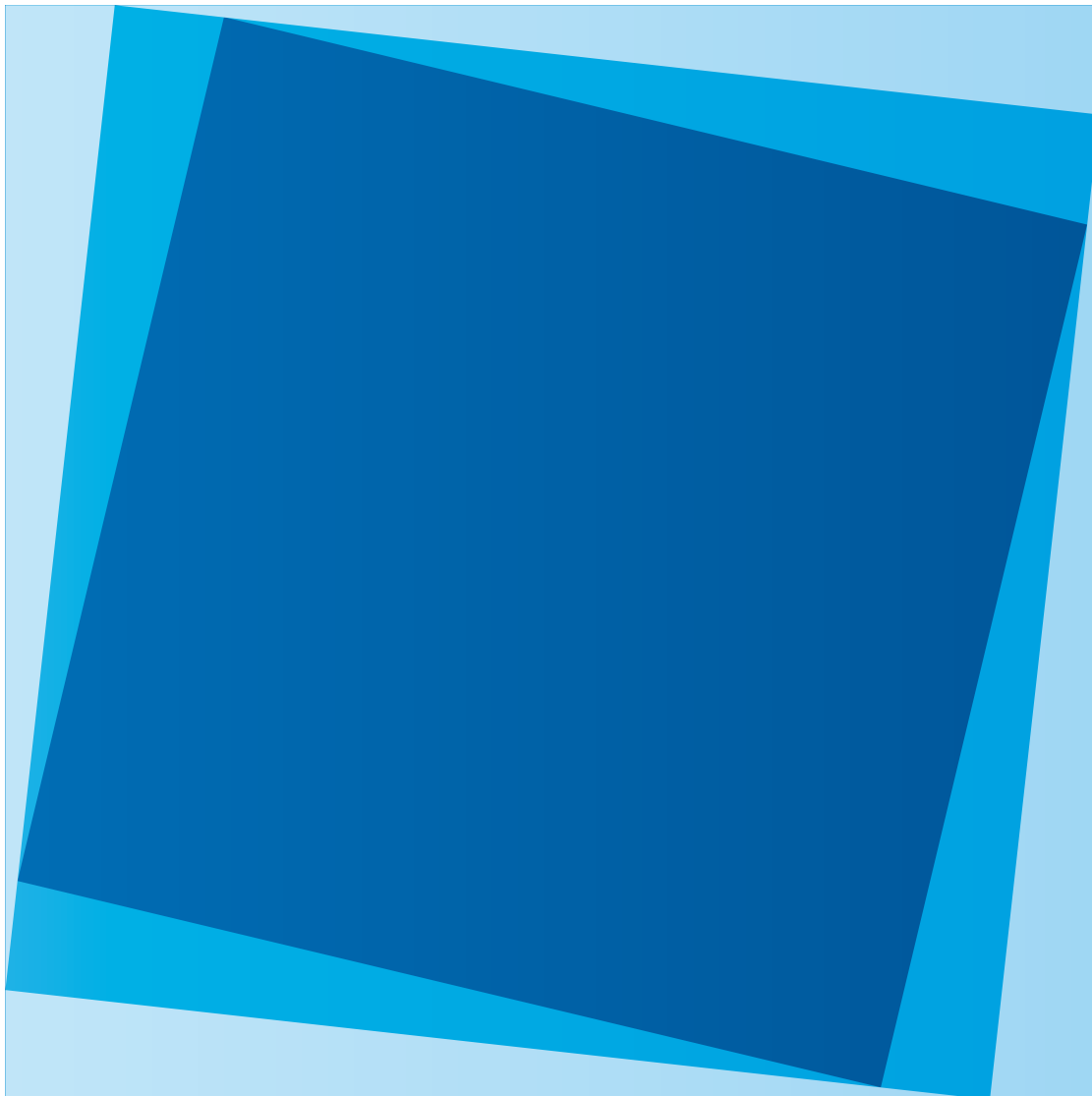


Towards a New Enlightenment? A Transcendent Decade



The Future of AI: Toward Truly Intelligent Artificial Intelligences

Ramón López de Mántaras



Ramón López de Mántaras
Institute of the Spanish
National Research Council

Ramón López de Mántaras is a Research Professor of the Spanish National Research Council (CSIC) and Director of the Artificial Intelligence Research Institute (IIIA). He has an MSc in Computer Science from the University of California Berkeley, a PhD in Physics (Automatic Control) from the University of Toulouse, and a PhD in Computer Science from the Technical University of Barcelona. López de Mántaras is a pioneer of artificial intelligence (AI) in Spain, with contributions, since 1976, in pattern recognition, approximate reasoning, expert systems, machine learning, case-based reasoning, autonomous robots, and AI and music. He is the author of nearly 300 papers and an invited plenary speaker at numerous international conferences. Former Editor-in-Chief of *AI Communications*, he is an editorial board member of several top international journals, an ECCAI Fellow, and the co-recipient of five best paper awards at international conferences. Among other awards, he has received the “City of Barcelona” Research Prize in 1981, the 2011 “American Association of Artificial Intelligence (AAAI) Robert S. Englemore Memorial Award,” the 2012 “Spanish National Computer Science Award” from the Spanish Computer Society, the 2016 “Distinguished Service Award” of the European Association of Artificial Intelligence (EurAI), and the 2017 “IJCAI Donald E. Walker Distinguished Service Award.” He is a member of the Institut d’Estudis Catalans (IEC). He serves on a variety of panels and advisory committees for public and private institutions based in the USA and Europe, such as the EC Joint Research Center High-Level Peer Group. He is presently working on case-based reasoning, machine learning, and AI applications to music.

Recommended books: *The Master Algorithm*, Pedro Domingos, Basic Books, 2015. *Inteligencia Artificial*, Ramón López de Mántaras and Pedro Meseguer, Los Libros de la Catarata, 2017.

This article contains some reflections about artificial intelligence (AI). First, the distinction between strong and weak AI and the related concepts of general and specific AI is made, making it clear that all existing manifestations of AI are weak and specific. The main existing models are briefly described, insisting on the importance of corporality as a key aspect to achieve AI of a general nature. Also discussed is the need to provide common-sense knowledge to the machines in order to move toward the ambitious goal of building general AI. The paper also looks at recent trends in AI based on the analysis of large amounts of data that have made it possible to achieve spectacular progress very recently, also mentioning the current difficulties of this approach to AI. The final part of the article discusses other issues that are and will continue to be vital in AI and closes with a brief reflection on the risks of AI.

Introduction

The final goal of artificial intelligence (AI)—that a machine can have a type of *general* intelligence similar to a human’s—is one of the most ambitious ever proposed by science. In terms of difficulty, it is comparable to other great scientific goals, such as explaining the origin of life or the Universe, or discovering the structure of matter. In recent centuries, this interest in building intelligent machines has led to the invention of models or metaphors of the human brain. In the seventeenth century, for example, Descartes wondered whether a complex mechanical system of gears, pulleys, and tubes could possibly emulate thought. Two centuries later, the metaphor had become telephone systems, as it seemed possible that their connections could be likened to a neural network. Today, the dominant model is computational and is based on the digital computer. Therefore, that is the model we will address in the present article.



