Artificial intelligence

From Wikipedia, the free encyclopedia

Jump to: navigation, search
"AI" redirects here. For other uses, see AI (disambiguation).

Artificial intelligence (**AI**) is the intelligence of machines and the branch of computer science which aims to create it. Major AI textbooks define the field as "the study and design of intelligent agents,"[1] where an intelligent agent is a system that perceives its environment and takes actions which maximize its chances of success.[2] John McCarthy, who coined the term in 1956,[3] defines it as "the science and engineering of making intelligent machines."[4]

The field was founded on the claim that a central property of human beings, intelligence—the sapience of *Homo sapiens*—can be so precisely described that it can be simulated by a machine.[5] This raises philosophical issues about the nature of the mind and limits of scientific hubris, issues which have been addressed by myth, fiction and philosophy since antiquity.[6] Artificial intelligence has been the subject of breathtaking optimism,[7] has suffered stunning setbacks[8] and, today, has become an essential part of the technology industry, providing the heavy lifting for many of the most difficult problems in computer science.

AI research is highly technical and specialized, so much so that some critics decry the "fragmentation" of the field.[9] Subfields of AI are organized around particular problems, the application of particular tools and around long standing theoretical differences of opinion. The central problems of AI include such traits as reasoning, knowledge, planning, learning, communication, perception and the ability to move and manipulate objects.[10] General intelligence (or "strong AI") is still a long term goal of (some) research.[11]

Contents

Perspectives on AI

AI in myth, fiction and speculation

Main articles: Artificial intelligence in fiction, Ethics of artificial intelligence, Transhumanism, and Technological singularity

Thinking machines and artificial beings appear in Greek myths, such as Talos of Crete, the golden robots of Hephaestus and Pygmalion's Galatea.[12] Human likenesses believed to have intelligence were built in many ancient societies; some of the earliest being the sacred statues worshipped in Egypt and Greece,[13][14] and including the machines of Yan Shi,[15] Hero of Alexandria,[16] Al-Jazari[17] or Wolfgang von Kempelen.[18] It was widely believed that artificial beings had been created by Geber,[19] Judah Loew[20] and Paracelsus.[21] Stories of these creatures and their fates discuss many of the same hopes, fears and ethical concerns that are presented by artificial intelligence.[6]

Mary Shelley's *Frankenstein*,[22] considers a key issue in the ethics of artificial intelligence: if a machine can be created that has intelligence, could it also *feel*? If it can feel, does it have the same rights as a human being? The idea also appears in modern science fiction: the film *Artificial Intelligence: A.I.* considers a machine in the form of a small boy which has been given the ability to feel human emotions, including, tragically, the capacity to suffer. This issue, now known as "robot rights", is currently being considered by, for example, California's Institute for the Future,[23] although many critics believe that the discussion is premature.[24]

Another issue explored by both science fiction writers and futurists is the impact of artificial intelligence on society. In fiction, AI has appeared as a servant (R2D2 in *Star Wars*), a comrade (Lt. Commander Data in *Star Trek*), a conqueror (*The Matrix*), a dictator (*With Folded Hands*), an exterminator (*Terminator*, *Battlestar Galactica*), an extension to human abilities (*Ghost in the Shell*) and the saviour of the human race (R. Daneel Olivaw in the *Foundation Series*). Academic sources have considered such consequences as: a decreased demand for human labor;[25] the enhancement of human ability or experience;[26] and a need for redefinition of human identity and basic values.[27]

Several futurists argue that artificial intelligence will transcend the limits of progress and fundamentally transform humanity. Ray Kurzweil has used Moore's law (which describes the relentless exponential improvement in digital technology with uncanny accuracy) to calculate that desktop computers will have the same processing power as human brains by the year 2029, and that by 2045 artificial intelligence will reach a point where it is able to improve *itself* at a rate that far exceeds anything conceivable in the past, a scenario that science fiction writer Vernor Vinge named the "technological singularity".[26] Edward Fredkin argues that "artificial intelligence is the next stage in evolution,"[28] an idea first proposed by Samuel Butler's *Darwin Among the Machines* (1863), and expanded upon by George Dyson in his book of the same name in 1998. Several futurists and science fiction writers have predicted that human beings and machines will merge in the future into cyborgs that are more capable and powerful than either. This idea, called transhumanism, which has roots in Aldous Huxley and Robert Ettinger, is now associated with robot designer Hans Moravec, cyberneticist Kevin Warwick and inventor Ray Kurzweil.[26] Transhumanism has been illustrated in fiction as well, for example in the manga *Ghost in the Shell* and the science fiction series *Dune*. Pamela McCorduck writes that these scenarios are expressions of an ancient human desire to, as she calls it, "forge the gods."[6]

History of AI research

Main articles: history of artificial intelligence and timeline of artificial intelligence

In the middle of the 20th century, a handful of scientists began a new approach to building intelligent machines, based on recent discoveries in neurology, a new mathematical theory of information, an understanding of control and stability called cybernetics, and above all, by the invention of the digital computer, a machine based on the abstract essence of mathematical reasoning.[29]

The field of modern AI research was founded at a conference on the campus of Dartmouth College in the summer of 1956.[30] Those who attended would become the leaders of AI research for many decades, especially John McCarthy, Marvin Minsky, Allen Newell and Herbert Simon, who founded AI laboratories at MIT, CMU and Stanford. They and their students wrote programs that were, to most people, simply astonishing:[31] computers were solving word problems in algebra, proving logical theorems and speaking English.[32] By the middle 60s their research was heavily funded by the U.S. Department of Defense[33] and they were optimistic about the future of the new field:

- 1965, H. A. Simon: "[M]achines will be capable, within twenty years, of doing any work a man can do"[34]
- 1967, Marvin Minsky: "Within a generation ... the problem of creating 'artificial intelligence' will substantially be solved."[35]

These predictions, and many like them, would not come true. They had failed to recognize the difficulty of some of the problems they faced.[36] In 1974, in response to the criticism of England's Sir James Lighthill and ongoing pressure from Congress to fund more productive projects, the U.S. and British governments cut off all undirected, exploratory research in AI. This was the first AI winter.[37]

In the early 80s, AI research was revived by the commercial success of expert systems[38] (a form of AI program that simulated the knowledge and analytical skills of one or more human experts). By 1985 the market for AI had reached more than a billion dollars and governments around the world poured money back into the field.[39] However, just a few years later, beginning with the collapse of the Lisp Machine market in 1987, AI once again fell into disrepute, and a second, more lasting AI winter began.[40]

In the 90s and early 21st century AI achieved its greatest successes, albeit somewhat behind the scenes. Artificial intelligence was adopted throughout the technology industry, providing the heavy lifting for logistics, data mining, medical diagnosis and many other areas.[41] The success was due to several factors: the incredible power of computers today (see Moore's law), a greater emphasis on solving specific subproblems, the creation of new ties between AI and other fields working on similar problems, and above all a new commitment by researchers to solid mathematical methods and rigorous scientific standards.[42]

Philosophy of AI

Mind and Brain portal

Main article: philosophy of artificial intelligence

Artificial intelligence, by claiming to be able to recreate the capabilities of the human mind, is both a challenge and an inspiration for philosophy. Are there limits to how intelligent machines can be? Is there an essential difference between human intelligence and artificial intelligence? Can a machine have a mind and consciousness? A few of the most influential answers to these questions are given below.[43]

Turing's "polite convention"

If a machine acts as intelligently as a human being, then it is as intelligent as a human being. Alan Turing theorized that, ultimately, we can only judge the intelligence of machine based on its behavior. This theory forms the basis of the Turing test.[44]

The Dartmouth proposal

"Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it." This assertion was printed in the proposal for the Dartmouth Conference of 1956, and represents the position of most working AI researchers.[5]

Newell and Simon's physical symbol system hypothesis

"A physical symbol system has the necessary and sufficient means of general intelligent action." This statement claims that the essence of intelligence is symbol manipulation.[45] Hubert Dreyfus argued that, on the contrary, human expertise depends on unconscious instinct rather than conscious symbol manipulation and on having a "feel" for the situation rather than explicit symbolic knowledge.[46][47]

Gödel's incompleteness theorem

A formal system (such as a computer program) can not prove all true statements. Roger Penrose is among those who claim that Gödel's theorem limits what machines can do.[48][49]

Searle's strong AI hypothesis

"The appropriately programmed computer with the right inputs and outputs would thereby have a mind in exactly the same sense human beings have minds."[50] Searle counters this assertion with his Chinese room argument, which asks us to look *inside* the computer and try to find where the "mind" might be.[51]

The artificial brain argument

The brain can be simulated. Hans Moravec, Ray Kurzweil and others have argued that it is technologically feasible to copy the brain directly into hardware and software, and that such a simulation will be essentially identical to the original. This argument combines the idea that a suitably powerful machine can simulate any process, with the materialist idea that the mind is the result of physical processes in the brain.[52]

AI research

In the 21st century, AI research has become highly specialized and technical. It is deeply divided into subfields that often fail to communicate with each other.[9] Subfields have grown up around particular institutions, the work of particular researchers, particular problems (listed below), long standing differences of opinion about how AI should be done (listed as "approaches" below) and the application of widely differing tools (see tools of AI, below).

