Artificial intelligence

From Wikipedia, the free encyclopedia

Artificial intelligence (**AI**) is intelligence exhibited by machines. In computer science, an ideal "intelligent" machine is a flexible rational agent that perceives its environment and takes actions that maximize its chance of success at some goal.[1] Colloquially, the term "artificial intelligence" is applied when a machine mimics "cognitive" functions that humans associate with other human minds, such as "learning" and "problem solving".[2] As machines become increasingly capable, facilities once thought to require intelligence are removed from the definition. For example, optical character recognition is no longer perceived as an exemplar of "artificial intelligence", having become a routine technology.[3] Capabilities currently classified as AI include successfully understanding human speech,[4] competing at a high level in strategic game systems (such as Chess and Go[5]), self-driving cars, and interpreting complex data. AI is also considered a danger to humanity if it progresses unabatedly.[6] AI research is divided into subfields[7] that focus on specific problems or on specific approaches or on the use of a particular tool or towards satisfying particular applications.

The central problems (or goals) of AI research include reasoning, knowledge, planning, learning, natural language processing (communication), perception and the ability to move and manipulate objects.[8] General intelligence is among the field's long-term goals.[9] Approaches include statistical methods, computational intelligence, soft computing (e.g. machine learning), and traditional symbolic AI. Many tools are used in AI, including versions of search and mathematical optimization, logic, methods based on probability and economics. The AI field draws upon computer science, mathematics, psychology, linguistics, philosophy, neuroscience and artificial psychology.

The field was founded on the claim that human intelligence "can be so precisely described that a machine can be made to simulate it."[10] This raises philosophical arguments about the nature of the mind and the ethics of creating artificial beings endowed with human-like intelligence, issues which have been explored by myth, fiction and philosophy since antiquity.[11] Attempts to create artificial intelligence have experienced many setbacks, including the ALPAC report of 1966, the abandonment of perceptrons in 1970, the Lighthill Report of 1973, the second AI winter 1987–1993 and the collapse of the Lisp machine market in 1987. In the twenty-first century AI techniques became an essential part of the technology industry, helping to solve many challenging problems in computer science.[12]

Contents

History

While the concept of thought capable artificial beings appeared as storytelling devices in antiquity,[13] the idea of actually trying to build a machine to perform useful reasoning may have begun with Ramon Llull (c. 1300 CE). With his Calculus ratiocinator, Gottfried Leibniz extended the concept of the calculating machine (Wilhelm Schickard engineered the first one around 1623), intending to perform operations on concepts rather than numbers.[14] Since the 19th century, artificial beings are common in fiction, as in Mary Shelley's *Frankenstein* or Karel Čapek's *R.U.R. (Rossum's Universal Robots)*.[15]

The study of mechanical or "formal" reasoning began with philosophers and mathematicians in antiquity. In the 19th century, George Boole refined those ideas into propositional logic and Gottlob Frege developed a notational system for mechanical reasoning (a "*predicate calculus*").[16] Around the 1940s, Alan Turing's theory of computation suggested that a machine, by shuffling symbols as simple as "0" and "1", could simulate any conceivable act of mathematical deduction. This insight, that digital computers can simulate any process of formal reasoning, is known as the Church–Turing thesis.[17] Along with concurrent discoveries in neurology, information theory and cybernetics, this led researchers to consider the possibility of building an electronic brain.[18] The first work that is now generally recognized as AI was McCullouch and Pitts' 1943 formal design for Turing-complete "artificial neurons".[14]

The field of AI research was founded at a conference at Dartmouth College in 1956.[19] The attendees, including John McCarthy, Marvin Minsky, Allen Newell, Arthur Samuel and Herbert Simon, became the leaders of AI research.[20] They and their students wrote programs that were, to most people, simply astonishing:[21] computers were winning at checkers, solving word problems in algebra, proving logical theorems and speaking English.[22] By the middle of the 1960s, research in the U.S. was heavily funded by the Department of Defense[23] and laboratories had been established around the world.[24] AI's founders were optimistic about the future: Herbert Simon predicted that "machines will be capable, within twenty years, of doing any work a man can do". Marvin Minsky agreed, writing that "within a generation ... the problem of creating 'artificial intelligence' will substantially be solved".[25]

They failed to recognize the difficulty of some of the remaining tasks. Progress slowed and in 1974, in response to the criticism of Sir James Lighthill[26] and ongoing pressure from the US Congress to fund more productive projects, both the U.S. and British governments cut off exploratory research in AI. The next few years would later be called an "AI winter",[27] a period when funding for AI projects was hard to find.

In the early 1980s, AI research was revived by the commercial success of expert systems,[28] a form of AI program that simulated the knowledge and analytical skills of human experts. By 1985 the market for AI had reached over a billion dollars. At the same time, Japan's fifth generation computer project inspired the U.S and British governments to restore funding for academic research.[29] However, beginning with the collapse of the Lisp Machine market in 1987, AI once again fell into disrepute, and a second, longer-lasting hiatus began.[30]

In the late 1990s and early 21st century, AI began to be used for logistics, data mining, medical diagnosis and other areas.[12] The success was due to increasing computational power (see Moore's law), greater emphasis on solving specific problems, new ties between AI and other fields and a commitment by researchers to mathematical methods and scientific standards.[31] Deep Blue became the first computer chess-playing system to beat a reigning world chess champion, Garry Kasparov on 11 May 1997.[32]