The Physical Symbol System Hypothesis: Weak AI Versus Strong AI

In a lecture that coincided with their reception of the prestigious Turing Prize in 1975, Allen Newell and Herbert Simon (Newell and Simon, 1976) formulated the “Physical Symbol System” hypothesis, according to which “a physical symbol system has the necessary and sufficient means for general intelligent action.” In that sense, given that human beings are able to display intelligent behavior in a general way, we, too, would be physical symbol systems. Let us clarify what Newell and Simon mean when they refer to a Physical Symbol System (PSS). A PSS consists of a set of entities called symbols that, through relations, can be combined to form larger structures—just as atoms combine to form molecules—and can be transformed by applying a set of processes. Those processes can create new symbols, create or modify relations among symbols, store symbols, detect whether two are the same or different, and so on. These symbols are physical in the sense that they have an underlying physical-electronic layer (in the case of computers) or a physical-biological one (in the case of human beings). In fact, in the case of computers, symbols are established through digital electronic circuits, whereas humans do so with neural networks. So, according to the PSS hypothesis, the nature of the underlying layer (electronic circuits or neural networks) is unimportant as long as it allows symbols to be processed. Keep in mind that this is a hypothesis, and should, therefore, be neither accepted nor rejected a priori. Either way, its validity or refutation must be verified according to the scientific method, with experimental testing. AI is precisely the scientific field dedicated to attempts to verify this hypothesis in the context of digital computers, that is, verifying whether a properly programmed computer is capable of general intelligent behavior.

Specifying that this must be general intelligence rather than specific intelligence is important, as human intelligence is also general. It is quite a different matter to exhibit specific intelligence. For example, computer programs capable of playing chess at Grand-Master levels are incapable of playing checkers, which is actually a much simpler game. In order for the same computer to play checkers, a different, independent program must be designed and executed. In other words, the computer cannot draw on its capacity to play chess as a means of adapting to the game of checkers. This is not the case, however, with humans, as any human chess player can take advantage of his knowledge of that game to play checkers perfectly in a matter of minutes. The design and application of artificial intelligences that can only behave intelligently in a very specific setting is related to what is known as *weak AI*, as opposed to *strong AI*. Newell, Simon, and the other founding fathers of AI refer to the



latter. Strictly speaking, the PSS hypothesis was formulated in 1975, but, in fact, it was implicit in the thinking of AI pioneers in the 1950s and even in Alan Turing's groundbreaking texts (Turing, 1948, 1950) on intelligent machines.