Problems of AI

The problem of simulating (or creating) intelligence has been broken down into a number of specific sub-problems. These consist of particular traits or capabilities that researchers would like an intelligent system to display. The traits described below have received the most attention.[10]

Deduction, reasoning, problem solving

Early AI researchers developed algorithms that imitated the step-by-step reasoning that human beings use when they solve puzzles, play board games or make logical deductions.[53] By the late 80s and 90s, AI research had also developed highly successful methods for dealing with uncertain or incomplete information, employing concepts from probability and economics.[54]

For difficult problems, most of these algorithms can require enormous computational resources — most experience a "combinatorial explosion": the amount of memory or computer time required becomes astronomical when the problem goes beyond a certain size. The search for more efficient problem solving algorithms is a high priority for AI research.[55]

Human beings solve most of their problems using fast, intuitive judgments rather than the conscious, step-by-step deduction that early AI research was able to model.[56] AI has made some progress at imitating this kind of "sub-symbolic" problem solving: embodied approaches emphasize the importance of sensorimotor skills to higher reasoning; neural net research attempts to simulate the structures inside human and animal brains that gives rise to this skill.

Knowledge representation

Main articles: knowledge representation and commonsense knowledge

Knowledge representation[57] and knowledge engineering[58] are central to AI research. Many of the problems machines are expected to solve will require extensive knowledge about the world. Among the things that AI needs to represent are: objects, properties, categories and relations between objects;[59] situations, events, states and time;[60] causes and effects;[61] knowledge about knowledge (what we know about what other people know);[62] and many other, less well researched domains. A complete representation of "what exists" is an ontology[63] (borrowing a word from traditional philosophy), of which the most general are called upper ontologies.

Among the most difficult problems in knowledge representation are:

Default reasoning and the qualification problem

Many of the things people know take the form of "working assumptions." For example, if a bird comes up in conversation, people typically picture an animal that is fist sized, sings, and flies. None of these things are true about all birds. John McCarthy identified this problem in 1969[64] as the qualification problem: for any commonsense rule that AI researchers care to represent, there tend to be a huge number of exceptions. Almost nothing is simply true or false in the way that abstract logic requires. AI research has explored a number of solutions to this problem.[65]

The breadth of commonsense knowledge

The number of atomic facts that the average person knows is astronomical. Research projects that attempt to build a complete knowledge base of commonsense knowledge (e.g., Cyc) require enormous amounts of laborious ontological engineering — they must be built, by hand, one complicated concept at a time.[66]

The subsymbolic form of some commonsense knowledge

Much of what people know isn't represented as "facts" or "statements" that they could actually say out loud. For example, a chess master will avoid a particular chess position because it "feels too exposed"[67] or an art critic can take one look at a statue and instantly realize that it is a fake.[68] These are intuitions or tendencies that are represented in the brain non-consciously and sub-symbolically. Knowledge like this informs, supports and provides a context for symbolic, conscious knowledge. As with the related problem of sub-symbolic reasoning, it is hoped that situated AI or computational intelligence will provide ways to represent this kind of knowledge.[69]

Planning

Main article: automated planning and scheduling

Intelligent agents must be able to set goals and achieve them.[70] They need a way to visualize the future (they must have a representation of the state of the world and be able to make predictions about how their actions will change it) and be able to make choices that maximize the utility (or "value") of the available choices.[71]

In some planning problems, the agent can assume that it is the only thing acting on the world and it can be certain what the consequences of its actions may be.[72] However, if this is not true, it must periodically check if the world matches its predictions and it must change its plan as this becomes necessary, requiring the agent to reason under uncertainty.[73]

Multi-agent planning uses the cooperation and competition of many agents to achieve a given goal. Emergent behavior such as this is used by evolutionary algorithms and swarm intelligence.[74]

Learning

Main article: machine learning

Machine learning[75] has been central to AI research from the beginning.[76] Unsupervised learning is the ability to find patterns in a stream of input. Supervised learning includes both classification (be able to determine what category something belongs in, after seeing a number of examples of things from several categories) and regression (given a set of numerical input/output examples, discover a continuous function that would generate the outputs from the inputs). In reinforcement learning[77] the agent is rewarded for good responses and punished for bad ones. These can be analyzed in terms of decision theory, using concepts like utility. The mathematical analysis of machine learning algorithms and their performance is a branch of theoretical computer science known as computational learning theory.

Natural language processing

Main article: natural language processing

Natural language processing[78] gives machines the ability to read and understand the languages that the human beings speak. Many researchers hope that a sufficiently powerful natural language processing system would be able to acquire knowledge on its own, by reading the existing text available over the internet. Some straightforward applications of natural language processing include information retrieval (or text mining) and machine translation.[79]

Motion and manipulation

ASIMO uses sensors and intelligent algorithms to avoid obstacles and navigate stairs.

Main article: robotics

The field of robotics[80] is closely related to AI. Intelligence is required for robots to be able to handle such tasks as object manipulation[81] and navigation, with sub-problems of localization (knowing where you are), mapping (learning what is around you) and motion planning (figuring out how to get there).[82]

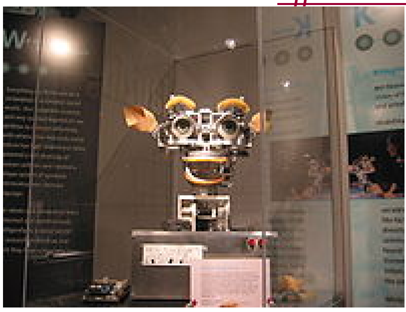
Perception

Main articles: machine perception, computer vision, and speech recognition

Machine perception[83] is the ability to use input from sensors (such as cameras, microphones, sonar and others more exotic) to deduce aspects of the world. Computer vision[84] is the ability to analyze visual input. A few selected subproblems are speech recognition,[85] facial recognition and object recognition.[86]

Social intelligence

Main article: affective computing



Kismet, a robot with rudimentary social skills.

Emotion and social skills play two roles for an intelligent agent:[87]

- It must be able to predict the actions of others, by understanding their motives and emotional states. (This involves elements of game theory, decision theory, as well as the ability to model human emotions and the perceptual skills to detect emotions.)
- For good human-computer interaction, an intelligent machine also needs to *display* emotions — at the very least it must appear polite and sensitive to the humans it interacts with. At best, it should appear to have normal emotions itself.

Creativity

Main article: computational creativity

A sub-field of AI addresses creativity both theoretically (from a philosophical and psychological perspective) and practically (via specific implementations of systems that generate outputs that can be considered creative).

General intelligence

Main articles: strong AI and AI-complete

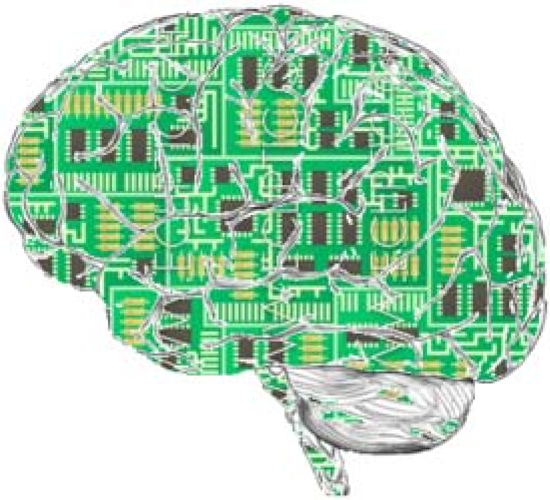
Most researchers hope that their work will eventually be incorporated into a machine with *general* intelligence (known as strong AI), combining all the skills above and exceeding human abilities at most or all of them.[11] A few believe that anthropomorphic features like artificial consciousness or an artificial brain may be required for such a project.[88]

Many of the problems above are considered AI-complete: to solve one problem, you must solve them all. For example, even a straightforward, specific task like machine translation requires that the machine follow the author's argument (reason), know what it's talking about (knowledge), and faithfully reproduce the author's intention (social intelligence). Machine translation, therefore, is believed to be AI-complete: it may require strong AI to be done as well as humans can do it.[89]

Approaches to AI

There is no established unifying theory or paradigm that guides AI research. Researchers disagree about many issues. A few of the most long standing questions that have remained unanswered are these: Can intelligence be reproduced using high-level symbols, similar to words and ideas? Or does it require "sub-symbolic" processing?[90] Should artificial intelligence simulate natural intelligence, by studying human psychology or animal neurobiology? Or is human biology as irrelevant to AI research as bird biology is to aeronautical engineering?[91] Can intelligent behavior be described using simple, elegant principles (such as logic or optimization)? Or does artificial intelligence necessarily require solving many unrelated problems?[92]

Cybernetics and brain simulation



The human brain provides inspiration for artificial intelligence researchers, however there is no consensus on how closely it should be simulated.