Advanced statistical techniques (loosely known as deep learning), access to large amounts of data and faster computers enabled advances in machine learning and perception.[33] By the mid 2010s, machine learning applications were used throughout the world.[34] In a *Jeopardy!* quiz show exhibition match, IBM's question answering system, Watson, defeated the two greatest Jeopardy champions, Brad Rutter and Ken Jennings, by a significant margin.[35] The Kinect, which provides a 3D body–motion interface for the Xbox 360 and the Xbox One use algorithms that emerged from lengthy AI research[36] as do intelligent personal assistants in smartphones.[37] In March 2016, AlphaGo won 4 out of 5 games of Go in a match with Go champion Lee Sedol, becoming the first computer Go-playing system to beat a professional Go player without handicaps.[5][38]

Research

Goals

The general problem of simulating (or creating) intelligence has been broken down into sub-problems. These consist of particular traits or capabilities that researchers expect an intelligent system to display. The traits described below have received the most attention.[8]

Deduction, reasoning, problem solving

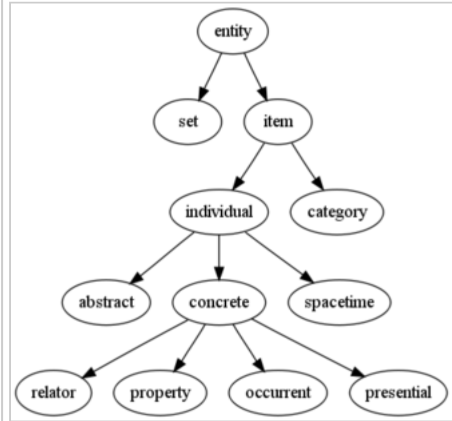
Early researchers developed algorithms that imitated step-by-step reasoning that humans use when they solve puzzles or make logical deductions (reason).[39] By the late 1980s and 1990s, AI research had developed methods for dealing with uncertain or incomplete information, employing concepts from probability and economics.[40]

For difficult problems, algorithms can require enormous computational resources—most experience a "combinatorial explosion": the amount of memory or computer time required becomes astronomical for problems of a certain size. The search for more efficient problem-solving algorithms is a high priority.[41]

Human beings ordinarily use fast, intuitive judgments rather than step-by-step deduction that early AI research was able to model.[42] AI has progressed using "sub-symbolic" problem solving: embodied agent approaches emphasize the importance of sensorimotor skills to higher reasoning; neural net research attempts to simulate the structures inside the brain that give rise to this skill; statistical approaches to AI mimic the human ability to guess.

Knowledge representation

Knowledge representation[43] and knowledge engineering[44] 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;[45] situations, events, states and time;[46] causes and effects;[47] knowledge about knowledge (what we know about what other people know);[48] and many other, less well researched domains. A representation of "what exists" is an ontology: the set of objects, relations, concepts and so on that the machine knows about. The most general are called upper ontologies, which attempt to provide a foundation for all other knowledge.[49]



An ontology represents knowledge as a set of concepts within a domain and the relationships between those concepts.

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[50] 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.[51]

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.[52] A major goal is to have the computer understand enough concepts to be able to learn by reading from sources like the Internet, and thus be able to add to its own ontology.

The subsymbolic form of some commonsense knowledge

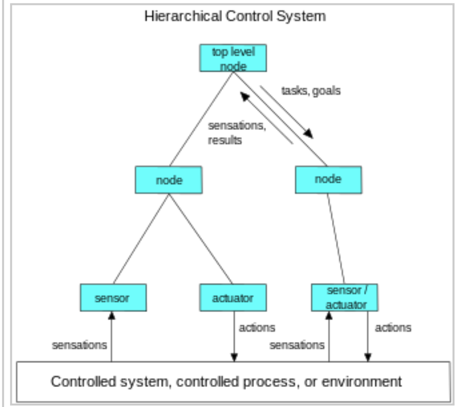
Much of what people know is not represented as "facts" or "statements" that they could express verbally. For example, a chess master will avoid a particular chess position because it "feels too exposed"[53] or an art critic can take one look at a statue and instantly realize that it is a fake.[54] These are intuitions or tendencies that are represented in the brain non-consciously and sub-symbolically.[55] 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, computational intelligence, or statistical AI will provide ways to represent this kind of knowledge.[55]

Planning

Intelligent agents must be able to set goals and achieve them.[56] 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.[57]

In classical 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.[58] However, if the agent is not the only actor, it must periodically ascertain whether the world matches its predictions and it must change its plan as this becomes necessary, requiring the agent to reason under uncertainty.[59]

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.[60]



A hierarchical control system is a form of control system in which a set of devices and governing software is arranged in a hierarchy.

Learning

Machine learning is the study of computer algorithms that improve automatically through experience[61][62] and has been central to AI research since the field's inception.[63]

Unsupervised learning is the ability to find patterns in a stream of input. Supervised learning includes both classification and numerical regression. Classification is used to determine what category something belongs in, after seeing a number of examples of things from several categories. Regression is the attempt to produce a function that describes the relationship between inputs and outputs and predicts how the outputs should change as the inputs change. In reinforcement learning[64] the agent is rewarded for good responses and punished for bad ones. The agent uses this sequence of rewards and punishments to form a strategy for operating in its problem space. These three types of learning 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.[65]

Within developmental robotics, developmental learning approaches were elaborated for lifelong cumulative acquisition of repertoires of novel skills by a robot, through autonomous self-exploration and social interaction with human teachers, and using guidance mechanisms such as active learning, maturation, motor synergies, and imitation.[66][67][68][69]

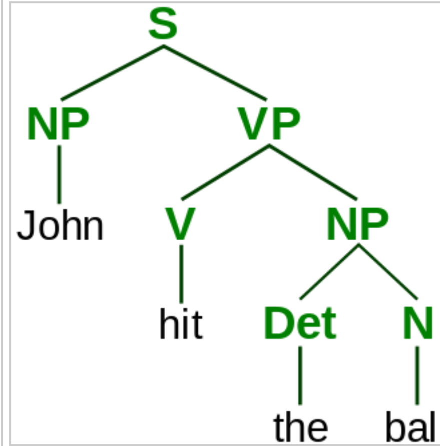
Natural language processing (communication)

Natural language processing[70] gives machines the ability to read and understand the languages that humans speak. A sufficiently powerful natural language processing system would enable natural language user interfaces and the acquisition of knowledge directly from human-written sources, such as newswire texts. Some straightforward applications of natural language processing include information retrieval, text mining, question answering[71] and machine translation.[72]

A common method of processing and extracting meaning from natural language is through semantic indexing. Increases in processing speeds and the drop in the cost of data storage makes indexing large volumes of abstractions of the user's input much more efficient.