This distinction between weak and strong AI was first introduced by philosopher John Searle in an article criticizing AI in 1980 (Searle, 1980), which provoked considerable discussion at the time, and still does today. Strong AI would imply that a properly designed computer does not simulate a mind but *actually is one*, and should, therefore, be capable of an intelligence equal, or even superior to human beings. In his article, Searle sought to demonstrate that strong AI is impossible, and, at this point, we should clarify that general AI is not the same as strong AI. Obviously they are connected, but only in one sense: all strong AI will necessarily be general, but there can be general AIs capable of multitasking but not strong in the sense that, while they can emulate the capacity to exhibit general intelligence similar to humans, they do not experience states of mind.

The final goal of AI—that a machine can have a type of general intelligence similar to a human's—is one of the most ambitious ever proposed by science. In terms of difficulty, it is comparable to other great scientific goals, such as explaining the origin of life or the Universe, or discovering the structure of matter

According to Searle, weak AI would involve constructing programs to carry out specific tasks, obviously without need for states of mind. Computers' capacity to carry out specific tasks, sometimes even better than humans, has been amply demonstrated. In certain areas, weak AI has become so advanced that it far outstrips human skill. Examples include solving logical formulas with many variables, playing chess or Go, medical diagnosis, and many others relating to decision-making. Weak AI is also associated with the formulation and testing of hypotheses about aspects of the mind (for example, the capacity for deductive reasoning, inductive learning, and so on) through the construction of programs that carry out those functions, even when they do so using processes totally unlike those of the human brain. As of today, absolutely all advances in the field of AI are manifestations of weak and specific AI.

The Principal Artificial Intelligence Models: Symbolic, Connectionist, Evolutionary, and Corporeal

The *symbolic* model that has dominated AI is rooted in the PSS model and, while it continues to be very important, is now considered classic (it is also known as GOFAI, that is, *Good Old-Fashioned AI*). This top-down model is based on logical reasoning and heuristic searching as the pillars of problem solving. It does not call for an intelligent system to be part of a body, or to be situated in a real setting. In other words, symbolic AI works with abstract representations of the real world that are modeled with representational languages based primarily on mathematical logic and its extensions. That is why the first intelligent systems mainly solved problems that did not require direct interaction with the environment, such as demonstrating simple mathematical theorems or playing chess—in fact, chess programs need neither visual perception for seeing the board, nor technology to actually move

the pieces. That does not mean that symbolic AI cannot be used, for example, to program the reasoning module of a physical robot situated in a real environment, but, during its first years, AI's pioneers had neither languages for representing knowledge nor programming that could do so efficiently. That is why the early intelligent systems were limited to solving problems that did not require direct interaction with the real world. Symbolic AI is still used today to demonstrate theorems and to play chess, but it is also a part of applications that require perceiving the environment and acting upon it, for example learning and decision-making in autonomous robots.

