Artificial intelligence

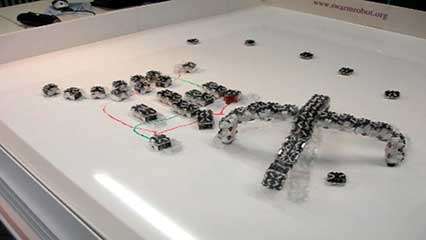
WRITTEN BY:

* [B.J. Copeland](https://www.britannica.com/contributor/BJ-Copeland/4511)

LAST UPDATED: 2-23-2018 [See Article History](https://www.britannica.com/technology/artificial-intelligence#accordion-article-history)

<https://www.britannica.com/technology/artificial-intelligence>

**Artificial intelligence (AI)**, the ability of a digital [computer](https://www.britannica.com/technology/computer) or computer-controlled [robot](https://www.britannica.com/technology/robot-technology) to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the [intellectual](https://www.merriam-webster.com/dictionary/intellectual) processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience. Since the development of the [digital computer](https://www.britannica.com/technology/digital-computer) in the 1940s, it has been demonstrated that computers can be programmed to carry out very complex tasks—as, for example, discovering proofs for mathematical theorems or playing [chess](https://www.britannica.com/topic/chess)—with great proficiency. Still, despite continuing advances in computer processing speed and memory capacity, there are as yet no programs that can match human flexibility over wider domains or in tasks requiring much everyday knowledge. On the other hand, some programs have attained the performance levels of human experts and professionals in performing certain specific tasks, so that artificial intelligence in this limited sense is found in applications as [diverse](https://www.merriam-webster.com/dictionary/diverse) as medical [diagnosis](https://www.merriam-webster.com/dictionary/diagnosis), computer [search engines](https://www.britannica.com/technology/search-engine), and voice or handwriting recognition.



**artificial intelligence**Overview of artificial intelligence.Contunico © ZDF Enterprises GmbH, Mainz

## What Is Intelligence?

All but the simplest [human behaviour](https://www.britannica.com/topic/human-behavior) is ascribed to intelligence, while even the most complicated [insect](https://www.britannica.com/animal/insect) behaviour is never taken as an indication of intelligence. What is the difference? Consider the behaviour of the digger [wasp](https://www.britannica.com/animal/wasp), Sphex ichneumoneus. When the female wasp returns to her burrow with food, she first deposits it on the [threshold](https://www.merriam-webster.com/dictionary/threshold), checks for intruders inside her burrow, and only then, if the coast is clear, carries her food inside. The real nature of the wasp’s [instinctual behaviour](https://www.britannica.com/animal/insect/Role-of-hormones#ref41274) is revealed if the food is moved a few inches away from the entrance to her burrow while she is inside: on emerging, she will repeat the whole procedure as often as the food is displaced. Intelligence—conspicuously absent in the case of Sphex—must include the ability to adapt to new circumstances.

[Psychologists](https://www.britannica.com/science/psychology) generally do not characterize [human intelligence](https://www.britannica.com/science/human-intelligence-psychology) by just one trait but by the combination of many diverse abilities. Research in AI has focused chiefly on the following components of intelligence: learning, reasoning, problem solving, [perception](https://www.britannica.com/topic/perception), and using language.

## [Learning](https://www.britannica.com/technology/machine-learning)

There are a number of different forms of learning as applied to artificial intelligence. The simplest is learning by trial and error. For example, a simple [computer](https://www.britannica.com/technology/computer) program for solving mate-in-one [chess](https://www.britannica.com/topic/chess)problems might try moves at random until mate is found. The program might then store the solution with the position so that the next time the computer encountered the same position it would recall the solution. This simple memorizing of individual items and procedures—known as rote learning—is relatively easy to [implement](https://www.merriam-webster.com/dictionary/implement) on a computer. More challenging is the problem of [implementing](https://www.merriam-webster.com/dictionary/implementing) what is called [generalization](https://www.britannica.com/topic/generalization). Generalization involves applying past experience to [analogous](https://www.merriam-webster.com/dictionary/analogous) new situations. For example, a program that learns the past tense of regular English verbs by rote will not be able to produce the past tense of a word such as jump unless it previously had been presented with jumped, whereas a program that is able to generalize can learn the “add ed” rule and so form the past tense of jump based on experience with similar verbs.

## Reasoning

To reason is to draw [inferences](https://www.merriam-webster.com/dictionary/inferences) appropriate to the situation. Inferences are classified as either [deductive](https://www.britannica.com/topic/deduction-reason) or [inductive](https://www.britannica.com/topic/induction-reason). An example of the former is, “Fred must be in either the museum or the café. He is not in the café; therefore he is in the museum,” and of the latter, “Previous accidents of this sort were caused by instrument failure; therefore this accident was caused by instrument failure.” The most significant difference between these forms of reasoning is that in the deductive case the truth of the [premises](https://www.merriam-webster.com/dictionary/premises) guarantees the truth of the conclusion, whereas in the inductive case the truth of the [premise](https://www.merriam-webster.com/dictionary/premise) lends support to the conclusion without giving absolute [assurance](https://www.merriam-webster.com/dictionary/assurance). Inductive reasoning is common in science, where data are collected and tentative models are developed to describe and predict future behaviour—until the appearance of anomalous data forces the model to be revised. Deductive reasoning is common in [mathematics](https://www.britannica.com/science/mathematics) and [logic](https://www.britannica.com/topic/logic), where elaborate structures of irrefutable theorems are built up from a small set of basic axioms and rules.

There has been considerable success in programming computers to draw inferences, especially deductive inferences. However, true reasoning involves more than just drawing inferences; it involves drawing inferences relevant to the solution of the particular task or situation. This is one of the hardest problems confronting AI.

## [Problem solving](https://www.britannica.com/science/problem-solving)

Problem solving, particularly in artificial intelligence, may be characterized as a systematic search through a range of possible actions in order to reach some predefined goal or solution. Problem-solving methods divide into special purpose and general purpose. A special-purpose method is tailor-made for a particular problem and often exploits very specific features of the situation in which the problem is embedded. In contrast, a general-purpose method is applicable to a wide variety of problems. One general-purpose technique used in AI is means-end analysis—a step-by-step, or [incremental](https://www.merriam-webster.com/dictionary/incremental), reduction of the difference between the current state and the final goal. The program selects actions from a list of means—in the case of a simple robot this might consist of PICKUP, PUTDOWN, MOVEFORWARD, MOVEBACK, MOVELEFT, and MOVERIGHT—until the goal is reached.

Many diverse problems have been solved by artificial intelligence programs. Some examples are finding the winning move (or sequence of moves) in a board game, devising mathematical proofs, and manipulating “virtual objects” in a computer-generated world.

## Perception

In perception the [environment](https://www.merriam-webster.com/dictionary/environment) is scanned by means of various sensory organs, real or artificial, and the scene is decomposed into separate objects in various spatial relationships. Analysis is complicated by the fact that an object may appear different depending on the angle from which it is viewed, the direction and intensity of illumination in the scene, and how much the object contrasts with the surrounding field.

At present, artificial perception is sufficiently well advanced to enable optical sensors to identify individuals, [autonomous](https://www.merriam-webster.com/dictionary/autonomous) vehicles to drive at moderate speeds on the open road, and robots to roam through buildings collecting empty soda cans. One of the earliest systems to [integrate](https://www.merriam-webster.com/dictionary/integrate) perception and action was FREDDY, a stationary robot with a moving television eye and a pincer hand, constructed at the [University of Edinburgh](https://www.britannica.com/topic/University-of-Edinburgh), Scotland, during the period 1966–73 under the direction of [Donald Michie](https://www.britannica.com/biography/Donald-Michie). FREDDY was able to recognize a variety of objects and could be instructed to assemble simple [artifacts](https://www.merriam-webster.com/dictionary/artifacts), such as a toy car, from a random heap of components.