In the 40s and 50s, a number of researchers explored the connection between neurology, information theory, and cybernetics. Some of them built machines that used electronic networks to exhibit rudimentary intelligence, such as W. Grey Walter's turtles and the Johns Hopkins Beast. Many of these researchers gathered for meetings of the Teleological Society at Princeton University and the Ratio Club in England.[29]

Traditional symbolic AI

When access to digital computers became possible in the middle 1950s, AI research began to explore the possibility that human intelligence could be reduced to symbol manipulation. The research was centered in three institutions: CMU, Stanford and MIT, and each one developed its own style of research. John Haugeland named these approaches to AI "good old fashioned AI" or "GOFAI".[93]

Cognitive simulation

Economist Herbert Simon and Alan Newell studied human problem solving skills and attempted to formalize them, and their work laid the foundations of the field of artificial intelligence, as well as cognitive science, operations research and management science. Their research team performed psychological experiments to demonstrate the similarities between human problem solving and the programs (such as their "General Problem Solver") they were developing. This tradition, centered at Carnegie Mellon University would eventually culminate in the development of the Soar architecture in the middle 80s.[94][95]

Logical AI
: Unlike Newell and Simon, John McCarthy felt that machines did not need to simulate human thought, but should instead try to find the essence of abstract reasoning and problem solving, regardless of whether people used the same algorithms.[96] His laboratory at Stanford (SAIL) focused on using formal logic to solve a wide variety of problems, including knowledge representation, planning and learning.[97] Logic was also focus of the work at the University of Edinburgh and elsewhere in Europe which led to the development of the programming language Prolog and the science of logic programming.[98]

"Scruffy" symbolic AI
: Researchers at MIT (such as Marvin Minsky and Seymour Papert) found that solving difficult problems in vision and natural language processing required ad-hoc solutions – they argued that there was no simple and general principle (like logic) that would capture all the aspects of intelligent behavior. Roger Schank described their "anti-logic" approaches as "scruffy" (as opposed to the "neat" paradigms at CMU and Stanford).[99][92] Commonsense knowledge bases (such as Doug Lenat's Cyc) are an example of "scruffy" AI, since they must be built by hand, one complicated concept at a time.[100]

Knowledge based AI
: When computers with large memories became available around 1970, researchers from all three traditions began to build knowledge into AI applications.[101] This "knowledge revolution" led to the development and deployment of expert systems (introduced by Edward Feigenbaum), the first truly successful form of AI software.[38] The knowledge revolution was also driven by the realization that truly enormous amounts of knowledge would be required by many simple AI applications.

Sub-symbolic AI

During the 1960s, symbolic approaches had achieved great success at simulating high-level thinking in small demonstration programs. Approaches based on cybernetics or neural networks were abandoned or pushed into the background.[102] By the 1980s, however, progress in symbolic AI seemed to stall and many believed that symbolic systems would never be able to imitate all the processes of human cognition, especially perception, robotics, learning and pattern recognition. A number of researchers began to look into "sub-symbolic" approaches to specific AI problems.[90]

Bottom-up, embodied, situated, behavior-based or nouvelle AI
: Researchers from the related field of robotics, such as Rodney Brooks, rejected symbolic AI and focussed on the basic engineering problems that would allow robots to move and survive.[103] Their work revived the non-symbolic viewpoint of the early cybernetics researchers of the 50s and reintroduced the use of control theory in AI. These approaches are also conceptually related to the embodied mind thesis.

Computational Intelligence
: Interest in neural networks and "connectionism" was revived by David Rumelhart and others in the middle 1980s.[104] These and other sub-symbolic approaches, such as

fuzzy systems and evolutionary computation, are now studied collectively by the emerging discipline of computational intelligence.[105]

Formalisation

In the 1990s, AI researchers developed sophisticated mathematical tools to solve specific subproblems. These tools are truly scientific, in the sense that their results are both measurable and verifiable, and they have been responsible for many of AI's recent successes. The shared mathematical language has also permitted a high level of collaboration with more established fields (like mathematics, economics or operations research). Russell & Norvig (2003) describe this movement as nothing less than a "revolution" and "the victory of the neats."[42]

Intelligent agent paradigm

The "intelligent agent" paradigm became widely accepted during the 1990s.[106] An intelligent agent is a system that perceives its environment and takes actions which maximizes its chances of success. The simplest intelligent agents are programs that solve specific problems. The most complicated intelligent agents are rational, thinking human beings.[107] The paradigm gives researchers license to study isolated problems and find solutions that are both verifiable and useful, without agreeing on one single approach. An agent that solves a specific problem can use any approach that works — some agents are symbolic and logical, some are sub-symbolic neural networks and others may use new approaches. The paradigm also gives researchers a common language to communicate with other fields—such as decision theory and economics—that also use concepts of abstract agents.

Integrating the approaches

An agent architecture or cognitive architecture allows researchers to build more versatile and intelligent systems out of interacting intelligent agents in a multi-agent system.[108] A system with both symbolic and sub-symbolic components is a hybrid intelligent system, and the study of such systems is artificial intelligence systems integration. A hierarchical control system provides a bridge between sub-symbolic AI at its lowest, reactive levels and traditional symbolic AI at its highest levels, where relaxed time constraints permit planning and world modelling.[109] Rodney Brooks' subsumption architecture was an early proposal for such a hierarchical system.

Tools of AI research

In the course of 50 years of research, AI has developed a large number of tools to solve the most difficult problems in computer science. A few of the most general of these methods are discussed below.

Search and optimization

Main articles: search algorithm, optimization (mathematics), and evolutionary computation

Many problems in AI can be solved in theory by intelligently searching through many possible solutions:[110] Reasoning can be reduced to performing a search. For example, logical proof can be viewed as searching for a path that leads from premises to conclusions, where each step is the application of an inference rule.[111] Planning algorithms search through trees of goals and subgoals, attempting to find a path to a target goal, a process called means-ends analysis.[112] Robotics algorithms for moving limbs and grasping objects use local searches in configuration space.[81] Many learning algorithms use search algorithms based on optimization.

Simple exhaustive searches[113] are rarely sufficient for most real world problems: the search space (the number of places to search) quickly grows to astronomical numbers. The result is a search that is too slow or never completes. The solution, for many problems, is to use "heuristics" or "rules of thumb" that eliminate choices that are unlikely to lead to the goal (called "pruning the search tree"). Heuristics supply the program with a "best guess" for what path the solution lies on.[114]

A very different kind of search came to prominence in the 1990s, based on the mathematical theory of optimization. For many problems, it is possible to begin the search with some form of a guess and then refine the guess incrementally until no more refinements can be made. These algorithms can be visualized as blind hill climbing: we begin the search at a random point on the landscape, and then, by jumps or steps, we keep moving our guess uphill, until we reach the top. Other optimization algorithms are simulated annealing, beam search and random optimization.[115]

Evolutionary computation uses a form of optimization search. For example, they may begin with a population of organisms (the guesses) and then allow them to mutate and recombine, selecting only the fittest to survive each generation (refining the guesses). Forms of evolutionary computation include swarm intelligence algorithms (such as ant colony or particle swarm optimization)[116] and evolutionary algorithms (such as genetic algorithms[117] and genetic programming[118][119]).

Logic

Main articles: logic programming and automated reasoning

Logic[120] was introduced into AI research by John McCarthy in his 1958 Advice Taker proposal. The most important technical development was J. Alan Robinson's discovery of the resolution and unification algorithm for logical deduction in 1963. This procedure is simple, complete and entirely algorithmic, and can easily be performed by digital computers.[121]

However, a naive implementation of the algorithm quickly leads to a combinatorial explosion or an infinite loop. In 1974, Robert Kowalski suggested representing logical expressions as Horn clauses (statements in the form of rules: "if *p* then *q*"), which reduced logical deduction to backward chaining or forward chaining. This greatly alleviated (but did not eliminate) the problem.[111][122]

Logic is used for knowledge representation and problem solving, but it can be applied to other problems as well. For example, the satplan algorithm uses logic for planning,[123] and inductive logic programming is a method for learning.[124] There are several different forms of logic used in AI research.

- Propositional or sentential logic[125] is the logic of statements which can be true or false.
- First-order logic[126] also allows the use of quantifiers and predicates, and can express facts about objects, their properties, and their relations with each other.
- Fuzzy logic, a version of first-order logic which allows the truth of a statement to be represented as a value between 0 and 1, rather than simply True (1) or False (0). Fuzzy systems can be used for uncertain reasoning and have been widely used in modern industrial and consumer product control systems.[127]
- Default logics, non-monotonic logics and circumscription are forms of logic designed to help with default reasoning and the qualification problem.[65]
- Several extensions of logic have been designed to handle specific domains of knowledge, such as: description logics;[59] situation calculus, event calculus and fluent calculus (for representing events and time);[60] causal calculus;[61] belief calculus; and modal logics.[62]

Probabilistic methods for uncertain reasoning

Main articles: Bayesian network, hidden Markov model, Kalman filter, decision theory, and utility theory

Many problems in AI (in reasoning, planning, learning, perception and robotics) require the agent to operate with incomplete or uncertain information. Starting in the late 80s and early 90s, Judea Pearl and others championed the use of methods drawn from probability theory and economics to devise a number of powerful tools to solve these problems.[128][129]

Bayesian networks[130] are a very general tool that can be used for a large number of problems: reasoning (using the Bayesian inference algorithm),[131] learning (using the expectation-maximization algorithm),[132] planning (using decision networks)[133] and perception (using dynamic Bayesian networks).[134]

Probabilistic algorithms can also be used for filtering, prediction, smoothing and finding explanations for streams of data, helping perception systems to analyze processes that occur over time[135] (e.g., hidden Markov models[136] and Kalman filters[137]).