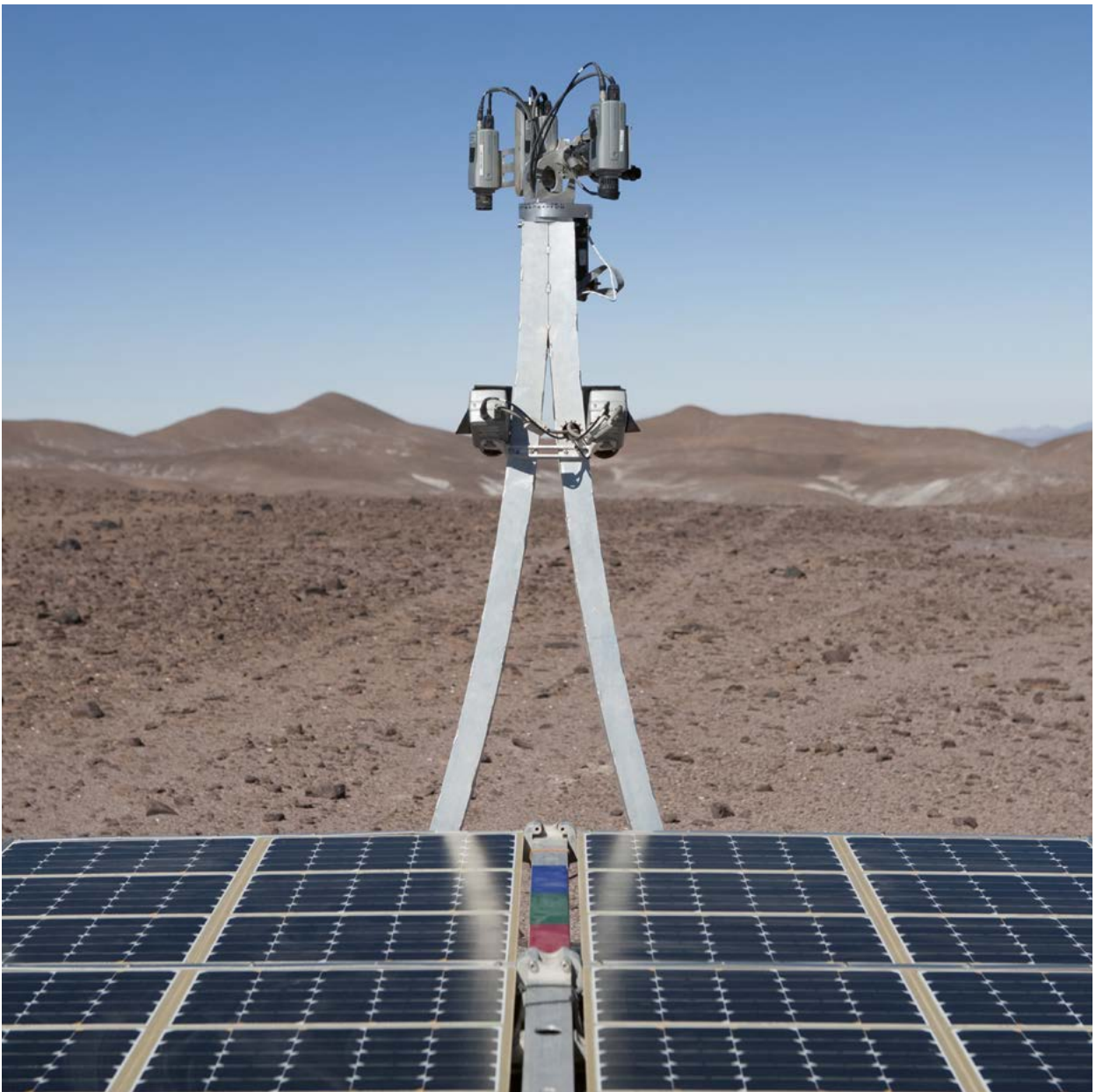


The symbolic model that has dominated AI is rooted in the PSS model and, while it continues to be very important, is now considered classic (it is also known as GOFAI, that is, Good Old-Fashioned AI). This top-down model is based on logical reasoning and heuristic searching as the pillars of problem solving

At the same time that symbolic AI was being developed, a biologically based approach called *connectionist* AI arose. Connectionist systems are not incompatible with the PSS hypothesis but, unlike symbolic AI, they are modeled from the bottom up, as their underlying hypothesis is that intelligence emerges from the distributed activity of a large number of interconnected units whose models closely resemble the electrical activity of biological neurons. In 1943, McCulloch and Pitts (1943) proposed a simplified model of the neuron based in the idea that it is essentially a logic unit. This model is a mathematical abstraction with inputs (dendrites) and outputs (axons). The output value is calculated according to the result of a weighted sum of the entries in such a way that if that sum surpasses a preestablished threshold, it functions as a “1,” otherwise it will be considered a “0.” Connecting the output of each neuron to the inputs of other neurons creates an artificial neural network. Based on what was then known about the reinforcement of synapses among biological neurons, scientists found that these artificial neural networks could be trained to learn functions that related inputs to outputs by adjusting the weights used to determine connections between neurons. These models were hence considered more conducive to learning, cognition, and memory than those based on symbolic AI. Nonetheless, like their symbolic counterparts, intelligent systems based on connectionism do not need to be part of a body, or situated in real surroundings. In that sense, they have the same limitations as symbolic systems. Moreover, real neurons have complex dendritic branching with truly significant electrical and chemical properties. They can contain ionic conductance that produces nonlinear effects. They can receive tens of thousands of synapses with varied positions, polarities, and magnitudes. Furthermore, most brain cells are not neurons, but rather *glial* cells that not only regulate neural functions but also possess electrical potentials, generate calcium waves, and communicate with others. This would seem to indicate that they play a very important role in cognitive processes, but no existing connectionist models include glial cells so they are, at best, extremely incomplete and, at worst, erroneous. In short, the enormous complexity of the brain is very far indeed from current models. And that very complexity also raises the idea of what has come to be known as *singularity*, that is, future artificial superintelligences based on replicas of the brain



Engineers at Carnegie Mellon University developed this robot named Zoe to detect life in apparently uninhabited environments. Zoe includes a cutting-edge system for detecting organic molecules which may help to find life on Mars. It is twenty times faster than the Mars explorer robots, Spirit and Opportunity. Atacama Desert, Chile, 2005





00

04:58

nico

電王戦



The design and application of artificial intelligences that can only behave intelligently in a very specific setting is related to what is known as weak AI, as opposed to strong AI

Masayuki Toyoshima, a professional shogi or Japanese chess player, plays against a YSS computer program that moves the pieces with a robotic arm. Osaka, March, 2014



but capable, in the coming twenty-five years, of far surpassing human intelligence. Such predictions have little scientific merit.

