. Nevertheless, McCulloch does not argue that the brain is a technical object. He does not use the word "machine" in its everyday language meaning. According to him, the machine is the model, the logical functioning abstractly described (Dupuy, 2009). This is why the brain is regarded as a "variety of computing machine" and why, in turn, it becomes possible to consider implementing the same functioning in artefacts.

What, then, of the alternative approach to AI, the symbolic approach, which takes the mind as its reference point rather than the brain? It ultimately reaches analogous conclusions, albeit via more tortuous paths. As posited by Allen Newell and Herbert Simon (1976), two of the most prominent figures in this line of reasoning, computer science has explained "at a rather basic level what symbols are" and thus led to a "scientific proposition about Nature" (p. 114). This proposition assumes that symbols are the fundamental units of intelligence. They can be considered to play a role analogous

[1] This section follows an exposition of the two major approaches in the AI field: the symbolic approach and the connectionist approach. It also follows the presentation of the foundational scientific workshop at Dartmouth College in 1956 and its stated objectives. It is within this context that the project guiding the development of AI systems is described here. It should be noted that AI has been driven by other desires, including, for example, the goal of developing general AI systems that would not be limited to a single task. However, even when such a goal is pursued, it is generally with the same epistemic foundations.

to cells in biological systems, tectonic plates in geology, or atoms in physics. This is the so-called hypothesis of "physical symbol systems."

Two exemplars of "physical symbol systems" are the human mind and the computer. We will not dwell on their "physical" dimension, that is to say, the body or the brain, in the first case, and the material machine, namely the hardware, in the second. This kind of *res extensa*, or "extended thing," as Descartes might have described it,[2] merely enables these systems to exist "in a world of objects wider than just these symbolic expressions themselves" (Newell & Simon, 1976, p. 116). A system of symbols, in other words, is not a self-contained entity; rather, its materiality anchors it in the world. However, intelligence lies elsewhere: the ability to store and manipulate the symbols according to logical rules. As apparently evidenced by a simple mathematical calculation, it can function in a relative abstraction from the world.

It is a commonly held view amongst (philosophical) commentaries on AI that this stance represents a continuation of a particular intellectual tradition within the West (Dreyfus & Dreyfus, 1988; Haugeland, 1989). It recalls the words of Hobbes, who, in the 17th century, presented reason as a "power" or "faculty" of computation, that is to say, logic.[3] It also recalls the endeavours of Gottfried Wilhelm Leibniz, who, contemporaneously, sought to construct a formal language that would enable the logical resolution of any controversy and could be used by machines.[4] Finally, it is directly inspired by the contributions of Gottlob Frege and, subsequently, Bertrand Russell to contemporary logic at the turn of the twentieth century. It is noteworthy that the first AI computer program, the *Logic Theorist*, aimed to demonstrate theorems in symbolic logic from Bertrand Russell and Alfred North Whitehead's *Principia Mathematica.*

The hypothesis of physical symbol systems thus establishes a fundamental link between intelligence and logical formalism: the latter expresses the rules governing the manipulation of symbols. It should be stressed that these rules do not derive directly from human cognition: the computer is a logical machine in the same way as the human mind, but it is not in the image of the latter. The two are functionally equivalent when, for example, they apply the same functions or algorithms: a calculation, a problem-solving procedure, and so on.

---

[2] The hypothesis of "physical symbol systems" should not be confused with substance dualism, as seen in Descartes' philosophy. Rather, the aforementioned hypothesis is based on a functionalist perspective, which means it is "officially neutral between materialism and dualism" (Levin, 2023). The parallel is intended to highlight the fact that for the hypothesis of "physical symbol systems," symbol systems exist in a relative abstraction of their physical dimension.
[3] The first chapter of Hobbes's De Corpore is called "*Computatio sive logica* [Computation or Logic]."
[4] Leibniz calls this language *characteristica universalis.* For a comparative study of how Hobbes and Leibniz link thinking and computation, see Triclot (2005).