## [Language](https://www.britannica.com/topic/language)

A [language](https://www.britannica.com/topic/language) is a system of signs having meaning by convention. In this sense, language need not be confined to the spoken word. Traffic signs, for example, form a minilanguage, it being a matter of convention that {hazard symbol} means “hazard ahead” in some countries. It is distinctive of languages that linguistic units possess meaning by convention, and linguistic meaning is very different from what is called natural meaning, exemplified in statements such as “Those clouds mean rain” and “The fall in pressure means the [valve](https://www.britannica.com/technology/valve-mechanics) is malfunctioning.”

An important characteristic of full-fledged human languages—in contrast to birdcalls and traffic signs—is their productivity. A productive language can formulate an unlimited variety of sentences.

It is relatively easy to write [computer programs](https://www.britannica.com/technology/computer-program) that seem able, in severely restricted [contexts](https://www.merriam-webster.com/dictionary/contexts), to respond fluently in a human language to questions and statements. Although none of these programs actually understands language, they may, in principle, reach the point where their command of a language is indistinguishable from that of a normal human. What, then, is involved in genuine understanding, if even a computer that uses language like a native human speaker is not acknowledged to understand? There is no universally agreed upon answer to this difficult question. According to one theory, whether or not one understands depends not only on one’s behaviour but also on one’s history: in order to be said to understand, one must have learned the language and have been trained to take one’s place in the linguistic [community](https://www.merriam-webster.com/dictionary/community) by means of interaction with other language users.

## Methods And Goals In AI

## Symbolic vs. connectionist approaches

AI research follows two distinct, and to some extent competing, methods, the symbolic (or “top-down”) approach, and the connectionist (or “bottom-up”) approach. The [top-down approach](https://www.britannica.com/technology/top-down-approach) seeks to replicate intelligence by analyzing [cognition](https://www.britannica.com/topic/cognition-thought-process) independent of the biological structure of the [brain](https://www.britannica.com/science/brain), in terms of the processing of symbols—whence the symbolic label. The [bottom-up approach](https://www.britannica.com/technology/bottom-up-approach), on the other hand, involves creating artificial [neural networks](https://www.britannica.com/technology/neural-network) in imitation of the brain’s structure—whence the connectionist label.

To illustrate the difference between these approaches, consider the task of building a system, equipped with an [optical scanner](https://www.britannica.com/technology/optical-scanner), that recognizes the letters of the alphabet. A bottom-up approach typically involves training an artificial [neural network](https://www.britannica.com/technology/neural-network) by presenting letters to it one by one, gradually improving performance by “tuning” the network. (Tuning adjusts the responsiveness of different neural pathways to different stimuli.) In contrast, a top-down approach typically involves writing a [computer program](https://www.britannica.com/technology/computer-program) that compares each letter with geometric descriptions. Simply put, neural activities are the basis of the bottom-up approach, while symbolic descriptions are the basis of the top-down approach.

In The Fundamentals of Learning (1932), [Edward Thorndike](https://www.britannica.com/biography/Edward-L-Thorndike), a psychologist at [Columbia University](https://www.britannica.com/topic/Columbia-University), [New York City](https://www.britannica.com/place/New-York-City), first suggested that human learning consists of some unknown property of connections between [neurons](https://www.britannica.com/science/neuron) in the brain. In The Organization of Behavior (1949), Donald Hebb, a psychologist at [McGill University](https://www.britannica.com/topic/McGill-University), Montreal, Canada, suggested that learning specifically involves strengthening certain patterns of neural activity by increasing the probability (weight) of induced [neuron](https://www.britannica.com/science/neuron) firing between the associated connections. The notion of weighted connections is described in a later section, [Connectionism](https://www.britannica.com/technology/artificial-intelligence/Expert-systems#ref219103).

In 1957 two vigorous advocates of symbolic AI—Allen Newell, a researcher at the [RAND Corporation](https://www.britannica.com/topic/RAND-Corporation), [Santa Monica](https://www.britannica.com/place/Santa-Monica), California, and [Herbert Simon](https://www.britannica.com/biography/Herbert-A-Simon), a [psychologist](https://www.britannica.com/science/psychology) and [computer scientist](https://www.britannica.com/science/computer-science) at [Carnegie Mellon University](https://www.britannica.com/topic/Carnegie-Mellon-University), Pittsburgh, Pennsylvania—summed up the top-down approach in what they called the physical symbol system hypothesis. This [hypothesis](https://www.merriam-webster.com/dictionary/hypothesis) states that processing structures of symbols is sufficient, in principle, to produce artificial intelligence in a digital computer and that, moreover, [human intelligence](https://www.britannica.com/science/human-intelligence-psychology) is the result of the same type of symbolic manipulations.

During the 1950s and ’60s the top-down and bottom-up approaches were pursued simultaneously, and both achieved noteworthy, if limited, results. During the 1970s, however, bottom-up AI was neglected, and it was not until the 1980s that this approach again became prominent. Nowadays both approaches are followed, and both are acknowledged as facing difficulties. Symbolic techniques work in simplified realms but typically break down when confronted with the real world; meanwhile, bottom-up researchers have been unable to replicate the nervous systems of even the simplest living things. Caenorhabditis elegans, a much-studied worm, has approximately 300 neurons whose pattern of interconnections is perfectly known. Yet connectionist models have failed to mimic even this worm. Evidently, the neurons of connectionist theory are gross oversimplifications of the real thing.

## Strong AI, applied AI, and cognitive simulation

Employing the methods outlined above, AI research attempts to reach one of three goals: strong AI, applied AI, or [cognitive](https://www.merriam-webster.com/dictionary/cognitive) simulation. [Strong AI](https://www.britannica.com/technology/strong-artificial-intelligence) aims to build machines that think. (The term strong AIwas introduced for this category of research in 1980 by the philosopher [John Searle](https://www.britannica.com/biography/John-Searle) of the [University of California](https://www.britannica.com/topic/University-of-California) at Berkeley.) The ultimate ambition of strong AI is to produce a machine whose overall intellectual ability is indistinguishable from that of a human being. As is described in the section [Early milestones in AI](https://www.britannica.com/technology/artificial-intelligence#ref219091), this goal generated great interest in the 1950s and ’60s, but such optimism has given way to an appreciation of the extreme difficulties involved. To date, progress has been meagre. Some critics doubt whether research will produce even a system with the overall intellectual ability of an [ant](https://www.britannica.com/animal/ant)in the forseeable future. Indeed, some researchers working in AI’s other two branches view strong AI as not worth pursuing.

Applied AI, also known as advanced information processing, aims to produce commercially viable “smart” systems—for example, “expert” medical diagnosis systems and stock-trading systems. Applied AI has enjoyed considerable success, as described in the section [Expert systems](https://www.britannica.com/technology/artificial-intelligence/Expert-systems#ref219098).

In cognitive simulation, computers are used to test theories about how the human [mind](https://www.britannica.com/topic/mind) works—for example, theories about how people recognize faces or recall memories. Cognitive simulation is already a powerful [tool](https://www.britannica.com/technology/tool) in both neuroscience and [cognitive psychology](https://www.britannica.com/science/cognitive-psychology).