Another biologically inspired but non-corporeal model that is also compatible with the PSS hypothesis is *evolutionary computation* (Holland, 1975). Biology's success at evolving complex organisms led some researchers from the early 1960s to consider the possibility of imitating evolution. Specifically, they wanted computer programs that could evolve, automatically improving solutions to the problems for which they had been programmed. The idea being that, thanks to mutation operators and crossed "chromosomes" modeled by those programs, they would produce new generations of modified programs whose solutions would be better than those offered by the previous ones. Since we can define AI's goal as the search for programs capable of producing intelligent behavior, researchers thought that evolutionary programming might be used to find those programs among all possible programs. The reality is much more complex, and this approach has many limitations although it has produced excellent results in the resolution of optimization problems.

The human brain is very far removed indeed from AI models, which suggests that so-called singularity—artificial superintelligences based on replicas of the brain that far surpass human intelligence—are a prediction with very little scientific merit

One of the strongest critiques of these non-corporeal models is based on the idea that an intelligent agent needs a body in order to have direct experiences of its surroundings (we would say that the agent is "situated" in its surroundings) rather than working from a programmer's abstract descriptions of those surroundings, codified in a language for representing that knowledge. Without a body, those abstract representations have no semantic content for the machine, whereas direct interaction with its surroundings allows the agent to relate signals perceived by its sensors to symbolic representations generated on the basis of what has been perceived. Some AI experts, particularly Rodney Brooks (1991), went so far as to affirm that it was not even necessary to generate those internal representations, that is, that an agent does not even need an internal representation of the world around it because the world itself is the best possible model of itself, and most intelligent behavior does not require reasoning, as it emerged directly from interaction between the agent and its surroundings. This idea generated considerable argument, and some years later, Brooks himself admitted that there are many situations in which an agent requires an internal representation of the world in order to make rational decisions.

In 1965, philosopher Hubert Dreyfus affirmed that AI's ultimate objective—strong AI of a general kind—was as unattainable as the seventeenth-century alchemists' goal of transforming lead into gold (Dreyfus, 1965). Dreyfus argued that the brain processes information in a global and continuous manner, while a computer uses a finite and discreet set of deterministic operations, that is, it applies rules to a finite body of data. In that sense, his argument resembles Searle's, but in later articles and books (Dreyfus, 1992), Dreyfus argued that the body plays a crucial role in intelligence. He was thus one of the first to advocate the need for intelligence to be part of a body that would allow it to interact with the world. The main idea is that living beings' intelligence derives from their situation in surroundings with which they



can interact through their bodies. In fact, this need for corporeality is based on Heidegger's phenomenology and its emphasis on the importance of the body, its needs, desires, pleasures, suffering, ways of moving and acting, and so on. According to Dreyfus, AI must model all of those aspects if it is to reach its ultimate objective of strong AI. So Dreyfus does not completely rule out the possibility of strong AI, but he does state that it is not possible with the classic methods of symbolic, non-corporeal AI. In other words, he considers the Physical Symbol System hypothesis incorrect. This is undoubtedly an interesting idea and today it is shared by many AI researchers. As a result, the corporeal approach with internal representation has been gaining ground in AI and many now consider it essential for advancing toward general intelligences. In fact, we base much of our intelligence on our sensory and motor capacities. That is, the body shapes intelligence and therefore, without a body general intelligence cannot exist. This is so because the body as hardware, especially the mechanisms of the sensory and motor systems, determines the type of interactions that an agent can carry out. At the same time, those interactions shape the agent's cognitive abilities, leading to what is known as *situated cognition*. In other words, as occurs with human beings, the machine is situated in real surroundings so that it can have interactive experiences that will eventually allow it to carry out something similar to what is proposed in Piaget's cognitive development theory (Inhelder and Piaget, 1958): a human being follows a process of mental maturity in stages and the different steps in this process may possibly work as a guide for designing intelligent machines. These ideas have led to a new sub-area of AI called *development robotics* (Weng et al., 2001).

Specialized AI's Successes

All of AI's research efforts have focused on constructing specialized artificial intelligences, and the results have been spectacular, especially over the last decade. This is thanks to the combination of two elements: the availability of huge amounts of data, and access to high-level computation for analyzing it. In fact, the success of systems such as AlphaGO (Silver et al., 2016), Watson (Ferrucci et al., 2013), and advances in autonomous vehicles or image-based medical diagnosis have been possible thanks to this capacity to analyze huge amounts of data and efficiently detect patterns. On the other hand, we have hardly advanced at all in the quest for general AI. In fact, we can affirm that current AI systems are examples of what Daniel Dennet called "competence without comprehension" (Dennet, 2018).