Perception

Machine perception[73] is the ability to use input from sensors (such as cameras, microphones, tactile sensors, sonar and others more exotic) to deduce aspects of the world. Computer vision[74] is the ability to analyze visual input. A few selected subproblems are speech recognition,[75] facial recognition and object recognition.[76]



A parse tree represents the syntactic structure of a sentence according to some formal grammar.

Motion and manipulation

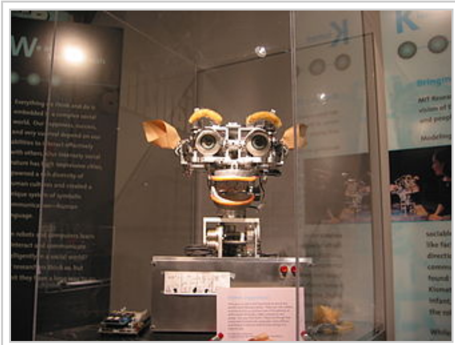
The field of robotics[77] is closely related to AI. Intelligence is required for robots to be able to handle such tasks as object manipulation[78] and navigation, with sub-problems of localization (knowing where you are, or finding out where other things are), mapping (learning what is around you, building a map of the environment), and motion planning (figuring out how to get there) or path planning (going from one point in space to another point, which may involve compliant motion – where the robot moves while maintaining physical contact with an object).[79][80]

Long-term goals

Among the long-term goals in the research pertaining to artificial intelligence are: (1) Social intelligence, (2) Creativity, and (3) General intelligence.

Social intelligence

Affective computing is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects.[82][83] It is an interdisciplinary field spanning computer sciences, psychology, and cognitive science.[84] While the origins of the field may be traced as far back as to early philosophical inquiries into emotion,[85] the more modern branch of computer science originated with Rosalind Picard's 1995 paper[86] on affective computing.[87][88] A motivation for the research is the ability to simulate empathy. The machine should interpret the emotional state of humans and adapt its behaviour to them, giving an appropriate response for those emotions.



Kismet, a robot with rudimentary social skills[81]

Emotion and social skills[89] play two roles for an intelligent agent. First, 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.) Also, in an effort to facilitate human-computer interaction, an intelligent machine might want to be able to *display* emotions—even if it does not actually experience them itself—in order to appear sensitive to the emotional dynamics of human interaction.

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, or systems that identify and assess creativity). Related areas of computational research are Artificial intuition and

Artificial thinking.

General intelligence

Many researchers think 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.[9] A few believe that anthropomorphic features like artificial consciousness or an artificial brain may be required for such a project.[90][91]

Many of the problems above may require general intelligence to be considered solved. For example, even a straightforward, specific task like machine translation requires that the machine read and write in both languages (NLP), follow the author's argument (reason), know what is being talked about (knowledge), and faithfully reproduce the author's intention (social intelligence). A problem like machine translation is considered "AI-complete". In order to solve this particular problem, one must solve all the problems.[92]

Approaches

There is no established unifying theory or paradigm that guides AI research. Researchers disagree about many issues.[93] A few of the most long standing questions that have remained unanswered are these: should artificial intelligence simulate natural intelligence by studying psychology or neurology? Or is human biology as irrelevant to AI research as bird biology is to aeronautical engineering?[94] Can intelligent behavior be described using simple, elegant principles (such as logic or optimization)? Or does it necessarily require solving a large number of completely unrelated problems?[95] Can intelligence be reproduced using high-level symbols, similar to words and ideas? Or does it require "sub-symbolic" processing?[96] John Haugeland, who coined the term GOFAI (Good Old-Fashioned Artificial Intelligence), also proposed that AI should more properly be referred to as synthetic intelligence,[97] a term which has since been adopted by some non-GOFAI researchers.[98][99]

Cybernetics and brain simulation

In the 1940s and 1950s, 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.[18] By 1960, this approach was largely abandoned, although elements of it would be revived in the 1980s.

Symbolic

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: Carnegie Mellon University, 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".[100] 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.[101] Researchers in the 1960s and the 1970s were convinced that symbolic approaches would eventually succeed in creating a machine with artificial general intelligence and considered this the goal of their field.

Cognitive simulation

Economist Herbert Simon and Allen 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 used the results of psychological experiments to develop programs that simulated the techniques that people used to solve problems. This tradition, centered at Carnegie Mellon University would eventually culminate in the development of the Soar architecture in the middle 1980s.[102][103]

Logic-based

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.[94] His laboratory at Stanford (SAIL) focused on using formal logic to solve a wide variety of problems, including knowledge representation, planning and learning.[104] Logic was also the 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.[105]

"Anti-logic" or "scruffy"

Researchers at MIT (such as Marvin Minsky and Seymour Papert)[106] 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).[95] 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.[107]

Knowledge-based

When computers with large memories became available around 1970, researchers from all three traditions began to build knowledge into AI applications.[108] This "knowledge revolution" led to the development and deployment of expert systems (introduced by Edward Feigenbaum), the first truly successful form of AI software.[28] The knowledge revolution was also driven by the realization that enormous amounts of knowledge would be required by many simple AI applications.