## [Alan Turing](https://www.britannica.com/biography/Alan-Turing) And The Beginning Of AI

## Theoretical work

The earliest substantial work in the field of artificial intelligence was done in the mid-20th century by the British logician and computer pioneer [Alan Mathison Turing](https://www.britannica.com/biography/Alan-Turing). In 1935 Turing described an abstract computing machine consisting of a limitless memory and a scanner that moves back and forth through the [memory](https://www.britannica.com/technology/computer-memory), symbol by symbol, reading what it finds and writing further symbols. The actions of the scanner are dictated by a program of instructions that also is stored in the memory in the form of symbols. This is Turing’s [stored-program](https://www.britannica.com/technology/stored-program-concept) concept, and [implicit](https://www.merriam-webster.com/dictionary/implicit) in it is the possibility of the machine operating on, and so modifying or improving, its own program. Turing’s [conception](https://www.merriam-webster.com/dictionary/conception) is now known simply as the universal [Turing machine](https://www.britannica.com/technology/Turing-machine). All modern computers are in essence universal Turing machines.

During [World War II](https://www.britannica.com/event/World-War-II), Turing was a leading cryptanalyst at the Government Code and Cypher School in [Bletchley Park](https://www.britannica.com/place/Bletchley-Park), Buckinghamshire, England. Turing could not turn to the project of building a stored-program electronic computing machine until the cessation of hostilities in Europe in 1945. Nevertheless, during the war he gave considerable thought to the issue of machine intelligence. One of Turing’s colleagues at Bletchley Park, Donald Michie (who later founded the Department of Machine Intelligence and Perception at the University of Edinburgh), later recalled that Turing often discussed how computers could learn from experience as well as solve new problems through the use of guiding principles—a process now known as heuristic problem solving.

Turing gave quite possibly the earliest public lecture (London, 1947) to mention computer intelligence, saying, “What we want is a machine that can learn from experience,” and that the “possibility of letting the machine alter its own instructions provides the [mechanism](https://www.britannica.com/technology/mechanism-machinery) for this.” In 1948 he introduced many of the central concepts of AI in a report entitled “Intelligent Machinery.” However, Turing did not publish this paper, and many of his ideas were later reinvented by others. For instance, one of Turing’s original ideas was to train a network of artificial [neurons](https://www.britannica.com/science/neuron) to perform specific tasks, an approach described in the section [Connectionism](https://www.britannica.com/technology/artificial-intelligence/Expert-systems#ref219103).

## [Chess](https://www.britannica.com/topic/chess)

At Bletchley Park, Turing illustrated his ideas on machine intelligence by reference to [chess](https://www.britannica.com/topic/chess)—a useful source of challenging and clearly defined problems against which proposed methods for problem solving could be tested. In principle, a chess-playing computer could play by searching exhaustively through all the available moves, but in practice this is impossible because it would involve examining an astronomically large number of moves. [Heuristics](https://www.merriam-webster.com/dictionary/Heuristics) are necessary to guide a narrower, more discriminative search. Although Turing experimented with designing chess programs, he had to content himself with theory in the absence of a computer to run his chess program. The first true AI programs had to await the arrival of [stored-program electronic digital computers](https://www.britannica.com/technology/computer/History-of-computing#ref216048).

In 1945 Turing predicted that computers would one day play very good chess, and just over 50 years later, in 1997, [Deep Blue](https://www.britannica.com/topic/Deep-Blue), a [chess computer](https://www.britannica.com/topic/chess/The-time-element-and-competition#ref80452) built by the [International Business Machines Corporation (IBM)](https://www.britannica.com/topic/International-Business-Machines-Corporation), beat the reigning world champion, [Garry Kasparov](https://www.britannica.com/biography/Garry-Kasparov), in a six-game match. While Turing’s prediction came true, his expectation that chess programming would contribute to the understanding of how human beings think did not. The huge improvement in computer chess since Turing’s day is attributable to advances in computer [engineering](https://www.britannica.com/technology/engineering) rather than advances in AI—Deep Blue’s 256 parallel processors enabled it to examine 200 million possible moves per second and to look ahead as many as 14 turns of play. Many agree with [Noam Chomsky](https://www.britannica.com/biography/Noam-Chomsky), a linguist at the [Massachusetts Institute of Technology (MIT)](https://www.britannica.com/topic/Massachusetts-Institute-of-Technology), who opined that a computer beating a grandmaster at chess is about as interesting as a [bulldozer](https://www.britannica.com/technology/bulldozer) winning an [Olympic](https://www.britannica.com/sports/Olympic-Games) [weightlifting](https://www.britannica.com/sports/weightlifting) competition.

## The [Turing test](https://www.britannica.com/technology/Turing-test)

In 1950 Turing sidestepped the traditional debate concerning the definition of intelligence, introducing a practical test for computer intelligence that is now known simply as the [Turing test](https://www.britannica.com/technology/Turing-test). The Turing test involves three participants: a computer, a human interrogator, and a human foil. The interrogator attempts to determine, by asking questions of the other two participants, which is the computer. All communication is via keyboard and display screen. The interrogator may ask questions as penetrating and wide-ranging as he or she likes, and the computer is permitted to do everything possible to force a wrong identification. (For instance, the computer might answer, “No,” in response to, “Are you a computer?” and might follow a request to multiply one large number by another with a long pause and an incorrect answer.) The foil must help the interrogator to make a correct identification. A number of different people play the roles of interrogator and foil, and, if a sufficient proportion of the interrogators are unable to distinguish the computer from the human being, then (according to proponents of Turing’s test) the computer is considered an intelligent, thinking entity.



**Searle, John: Chinese room argument**Learn about John Searle's Chinese room argument, a critique of the Turing test.© Open University

In 1991 the American philanthropist Hugh Loebner started the annual Loebner Prize competition, promising a $100,000 payout to the first computer to pass the Turing test and awarding $2,000 each year to the best effort. However, no AI program has come close to passing an undiluted Turing test.

## Early Milestones In AI

## The first AI programs

The earliest successful AI program was written in 1951 by Christopher Strachey, later director of the Programming Research Group at the [University of Oxford](https://www.britannica.com/topic/University-of-Oxford). Strachey’s [checkers](https://www.britannica.com/topic/checkers) (draughts) program ran on the [Ferranti Mark I](https://www.britannica.com/technology/computer/History-of-computing#ref216048) computer at the [University of Manchester](https://www.britannica.com/topic/University-of-Manchester), England. By the summer of 1952 this program could play a complete game of checkers at a reasonable speed.

Information about the earliest successful demonstration of [machine learning](https://www.britannica.com/technology/machine-learning) was published in 1952. Shopper, written by Anthony Oettinger at the [University of Cambridge](https://www.britannica.com/topic/University-of-Cambridge), ran on the [EDSAC](https://www.britannica.com/technology/computer/History-of-computing#ref216048) computer. Shopper’s simulated world was a mall of eight shops. When instructed to purchase an item, Shopper would search for it, visiting shops at random until the item was found. While searching, Shopper would memorize a few of the items stocked in each shop visited (just as a human shopper might). The next time Shopper was sent out for the same item, or for some other item that it had already located, it would go to the right shop straight away. This simple form of learning, as is pointed out in the introductory section [What is intelligence?](https://www.britannica.com/technology/artificial-intelligence#ref219078), is called rote learning.

The first AI program to run in the United States also was a checkers program, written in 1952 by Arthur Samuel for the [prototype](https://www.merriam-webster.com/dictionary/prototype) of the IBM 701. Samuel took over the essentials of Strachey’s checkers program and over a period of years considerably extended it. In 1955 he added features that enabled the program to learn from experience. Samuel included mechanisms for both rote learning and generalization, enhancements that eventually led to his program’s winning one game against a former Connecticut [checkers](https://www.britannica.com/topic/checkers) champion in 1962.

## Evolutionary computing

Samuel’s checkers program was also notable for being one of the first efforts at evolutionary computing. (His program “evolved” by pitting a modified copy against the current best version of his program, with the winner becoming the new standard.) Evolutionary computing typically involves the use of some automatic method of generating and evaluating successive “generations” of a program, until a highly proficient solution evolves.