All of AI's research efforts have focused on constructing specialized artificial intelligences, and the results have been spectacular, especially over the last decade. This is thanks to the combination of two elements: the availability of huge amounts of data, and access to high-level computation for analyzing it

Perhaps the most important lesson we have learned over the last sixty years of AI is that what seemed most difficult (diagnosing illnesses, playing chess or Go at the highest level) have turned out to be relatively easy, while what seemed easiest has turned out to be the most



Porior millenem exceaue corat et rempore officatemqui
con nonsedit aut que repel et eic to iust, consequid
quundis doluptur, ullat hicilitio eum recte est ut aut lab
id ex et dolupta tioria deni re, oditis invero nsent, susam
remoles diaestem voloreh endaect inciam, conse pro





difficult of all. The explanation of this apparent contradiction may be found in the difficulty of equipping machines with the knowledge that constitutes “common sense.” without that knowledge, among other limitations, it is impossible to obtain a deep understanding of language or a profound interpretation of what a visual perception system captures. Common-sense knowledge is the result of our lived experiences. Examples include: “water always flows downward;” “to drag an object tied to a string, you have to pull on the string, not push it;” “a glass can be stored in a cupboard, but a cupboard cannot be stored in a glass;” and so on. Humans easily handle millions of such common-sense data that allow us to understand the world we inhabit. A possible line of research that might generate interesting results about the acquisition of common-sense knowledge is the development robotics mentioned above. Another interesting area explores the mathematical modeling and learning of cause-and-effect relations, that is, the learning of causal, and thus asymmetrical, models of the world. Current systems based on deep learning are capable of learning symmetrical mathematical functions, but unable to learn asymmetrical relations. They are, therefore, unable to distinguish cause from effects, such as the idea that the rising sun causes a rooster to crow, but not vice versa (Pearl and Mackenzie, 2018; Lake et al., 2016).

The Future: Toward Truly Intelligent Artificial Intelligences

The most complicated capacities to achieve are those that require interacting with unrestricted and not previously prepared surroundings. Designing systems with these capabilities requires the integration of development in many areas of AI. We particularly need knowledge-representation languages that codify information about many different types of objects, situations, actions, and so on, as well as about their properties and the relations among them—especially, cause-and-effect relations. We also need new algorithms that can use these representations in a robust and efficient manner to resolve problems and answer questions on almost any subject. Finally, given that they will need to acquire an almost unlimited amount of knowledge, those systems will have to be able to learn continuously throughout their existence. In sum, it is essential to design systems that combine perception, representation, reasoning, action, and learning. This is a very important AI problem as we still do not know how to integrate all of these components of intelligence. We need cognitive architectures (Forbus, 2012) that integrate these components adequately. Integrated systems are a fundamental first step in someday achieving general AI.

The most complicated capacities to achieve are those that require interacting with unrestricted and not previously prepared surroundings. Designing systems with these capabilities requires the integration of development in many areas of AI

Among future activities, we believe that the most important research areas will be hybrid systems that combine the advantages of systems capable of reasoning on the basis of knowledge and memory use (Graves et al., 2016) with those of AI based on the analysis of massive amounts of data, that is, deep learning (Bengio, 2009). Today, deep-learning systems are significantly limited by what is known as “catastrophic forgetting.” This means that if they



have been trained to carry out one task (playing Go, for example) and are then trained to do something different (distinguishing between images of dogs and cats, for example) they completely forget what they learned for the previous task (in this case, playing Go). This limitation is powerful proof that those systems do not learn anything, at least in the human sense of learning. Another important limitation of these systems is that they are “black boxes” with no capacity to explain. It would, therefore, be interesting to research how to endow deep-learning systems with an explicative capacity by adding modules that allow them to explain how they reached the proposed results and conclusion, as the capacity to explain is an essential characteristic of any intelligent system. It is also necessary to develop new learning algorithms that do not require enormous amounts of data to be trained, as well as much more energy-efficient hardware to implement them, as energy consumption could end up being one of the main barriers to AI development. Comparatively, the brain is various orders of magnitude more efficient than the hardware currently necessary to implement the most sophisticated AI algorithms. One possible path to explore is memristor-based neuromorphic computing (Saxena et al., 2018).

Other more classic AI techniques that will continue to be extensively researched are multi-agent systems, action planning, experience-based reasoning, artificial vision, multimodal person-machine communication, humanoid robotics, and particularly, new trends in *development robotics*, which may provide the key to endowing machines with common sense, especially the capacity to learn the relations between their actions and the effects these produce on their surroundings. We will also see significant progress in biomimetic approaches to reproducing animal behavior in machines. This is not simply a matter of reproducing an animal’s behavior, it also involves understanding how the brain that produces that behavior actually works. This involves building and programming electronic circuits that reproduce the cerebral activity responsible for this behavior. Some biologists are interested in efforts to create the most complex possible artificial brain because they consider it a means of better understanding that organ. In that context, engineers are seeking biological information that makes designs more efficient. Molecular biology and recent advances in optogenetics will make it possible to identify which genes and neurons play key roles in different cognitive activities.