Sub-symbolic

By the 1980s 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.[96] Sub-symbolic methods manage to approach intelligence without specific representations of knowledge.

Bottom-up, embodied, situated, behavior-based or nouvelle AI

Researchers from the related field of robotics, such as Rodney Brooks, rejected symbolic AI and focused on the basic engineering problems that would allow robots to move and survive.[109] Their work revived the non-symbolic viewpoint of the early cybernetics researchers of the 1950s and reintroduced the use of control theory in AI. This coincided with the development of the embodied mind thesis in the related field of cognitive science: the idea that aspects of the body (such as movement, perception and visualization) are required for higher intelligence.

Computational intelligence and soft computing

Interest in neural networks and "connectionism" was revived by David Rumelhart and others in the middle of 1980s.[110] Neural networks are an example of soft computing --- they are solutions to problems which cannot be solved with complete logical certainty, and where an approximate solution is often enough. Other soft computing approaches to AI include fuzzy systems, evolutionary computation and many statistical tools.

The application of soft computing to AI is studied collectively by the emerging discipline of computational intelligence.[111]

Statistical

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). Stuart Russell and Peter Norvig describe this movement as nothing less than a "revolution" and "the victory of the neats."[31] Critics argue that these techniques (with few exceptions[112]) are too focused on particular problems and have failed to address the long-term goal of general intelligence.[113] There is an ongoing debate about the relevance and validity of statistical approaches in AI, exemplified in part by exchanges between Peter Norvig and Noam Chomsky.[114][115]

Integrating the approaches

Intelligent agent paradigm

An intelligent agent is a system that perceives its environment and takes actions which maximize its chances of success. The simplest intelligent agents are programs that solve specific problems. More complicated agents include human beings and organizations of human beings (such as firms). 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. The intelligent agent paradigm became widely accepted during the 1990s.[1]

Agent architectures and cognitive architectures

Researchers have designed systems to build intelligent systems out of interacting intelligent agents in a multi-agent system.[116] 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.[117] Rodney Brooks' subsumption architecture was an early proposal for such a hierarchical system.[118]

Tools

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

Many problems in AI can be solved in theory by intelligently searching through many possible solutions:[119] 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.[120] Planning algorithms search through trees of goals and subgoals, attempting to find a path to a target goal, a process called means-ends analysis.[121] Robotics algorithms for moving limbs and grasping objects use local searches in configuration space.[78] Many learning algorithms use search algorithms based on optimization.

Simple exhaustive searches[122] 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 the path on which the solution lies.[123] Heuristics limit the search for solutions into a smaller sample size.[79]

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.[124]

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)[125] and evolutionary algorithms (such as genetic algorithms, gene expression programming, and genetic programming).[126]

Logic

Logic[127] 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[128] and inductive logic programming is a method for learning.[129]

Several different forms of logic are used in AI research. Propositional or sentential logic[130] is the logic of statements which can be true or false. First-order logic[131] also allows the use of quantifiers and predicates, and can express facts about objects, their properties, and their relations with each other. Fuzzy logic,[132] is 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. Subjective logic[133] models uncertainty in a different and more explicit manner than fuzzy-logic: a given binomial opinion satisfies belief + disbelief + uncertainty = 1 within a Beta distribution. By this method, ignorance can be distinguished from probabilistic statements that an agent makes with high confidence.

Default logics, non-monotonic logics and circumscription[51] are forms of logic designed to help with default reasoning and the qualification problem. Several extensions of logic have been designed to handle specific domains of knowledge, such as: description logics;[45] situation calculus, event calculus and fluent calculus (for representing events and time);[46] causal calculus;[47] belief calculus;[134] and modal logics.[48]

Probabilistic methods for uncertain reasoning

Many problems in AI (in reasoning, planning, learning, perception and robotics) require the agent to operate with incomplete or uncertain information. AI researchers have devised a number of powerful tools to solve these problems using methods from probability theory and economics.[135]

Bayesian networks[136] are a very general tool that can be used for a large number of problems: reasoning (using the Bayesian inference algorithm),[137] learning (using the expectation-maximization algorithm),[138] planning (using decision networks)[139] and perception (using dynamic Bayesian networks).[140] 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 (e.g., hidden Markov models or Kalman filters).[140]

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,[141] and information value theory.[57] These tools include models such as Markov decision processes,[142] dynamic decision networks,[140] game theory and mechanism design.[143]

Classifiers and statistical learning methods

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 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.[144]

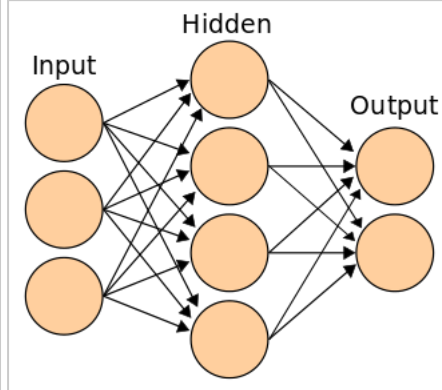
A classifier can be trained in various ways; there are many statistical and machine learning approaches. The most widely used classifiers are the neural network,[145] kernel methods such as the support vector machine,[146] k-nearest neighbor algorithm,[147] Gaussian mixture model,[148] naive Bayes classifier,[149] and decision tree.[150] The performance of these classifiers have been compared over a wide range of tasks. 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. Determining a suitable classifier for a given problem is still more an art than science.[151]

Neural networks

The study of non-learning artificial neural networks[145] began in the decade before the field of AI research was founded, in the work of Walter Pitts and Warren McCullouch. Frank Rosenblatt invented the perceptron, a learning network with a single layer, similar to the old concept of linear regression. Early pioneers also include Alexey Grigorevich Ivakhnenko, Teuvo Kohonen, Stephen Grossberg, Kunihiko Fukushima, Christoph von der Malsburg, David Willshaw, Shun-Ichi Amari, Bernard Widrow, John Hopfield, Eduardo R. Caianiello, and others.