A key concept from the science of economics is "utility": a measure of how valuable something is to an intelligent agent. Precise mathematical tools have been developed that analyze how an agent can make choices and plan, using decision theory, decision analysis,[138] information value theory.[71] These tools include models such as Markov decision processes,[139] dynamic decision networks,[139] game theory and mechanism design[140]

Classifiers and statistical learning methods

Main articles: classifier (mathematics), statistical classification, and machine learning

The simplest AI applications can be divided into two types: classifiers ("if shiny then diamond") and controllers ("if shiny then pick up"). Controllers do however also classify conditions before inferring actions, and therefore classification forms a central part of many AI systems.

Classifiers[141] are functions that use pattern matching to determine a closest match. They can be tuned according to examples, making them very attractive for use in AI. These examples are known as observations or patterns. In supervised learning, each pattern belongs to a certain predefined class. A class can be seen as a decision that has to be made. All the observations combined with their class labels are known as a data set.

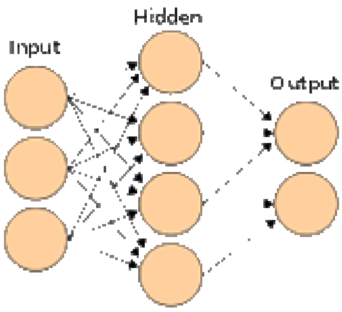
When a new observation is received, that observation is classified based on previous experience. A classifier can be trained in various ways; there are many statistical and machine learning approaches.

A wide range of classifiers are available, each with its strengths and weaknesses. Classifier performance depends greatly on the characteristics of the data to be classified. There is no single classifier that works best on all given problems; this is also referred to as the "no free lunch" theorem. Various empirical tests have been performed to compare classifier performance and to find the characteristics of data that determine classifier performance. Determining a suitable classifier for a given problem is however still more an art than science.

The most widely used classifiers are the neural network,[142] kernel methods such as the support vector machine,[143] k-nearest neighbor algorithm,[144] Gaussian mixture model,[145] naive Bayes classifier,[146] and decision tree.[147] The performance of these classifiers have been compared over a wide range of classification tasks[148] in order to find data characteristics that determine classifier performance.

Neural networks

Main articles: neural networks and connectionism



A neural network is an interconnected group of nodes, akin to the vast network of neurons in the human brain.

The study of artificial neural networks[142] began in the decade before the field AI research was founded. In the 1960s Frank Rosenblatt developed an important early version, the perceptron.[149] Paul Werbos developed the backpropagation algorithm for multilayer perceptrons in 1974,[150] which led to a renaissance in neural network research and connectionism in general in the middle 1980s. The Hopfield net, a form of attractor network, was first described by John Hopfield in 1982.

Common network architectures which have been developed include the feedforward neural network, the radial basis network, the Kohonen self-organizing map and various recurrent neural networks.[*citation needed*] Neural networks are applied to the problem of learning, using such techniques as Hebbian learning, competitive learning[151] and the relatively new architectures of Hierarchical Temporal Memory and Deep Belief Networks.

Control theory

Main article: intelligent control

Control theory, the grandchild of cybernetics, has many important applications, especially in robotics.[152]

Specialized languages

AI researchers have developed several specialized languages for AI research:

- IPL[153] includes features intended to support programs that could perform general problem solving, including lists, associations, schemas (frames), dynamic memory allocation, data types, recursion, associative retrieval, functions as arguments, generators (streams), and cooperative multitasking.

- Lisp[154][155] is a practical mathematical notation for computer programs based on lambda calculus. Linked lists are one of Lisp languages' major data structures, and Lisp source code is itself made up of lists. As a result, Lisp programs can manipulate source code as a data structure, giving rise to the macro systems that allow programmers to create new syntax or even new domain-specific programming languages embedded in Lisp. There are many dialects of Lisp in use today.
- Prolog[156][122] is a declarative language where programs are expressed in terms of relations, and execution occurs by running *queries* over these relations. Prolog is particularly useful for symbolic reasoning, database and language parsing applications. Prolog is widely used in AI today.
- STRIPS is a language for expressing automated planning problem instances. It expresses an initial state, the goal states, and a set of actions. For each action preconditions (what must be established before the action is performed) and postconditions (what is established after the action is performed) are specified.
- Planner is a hybrid between procedural and logical languages. It gives a procedural interpretation to logical sentences where implications are interpreted with pattern-directed inference.

AI applications are also often written in standard languages like C++ and languages designed for mathematics, such as Matlab and Lush.

Evaluating artificial intelligence

Main article: Progress in artificial intelligence

How can one determine if an agent is intelligent? In 1950, Alan Turing proposed a general procedure to test the intelligence of an agent now known as the Turing test. This procedure allows almost all the major problems of artificial intelligence to be tested. However, it is a very difficult challenge and at present all agents fail.

Artificial intelligence can also be evaluated on specific problems such as small problems in chemistry, hand-writing recognition and game-playing. Such tests have been termed subject matter expert Turing tests. Smaller problems provide more achievable goals and there are an ever-increasing number of positive results.

The broad classes of outcome for an AI test are:

- **optimal**: it is not possible to perform better
- **strong super-human**: performs better than all humans
- **super-human**: performs better than most humans
- **sub-human**: performs worse than most humans

For example, performance at checkers (draughts) is optimal,[157] performance at chess is super-human and nearing strong super-human,[158] and performance at many everyday tasks performed by humans is sub-human.

Competitions and prizes

Main article: Competitions and prizes in artificial intelligence

There are a number of competitions and prizes to promote research in artificial intelligence. The main areas promoted are: general machine intelligence, conversational behaviour, data-mining, driverless cars, robot soccer and games.

Applications of artificial intelligence

Main article: Applications of artificial intelligence

Artificial intelligence has successfully been used in a wide range of fields including medical diagnosis, stock trading, robot control, law, scientific discovery, video games and toys. Frequently, when a technique reaches mainstream use it is no longer considered artificial intelligence, sometimes described as the AI effect.[159] It may also become integrated into artificial life.

See also

- List of AI projects
- List of AI researchers
- List of emerging technologies
- List of basic artificial intelligence topics
- List of important AI publications

Notes

1. ^ Poole, Mackworth & Goebel 1998, p. 1 (who use the term "computational intelligence" as a synonym for artificial intelligence). Other textbooks that define AI this way include Nilsson (1998), and Russell & Norvig (2003) (who prefer the term "rational agent") and write "The whole-agent view is now widely accepted in the field" Template:Harvy
2. ^ This definition, in terms of goals, actions, perception and environment, is due to Russell & Norvig (2003). Other definitions also include knowledge and learning as additional criteria.
3. ^ Although there is some controversy on this point (see Crevier 1993, p. 50), McCarthy states unequivocally "I came up with the term" in a c|net interview. (See Getting Machines to Think Like Us.)
4. ^ See John McCarthy, What is Artificial Intelligence?
5. ^ *a* *b* Dartmouth proposal:
    - McCarthy et al. 1955
6. ^ *a* *b* *c* This is a central idea of Pamela McCorduck's *Machines That Think*. She writes: "I like to think of artificial intelligence as the scientific apotheosis of a veneralbe cultural tradition." *(McCorduck 2004, p. 34)* "Artificial intelligence in one form or another is an idea that has pervaded Western intellectual history, a dream in urgent need of being realized." *(McCorduck 2004, p. xviii)* "Our history is full of attempts—nutty, eerie, comical, earnest, legendary and real—to make artificial intelligences, to reproduce what is the essential us—bypassing the ordinary means. Back and forth between myth and reality, our imaginations supplying what our workshops couldn't, we have engaged for a long time in this odd form of self-reproduction." *(McCorduck 2004, p. 3)* She traces the desire back to its Hellenistic roots and calls it the urge to "forge the Gods." *(McCorduck 2004, p. 340-400)*
7. ^ The optimism referred to includes the predictions of early AI researchers (see optimism in the history of AI) as well as the ideas of modern transhumanists such as Ray Kurzweil.