Development robotics may provide the key to endowing machines with common sense, especially the capacity to learn the relations between their actions and the effects these produce on their surroundings

As to applications: some of the most important will continue to be those related to the Web, video-games, personal assistants, and autonomous robots (especially autonomous vehicles, social robots, robots for planetary exploration, and so on). Environmental and energy-saving applications will also be important, as well as those designed for economics and sociology. Finally, AI applications for the arts (visual arts, music, dance, narrative) will lead to important changes in the nature of the creative process. Today, computers are no longer simply aids to creation; they have begun to be creative agents themselves. This has led to a new and very promising AI field known as *computational creativity* which is producing very interesting results (Colton et al., 2009, 2015; López de Mántaras, 2016) in chess, music, the visual arts, and narrative, among other creative activities.

Some Final Thoughts

No matter how intelligent future artificial intelligences become—even general ones—they will never be the same as human intelligences. As we have argued, the mental development needed for all complex intelligence depends on interactions with the environment and those interactions depend, in turn, on the body—especially the perceptive and motor systems. This, along with the fact that machines will not follow the same socialization and culture-acquisition processes as ours, further reinforces the conclusion that, no matter how sophisticated they become, these intelligences will be different from ours. The existence of intelligences unlike ours, and therefore alien to our values and human needs, calls for reflection on the possible ethical limitations of developing AI. Specifically, we agree with Weizenbaum's affirmation (Weizenbaum, 1976) that no machine should ever make entirely autonomous decisions or give advice that call for, among other things, wisdom born of human experiences, and the recognition of human values.



No matter how intelligent future artificial intelligences become, they will never be the same as human intelligence: the mental development needed for all complex intelligence depends on interactions with the environment and those interactions depend, in turn, on the body—especially the perceptive and motor systems

The true danger of AI is not the highly improbable technological singularity produced by the existence of hypothetical future artificial superintelligences; the true dangers are already here. Today, the algorithms driving Internet search engines or the recommendation and personal-assistant systems on our cellphones, already have quite adequate knowledge of what we do, our preferences and tastes. They can even infer what we think about and how we feel. Access to massive amounts of data that we generate voluntarily is fundamental for this, as the analysis of such data from a variety of sources reveals relations and patterns that could not be detected without AI techniques. The result is an alarming loss of privacy. To avoid this, we should have the right to own a copy of all the personal data we generate, to control its use, and to decide who will have access to it and under what conditions, rather than it being in the hands of large corporations without knowing what they are really doing with our data.

AI is based on complex programming, and that means there will inevitably be errors. But even if it were possible to develop absolutely dependable software, there are ethical dilemmas that software developers need to keep in mind when designing it. For example, an autonomous vehicle could decide to run over a pedestrian in order to avoid a collision that could harm its occupants. Outfitting companies with advanced AI systems that make management and production more efficient will require fewer human employees and thus generate more unemployment. These ethical dilemmas are leading many AI experts to point out the need to regulate its development. In some cases, its use should even be prohibited. One clear example is autonomous weapons. The three basic principles that govern armed conflict: discrimination (the need to distinguish between combatants and civilians, or between

a combatant who is surrendering and one who is preparing to attack), proportionality (avoiding the disproportionate use of force), and precaution (minimizing the number of victims and material damage) are extraordinarily difficult to evaluate and it is therefore almost impossible for the AI systems in autonomous weapons to obey them. But even if, in the very long term, machines were to attain this capacity, it would be indecent to delegate the decision to kill to a machine. Beyond this kind of regulation, it is imperative to educate the citizenry as to the risks of intelligent technologies, and to insure that they have the necessary competence for controlling them, rather than being controlled *by* them. Our future citizens need to be much more informed, with a greater capacity to evaluate technological risks, with a greater critical sense and a willingness to exercise their rights. This training process must begin at school and continue at a university level. It is particularly necessary for science and engineering students to receive training in ethics that will allow them to better grasp the social implications of the technologies they will very likely be developing. Only when we invest in education will we achieve a society that can enjoy the advantages of intelligent technology while minimizing the risks. AI unquestionably has extraordinary potential to benefit society, as long as we use it properly and prudently. It is necessary to increase awareness of AI's limitations, as well as to act collectively to guarantee that AI is used for the common good, in a safe, dependable, and responsible manner.

The road to truly intelligent AI will continue to be long and difficult. After all, this field is barely sixty years old, and, as Carl Sagan would have observed, sixty years are barely the blink of an eye on a cosmic time scale. Gabriel García Márquez put it more poetically in a 1936 speech (“The Cataclysm of Damocles”): “Since the appearance of visible life on Earth, 380 million years had to elapse in order for a butterfly to learn how to fly, 180 million years to create a rose with no other commitment than to be beautiful, and four geological eras in order for us human beings to be able to sing better than birds, and to be able to die from love.”