A leading proponent of evolutionary computing, [John Holland](https://www.britannica.com/biography/John-Henry-Holland), also wrote test [software](https://www.britannica.com/technology/software) for the prototype of the IBM 701 computer. In particular, he helped design a [neural-network](https://www.britannica.com/technology/neural-network) “virtual” rat that could be trained to navigate through a maze. This work convinced Holland of the [efficacy](https://www.merriam-webster.com/dictionary/efficacy) of the bottom-up approach. While continuing to consult for [IBM](https://www.britannica.com/topic/International-Business-Machines-Corporation), Holland moved to the [University of Michigan](https://www.britannica.com/topic/University-of-Michigan) in 1952 to pursue a doctorate in [mathematics](https://www.britannica.com/science/mathematics). He soon switched, however, to a new interdisciplinary program in [computers](https://www.britannica.com/technology/computer) and information processing (later known as communications science) created by Arthur Burks, one of the builders of [ENIAC](https://www.britannica.com/technology/ENIAC) and its successor EDVAC. In his 1959 dissertation, for most likely the world’s first [computer science](https://www.britannica.com/science/computer-science) Ph.D., Holland proposed a new type of computer—a [multiprocessor](https://www.britannica.com/technology/multiprocessing) computer—that would assign each artificial neuron in a network to a separate processor. (In 1985 [Daniel Hillis](https://www.britannica.com/biography/Danny-Hillis) solved the engineering difficulties to build the first such computer, the 65,536-processor Thinking Machines Corporation [supercomputer](https://www.britannica.com/technology/supercomputer#ref93020).)

Holland joined the faculty at Michigan after graduation and over the next four decades directed much of the research into methods of automating evolutionary computing, a process now known by the term [*genetic algorithms*](https://www.britannica.com/technology/genetic-algorithm). Systems [implemented](https://www.merriam-webster.com/dictionary/implemented) in Holland’s laboratory included a [chess](https://www.britannica.com/topic/chess)program, models of single-[cell](https://www.britannica.com/science/cell-biology) biological organisms, and a classifier system for controlling a simulated gas-pipeline network. Genetic [algorithms](https://www.merriam-webster.com/dictionary/algorithms) are no longer restricted to “academic” demonstrations, however; in one important practical application, a [genetic algorithm](https://www.britannica.com/technology/genetic-algorithm) cooperates with a witness to a crime in order to generate a portrait of the criminal.

## [Logical](https://www.britannica.com/topic/logic) reasoning and problem solving

The ability to reason logically is an important aspect of intelligence and has always been a major focus of AI research. An important landmark in this area was a theorem-proving program written in 1955–56 by [Allen Newell](https://www.britannica.com/biography/Allen-Newell) and J. Clifford Shaw of the [RAND Corporation](https://www.britannica.com/topic/RAND-Corporation) and [Herbert Simon](https://www.britannica.com/biography/Herbert-A-Simon) of the Carnegie Mellon University. The [Logic Theorist](https://www.britannica.com/technology/Logic-Theorist), as the program became known, was designed to prove theorems from Principia Mathematica (1910–13), a three-volume work by the British philosopher-mathematicians [Alfred North Whitehead](https://www.britannica.com/biography/Alfred-North-Whitehead) and [Bertrand Russell](https://www.britannica.com/biography/Bertrand-Russell). In one instance, a proof devised by the program was more elegant than the proof given in the books.

Newell, Simon, and Shaw went on to write a more powerful program, the [General Problem Solver](https://www.britannica.com/science/General-Problem-Solver), or GPS. The first version of GPS ran in 1957, and work continued on the project for about a decade. GPS could solve an impressive variety of puzzles using a trial and error approach. However, one [criticism](https://www.merriam-webster.com/dictionary/criticism) of GPS, and similar programs that lack any learning capability, is that the program’s intelligence is entirely secondhand, coming from whatever information the programmer explicitly includes.

## English dialogue

Two of the best-known early AI programs, Eliza and [Parry](https://www.britannica.com/topic/Parry), gave an eerie semblance of intelligent conversation. (Details of both were first published in 1966.) Eliza, written by [Joseph Weizenbaum](https://www.britannica.com/biography/Joseph-Weizenbaum) of MIT’s AI Laboratory, simulated a human therapist. Parry, written by [Stanford University](https://www.britannica.com/topic/Stanford-University) psychiatrist Kenneth Colby, simulated a human [paranoiac](https://www.britannica.com/science/paranoia). Psychiatrists who were asked to decide whether they were communicating with Parry or a human paranoiac were often unable to tell. Nevertheless, neither Parry nor Eliza could reasonably be described as intelligent. Parry’s contributions to the conversation were canned—constructed in advance by the programmer and stored away in the [computer’s memory](https://www.britannica.com/technology/computer-memory). Eliza, too, relied on canned sentences and simple programming tricks.

## AI [programming languages](https://www.britannica.com/technology/computer-programming-language)

In the course of their work on the Logic Theorist and GPS, Newell, Simon, and Shaw developed their [Information Processing Language](https://www.britannica.com/technology/Information-Processing-Language) (IPL), a computer language tailored for AI programming. At the heart of IPL was a highly flexible [data structure](https://www.britannica.com/technology/data-structure) that they called a list. A list is simply an ordered sequence of items of data. Some or all of the items in a list may themselves be lists. This scheme leads to richly branching structures.

In 1960 [John McCarthy](https://www.britannica.com/biography/John-McCarthy) combined elements of IPL with the [lambda calculus](https://www.britannica.com/topic/lambda-calculus) (a formal mathematical-logical system) to produce the programming language [LISP](https://www.britannica.com/technology/LISP-computer-language) (List Processor), which remains the principal language for AI work in the United States. (The lambda calculus itself was invented in 1936 by the Princeton logician [Alonzo Church](https://www.britannica.com/biography/Alonzo-Church) while he was investigating the abstract Entscheidungsproblem, or “decision problem,” for [predicate](https://www.merriam-webster.com/dictionary/predicate) logic—the same problem that [Turing](https://www.britannica.com/biography/Alan-Turing)had been attacking when he invented the universal [Turing machine](https://www.britannica.com/technology/Turing-machine).)

The logic programming language [PROLOG](https://www.britannica.com/technology/PROLOG) (Programmation en Logique) was conceived by Alain Colmerauer at the University of Aix-Marseille, France, where the language was first implemented in 1973. PROLOG was further developed by the logician Robert Kowalski, a member of the AI group at the University of Edinburgh. This language makes use of a powerful theorem-proving technique known as [resolution](https://www.britannica.com/technology/resolution-computer-logic), invented in 1963 at the U.S. [Atomic Energy Commission’s](https://www.britannica.com/topic/Atomic-Energy-Commission-United-States-organization) [Argonne National Laboratory](https://www.britannica.com/topic/Argonne-National-Laboratory) in Illinois by the British logician Alan Robinson. PROLOG can determine whether or not a given statement follows logically from other given statements. For example, given the statements “All logicians are rational” and “Robinson is a logician,” a PROLOG program responds in the [affirmative](https://www.merriam-webster.com/dictionary/affirmative) to the query “Robinson is rational?” PROLOG is widely used for AI work, especially in Europe and Japan.

Researchers at the Institute for New Generation Computer Technology in Tokyo have used PROLOG as the basis for sophisticated logic programming languages. Known as [fifth-generation languages](https://www.britannica.com/technology/fifth-generation-language), these are in use on nonnumerical parallel computers developed at the Institute.

Other recent work includes the development of languages for reasoning about time-dependent data such as “the account was paid yesterday.” These languages are based on [tense logic](https://www.britannica.com/topic/temporal-logic), which permits statements to be located in the flow of time. (Tense logic was invented in 1953 by the philosopher Arthur Prior at the University of Canterbury, Christchurch, New Zealand.)