8. ^ The "setbacks" referred to include the ALPAC report of 1966, the abandonment of perceptrons in 1970, the the Lighthill Report of 1973 and the collapse of the lisp machine market in 1987.
9. ^ [a] [b] Fractioning of AI into subfields:
    - McCorduck 2004, pp. 421-425
10. ^ [a] [b] This list of intelligent traits is based on the topics covered by the major AI textbooks, including:
    - Russell & Norvig 2003
    - Luger & Stubblefield 2004
    - Poole, Mackworth & Goebel 1998
    - Nilsson 1998.
11. ^ [a] [b] General intelligence (strong AI) is discussed in popular introductions to AI:
    - Kurzweil 1999 and Kurzweil 2005
12. ^ AI in Myth:
    - McCorduck 2004, p. 4-5
    - Russell & Norvig 2003, p. 939
13. ^ Sacred statues as artificial intelligence:
    - Crevier (1993, p. 1) (statue of Amun)
    - McCorduck (2004, pp. 6-9)
14. ^ These were the first machines to be believed to have true intelligence and consciousness. Hermes Trismegistus expressed the common belief that with these statues, craftsman had reproduced "the true nature of the gods", their *sensus* and *spiritus*. McCorduck makes the connection between sacred automatons and Mosaic law (developed around the same time), which expressly forbids the worship of robots *(McCorduck 2004, pp. 6-9)*
15. ^ Needham 1986, p. 53
16. ^ McCorduck 2004, p. 6
17. ^ A Thirteenth Century Programmable Robot
18. ^ McCorduck 2004, p. 17
19. ^ Takwin: O'Connor, Kathleen Malone. "*The alchemical creation of life (takwin) and other concepts of Genesis in medieval Islam*". University of Pennsylvania. Retrieved on 2007-01-10.
20. ^ Golem: McCorduck 2004, p. 15-16, Buchanan 2005, p. 50
21. ^ McCorduck 2004, p. 13-14
22. ^ McCorduck (2004, p. 190-25) discusses *Frankenstein* and identifies the key ethical issues as scientific hubris and the suffering of the monster, i.e. robot rights.
23. ^ Robot rights:
    - Russell & Norvig 2003, p. 964
    - Robots could demand legal rights
24. ^ See the Times Online, Human rights for robots? We're getting carried away
25. ^ Russell & Norvig (2003, p. 960-961)
26. ^ [a] [b] [c] Singularity, transhumanism:
    - Kurzweil 2005
    - Russell & Norvig 2003, p. 963
27. ^ Joseph Weizenbaum's critique of AI:
    - Weizenbaum 1976
    - Crevier 1993, pp. 132−144
    - McCorduck 2004, pp. 356-373
    - Russell & Norvig 2003, p. 961

Weizenbaum (the AI researcher who developed the first chatterbot program, ELIZA) argued in 1976 that the misuse of artificial intelligence has the potential to devalue human life.

28. ^ Quoted in McCorduck (2004, p. 401)
29. ^ a b AI's immediate precursors:
    - McCorduck 2004, pp. 51-107
    - Crevier 1993, pp. 27-32
    - Russell & Norvig 2003, pp. 15,940
    - Moravec 1988, p. 3

    Among the researchers who laid the foundations of the theory of computation, cybernetics, information theory and neural networks were Alan Turing, John Von Neumann, Norbert Weiner, Claude Shannon, Warren McCullough, Walter Pitts and Donald Hebb

30. ^ Dartmouth conference:
    - McCorduck, pp. 111-136
    - Crevier 1993, pp. 47-49
    - Russell & Norvig 2003, p. 17
    - NRC 1999, pp. 200-201
31. ^ Russell and Norvig write "it was astonishing whenever a computer did anything kind of smartish." Russell & Norvig 2003, p. 18
32. ^ "Golden years" of AI (successful symbolic reasoning programs 1956-1973):
    - McCorduck, pp. 243-252
    - Crevier 1993, pp. 52-107
    - Moravec 1988, p. 9
    - Russell & Norvig 2003, p. 18-21

    The programs described are Daniel Bobrow's STUDENT, Newell and Simon's Logic Theorist and Terry Winograd's SHRDLU.

33. ^ DARPA pours money into undirected pure research into AI during the 1960s:
    - McCorduck 2005, pp. 131
    - Crevier 1993, pp. 51, 64-65
    - NRC 1999, pp. 204-205
34. ^ Simon 1965, p. 96 quoted in Crevier 1993, p. 109
35. ^ Minsky 1967, p. 2 quoted in Crevier 1993, p. 109
36. ^ See History of artificial intelligence — the problems.
37. ^ First AI Winter:
    - Crevier 1993, pp. 115-117
    - Russell & Norvig 2003, p. 22
    - NRC 1999, pp. 212-213
    - Howe 1994
38. ^ a b Expert systems:
    - ACM 1998, I.2.1,
    - Russell & Norvig 2003, pp. 22−24
    - Luger & Stubblefield 2004, pp. 227-331,
    - Nilsson 1998, chpt. 17.4
    - McCorduck 2004, pp. 327-335, 434-435
    - Crevier 1993, pp. 145-62, 197−203

39. ^ Boom of the 1980s: rise of expert systems, Fifth Generation Project, Alvey, MCC, SCI:
    - McCorduck 2004, pp. 426-441
    - Crevier 1993, pp. 161-162,197-203, 211, 240
    - Russell & Norvig 2003, p. 24
    - NRC 1999, pp. 210-211
40. ^ Second AI Winter:
    - McCorduck 2004, pp. 430-435
    - Crevier 1993, pp. 209-210
    - NRC 1999, pp. 214-216
41. ^ AI applications widely used behind the scenes:
    - Russell & Norvig 2003, p. 28
    - Kurzweil 2005, p. 265
    - NRC 1999, pp. 216-222
42. ^ a b Formal methods are now preferred ("Victory of the neats"):
    - Russell & Norvig 2003, pp. 25-26
    - McCorduck 2004, pp. 486-487
43. ^ All of these positions below are mentioned in standard discussions of the subject, such as:
    - Russell & Norvig 2003, pp. 947-960
    - Fearn 2007, pp. 38-55
44. ^ Philosophical implications of the Turing test:
    - Turing 1950,
    - Haugeland 1985, pp. 6-9,
    - Crevier 1993, p. 24,
    - Russell & Norvig 2003, pp. 2-3 and 948
45. ^ The physical symbol systems hypothesis:
    - Newell & Simon 1976, p. 116
    - Russell & Norvig 2003, p. 18
46. ^ Dreyfus criticized the necessary condition of the physical symbol system hypothesis, which he called the "psychological assumption": "The mind can be viewed as a device operating on bits of information according to formal rules". *(Dreyfus 1992, p. 156)*
47. ^ Dreyfus' Critique of AI:
    - Dreyfus 1972,
    - Dreyfus & Dreyfus 1986,
    - Russell & Norvig 2003, pp. 950-952,
    - Crevier 1993, pp. 120-132 and
48. ^ This is a paraphrase of the important implication of Gödel's theorems.
49. ^ The Mathematical Objection:
    - Russell & Norvig 2003, p. 949
    - McCorduck 2004, p. 448-449

    Refuting Mathematical Objection:
    - Turing 1950 under "(2) The Mathematical Objection"
    - Hofstadter 1979,

    Making the Mathematical Objection:
    - Lucas 1961,
    - Penrose 1989.

Background:

- Gödel 1931, Church 1936, Kleene 1935, Turing 1937,

50. ^ This version is from Searle (1999), and is also quoted in Dennett 1991, p. 435. Searle's original formulation was "The appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states." *(Searle 1980, p. 1)*. Strong AI is defined similarly by Russell & Norvig (2003, p. 947): "The assertion that machines could possibly act intelligently (or, perhaps better, act as if they were intelligent) is called the 'weak AI' hypothesis by philosophers, and the assertion that machines that do so are actually thinking (as opposed to simulating thinking) is called the 'strong AI' hypothesis."
51. ^ Searle's Chinese Room argument:
    - Searle 1980, Searle 1991
    - Russell & Norvig 2003, pp. 958-960
    - McCorduck 2004, pp. 443-445
    - Crevier 1993, pp. 269-271
52. ^ Artificial brain:
    - Moravec 1988
    - Kurzweil 2005, p. 262
    - Russell Norvig, p. 957
    - Crevier 1993, pp. 271 and 279

    The most extreme form of this argument (the brain replacement scenario) was put forward by Clark Glymour in the mid-70s and was touched on by Zenon Pylyshyn and John Searle in 1980. Daniel Dennett sees human consciousness as multiple functional thought patterns; see "Consciousness Explained."