From a philosophical perspective, however, the hypothesis of physical symbol systems entails a heavier and more ambiguous assumption. Claiming the "empirical character" of their hypothesis, Newell and Simon (1976) assert that it must be verified by empirical enquiry, and to this end, they propose different paths. The observation of human symbolic behaviour (i.e., our ways of reasoning) is one; the creation of computer programs that perform tasks deemed "intelligent" (i.e., the construction of other symbol systems) is another. The authors contend that these parallel investigations will strengthen the functional equivalence between the mind and the computer.

Nevertheless, the testing of a hypothesis through experience does not suffice to render it empirical. In philosophical parlance, Simon and Newell's hypothesis can be considered transcendental; it implies a priori conditions of intelligence (Proust, 1987). Supposing that all intelligent systems share "the ability to store and manipulate symbols" (Newell & Simon, 1976, p. 115), it ultimately refers to another logical machine: the Turing machine.

In a seminal article, the British mathematician Alan Turing (1937) introduced this purely imaginary device. His description is particularly unsettling in that it portrays a purely ideal machine, a model of computation, while evoking tangible technical objects. One should imagine it furnished with a potentially infinite "tape," divided into "squares," each bearing a symbol. To facilitate comprehension, the metaphor is often expanded. The Turing machine is commonly said to have a "head" (capable of moving along the tape and reading, writing, and modifying it), a state register, and a table of possible actions or instructions. Putting metaphors aside, we find here the conditions of intelligence postulated by the physical symbol systems hypothesis: the ability to store and manipulate symbols.

What is the purpose of the Turing machine? To solve any computational problem, it is necessary to define a non-ambiguous series of steps to be strictly followed.[5] A minimal example of such a procedure could be: If a square contains an "a," replace it with a "b," then move on to the next square and repeat the operation. In other words, the Turing machine is merely a formal model of the notion of algorithm. There is a tendency to subsume the functioning of computers under this model; the hypothesis of physical systems of symbols also subsumes the functioning of the mind under it.

While connectionist AI views the brain as a machine (a network of formal neurons), symbolic AI views the mind as a machine (a Turing machine). Despite their

[5] Within the limits of calculability as stated by Church-Turing's thesis, that is, the class of recursive functions, which is equivalent to the class of lambda-definable functions.

initial divergence, these approaches ultimately converge. According to Warren McCulloch (1965), neural networks are, in the final analysis, equivalent to Turing machines. Anything one can do, the others could do too: "any Turing machine could be made out of neurons" (p. 159).

In light of the foundational insights that have shaped the field of AI, whether they focus on the brain or the mind, the human being is a machine.

Serving AI: Data Capture and Work

Microblogging, collaborative encyclopaedias, photo and video sharing, sport performance tracking… The advent of platforms and social media fed by user-generated content, also known as participatory web, has provided AI[6] with resources it barely dreamed of. The phenomenon has reached the scale we have come to know as a result of the growth of ubiquitous computing, which includes smartphones, tablets, and "personal digital assistants," as well as smart TVs, on-board computers, smart objects, and other devices that are constantly creating data of all kinds.

The idea of creating large databases from such a fountain of data rapidly emerged. ImageNet is a canonical example, as it is widely claimed to have paved the way for the current triumph of neural networks. Initiated in 2006 by Fei-Fei Li at Princeton University, the project consists of the accumulation of annotated images. In addition to providing an unprecedented resource for training neural networks, it also offers the opportunity for an annual computer vision competition from 2010 to 2017. The objective is to recognise as many categories of objects as possible in a large number of images. In 2012, the competition was won outright by AlexNet, a neural network trained on ImageNet and developed under the supervision of Geoffrey Hinton. Its error rate is almost 11 %, lower than that of the methods still used by the competition. The "revenge of the neurons" (Cardon et al., 2018) has just begun.

The creation of ImageNet was made possible by another database, WordNet, which was also created at Princeton University more than twenty years earlier. It brings together *synsets*, sets of English words grouped according to a common meaning.[7] Ironically, this type of semantic network was highly prized by symbolic AI expert systems at the time of its creation. In the context of ImageNet, synsets were used to launch automated

---

[6] In contrast to the previous part, this paragraph focuses more on contemporary AI systems based on neural networks, that is, connectionist systems.