The main categories of networks are acyclic or feedforward neural networks (where the signal passes in only one direction) and recurrent neural networks (which allow feedback and short-term memories of previous input events). Among the most popular feedforward networks are perceptrons, multi-layer perceptrons and radial basis networks.[152] Neural networks can be applied to the problem of intelligent control (for robotics) or learning, using such techniques as Hebbian learning, GMDH or competitive learning.[153]

Today, neural networks are often trained by the backpropagation algorithm, which had been around since 1970 as the reverse mode of automatic differentiation published by Seppo Linnainmaa,[154][155] and was introduced to neural networks by Paul Werbos.[156][157][158]



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

Hierarchical temporal memory is an approach that models some of the structural and algorithmic properties of the neocortex.[159]

Deep feedforward neural networks

Deep learning in artificial neural networks with many layers has transformed many important subfields of artificial intelligence, including computer vision, speech recognition, natural language processing and others.[160][161][162]

According to a survey,[163] the expression "Deep Learning" was introduced to the Machine Learning community by Rina Dechter in 1986[164] and gained traction after Igor Aizenberg and colleagues introduced it to Artificial Neural Networks in 2000.[165] The first functional Deep Learning networks were published by Alexey Grigorevich Ivakhnenko and V. G. Lapa in 1965.[166] These networks are trained one layer at a time. Ivakhnenko's 1971 paper[167] describes the learning of a deep feedforward multilayer perceptron with eight layers, already much deeper than many later networks. In 2006, a publication by Geoffrey Hinton and Ruslan Salakhutdinov introduced another way of pre-training many-layered feedforward neural networks (FNNs) one layer at a time, treating each layer in turn as an unsupervised restricted Boltzmann machine, then using supervised backpropagation for fine-tuning.[168] Similar to shallow artificial neural networks, deep neural networks can model complex non-linear relationships. Over the last few years, advances in both machine learning algorithms and computer hardware have led to more efficient methods for training deep neural networks that contain many layers of non-linear hidden units and a very large output layer.[169]

Deep learning often uses convolutional neural networks (CNNs), whose origins can be traced back to the Neocognitron introduced by Kunihiko Fukushima in 1980.[170] In 1989, Yann LeCun and colleagues applied backpropagation to such an architecture. In the early 2000s, in an industrial application CNNs already processed an estimated 10% to 20% of all the checks written in the US.[171] Since 2011, fast implementations of CNNs on GPUs have won many visual pattern recognition competitions.[162]

Deep feedforward neural networks were used in conjunction with reinforcement learning by AlphaGo, Google Deepmind's program that was the first to beat a professional human player.[172]

Deep recurrent neural networks

Early on, deep learning was also applied to sequence learning with recurrent neural networks (RNNs)[173] which are general computers and can run arbitrary programs to process arbitrary sequences of inputs. The depth of an RNN is unlimited and depends on the length of its input sequence.[162] RNNs can be trained by gradient descent[174][175][176] but suffer from the vanishing gradient problem.[160][177] In 1992, it was shown that unsupervised pre-training of a stack of recurrent neural networks can speed up subsequent supervised learning of deep sequential problems.[178]

Numerous researchers now use variants of a deep learning recurrent NN called the long short-term memory (LSTM) network published by Hochreiter & Schmidhuber in 1997.[179] LSTM is often trained by Connectionist Temporal Classification (CTC).[180] At Google, Microsoft and Baidu this approach has revolutionised speech recognition.[181][182][183] For example, in 2015, Google's speech recognition experienced a dramatic performance jump of 49% through CTC-trained LSTM, which is now available through Google Voice to billions of smartphone users.[184] Google also used LSTM to improve machine translation,[185] Language Modeling[186] and Multilingual Language Processing.[187] LSTM combined with CNNs also improved automatic image captioning[188] and a plethora of other applications.

Control theory

Control theory, the grandchild of cybernetics, has many important applications, especially in robotics.[189]

Languages

AI researchers have developed several specialized languages for AI research, including Lisp[190] and Prolog.[191]

Evaluating progress

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.[192]

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.[193]

One classification for outcomes of an AI test is:[194]

1. Optimal: it is not possible to perform better.
2. Strong super-human: performs better than all humans.
3. Super-human: performs better than most humans.
4. Sub-human: performs worse than most humans.

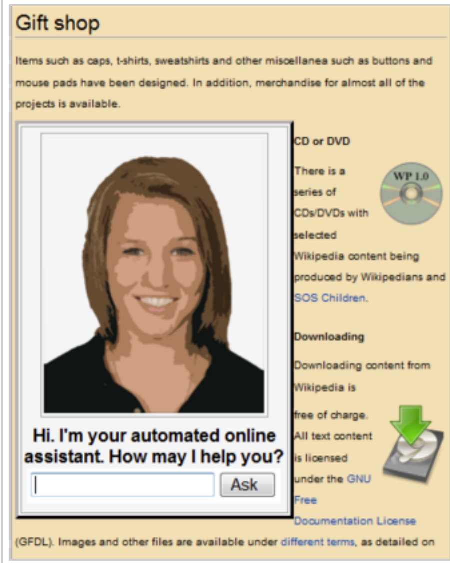
For example, performance at draughts (i.e. checkers) is optimal,[195] performance at chess is super-human and nearing strong super-human (see computer chess: computers versus human) and performance at many everyday tasks (such as recognizing a face or crossing a room without bumping into something) is sub-human.