## Microworld programs

To cope with the bewildering complexity of the real world, scientists often ignore less relevant details; for instance, physicists often ignore [friction](https://www.britannica.com/science/friction) and [elasticity](https://www.britannica.com/science/elasticity-physics) in their models. In 1970 [Marvin Minsky](https://www.britannica.com/biography/Marvin-Lee-Minsky) and [Seymour Papert](https://www.britannica.com/biography/Seymour-Papert) of the MIT AI Laboratory proposed that likewise AI research should focus on developing programs capable of intelligent behaviour in simpler artificial [environments](https://www.merriam-webster.com/dictionary/environments) known as microworlds. Much research has focused on the so-called blocks world, which consists of coloured blocks of various shapes and sizes arrayed on a flat surface.

An early success of the microworld approach was [SHRDLU](https://www.britannica.com/technology/SHRDLU), written by Terry Winograd of MIT. (Details of the program were published in 1972.) SHRDLU controlled a robot arm that operated above a flat surface strewn with play blocks. Both the arm and the blocks were virtual. SHRDLU would respond to commands typed in natural English, such as “Will you please stack up both of the red blocks and either a green cube or a pyramid.” The program could also answer questions about its own actions.Although SHRDLU was initially hailed as a major breakthrough, Winograd soon announced that the program was, in fact, a dead end. The techniques pioneered in the program proved unsuitable for application in wider, more interesting worlds. Moreover, the appearance that SHRDLU gave of understanding the blocks microworld, and English statements concerning it, was in fact an [illusion](https://www.merriam-webster.com/dictionary/illusion). SHRDLU had no [idea](https://www.britannica.com/topic/idea) what a green block was.

Another product of the microworld approach was [Shakey](https://www.britannica.com/topic/Shakey), a mobile [robot](https://www.britannica.com/technology/robot-technology) developed at the Stanford Research Institute by Bertram Raphael, Nils Nilsson, and others during the period 1968–72. The robot occupied a specially built microworld consisting of walls, doorways, and a few simply shaped wooden blocks. Each wall had a carefully painted baseboard to enable the robot to “see” where the wall met the floor (a simplification of reality that is typical of the microworld approach). Shakey had about a dozen basic abilities, such as TURN, PUSH, and CLIMB-RAMP.



Shakey, the robotShakey was developed (1966–72) at the Stanford Research Institute, Menlo Park, California.The robot is equipped with of a television camera, a range finder, and collision sensors that enable a minicomputer to control its actions remotely. Shakey can perform a few basic actions, such as go forward, turn, and push, albeit at a very slow pace. Contrasting colours, particularly the dark baseboard on each wall, help the robot to distinguish separate surfaces.Courtesy of SRI International

Critics pointed out the highly simplified nature of Shakey’s environment and emphasized that, despite these simplifications, Shakey operated excruciatingly slowly; a series of actions that a human could plan out and execute in minutes took Shakey days.

The greatest success of the microworld approach is a type of program known as an [expert system](https://www.britannica.com/technology/expert-system), described in the next section.

## [Expert Systems](https://www.britannica.com/technology/expert-system)

Expert systems occupy a type of microworld—for example, a model of a ship’s hold and its cargo—that is self-contained and relatively uncomplicated. For such AI systems every effort is made to incorporate all the information about some narrow field that an expert (or group of experts) would know, so that a good [expert system](https://www.britannica.com/technology/expert-system) can often outperform any single human expert. There are many commercial expert systems, including programs for medical [diagnosis](https://www.merriam-webster.com/dictionary/diagnosis), [chemical analysis](https://www.britannica.com/science/chemical-analysis), credit authorization, financial management, corporate planning, financial document routing, [oil](https://www.britannica.com/science/oil-chemical-compound) and [mineral](https://www.britannica.com/science/mineral-chemical-compound) prospecting, [genetic engineering](https://www.britannica.com/science/genetic-engineering), [automobile](https://www.britannica.com/technology/automobile) design and manufacture, [camera](https://www.britannica.com/technology/camera) lens design, [computer](https://www.britannica.com/technology/computer) installation design, airline scheduling, cargo placement, and automatic help services for home computer owners.

## Knowledge and inference

The basic components of an expert system are a [knowledge base](https://www.britannica.com/technology/knowledge-base), or KB, and an [inference](https://www.merriam-webster.com/dictionary/inference) engine. The information to be stored in the KB is obtained by interviewing people who are expert in the area in question. The interviewer, or knowledge engineer, organizes the information elicited from the experts into a collection of rules, typically of an “if-then” structure. Rules of this type are called production rules. The [inference engine](https://www.britannica.com/technology/inference-engine) enables the expert system to draw deductions from the rules in the KB. For example, if the KB contains the production rules “if x, then y” and “if y, then z,” the inference engine is able to deduce “if x, then z.” The expert system might then query its user, “Is x true in the situation that we are considering?” If the answer is [affirmative](https://www.merriam-webster.com/dictionary/affirmative), the system will proceed to infer z.

Some expert systems use [fuzzy logic](https://www.britannica.com/science/fuzzy-logic). In standard logic there are only two truth values, true and false. This absolute precision makes vague attributes or situations difficult to characterize. (When, precisely, does a thinning head of hair become a bald head?) Often the rules that human experts use contain vague expressions, and so it is useful for an expert system’s inference engine to employ fuzzy logic.

## [DENDRAL](https://www.britannica.com/technology/DENDRAL)

In 1965 the AI researcher Edward Feigenbaum and the geneticist [Joshua Lederberg](https://www.britannica.com/biography/Joshua-Lederberg), both of [Stanford University](https://www.britannica.com/topic/Stanford-University), began work on [Heuristic](https://www.merriam-webster.com/dictionary/Heuristic) DENDRAL (later shortened to DENDRAL), a chemical-analysis expert system. The substance to be analyzed might, for example, be a complicated [compound](https://www.britannica.com/science/chemical-compound) of [carbon](https://www.britannica.com/science/carbon-chemical-element), [hydrogen](https://www.britannica.com/science/hydrogen), and [nitrogen](https://www.britannica.com/science/nitrogen). Starting from spectrographic data obtained from the substance, DENDRAL would hypothesize the substance’s [molecular](https://www.britannica.com/science/molecule) structure. DENDRAL’s performance rivaled that of [chemists](https://www.britannica.com/science/chemistry) expert at this task, and the program was used in industry and in [academia](https://www.merriam-webster.com/dictionary/academia).

## [MYCIN](https://www.britannica.com/technology/MYCIN)

Work on MYCIN, an expert system for treating [blood](https://www.britannica.com/science/blood-biochemistry) infections, began at [Stanford University](https://www.britannica.com/topic/Stanford-University) in 1972. MYCIN would attempt to diagnose patients based on reported symptoms and medical test results. The program could request further information concerning the patient, as well as suggest additional laboratory tests, to arrive at a probable diagnosis, after which it would recommend a course of treatment. If requested, MYCIN would explain the reasoning that led to its diagnosis and recommendation. Using about 500 production rules, MYCIN operated at roughly the same level of competence as human specialists in blood infections and rather better than general practitioners.

Nevertheless, expert systems have no common sense or understanding of the limits of their expertise. For instance, if MYCIN were told that a patient who had received a gunshot wound was bleeding to death, the program would attempt to diagnose a [bacterial](https://www.britannica.com/science/bacteria) cause for the patient’s symptoms. Expert systems can also act on absurd clerical errors, such as prescribing an obviously incorrect dosage of a drug for a patient whose weight and age data were accidentally transposed.