[7] For example: "book, volume, product (physical objects consisting of a number of pages bound together)" (Princeton University, n.d.).

queries to various search engines in order to retrieve links to images that were supposed to correspond to them. This is known as web scraping and involves extracting data from the web, not by searching for it individually, but by programming a script to do it for us. The match between images and words was then checked by humans.

To be more precise, the creation of the ImageNet database has relied on crowd-sourcing (a term derived from the words "crowd" and "outsourcing") in at least two ways. First, the aforementioned database contains links (ImageNet, 2021) to images (for which copyright has not been acquired) that were not created with the specific purpose of being included in the database. A considerable proportion of this content has been sourced from photo-sharing platforms such as Flickr. Second, the tagging and classification verification have been directly outsourced through Mechanical Turk (MTurk), which is Amazon's micro-labour marketplace.

The reference contained in the name of this service is clear. The Mechanical Turk was an 18th-century automaton designed to appear to play chess. Its construction and the placement of mirrors enabled it to conceal a chess master. Giving a new twist to this story, Amazon offers assistance to an "artificial artificial intelligence" (Pontin, 2007) that no longer needs to lock anyone in a chest. Its platform connects "requesters," who offer menial tasks in large numbers (data validation, content description, online survey, etc.), with "turkers," who receive a micro-payment for each completed task. Thanks to the Internet, it mobilises a plethora of hands whose gestures crystallise in the training of AI systems. They produce or prepare the data, sort it, correct potential labels, etc.

Framing effects are used to enhance efficiency. The resulting interface prompts turkers to validate, reject, check, correct, or answer questions using the minimum number of clicks and keystrokes. In the case of the ImageNet database, the mean rate was approximately one image per second per turker (Markoff, 2012). It is therefore not productive to assume that a question has been poorly formulated or that a label is ambiguous. For the time spent on the platform to be profitable, interactions with it need to be as frictionless as possible.

MTurk reveals a new mirror here: the conditioning of automatic machines entails a conditioned automation of human interaction with the machine. Unlike the chess master of the old Mechanical Turk, whose exceptional play was the key to the show, the turkers perform tasks considered human because they generally involve an extremely low degree of specialisation.[8]

[8] Amazon calls them *Human Intelligence Tasks* (HIT).

The magic of automation offered by AI systems trained in this way operates within social conditions. ImageNet provides an illustrative example of two practices that can be subjected to such an analysis. One is poorly remunerated, the other not at all, but both take place on the margins of traditional production sites. They are as essential to the system as they are inconspicuous. Computer vision relies on invisible work.

The ImageNet database is not an isolated case. The methodology employed in its construction has inspired the creation of numerous other databases. Examples include MusicNet (which collates classical music recordings with labels indicating the time of each note, the instrument that plays it, etc.) and ActivityNet (which matches videos with the human activities they represent).

Beyond Amazon Mechanical Turk, other systems have also been developed to delegate the task of labelling. Everyone has, at least once, had to pass the reverse Turing test of Google's reCAPTCHA. It is now famous that "I am not a robot" means "I am capable of doing what machines are not yet capable of doing properly." By writing the characters or words on the screen or clicking on images of cats, cars, traffic lights, and so on, users contribute to training AI systems.[9]

The identification and description of such forms of invisible work have become a significant theoretical issue and have raised claims. This has resulted in the emergence of numerous concepts, including "digital labour" (in contrast to the digital work of individuals employed in the IT sector), "ghost work," "fauxtomation," and "heteromation." (Casilli, 2019; Ekbia & Bonnie, 2017; Gray & Suri, 2019; Taylor, 2018) None of these concepts align with the mantra of the end of work through automation. That said, automation does not leave the nature of work, its conduct, or organisation untouched. It may rather contribute to a steady casualisation of labour.