A quite different approach measures machine intelligence through tests which are developed from *mathematical* definitions of intelligence. Examples of these kinds of tests start in the late nineties devising intelligence tests using notions from Kolmogorov complexity and data compression.[196] Two major advantages of mathematical definitions are their applicability to nonhuman intelligences and their absence of a requirement for human testers.

A derivative of the Turing test is the Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA). As the name implies, this helps to determine that a user is an actual person and not a computer posing as a human. In contrast to the standard Turing test, CAPTCHA administered by a machine and targeted to a human as opposed to being administered by a human and targeted to a machine. A computer asks a user to complete a simple test then generates a grade for that test. Computers are unable to solve the problem, so correct solutions are deemed to be the result of a person taking the test. A common type of CAPTCHA is the test that requires the typing of distorted letters, numbers or symbols that appear in an image undecipherable by a computer.[197]

Applications

AI is relevant to any intellectual task.[198] Modern artificial intelligence techniques are pervasive and are too numerous to list here. Frequently, when a technique reaches mainstream use, it is no longer considered artificial intelligence; this phenomenon is described as the AI effect.[199]



An automated online assistant providing customer service on a web page – one of many very primitive applications of artificial intelligence.

High-profile examples of AI include autonomous vehicles (such as drones and self-driving cars), medical diagnosis, creating art (such as poetry), proving mathematical theorems, playing games (such as Chess or Go), search engines (such as Google search), online assistants (such as Siri), image recognition in photographs, spam filtering, prediction of judicial decisions [200] and targeting online advertisements.[198][201][202]

With social media sites overtaking TV as a source for news for young people and news organisations increasingly reliant on social media platforms for generating distribution,[203] major publishers now use artificial intelligence (AI) technology to post stories more effectively and generate higher volumes of traffic.[204]

Competitions and prizes

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

Platforms

A platform (or "computing platform") is defined as "some sort of hardware architecture or software framework (including application frameworks), that allows software to run." As Rodney Brooks pointed out many years ago,[205] it is not just the artificial intelligence software that defines the AI features of the platform, but rather the actual platform itself that affects the AI that results, i.e., there needs to be work in AI problems on real-world platforms rather than in isolation.

A wide variety of platforms has allowed different aspects of AI to develop, ranging from expert systems such as Cyc to deep-learning frameworks to robot platforms such as the Roomba with open interface.[206] Recent advances in deep artificial neural networks and distributed computing have led to a proliferation of software libraries, including Deeplearning4j, TensorFlow, Theano and Torch.

Philosophy and ethics

There are three philosophical questions related to AI:

1. Is artificial general intelligence possible? Can a machine solve any problem that a human being can solve using intelligence? Or are there hard limits to what a machine can accomplish?
2. Are intelligent machines dangerous? How can we ensure that machines behave ethically and that they are used ethically?
3. Can a machine have a mind, consciousness and mental states in exactly the same sense that human beings do? Can a machine be sentient, and thus deserve certain rights? Can a machine intentionally cause harm?

The limits of artificial general intelligence

Can a machine be intelligent? Can it "think"?

Turing's "polite convention"
: We need not decide if a machine can "think"; we need only decide if a machine can act as intelligently as a human being. This approach to the philosophical problems associated with artificial intelligence forms the basis of the Turing test.[192]

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 conjecture was printed in the proposal for the Dartmouth Conference of 1956, and represents the position of most working AI researchers.[207]

Newell and Simon's physical symbol system hypothesis
: "A physical symbol system has the necessary and sufficient means of general intelligent action." Newell and Simon argue that intelligence consists of formal operations on symbols.[208] 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. (See Dreyfus' critique of AI.)[209][210]

Gödelian arguments
: Gödel himself,[211] John Lucas (in 1961) and Roger Penrose (in a more detailed argument from 1989 onwards) made highly technical arguments that human mathematicians can consistently see the truth of their own "Gödel statements" and therefore have computational abilities beyond that of mechanical Turing machines.[212] However, the modern consensus in the scientific and mathematical community is that these "Gödelian arguments" fail.[213][214][215]

The artificial brain argument
: The brain can be simulated by machines and because brains are intelligent, simulated brains must also be intelligent; thus machines can be intelligent. 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.[91]

The AI effect
: Machines are *already* intelligent, but observers have failed to recognize it. When Deep Blue beat Garry Kasparov in chess, the machine was acting intelligently. However, onlookers commonly discount the behavior of an artificial intelligence program by arguing that it is not "real" intelligence after all; thus "real" intelligence is whatever intelligent behavior people can do that machines still can not. This is known as the AI Effect: "AI is whatever hasn't been done yet."

Intelligent behaviour and machine ethics

As a minimum, an AI system must be able to reproduce aspects of human intelligence. This raises the issue of how ethically the machine should behave towards both humans and other AI agents. This issue was addressed by Wendell Wallach in his book titled *Moral Machines* in which he introduced the concept of artificial moral agents (AMA).[216] For Wallach, AMAs have become a part of the research landscape of artificial intelligence as guided by its two central questions which he identifies as "Does Humanity Want Computers Making Moral Decisions"[217] and "Can (Ro)bots Really Be Moral".[218] For Wallach the question is not centered on the issue of *whether* machines can demonstrate the equivalent of moral behavior in contrast to the *constraints* which society may place on the development of AMAs.[219]