53. ^ Problem solving, puzzle solving, game playing and deduction:
    - Russell & Norvig 2003, chpt. 3-9,
    - Poole et al. chpt. 2,3,7,9,
    - Luger & Stubblefield 2004, chpt. 3,4,6,8,
    - Nilsson, chpt. 7-12.
54. ^ Uncertain reasoning:
    - Russell & Norvig 2003, pp. 452-644,
    - Poole, Mackworth & Goebel 1998, pp. 345-395,
    - Luger & Stubblefield 2004, pp. 333-381,
    - Nilsson 1998, chpt. 19
55. ^ Intractability and efficiency and the combinatorial explosion:
    - Russell & Norvig 2003, pp. 9, 21-22
56. ^ Cognitive science has provided several famous examples:
    - Wason (1966) showed that people do poorly on completely abstract problems, but if the problem is restated to allow the use of intuitive social intelligence, performance dramatically improves. (See Wason selection task)
    - Tversky, Slovic & Kahnemann (1982) have shown that people are terrible at elementary problems that involve uncertain reasoning. (See list of cognitive biases for several examples).
    - Lakoff & Núñez (2000) have controversially argued that even our skills at mathematics depend on knowledge and skills that come from "the body", i.e. sensorimotor and perceptual skills. (See Where Mathematics Comes From)
    -

57. ^ Knowledge representation:
    - ACM 1998, I.2.4,
    - Russell & Norvig 2003, pp. 320-363,
    - Poole, Mackworth & Goebel 1998, pp. 23-46, 69-81, 169-196, 235-277, 281-298, 319-345,
    - Luger & Stubblefield 2004, pp. 227-243,
    - Nilsson 1998, chpt. 18
58. ^ Knowledge engineering:
    - Russell & Norvig 2003, pp. 260-266,
    - Poole, Mackworth & Goebel 1998, pp. 199-233,
    - Nilsson 1998, chpt. ~17.1-17.4
59. ^ a b Representing categories and relations: Semantic networks, description logics, inheritance (including frames and scripts):
    - Russell & Norvig 2003, pp. 349-354,
    - Poole, Mackworth & Goebel 1998, pp. 174-177,
    - Luger & Stubblefield 2004, pp. 248-258,
    - Nilsson 1998, chpt. 18.3
60. ^ a b Representing events and time:Situation calculus, event calculus, fluent calculus (including solving the frame problem):
    - Russell & Norvig 2003, pp. 328-341,
    - Poole, Mackworth & Goebel 1998, pp. 281-298,
    - Nilsson 1998, chpt. 18.2
61. ^ a b Causal calculus:
    - Poole, Mackworth & Goebel 1998, pp. 335-337
62. ^ a b Representing knowledge about knowledge: Belief calculus, modal logics:
    - Russell & Norvig 2003, pp. 341-344,
    - Poole, Mackworth & Goebel 1998, pp. 275-277
63. ^ Ontology:
    - Russell & Norvig 2003, pp. 320-328
64. ^ McCarthy & Hayes 1969. While McCarthy was primarily concerned with issues in the logical representation of actions, Russell & Norvig 2003 apply the term to the more general issue of default reasoning in the vast network of assumptions underlying all our commonsense knowledge.
65. ^ a b Default reasoning and default logic, non-monotonic logics, circumscription, closed world assumption, abduction (Poole *et al.* places abduction under "default reasoning". Luger *et al.* places this under "uncertain reasoning"):
    - Russell & Norvig 2003, pp. 354-360,
    - Poole, Mackworth & Goebel 1998, pp. 248-256, 323-335,
    - Luger & Stubblefield 2004, pp. 335-363,
    - Nilsson 1998, ~18.3.3
66. ^ Breadth of commonsense knowledge:
    - Russell & Norvig 2003, p. 21,
    - Crevier 1993, pp. 113-114,
    - Moravec 1988, p. 13,
    - Lenat & Guha 1989 (Introduction)
67. ^ Dreyfus & Dreyfus 1986
68. ^ Gladwell 2005
69. ^ Expert knowledge as embodied intuition:

- Dreyfus & Dreyfus 1986 (Hubert Dreyfus is a philosopher and critic of AI who was among the first to argue that most useful human knowledge was encoded sub-symbolically.)
- Gladwell 2005 (Gladwell's *Blink* is a popular introduction to sub-symbolic reasoning and knowledge.)
- Hawkins 2005 (Hawkins argues that sub-symbolic knowledge should be the primary focus of AI research.)

70. ^ Planning:
    - ACM 1998, ~I.2.8,
    - Russell & Norvig 2003, pp. 375-459,
    - Poole, Mackworth & Goebel 1998, pp. 281-316,
    - Luger & Stubblefield 2004, pp. 314-329,
    - Nilsson 1998, chpt. 10.1-2, 22
71. ^ *a* *b* Information value theory:
    - Russell & Norvig 2003, pp. 600-604
72. ^ Classical planning:
    - Russell & Norvig 2003, pp. 375-430,
    - Poole, Mackworth & Goebel 1998, pp. 281-315,
    - Luger & Stubblefield 2004, pp. 314-329,
    - Nilsson 1998, chpt. 10.1-2, 22
73. ^ Planning and acting in non-deterministic domains: conditional planning, execution monitoring, replanning and continuous planning:
    - Russell & Norvig 2003, pp. 430-449
74. ^ Multi-agent planning and emergent behavior:
    - Russell & Norvig 2003, pp. 449-455
75. ^ Learning:
    - ACM 1998, I.2.6,
    - Russell & Norvig 2003, pp. 649-788,
    - Poole, Mackworth & Goebel 1998, pp. 397-438,
    - Luger & Stubblefield 2004, pp. 385-542,
    - Nilsson 1998, chpt. 3.3 , 10.3, 17.5, 20
76. ^ Alan Turing discussed the centrality of learning as early as 1950, in his classic paper Computing Machinery and Intelligence. *(Turing 1950)*
77. ^ Reinforcement learning:
    - Russell & Norvig 2003, pp. 763-788
    - Luger & Stubblefield 2004, pp. 442-449
78. ^ Natural language processing:
    - ACM 1998, I.2.7
    - Russell & Norvig 2003, pp. 790-831
    - Poole, Mackworth & Goebel 1998, pp. 91-104
    - Luger & Stubblefield 2004, pp. 591-632
79. ^ Applications of natural language processing, including information retrieval (i.e. text mining) and machine translation:
    - Russell & Norvig 2003, pp. 840-857,
    - Luger & Stubblefield 2004, pp. 623-630
80. ^ Robotics:
    - ACM 1998, I.2.9,
    - Russell & Norvig 2003, pp. 901-942,
    - Poole, Mackworth & Goebel 1998, pp. 443-460
81. ^ *a* *b* Moving and configuration space:

- Russell & Norvig 2003, pp. 916-932

82. ^ Robotic mapping (localization, etc):
    - Russell & Norvig 2003, pp. 908-915
83. ^ Machine perception: Russell & Norvig 2003, pp. 537-581, 863-898, Nilsson 1998, ~chpt. 6
84. ^ Computer vision:
    - ACM 1998, I.2.10
    - Russell & Norvig 2003, pp. 863-898
    - Nilsson 1998, chpt. 6
85. ^ Speech recognition:
    - ACM 1998, ~I.2.7
    - Russell & Norvig 2003, pp. 568-578
86. ^ Object recognition:
    - Russell & Norvig 2003, pp. 885-892
87. ^ Emotion and affective computing:
    - Minsky 2007
    - Picard 1997
88. ^ Gerald Edelman, Igor Aleksander and others have both argued that artificial consciousness is required for strong AI. CITATION IN PROGRESS Ray Kurzweil, Jeff Hawkins and others have argued that strong AI requires a simulation of the operation of the human brain. CITATION IN PROGRESS
89. ^ AI complete:
    - Shapiro 1992, p. 9
90. ^ [a] [b] Nilsson (1998, p. 7) characterizes newer approaches to AI as "sub-symbolic".
91. ^ The analogy with aeronautical engineering is due to Russell & Norvig (2003, p. 3).
92. ^ [a] [b] Neats vs. scruffies:
    - McCorduck 2004, pp. 421-424, 486-489
    - Crevier 1993, pp. 168
93. ^ Haugeland 1985, pp. 112-117
94. ^ Cognitive simulation, Newell and Simon, AI at CMU (then called Carnegie Tech):
    - McCorduck 2004, pp. 139-179, 245-250, 322-323 (EPAM)
    - Crevier 2004, pp. 145-149
95. ^ Soar (history):
    - McCorduck 2004, pp. 450-451
    - Crevier 1993, pp. 258-263
96. ^ McCarthy's opposition to "cognitive simulation":
    - Science at Google Books
    - McCarthy's presentation at AI@50
97. ^ McCarthy and AI research at SAIL and SRI:
    - McCorduck 2004, pp. 251-259
    - Crevier 1993, pp. Check
98. ^ AI research at Edinburgh and in France, birth of Prolog:
    - Crevier 1993, pp. 193-196
    - Howe 1994
99. ^ AI at MIT under Marvin Minsky in the 1960s :
    - McCorduck 2004, pp. 259-305
    - Crevier 1993, pp. 83-102, 163-176
    - Russell & Norvig 2003, p. 19
100. ^ Cyc:
    - McCorduck 2004, p. 489, who calls it "a determinedly scruffy enterprise"

- Crevier 1993, pp. 239−243
- Russell & Norvig 2003, p. 363−365
- Lenat & Guha 1989

101. ^ Knowledge revolution:
- McCorduck 2004, pp. 266-276, 298-300, 314, 421
- Russell & Norvig 2003, pp. 22-23