The historian Lewis Mumford, a few decades earlier, had already highlighted the connection between certain social and technical forms. He showed how, over and above tools and machines, the aggregation of masses of people and the hierarchical organisation of their work had enabled, among other things, the construction of the pyramids

[9] It would be inaccurate to assert that Google reCAPTCHA is wholly analogous to Amazon Mechanical Turk. Unlike Amazon, Google does not act as an intermediary for individuals requiring labelled data; instead, it acts on its own behalf. Moreover, it is not based on an economic incentive model; rather, it blocks access to a website or service if the test is not deemed conclusive. The primary objective of the system, which is offered free of charge to webmasters to a certain extent, is to prevent massive requests from malicious robots or crawlers, potentially including those attempting to retrieve data for the purpose of training neural networks. It is noteworthy that in its latest versions, reCAPTCHA minimised the appearance of the tests that made it famous by conducting a risk analysis that relies heavily on data supplied by the web browser about hardware, software, and browser-based user behaviour. Those who protect their online privacy are more likely to fall into the non-human category.

several thousand years ago. Mumford (1970) described this phenomenon as a "megamachine," an "invisible machine," and predicted that its modern analogue would "escape spatial and temporal limitations… with its functioning parts operating as a whole through instant communication" (p. 258).

References

Cardon, D., Cointet, J.-P., & Mazières, A. (2018). La Revanche des neurones. *Réseaux, 5*(211), 173–220. https://doi.org/10.3917/res.211.0173.

Casilli, A. A. (2019). *En attendant les robots : Enquête sur le travail du clic.* Seuil.

Dreyfus, H. L., & Dreyfus, S. E. (1988). Making a Mind versus Modeling the Brain: Artificial Intelligence Back at a Branchpoint. *Daedalus, 117*(1), 15–43.

Dupuy, J.-P. (2009). *On the Origins of Cognitive Science: The Mechanization of the Mind* (M. B. DeBevoise, Trans.). MIT Press.

Ekbia, H. R., & Bonnie, N. (2017). *Heteromation, and Other Stories of Computing and Capitalism.* The MIT Press. https://doi.org/10.7551/mitpress/10767.001.0001

Gray, M. L., & Suri, S. (2019). *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt.

Haugeland, J. (1989). *Artificial Intelligence: The Very Idea.* MIT Press. https://doi.org/10.7551/mitpress/1170.001.0001

ImageNet. (2021, March 11). *About ImageNet.* https://image-net.org/about.php

Levin, J. (2023). *Functionalism.* In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Summer 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2023/entries/functionalism/

Markoff, J. (2012). Seeking a Better Way to Find Web Images. *The New York Times.* https://www.nytimes.com/2012/11/20/science/for-web-images-creating-new-technology-to-seek-and-find.html

McCulloch, W. S. (1965). *Mysterium iniquitatis of sinful man aspiring into the place of god.* In *Embodiments of Mind* (pp. 157–164). MIT Press. https://archive.org/details/embodimentsofmin0000mccu/page/n13/mode/

Mumford, L. (1970). *The Myth of the Machine: The Pentagon of Power* (Vol. 2). Harcourt Brace Jovannich.

Newell, A., & Simon, H. A. (1976). Computer Science as Empirical Inquiry. *Communications of the ACM*, 19(3), 113–126. https://doi.org/10.1145/360018.360022

Pontin, J. (2007). Artificial Intelligence, With Help From the Humans. *The New York Times*. https://www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html

Princeton University. (n.d.). *WordNet 3.1 dictionary files* [Data set]. https://wordnet-code.princeton.edu/wn3.1.dict.tar.gz

Proust, J. (1987). L'Intelligence artificielle comme philosophie. *Le Débat, 47*(5), 88. https://doi.org/10.3917/deba.047.0088

Taylor, A. (2018). The Automation Charade. *Logic(S)*, (5). https://logicmag.io/failure/the-automation-charade/

Triclot, M. (2005). *Penser c'est calculer : éléments pour une préhistoire de l'informatique.* https://hal.science/hal-01526653/document

Turing, A. M. (1937). On Computable Numbers, With an Application to the Entscheidungsproblem. *Proceedings Of The London Mathematical Society, s2-42*(1), 230–265. https://doi.org/10.1112/plms/s2-42.1.230k