Machine ethics

The field of machine ethics is concerned with giving machines ethical principles, or a procedure for discovering a way to resolve the ethical dilemmas they might encounter, enabling them to function in an ethically responsible manner through their own ethical decision making.[220] The field was delineated in the AAAI Fall 2005 Symposium on Machine Ethics: "Past research concerning the relationship between technology and ethics has largely focused on responsible and irresponsible use of technology by human beings, with a few people being interested in how human beings ought to treat machines. In all cases, only human beings have engaged in ethical reasoning. The time has come for adding an ethical dimension to at least some machines. Recognition of the ethical ramifications of behavior involving machines, as well as recent and potential developments in machine autonomy, necessitate this. In contrast to computer hacking, software property issues, privacy issues and other topics normally ascribed to computer ethics, machine ethics is concerned with the behavior of machines towards human users and other machines. Research in machine ethics is key to alleviating concerns with autonomous systems—it could be argued that the notion of autonomous machines without such a dimension is at the root of all fear concerning machine intelligence. Further, investigation of machine ethics could enable the discovery of problems with current ethical theories, advancing our thinking about Ethics."[221] Machine ethics is sometimes referred to as machine morality, computational ethics or computational morality. A variety of perspectives of this nascent field can be found in the collected edition "Machine Ethics"[220] that stems from the AAAI Fall 2005 Symposium on Machine Ethics.[221] Some suggest that to ensure that AI-equipped machines (sometimes called "smart machines") will act ethically requires a new kind of AI. This AI would be able to monitor, supervise, and if need be, correct the first order AI.[222]

Malevolent and friendly AI

Political scientist Charles T. Rubin believes that AI can be neither designed nor guaranteed to be benevolent.[223] He argues that "any sufficiently advanced benevolence may be indistinguishable from malevolence." Humans should not assume machines or robots would treat us favorably, because there is no *a priori* reason to believe that they would be sympathetic to our system of morality, which has evolved along with our particular biology (which AIs would not share). Hyper-intelligent software may not necessarily decide to support the continued existence of mankind, and would be extremely difficult to stop. This topic has also recently begun to be discussed in academic publications as a real source of risks to civilization, humans, and planet Earth.

Physicist Stephen Hawking, Microsoft founder Bill Gates and SpaceX founder Elon Musk have expressed concerns about the possibility that AI could evolve to the point that humans could not control it, with Hawking theorizing that this could "spell the end of the human race".[224]

One proposal to deal with this is to ensure that the first generally intelligent AI is 'Friendly AI', and will then be able to control subsequently developed AIs. Some question whether this kind of check could really remain in place.

Leading AI researcher Rodney Brooks writes, "I think it is a mistake to be worrying about us developing malevolent AI anytime in the next few hundred years. I think the worry stems from a fundamental error in not distinguishing the difference between the very real recent advances in a particular aspect of AI, and the enormity and complexity of building sentient volitional intelligence."[225]

Devaluation of humanity

Joseph Weizenbaum wrote that AI applications can not, by definition, successfully simulate genuine human empathy and that the use of AI technology in fields such as customer service or psychotherapy[226] was deeply misguided. Weizenbaum was also bothered that AI researchers (and some philosophers) were willing to view the human mind as nothing more than a computer program (a position now known as computationalism). To Weizenbaum these points suggest that AI research devalues human life.[227]

Decrease in demand for human labor

Martin Ford, author of *The Lights in the Tunnel: Automation, Accelerating Technology and the Economy of the Future*,[228] and others argue that specialized artificial intelligence applications, robotics and other forms of automation will ultimately result in significant unemployment as machines begin to match and exceed the capability of workers to perform most routine and repetitive jobs. Ford predicts that many knowledge-based occupations—and in particular entry level jobs—will be increasingly susceptible to automation via expert systems, machine learning[229] and other AI-enhanced applications. AI-based applications may also be used to amplify the capabilities of low-wage offshore workers, making it more feasible to outsource knowledge work.[230]

Machine consciousness, sentience and mind

If an AI system replicates all key aspects of human intelligence, will that system also be sentient – will it have a mind which has conscious experiences? This question is closely related to the philosophical problem as to the nature of human consciousness, generally referred to as the hard problem of consciousness.

Consciousness

Computationalism

Computationalism is the position in the philosophy of mind that the human mind or the human brain (or both) is an information processing system and that thinking is a form of computing.[231] Computationalism argues that the relationship between mind and body is similar or identical to the relationship between software and hardware and thus may be a solution to the mind-body problem. This philosophical position was inspired by the work of AI researchers and cognitive scientists in the 1960s and was originally proposed by philosophers Jerry Fodor and Hilary Putnam.

Strong AI hypothesis

The philosophical position that John Searle has named "strong AI" states: "The appropriately programmed computer with the right inputs and outputs would thereby have a mind in exactly the same sense human beings have minds."[232] 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.[233]

Robot rights

Mary Shelley's *Frankenstein* 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? The idea also appears in modern science fiction, such as the film *A.I.: Artificial Intelligence*, in which humanoid machines have the ability to feel emotions. This issue, now known as "robot rights", is currently being considered by, for example, California's Institute for the Future, although many critics believe that the discussion is premature.[234] The subject is profoundly discussed in the 2010 documentary film *Plug & Pray*.[235]

Superintelligence

Are there limits to how intelligent machines – or human-machine hybrids – can be? A superintelligence, hyperintelligence, or superhuman intelligence is a hypothetical agent that would possess intelligence far surpassing that of the brightest and most gifted human mind. ‘’Superintelligence’’ may also refer to the form or degree of intelligence possessed by such an agent.