102. ^ The most dramatic case of sub-symbolic AI being pushed into the background was the devastating critique of perceptrons by Marvin Minsky and Seymour Papert in 1969. See History of AI, AI winter, or Frank Rosenblatt.
103. ^ Embodied approaches to AI:
- McCorduck 2004, pp. 454-462
- Brooks 1990
- Moravec 1988

104. ^ Revival of connectionism:
- Crevier 1993, pp. 214-215
- Russell & Norvig 2003, p. 25

105. ^ See IEEE Computational Intelligence Society
106. ^ "The whole-agent view is now widely accepted in the field" Russell & Norvig 2003, p. 55.
107. ^ The intelligent agent paradigm:
- Russell & Norvig 2003, pp. 27, 32-58, 968-972,
- Poole, Mackworth & Goebel 1998, pp. 7-21,
- Luger & Stubblefield 2004, pp. 235-240

108. ^ Agent architectures, hybrid intelligent systems:
- Russell & Norvig (1998, pp. 27, 932, 970-972)
- Nilsson (1998, chpt. 25)

109. ^ Albus, J. S. 4-D/RCS reference model architecture for unmanned ground vehicles. In G Gerhart, R Gunderson, and C Shoemaker, editors, Proceedings of the SPIE AeroSense Session on Unmanned Ground Vehicle Technology, volume 3693, pages 11—20
110. ^ Search algorithms:
- Russell & Norvig 2003, pp. 59-189
- Poole, Mackworth & Goebel 1998, pp. 113-163
- Luger & Stubblefield 2004, pp. 79-164, 193-219
- Nilsson 1998, chpt. 7-12

111. ^ a b Forward chaining, backward chaining, Horn clauses, and logical deduction as search:
- Russell & Norvig 2003, pp. 217-225, 280-294
- Poole, Mackworth & Goebel 1998, pp. ~46-52
- Luger & Stubblefield 2004, pp. 62-73
- Nilsson 1998, chpt. 4.2, 7.2

112. ^ State space search and planning:
- Russell & Norvig 2003, pp. 382-387
- Poole, Mackworth & Goebel 1998, pp. 298-305
- Nilsson 1998, chpt. 10.1-2

113. ^ Uninformed searches (breadth first search, depth first search and general state space search):
- Russell & Norvig 2003, pp. 59-93
- Poole, Mackworth & Goebel 1998, pp. 113-132
- Luger & Stubblefield 2004, pp. 79-121

- Nilsson 1998, chpt. 8

114. ^ Heuristic or informed searches (e.g., greedy best first and A*):
- Russell & Norvig 2003, pp. 94-109,
- Poole, Mackworth & Goebel 1998, pp. pp. 132-147,
- Luger & Stubblefield 2004, pp. 133-150,
- Nilsson 1998, chpt. 9

115. ^ Optimization searches:
- Russell & Norvig 2003, pp. 110-116,120-129
- Poole, Mackworth & Goebel 1998, pp. 56-163
- Luger & Stubblefield 2004, pp. 127-133

116. ^ Artificial life and society based learning:
- Luger & Stubblefield 2004, pp. 530-541

117. ^ Genetic algorithms for learning:
- Luger & Stubblefield 2004, pp. 509-530,
- Nilsson 1998, chpt. 4.2.

See also: Holland, John H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press. ISBN 0262581116.

118. ^ Koza, John R. (1992). *Genetic Programming*. MIT Press.

119. ^ Poli, R., Langdon, W. B., McPhee, N. F. (2008). *A Field Guide to Genetic Programming*. Lulu.com, freely available from http://www.gp-field-guide.org.uk/. ISBN 978-1-4092-0073-4.

120. ^ Logic:
- ACM 1998, ~I.2.3,
- Russell & Norvig 2003, pp. 194-310,
- Luger & Stubblefield 2004, pp. 35-77,
- Nilsson 1998, chpt. 13-16

121. ^ Resolution and unification:
- Russell & Norvig 2003, pp. 213-217, 275-280, 295-306,
- Poole, Mackworth & Goebel 1998, pp. 56-58,
- Luger & Stubblefield 2004, pp. 554-575,
- Nilsson 1998, chpt. 14 & 16

122. ^ [a] [b] History of logic programming:
- Crevier 1993, pp. 190-196.
- Howe 1994

Advice Taker:

- McCorduck 2004, p. 51,
- Russell & Norvig 2003, pp. 19

123. ^ Satplan:
- Russell & Norvig 2003, pp. 402-407,
- Poole, Mackworth & Goebel 1998, pp. 300-301,
- Nilsson 1998, chpt. 21

124. ^ Explanation based learning, relevance based learning, inductive logic programming, case based reasoning:
- Russell & Norvig 2003, pp. 678-710,
- Poole, Mackworth & Goebel 1998, pp. 414-416,
- Luger & Stubblefield 2004, pp. ~422-442,

- Nilsson 1998, chpt. 10.3, 17.5

125. ^ Propositional logic:
    - Russell & Norvig 2003, pp. 204-233,
    - Luger & Stubblefield 2004, pp. 45-50
    - Nilsson 1998, chpt. 13
126. ^ First-order logic and features such as equality:
    - ACM 1998, ~I.2.4,
    - Russell & Norvig 2003, pp. 240-310,
    - Poole, Mackworth & Goebel 1998, pp. 268-275,
    - Luger & Stubblefield 2004, pp. 50-62,
    - Nilsson 1998, chpt. 15
127. ^ Fuzzy logic:
    - Russell & Norvig 2003, pp. 526-527
128. ^ Judea Pearl's contribution to AI:
    - Russell & Norvig 2003, pp. 25-26
129. ^ Stochastic methods for uncertain reasoning:
    - ACM 1998, ~I.2.3,
    - Russell & Norvig 2003, pp. 462-644,
    - Poole, Mackworth & Goebel 1998, pp. 345-395,
    - Luger & Stubblefield 2004, pp. 165-191, 333-381,
    - Nilsson 1998, chpt. 19
130. ^ Bayesian networks:
    - Russell & Norvig 2003, pp. 492-523,
    - Poole, Mackworth & Goebel 1998, pp. 361-381,
    - Luger & Stubblefield 2004, pp. ~182-190, ~363-379,
    - Nilsson 1998, chpt. 19.3-4
131. ^ Bayesian inference algorithm:
    - Russell & Norvig 2003, pp. 504-519,
    - Poole, Mackworth & Goebel 1998, pp. 361-381,
    - Luger & Stubblefield 2004, pp. ~363-379,
    - Nilsson 1998, chpt. 19.4 & 7
132. ^ Bayesian learning and the expectation-maximization algorithm:
    - Russell & Norvig 2003, pp. 712-724,
    - Poole, Mackworth & Goebel 1998, pp. 424-433,
    - Nilsson 1998, chpt. 20
133. ^ Bayesian decision networks:
    - Russell & Norvig 2003, pp. 597-600
134. ^ Dynamic Bayesian network:
    - Russell & Norvig 2003, pp. 551-557
135. ^ Stochastic temporal models: Russell & Norvig 2003, pp. 537-581
136. ^ Hidden Markov model:
    - Russell & Norvig 2003, pp. 549-551
137. ^ Kalman filter:
    - Russell & Norvig 2003, pp. 551-557
138. ^ decision theory and decision analysis:
    - Russell & Norvig 2003, pp. 584-597,
    - Poole, Mackworth & Goebel 1998, pp. 381-394
139. ^ [a] [b] Markov decision processes and dynamic decision networks:
    - Russell & Norvig 2003, pp. 613-631
140. ^ Game theory and mechanism design:

- Russell & Norvig 2003, pp. 631-643

141. ^ Statistical learning methods and classifiers:
    - Russell & Norvig 2003, pp. 712-754,
    - Luger & Stubblefield 2004, pp. 453-541
142. ^ a b Neural networks and connectionism:
    - Russell & Norvig 2003, pp. 736-748,
    - Poole, Mackworth & Goebel 1998, pp. 408-414,
    - Luger & Stubblefield 2004, pp. 453-505,
    - Nilsson 1998, chpt. 3
143. ^ Kernel methods:
    - Russell & Norvig 2003, pp. 749-752
144. ^ K-nearest neighbor algorithm:
    - Russell & Norvig 2003, pp. 733-736
145. ^ Gaussian mixture model:
    - Russell & Norvig 2003, pp. 725-727
146. ^ Naive Bayes classifier:
    - Russell & Norvig 2003, pp. 718
147. ^ Decision tree:
    - Russell & Norvig 2003, pp. 653-664,
    - Poole, Mackworth & Goebel 1998, pp. 403-408,
    - Luger & Stubblefield 2004, pp. 408-417
148. ^ van der Walt, Christiaan. "Data characteristics that determine classifier performance" (PDF). http://www.patternrecognition.co.za/publications/cvdwalt_data_characteristics_classifiers.pdf.
149. ^ Perceptrons:
    - Russell & Norvig 2003, pp. 740-743,
    - Luger & Stubblefield 2004, pp. 458-467
150. ^ Backpropagation:
    - Russell & Norvig 2003, pp. 744-748,
    - Luger & Stubblefield 2004, pp. 467-474,
    - Nilsson 1998, chpt. 3.3
151. ^ Competitive learning, Hebbian coincidence learning, Hopfield networks and attractor networks:
    - Luger & Stubblefield 2004, pp. 474-505.
152. ^ Control theory:
    - ACM 1998, ~I.2.8,
    - Russell & Norvig 2003, pp. 926-932
153. ^ Crevier 1993, p. 46-48
154. ^ Lisp:
    - Luger & Stubblefield 2004, pp. 723-821
155. ^ Crevier 1993, pp. 59-62, Russell & Norvig 2003, p. 18
156. ^ Prolog:
    - Poole, Mackworth & Goebel 1998, pp. 477-491,
    - Luger & Stubblefield 2004, pp. 641-676, 575-581
157. ^ Schaeffer, Jonathan (2007-07-19). "Checkers Is Solved". Science. http://www.sciencemag.org/cgi/content/abstract/1144079. Retrieved on 2007-07-20.
158. ^ Computer Chess#Computers versus humans
159. ^ "AI set to exceed human brain power" (web article). CNN.com. 2006-07-26. http://www.cnn.com/2006/TECH/science/07/24/ai.bostrom/. Retrieved on 2008-02-26.