## The [CYC project](https://www.britannica.com/topic/CYC)

CYC is a large experiment in [symbolic AI](https://www.britannica.com/technology/top-down-approach). The project began in 1984 under the [auspices](https://www.merriam-webster.com/dictionary/auspices) of the Microelectronics and Computer Technology Corporation, a [consortium](https://www.merriam-webster.com/dictionary/consortium) of computer, [semiconductor](https://www.britannica.com/science/semiconductor), and [electronics](https://www.britannica.com/technology/electronics) manufacturers. In 1995 Douglas Lenat, the CYC project director, spun off the project as Cycorp, Inc., based in Austin, Texas. The most ambitious goal of Cycorp was to build a KB containing a significant percentage of the commonsense knowledge of a [human being](https://www.britannica.com/topic/human-being). Millions of commonsense assertions, or rules, were coded into CYC. The expectation was that this “critical mass” would allow the system itself to extract further rules directly from ordinary prose and eventually serve as the foundation for future generations of expert systems.

With only a fraction of its commonsense KB compiled, CYC could draw [inferences](https://www.merriam-webster.com/dictionary/inferences) that would defeat simpler systems. For example, CYC could infer, “Garcia is wet,” from the statement, “Garcia is finishing a marathon run,” by employing its rules that running a [marathon](https://www.britannica.com/sports/marathon-race) entails high exertion, that people sweat at high levels of exertion, and that when something sweats it is wet. Among the outstanding remaining problems are issues in searching and problem solving—for example, how to search the KB automatically for information that is relevant to a given problem. AI researchers call the problem of updating, searching, and otherwise manipulating a large structure of symbols in realistic amounts of time the frame problem. Some critics of symbolic AI believe that the frame problem is largely unsolvable and so maintain that the symbolic approach will never yield genuinely intelligent systems. It is possible that CYC, for example, will [succumb](https://www.merriam-webster.com/dictionary/succumb) to the frame problem long before the system achieves human levels of knowledge.

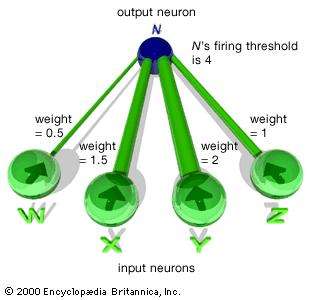
## [Connectionism](https://www.britannica.com/technology/connectionism-artificial-intelligence)

Connectionism, or neuronlike computing, developed out of attempts to understand how the human [brain](https://www.britannica.com/science/brain) works at the neural level and, in particular, how people learn and remember. In 1943 the neurophysiologist Warren McCulloch of the [University of Illinois](https://www.britannica.com/topic/University-of-Illinois) and the mathematician Walter Pitts of the [University of Chicago](https://www.britannica.com/topic/University-of-Chicago) published an influential [treatise](https://www.merriam-webster.com/dictionary/treatise) on [neural nets and automatons](https://www.britannica.com/topic/automata-theory#ref21498), according to which each [neuron](https://www.britannica.com/science/neuron) in the brain is a simple digital processor and the brain as a whole is a form of computing machine. As McCulloch put it subsequently, “What we thought we were doing (and I think we succeeded fairly well) was treating the brain as a [Turing machine](https://www.britannica.com/technology/Turing-machine).”

## Creating an artificial [neural network](https://www.britannica.com/technology/neural-network)

It was not until 1954, however, that Belmont Farley and Wesley Clark of [MIT](https://www.britannica.com/topic/Massachusetts-Institute-of-Technology) succeeded in running the first artificial [neural network](https://www.britannica.com/technology/neural-network)—albeit limited by [computer memory](https://www.britannica.com/technology/computer-memory) to no more than 128 [neurons](https://www.britannica.com/science/neuron). They were able to train their networks to [recognize simple patterns](https://www.britannica.com/technology/pattern-recognition-computer-science). In addition, they discovered that the random destruction of up to 10 percent of the neurons in a trained network did not affect the network’s performance—a feature that is reminiscent of the brain’s ability to tolerate limited damage inflicted by surgery, accident, or disease.

The simple neural network depicted in the figure illustrates the central ideas of connectionism. Four of the network’s five neurons are for input, and the fifth—to which each of the others is connected—is for output. Each of the neurons is either firing (1) or not firing (0). Each connection leading to N, the output neuron, has a “weight.” What is called the total weighted input into N is calculated by adding up the weights of all the connections leading to N from neurons that are firing. For example, suppose that only two of the input neurons, X and Y, are firing. Since the weight of the connection from X to Nis 1.5 and the weight of the connection from Y to N is 2, it follows that the total weighted input to N is 3.5. As shown in the figure, N has a firing [threshold](https://www.merriam-webster.com/dictionary/threshold) of 4. That is to say, if N’s total weighted input equals or exceeds 4, then N fires; otherwise, N does not fire. So, for example, N does not fire if the only input neurons to fire are X and Y, but N does fire if X, Y, and Z all fire.



A section of an artificial neural networkIn the figure the weight, or strength, of each input is indicated by the relative size of its connection. The firing threshold for the output neuron, N, is 4 in this example. Hence, N is quiescent unless a combination of input signals is received from W, X, Y, and Z that exceeds a weight of 4.Encyclopædia Britannica, Inc.

Training the network involves two steps. First, the external [agent](https://www.britannica.com/technology/agent) inputs a pattern and observes the behaviour of N. Second, the agent adjusts the connection weights in accordance with the rules:

1. If the actual output is 0 and the desired output is 1, increase by a small fixed amount the weight of each connection leading to N from neurons that are firing (thus making it more likely that N will fire the next time the network is given the same pattern);
2. If the actual output is 1 and the desired output is 0, decrease by that same small amount the weight of each connection leading to the output neuron from neurons that are firing (thus making it less likely that the output neuron will fire the next time the network is given that pattern as input).

The external agent—actually a computer program—goes through this two-step procedure with each pattern in a training sample, which is then repeated a number of times. During these many repetitions, a pattern of connection weights is forged that enables the network to respond correctly to each pattern. The striking thing is that the [learning](https://www.britannica.com/science/learning) process is entirely mechanical and requires no human intervention or adjustment. The connection weights are increased or decreased automatically by a constant amount, and exactly the same learning procedure applies to different tasks.

## [Perceptrons](https://www.britannica.com/technology/perceptrons)

In 1957 [Frank Rosenblatt](https://www.britannica.com/biography/Frank-Rosenblatt) of the Cornell Aeronautical Laboratory at [Cornell University](https://www.britannica.com/topic/Cornell-University) in Ithaca, New York, began investigating artificial neural networks that he called perceptrons. He made major contributions to the field of AI, both through experimental investigations of the properties of neural networks (using computer simulations) and through detailed mathematical analysis. Rosenblatt was a [charismatic](https://www.merriam-webster.com/dictionary/charismatic) communicator, and there were soon many research groups in the United States studying perceptrons. Rosenblatt and his followers called their approach connectionist to emphasize the importance in learning of the creation and modification of connections between neurons. Modern researchers have adopted this term.

One of Rosenblatt’s contributions was to generalize the training procedure that Farley and Clark had applied to only two-layer networks so that the procedure could be applied to multilayer networks. Rosenblatt used the phrase “back-propagating error correction” to describe his method. The method, with substantial improvements and extensions by numerous scientists, and the term [*back-propagation*](https://www.britannica.com/technology/back-propagation-algorithm) are now in everyday use in connectionism.