Select Bibliography

- Bengio, Y. 2009. "Learning deep architectures for AI." *Foundations and Trends in Machine Learning* 2(1): 1–127.
- Brooks, R. A. 1991. "Intelligence without reason." *IJCAI-91 Proceedings of the Twelfth International Joint Conference on Artificial Intelligence* 1: 569–595.
- Colton, S., Lopez de Mantaras, R., and Stock, O. 2009. "Computational creativity: Coming of age." *AI Magazine* 30(3): 11–14.
- Colton, S., Halskov, J., Ventura, D., Gouldstone, I., Cook, M., and Pérez-Ferrer, B. 2015. "The Painting Fool sees! New projects with the automated painter." *International Conference on Computational Creativity (ICCC 2015)*: 189–196.
- Dennet, D. C. 2018. *From Bacteria to Bach and Back: The Evolution of Minds*. London: Penguin.
- Dreyfus, H. 1965. *Alchemy and Artificial Intelligence*. Santa Monica: Rand Corporation.
- Dreyfus, H. 1992. *What Computers Still Can't Do*. New York: MIT Press.
- Ferrucci, D. A., Levas, A., Bagchi, S., Gondek, D., and Mueller, E. T. 2013. "Watson: Beyond jeopardy!" *Artificial Intelligence* 199: 93–105.
- Forbus, K. D. 2012. "How minds will be built." *Advances in Cognitive Systems* 1: 47–58.
- Graves, A., Wayne, G., Reynolds, M., Harley, T., Danihelka, I., Grabska-Barwińska, A., Gómez-Colmenarejo, S., Grefenstette, E., Ramalho, T., Agapiou, J., Puigdomènech-Badia, A., Hermann, K. M., Zwols, Y., Ostrovski, G., Cain, A., King, H., Summerfield, C., Blunsom, P., Kavukcuoglu, K., and Hassabis, D. 2016. "Hybrid computing using a neural network with dynamic external memory." *Nature* 538: 471–476.
- Holland, J. H. 1975. *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.
- Inhelder, B., and Piaget, J. 1958. *The Growth of Logical Thinking from Childhood to Adolescence*. New York: Basic Books.
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J. 2017. "Building machines that learn and think like people." *Behavioral and Brain Sciences* 40:e253.
- López de Mántaras, R. 2016. "Artificial intelligence and the arts: Toward computational creativity." In *The Next Step: Exponential Life*. Madrid: BBVA Open Mind, 100–125.
- McCulloch, W. S., and Pitts, W. 1943. "A logical calculus of ideas immanent in nervous activity." *Bulletin of Mathematical Biophysics* 5: 115–133.
- Newell, A., and Simon, H. A. 1976. "Computer science as empirical inquiry: Symbols and search." *Communications of the ACM* 19(3): 113–126.
- Pearl, J., and Mackenzie, D. 2018. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books.
- Saxena, V., Wu, X., Srivastava, I., and Zhu, K. 2018. "Towards neuromorphic learning machines using emerging memory devices with brain-like energy efficiency." Preprints: <https://www.preprints.org/manuscript/201807.0362/v1>.
- Searle, J. R. 1980. "Minds, brains, and programs." *Behavioral and Brain Sciences* 3(3): 417–457.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D. 2016. "Mastering the game of Go with deep neural networks and tree search." *Nature* 529(7587): 484–489.
- Turing, A. M. 1948. "Intelligent machinery." National Physical Laboratory Report. Reprinted in: *Machine Intelligence* 5, B. Meltzer and D. Michie (eds.). Edinburgh: Edinburgh University Press, 1969.
- Turing, A. M. 1950. "Computing machinery and intelligence." *Mind* LIX(236): 433–460.
- Weizenbaum, J. 1976. *Computer Power and Human Reasoning: From Judgment to Calculation*. San Francisco: W. H. Freeman and Co.
- Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., and Thelen, E. 2001. "Autonomous mental development by robots and animals." *Science* 291: 599–600.



LISTEN TO THE ARTICLE HERE



READ THE FULL BOOK

- ¿Hacia una nueva Ilustración? Una década trascendente
- Towards A New Enlightenment? A Tascendent Decade

ACCESS THE ARTICLE IN SPANISH

El futuro de la IA: hacia inteligencias artificiales realmente inteligentes

HOW TO CITE THIS ARTICLE

López de Mántaras, Ramón. "The Future of AI: Toward Truly Intelligent Artificial Intelligences." In Towards a New Enlightenment? A Transcendent Decade. Madrid: BBVA, 2018.

RELATED ARTICLES

MORE ABOUT #TECHNOLOGY #ARTIFFICIAL-INTELLIGENCE

- The Future of Human-Machine Communications: The Turing Test
The Next Step: Exponential Life, Huma Shah y Kevin Warwick
- The Future of Artificial Intelligence and Cybernetics
There's a Future: Visions for a Better World, Kevin Warwick
- Artificial Intelligence and the Arts: Toward Computational Creativity
The Next Step: Exponential Life, Ramón López de Mántaras

ALL THE OPENMIND COLLECTION TITLES