References

Major AI textbooks

See also A.I. Textbook survey

- Luger, George; Stubblefield, William (2004), *Artificial Intelligence: Structures and Strategies for Complex Problem Solving* (5th ed.), The Benjamin/Cummings Publishing Company, Inc., ISBN 0-8053-4780-1, http://www.cs.unm.edu/~luger/ai-final/tocfull.html
- Nilsson, Nils (1998), *Artificial Intelligence: A New Synthesis*, Morgan Kaufmann Publishers, ISBN 978-1-55860-467-4
- Russell, Stuart J.; Norvig, Peter (2003), *Artificial Intelligence: A Modern Approach* (2nd ed.), Upper Saddle River, NJ: Prentice Hall, ISBN 0-13-790395-2, http://aima.cs.berkeley.edu/
- Poole, David; Mackworth, Alan; Goebel, Randy (1998), *Computational Intelligence: A Logical Approach*, New York: Oxford University Press, http://www.cs.ubc.ca/spider/poole/ci.html
- Winston, Patrick Henry (1984), *Artificial Intelligence*, Reading, Massachusetts: Addison-Wesley, ISBN 0201082594

History of AI

- Crevier, Daniel (1993), *AI: The Tumultuous Search for Artificial Intelligence*, New York, NY: BasicBooks, ISBN 0-465-02997-3
- McCorduck, Pamela (2004), *Machines Who Think* (2nd ed.), Natick, MA: A. K. Peters, Ltd., ISBN 1-56881-205-1, http://www.pamelamc.com/html/machines_who_think.html

Other sources

- ACM, (Association of Computing Machinery) (1998), *ACM Computing Classification System: Artificial intelligence*, http://www.acm.org/class/1998/I.2.html
- Brooks, Rodney (1990), "Elephants Don't Play Chess" (PDF), *Robotics and Autonomous Systems* **6**: 3–15, doi:10.1016/S0921-8890(05)80025-9, http://people.csail.mit.edu/brooks/papers/elephants.pdf, retrieved on 2007-08-30
- Buchanan, Bruce G. (Winter 2005), "A (Very) Brief History of Artificial Intelligence" (PDF), *AI Magazine*: 53–60, http://www.aaai.org/AITopics/assets/PDF/AIMag26-04-016.pdf, retrieved on 2007-08-30
- Dreyfus, Hubert (1972), *What Computers Can't Do*, New York: MIT Press, ISBN 0060110821
- Dreyfus, Hubert (1979), *What Computers* Still *Can't Do*, New York: MIT Press.
- Dreyfus, Hubert; Dreyfus, Stuart (1986), *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*, Oxford, UK: Blackwell.
- Gladwell, Malcolm (2005), *Blink*, New York: Little, Brown and Co., ISBN 0-316-17232-4.
- Haugeland, John (1985), *Artificial Intelligence: The Very Idea*, Cambridge, Mass.: MIT Press, ISBN 0-262-08153-9.
- Hawkins, Jeff; Blakeslee, Sandra (2004), *On Intelligence*, New York, NY: Owl Books, ISBN 0-8050-7853-3.
- Hofstadter, Douglas (1979), *Gödel, Escher, Bach: an Eternal Golden Braid*.

- Howe, J. (November 1994), *Artificial Intelligence at Edinburgh University: a Perspective*, http://www.inf.ed.ac.uk/about/AIhistory.html.
- Kahneman, Daniel; Slovic, D.; Tversky, Amos (1982), *Judgment under uncertainty: Heuristics and biases*, New York: Cambridge University Press.
- Kurzweil, Ray (1999), *The Age of Spiritual Machines*, Penguin Books, ISBN 0-670-88217-8
- Kurzweil, Ray (2005), *The Singularity is Near*, Penguin Books, ISBN 0-670-03384-7
- Lakoff, George (1987), *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*, University of Chicago Press., ISBN 0-226-46804-6
- Lakoff, George; Núñez, Rafael E. (2000), *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*, Basic Books, ISBN 0-465-03771-2.
- Lenat, Douglas; Guha, R. V. (1989), *Building Large Knowledge-Based Systems*, Addison-Wesley
- Lighthill, Professor Sir James (1973), "Artificial Intelligence: A General Survey", *Artificial Intelligence: a paper symposium*, Science Research Council
- Lucas, John (1961), "Minds, Machines and Gödel", in Anderson, A.R., *Minds and Machines*, http://users.ox.ac.uk/~jrlucas/Godel/mmg.html.
- McCarthy, John; Minsky, Marvin; Rochester, Nathan; Shannon, Claude (1955), *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html.
- McCarthy, John; Hayes, P. J. (1969), "Some philosophical problems from the standpoint of artificial intelligence", *Machine Intelligence* **4**: 463–502, http://www-formal.stanford.edu/jmc/mcchay69.html
- Minsky, Marvin (1967), *Computation: Finite and Infinite Machines*, Englewood Cliffs, N.J.: Prentice-Hall
- Minsky, Marvin (2006), *The Emotion Machine*, New York, NY: Simon & Schusterl, ISBN 0-7432-7663-9
- Moravec, Hans (1976), *The Role of Raw Power in Intelligence*, http://www.frc.ri.cmu.edu/users/hpm/project.archive/general.articles/1975/Raw.Power.html
- Moravec, Hans (1988), *Mind Children*, Harvard University Press
- NRC (1999), "Developments in Artificial Intelligence", *Funding a Revolution: Government Support for Computing Research*, National Academy Press
- Needham, Joseph (1986), *Science and Civilization in China: Volume 2*, Caves Books Ltd.
- Newell, Allen; Simon, H. A. (1963), "GPS: A Program that Simulates Human Thought", in Feigenbaum, E.A.; Feldman, J., *Computers and Thought*, McGraw-Hill
- Newell, Allen; Simon, H. A. (1976), "Computer Science as Empirical Inquiry: Symbols and Search", *Communications of the ACM*, **19**, http://www.rci.rutgers.edu/~cfs/472_html/AI_SEARCH/PSS/PSSH4.html
- Searle, John (1980), "Minds, Brains and Programs", *Behavioral and Brain Sciences* **3** (3): 417–457, http://www.bbsonline.org/documents/a/00/00/04/84/bbs00000484-00/bbs.searle2.html
- Searle, John (1999), *Mind, language and society*, New York, NY: Basic Books, ISBN 0465045219, OCLC 231867665 43689264
- Shapiro, Stuart C. (1992), "Artificial Intelligence", in Shapiro, Stuart C., *Encyclopedia of Artificial Intelligence* (2nd ed.), New York: John Wiley, pp. 54–57, http://www.cse.buffalo.edu/~shapiro/Papers/ai.ps.
- Simon, H. A. (1965), *The Shape of Automation for Men and Management*, New York: Harper & Row

- Turing, Alan (October 1950), "Computing Machinery and Intelligence", *Mind* **LIX** (236): 433–460, doi:10.1093/mind/LIX.236.433, ISSN 0026-4423, http://loebner.net/Prizef/TuringArticle.html, retrieved on 2008-08-18
- Wason, P. C.; Shapiro, D. (1966), "Reasoning", in Foss, B. M., *New horizons in psychology*, Harmondsworth: Penguin
- Weizenbaum, Joseph (1976), *Computer Power and Human Reason*, San Francisco: W.H. Freeman & Company, ISBN 0716704641

Further reading

- R. Sun & L. Bookman, (eds.), *Computational Architectures: Integrating Neural and Symbolic Processes*. Kluwer Academic Publishers, Needham, MA. 1994.
- Margaret Boden, Mind As Machine, Oxford University Press, 2006
- John Johnston, (2008) "The Allure of Machinic Life: Cybernetics, Artificial Life, and the New AI", MIT Press
- Michaela Ong, pioneer of Artificial Intelligence. "Origin of AI through cs191"