## Conjugating [verbs](https://www.britannica.com/topic/verb)

In one famous connectionist experiment conducted at the University of California at San Diego (published in 1986), David Rumelhart and James McClelland trained a network of 920 artificial neurons, arranged in two layers of 460 neurons, to form the past tenses of English verbs. Root forms of verbs—such as come, look, and sleep—were presented to one layer of neurons, the input layer. A supervisory [computer program](https://www.britannica.com/technology/computer-program) observed the difference between the actual response at the layer of output neurons and the desired response—came, say—and then mechanically adjusted the connections throughout the network in accordance with the procedure described above to give the network a slight push in the direction of the correct response. About 400 different verbs were presented one by one to the network, and the connections were adjusted after each presentation. This whole procedure was repeated about 200 times using the same verbs, after which the network could correctly form the past tense of many unfamiliar verbs as well as of the original verbs. For example, when presented for the first time with guard, the network responded guarded; with weep, wept; with cling, clung; and with drip, dripped (complete with double p). This is a striking example of learning involving generalization. (Sometimes, though, the peculiarities of English were too much for the network, and it formed squawked from squat, shipped from shape, and membled from mail.)

Another name for connectionism is [*parallel distributed processing*](https://www.britannica.com/science/parallel-distributed-processing), which emphasizes two important features. First, a large number of relatively simple processors—the neurons—operate in parallel. Second, neural networks store information in a distributed fashion, with each individual connection participating in the storage of many different items of information. The know-how that enabled the past-tense network to form wept from weep, for example, was not stored in one specific location in the network but was spread throughout the entire pattern of connection weights that was forged during training. The human brain also appears to store information in a distributed fashion, and connectionist research is contributing to attempts to understand how it does so.

## Other neural networks

Other work on neuronlike computing includes the following:

* Visual perception. Networks can recognize faces and other objects from visual data. A neural network designed by John Hummel and Irving Biederman at the [University of Minnesota](https://www.britannica.com/topic/University-of-Minnesota) can identify about 10 objects from simple line drawings. The network is able to recognize the objects—which include a mug and a frying pan—even when they are drawn from different angles. Networks investigated by Tomaso Poggio of MIT are able to recognize bent-wire shapes drawn from different angles, faces photographed from different angles and showing different expressions, and objects from cartoon drawings with gray-scale shading indicating depth and orientation.
* Language processing. Neural networks are able to convert handwritten and typewritten material to electronic text. The U.S. Internal Revenue Service has commissioned a neuronlike system that will automatically read tax returns and correspondence. Neural networks also convert speech to printed text and printed text to speech.
* Financial analysis. Neural networks are being used increasingly for loan risk [assessment](https://www.merriam-webster.com/dictionary/assessment), real estate valuation, bankruptcy prediction, share price prediction, and other business applications.
* Medicine. Medical applications include detecting [lung](https://www.britannica.com/science/lung) nodules and [heart](https://www.britannica.com/science/heart) [arrhythmias](https://www.britannica.com/science/arrhythmia) and predicting adverse drug reactions.
* Telecommunications. [Telecommunications](https://www.britannica.com/technology/telecommunications-network) applications of neural networks include [control](https://www.britannica.com/technology/control-system) of [telephone](https://www.britannica.com/technology/telephone)switching networks and echo cancellation in [modems](https://www.britannica.com/technology/modem) and on [satellite](https://www.britannica.com/science/satellite) links.

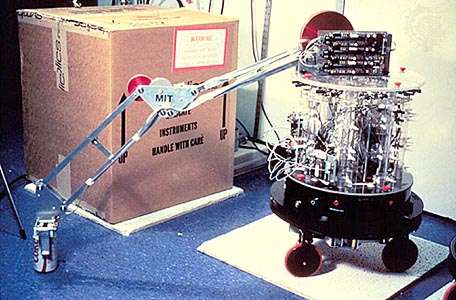
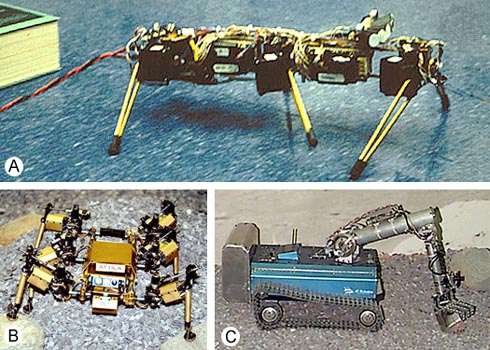
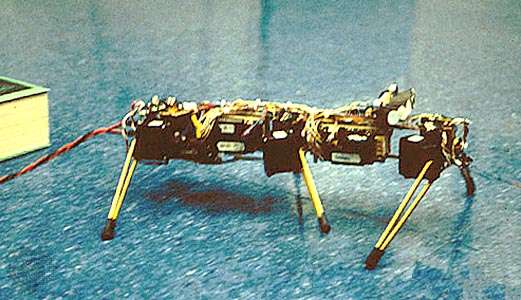
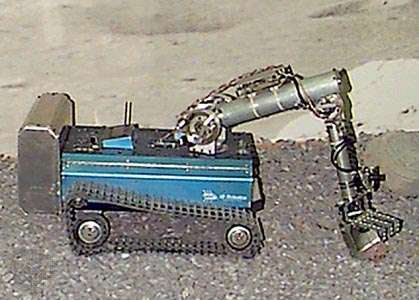
## [Nouvelle AI](https://www.britannica.com/technology/nouvelle-artificial-intelligence)

## New foundations

The approach now known as nouvelle AI was pioneered at the MIT AI Laboratory by the Australian [Rodney Brooks](https://www.britannica.com/biography/Rodney-Allen-Brooks) during the latter half of the 1980s. Nouvelle AI distances itself from strong AI, with its emphasis on human-level performance, in favour of the relatively modest aim of [insect](https://www.britannica.com/animal/insect)-level performance. At a very fundamental level, nouvelle AI rejects symbolic AI’s reliance upon constructing internal models of reality, such as those described in the section [Microworld programs](https://www.britannica.com/technology/artificial-intelligence" \l "ref219097). Practitioners of nouvelle AI assert that true intelligence involves the ability to function in a real-world [environment](https://www.merriam-webster.com/dictionary/environment).

A central [idea](https://www.britannica.com/topic/idea) of nouvelle AI is that intelligence, as expressed by complex behaviour, “emerges” from the interaction of a few simple behaviours. For example, a [robot](https://www.britannica.com/technology/robot-technology) whose simple behaviours include collision avoidance and motion toward a moving object will appear to stalk the object, pausing whenever it gets too close.

One famous example of nouvelle AI is Brooks’s robot [Herbert](https://www.britannica.com/topic/Herbert-robot) (named after [Herbert Simon](https://www.britannica.com/biography/Herbert-A-Simon)), whose environment is the busy offices of the MIT AI Laboratory. Herbert searches desks and tables for empty soda cans, which it picks up and carries away. The robot’s seemingly goal-directed behaviour emerges from the interaction of about 15 simple behaviours. More recently, Brooks has constructed [prototypes](https://www.merriam-webster.com/dictionary/prototypes) of mobile robots for exploring the surface of [Mars](https://www.britannica.com/place/Mars-planet). (See the photographs and an interview with Rodney Brooks.)

* 
* 
* 
* 
* 

Herbert, the robot, c. 1987Designed by Rodney Brooks and affectionately named for artificial intelligence pioneer Herbert Simon, Herbert employed 30 infrared sensors, a laser scanner, and a magnetic compass to locate soft-drink cans and keep itself oriented as it wandered throughout the MIT Artificial Intelligence Laboratory. After collecting an empty can with its robotic arm, Herbert would return it to a recycling bin.© MIT, Artificial Intelligence Laboratory

The Mars Rover Research ProjectThree stages (A, Genghis; B, Attila; C, Pebbles) are displayed in MIT's development of a mobile robot to reconnoitre the Martian surface. To see a larger image and obtain information on each robot, click on the individual photograph.© MIT, Artificial Intelligence Laboratory

Genghis, the robotGenghis was built at MIT in the mid-1980s to demonstrate the efficacy of using numerous small, light, mobile robots to reconnoitre the Martian surface. Genghis was the prototype for the later autonomous “spider” robots Attila and Hannibal. Genghis weighs about 1 kilogram (2.2 pounds), contains 6 pyroelectric sensors for detecting animal life, and employs 12 motors to power its 6 independently operating legs. Genghis is now located in the National Air and Space Museum, Washington, D.C.© MIT, Artificial Intelligence Laboratory

Attila, the robotAttila, along with its twin, Hannibal, was built at MIT (1989–91) as part of a research project to develop autonomous robots for planetary exploration. Attila, like its predecessor Genghis, is a small, six-legged robot, but, whereas Genghis has no independent power source, Attila was equipped with solar cells to recharge its batteries.© MIT Artificial Intelligence Laboratory

Pebbles, the robot. This tractorlike robot utilizes a vision-based control system developed during the late 1990s as part of MIT's Mars Rover Research Project. Pebbles, which is about the size of a domestic cat, negotiates around obstacles with the aid of a single camera, the robot's only sensor. With its arm attached, Pebbles can collect samples or handle dangerous objects.© MIT, Artificial Intelligence Laboratory

Nouvelle AI sidesteps the frame problem discussed in the section [The CYC project](https://www.britannica.com/technology/artificial-intelligence/Expert-systems#ref219102). Nouvelle systems do not contain a complicated symbolic model of their environment. Instead, information is left “out in the world” until such time as the system needs it. A nouvelle system refers continuously to its sensors rather than to an internal model of the world: it “reads off” the external world whatever information it needs at precisely the time it needs it. (As Brooks insisted, the world is its own best model—always exactly up-to-date and complete in every detail.)

## The [situated approach](https://www.britannica.com/technology/artificial-intelligence-situated-approach)

Traditional AI has by and large attempted to build disembodied intelligences whose only interaction with the world has been indirect (CYC, for example). Nouvelle AI, on the other hand, attempts to build embodied intelligences situated in the real world—a method that has come to be known as the situated approach. Brooks quoted approvingly from the brief sketches that [Turing](https://www.britannica.com/biography/Alan-Turing) gave in 1948 and 1950 of the situated approach. By equipping a machine “with the best sense organs that money can buy,” Turing wrote, the machine might be taught “to understand and speak English” by a process that would “follow the normal teaching of a child.” Turing contrasted this with the approach to AI that focuses on abstract activities, such as the playing of [chess](https://www.britannica.com/topic/chess). He advocated that both approaches be pursued, but until recently little attention has been paid to the situated approach.

The situated approach was also anticipated in the writings of the philosopher [Bert Dreyfus](https://www.britannica.com/biography/Bert-Dreyfus) of the [University of California](https://www.britannica.com/topic/University-of-California) at Berkeley. Beginning in the early 1960s, Dreyfus opposed the physical symbol system [hypothesis](https://www.merriam-webster.com/dictionary/hypothesis), arguing that intelligent behaviour cannot be completely captured by symbolic descriptions. As an [alternative](https://www.merriam-webster.com/dictionary/alternative), Dreyfus advocated a view of intelligence that stressed the need for a body that could move about, interacting directly with [tangible](https://www.merriam-webster.com/dictionary/tangible) physical objects. Once reviled by advocates of AI, Dreyfus is now regarded as a prophet of the situated approach.

Critics of nouvelle AI point out the failure to produce a system exhibiting anything like the complexity of behaviour found in real insects. Suggestions by researchers that their nouvelle systems may soon be conscious and possess language seem entirely premature.

## Is [Strong AI](https://www.britannica.com/technology/strong-artificial-intelligence) Possible?

The ongoing success of applied AI and of [cognitive](https://www.merriam-webster.com/dictionary/cognitive) simulation, as described in the preceding sections of this article, seems assured. However, strong AI—that is, artificial intelligence that aims to duplicate human [intellectual](https://www.merriam-webster.com/dictionary/intellectual) abilities—remains controversial. Exaggerated claims of success, in professional journals as well as the popular press, have damaged its reputation. At the present time even an embodied system displaying the overall intelligence of a [cockroach](https://www.britannica.com/animal/cockroach-insect) is proving [elusive](https://www.merriam-webster.com/dictionary/elusive), let alone a system that can rival a human being. The difficulty of scaling up AI’s modest achievements cannot be overstated. Five decades of research in symbolic AI have failed to produce any firm evidence that a symbol system can [manifest](https://www.merriam-webster.com/dictionary/manifest) human levels of general intelligence; connectionists are unable to model the nervous systems of even the simplest [invertebrates](https://www.britannica.com/animal/invertebrate); and critics of nouvelle AI regard as simply mystical the view that high-level behaviours involving language understanding, planning, and reasoning will somehow emerge from the interaction of basic behaviours such as obstacle avoidance, gaze control, and object manipulation.

However, this lack of substantial progress may simply be testimony to the difficulty of strong AI, not to its impossibility. Let us turn to the very idea of strong artificial intelligence. Can a computer possibly think? [Noam Chomsky](https://www.britannica.com/biography/Noam-Chomsky) suggests that debating this question is pointless, for it is an essentially arbitrary decision whether to extend common usage of the word think to include machines. There is, Chomsky claims, no factual question as to whether any such decision is right or wrong—just as there is no question as to whether our decision to say that airplanes fly is right, or our decision not to say that ships swim is wrong. However, this seems to oversimplify matters. The important question is, Could it ever be appropriate to say that computers think, and, if so, what conditions must a computer satisfy in order to be so described?

Some authors offer the [Turing test](https://www.britannica.com/technology/Turing-test) as a definition of intelligence. However, Turing himself pointed out that a computer that ought to be described as intelligent might nevertheless fail his test if it were incapable of successfully imitating a human being. For example, why should an intelligent robot designed to oversee mining on the [Moon](https://www.britannica.com/place/Moon) necessarily be able to pass itself off in conversation as a human being? If an intelligent entity can fail the test, then the test cannot function as a definition of intelligence. It is even questionable whether passing the test would actually show that a computer is intelligent, as the information theorist [Claude Shannon](https://www.britannica.com/biography/Claude-Shannon) and the AI pioneer [John McCarthy](https://www.britannica.com/biography/John-McCarthy) pointed out in 1956. Shannon and McCarthy argued that it is possible, in principle, to design a machine containing a complete set of canned responses to all the questions that an interrogator could possibly ask during the fixed time span of the test. Like Parry, this machine would produce answers to the interviewer’s questions by looking up appropriate responses in a giant table. This objection seems to show that in principle a system with no intelligence at all could pass the Turing test.

In fact, AI has no real definition of intelligence to offer, not even in the subhuman case. [Rats](https://www.britannica.com/animal/rat) are intelligent, but what exactly must an artificial intelligence achieve before researchers can claim this level of success? In the absence of a reasonably precise [criterion](https://www.merriam-webster.com/dictionary/criterion) for when an artificial system counts as intelligent, there is no objective way of telling whether an AI research program has succeeded or failed. One result of AI’s failure to produce a satisfactory criterion of intelligence is that, whenever researchers achieve one of AI’s goals—for example, a program that can summarize newspaper articles or beat the world chess champion—critics are able to say “That’s not intelligence!” [Marvin Minsky’s](https://www.britannica.com/biography/Marvin-Lee-Minsky)response to the problem of defining intelligence is to maintain—like Turing before him—that intelligence is simply our name for any problem-solving mental process that we do not yet understand. Minsky likens intelligence to the concept “unexplored regions of Africa”: it disappears as soon as we discover it.

[*B.J. Copeland*](https://www.britannica.com/contributor/BJ-Copeland/4511)