Technological singularity

If research into Strong AI produced sufficiently intelligent software, it might be able to reprogram and improve itself. The improved software would be even better at improving itself, leading to recursive self-improvement.[236] The new intelligence could thus increase exponentially and dramatically surpass humans. Science fiction writer Vernor Vinge named this scenario "singularity".[237] Technological singularity is when accelerating progress in technologies will cause a runaway effect wherein artificial intelligence will exceed human intellectual capacity and control, thus radically changing or even ending civilization. Because the capabilities of such an intelligence may be impossible to comprehend, the technological singularity is an occurrence beyond which events are unpredictable or even unfathomable.[237]

Ray Kurzweil has used Moore's law (which describes the relentless exponential improvement in digital technology) to calculate that desktop computers will have the same processing power as human brains by the year 2029, and predicts that the singularity will occur in 2045.[237]

Transhumanism

> You awake one morning to find your brain has another lobe functioning. Invisible, this auxiliary lobe answers your questions with information beyond the realm of your own memory, suggests plausible courses of action, and asks questions that help bring out relevant facts. You quickly come to rely on the new lobe so much that you stop wondering how it works. You just use it. This is the dream of artificial intelligence.
>
> — *BYTE*, April 1985[238]

Robot designer Hans Moravec, cyberneticist Kevin Warwick and inventor Ray Kurzweil have predicted that humans and machines will merge in the future into cyborgs that are more capable and powerful than either.[239] This idea, called transhumanism, which has roots in Aldous Huxley and Robert Ettinger, has been illustrated in fiction as well, for example in the manga *Ghost in the Shell* and the science-fiction series *Dune*.

In the 1980s artist Hajime Sorayama's Sexy Robots series were painted and published in Japan depicting the actual organic human form with lifelike muscular metallic skins and later "the Gynoids" book followed that was used by or influenced movie makers including George Lucas and other creatives. Sorayama never considered these organic robots to be real part of nature but always unnatural product of the human mind, a fantasy existing in the mind even when realized in actual form.

Edward Fredkin argues that "artificial intelligence is the next stage in evolution", 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.[240]

Existential risk

> The development of full artificial intelligence could spell the end of the human race. Once humans develop artificial intelligence, it will take off on its own and redesign itself at an ever-increasing rate. Humans, who are limited by slow biological evolution, couldn't compete and would be superseded.
>
> — Stephen Hawking[241]

A common concern about the development of artificial intelligence is the potential threat it could pose to mankind. This concern has recently gained attention after mentions by celebrities including Stephen Hawking, Bill Gates,[242] and Elon Musk.[243] A group of prominent tech titans including Peter Thiel, Amazon Web Services and Musk have committed $1billion to OpenAI a nonprofit company aimed at championing responsible AI development.[244] The opinion of experts within the field of artificial intelligence is mixed, with sizable fractions both concerned and unconcerned by risk from eventual superhumanly-capable AI.[245]

In his book *Superintelligence*, Nick Bostrom provides an argument that artificial intelligence will pose a threat to mankind. He argues that sufficiently intelligent AI, if it chooses actions based on achieving some goal, will exhibit convergent behavior such as acquiring resources or protecting itself from being shut down. If this AI's goals do not reflect humanity's - one example is an AI told to compute as many digits of pi as possible - it might harm humanity in order to acquire more resources or prevent itself from being shut down, ultimately to better achieve its goal.

For this danger to be realized, the hypothetical AI would have to overpower or out-think all of humanity, which a minority of experts argue is a possibility far enough in the future to not be worth researching.[246][247] Other counterarguments revolve around humans being either intrinsically or convergently valuable from the perspective of an artificial intelligence.[248]

Concern over risk from artificial intelligence has led to some high-profile donations and investments. In January 2015, Elon Musk donated ten million dollars to the Future of Life Institute to fund research on understanding AI decision making. The goal of the institute is to "grow wisdom with which we manage" the growing power of technology. Musk also funds companies developing artificial intelligence such as Google DeepMind and Vicarious to "just keep an eye on what's going on with artificial intelligence.[249] I think there is potentially a dangerous outcome there."[250][251]

Development of militarized artificial intelligence is a related concern. Currently, 50+ countries are researching battlefield robots, including the United States, China, Russia, and the United Kingdom. Many people concerned about risk from superintelligent AI also want to limit the use of artificial soldiers.[252]

Moral decision-making

To keep AI ethical, some have suggested teaching new technologies equipped with AI, such as driver-less cars, to render moral decisions on their own.[253] Others argued that these technologies could learn to act ethically the way children do—by interacting with adults, in particular, with ethicists.[254] Still others suggest these smart technologies can determine the moral preferences of those who use them (just the way one learns about consumer preferences) and then be programmed to heed these preferences.[255]

In fiction

The implications of artificial intelligence have been a persistent theme in science fiction. Early stories typically revolved around intelligent robots. The word "robot" itself was coined by Karel Čapek in his 1921 play *R.U.R.*, the title standing for "Rossum's Universal Robots". Later, the SF writer Isaac Asimov developed the Three Laws of Robotics which he subsequently explored in a long series of robot stories. These laws have since gained some traction in genuine AI research.

Other influential fictional intelligences include HAL, the computer in charge of the spaceship in *2001: A Space Odyssey*, released as both a film and a book in 1968 and written by Arthur C. Clarke.

AI has since become firmly rooted in popular culture and is in many films, such as *The Terminator* (1984) and *A.I. Artificial Intelligence* (2001).

See also

- Glossary of artificial intelligence
- Abductive reasoning
- Case-based reasoning
- Commonsense reasoning
- Soft computing
  - Machine learning
  - Evolutionary computing

Notes

1. 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
   - Hutter 2005, pp. 125–126

   The definition used in this article, in terms of goals, actions, perception and environment, is due to Russell & Norvig (2003). Other definitions also include knowledge and learning as additional criteria.
2. Russell & Norvig 2009, p. 2.
3. Schank, Roger C. (1991). "Where's the AI". AI magazine. p. 38.
4. Russell & Norvig 2009.
5. "AlphaGo - Google DeepMind".
6. http://betanews.com/2016/10/21/artificial-intelligence-stephen-hawking/
7. Pamela McCorduck (2004, pp. 424) writes of "the rough shattering of AI in subfields—vision, natural language, decision theory, genetic algorithms, robotics ... and these with own sub-subfield—that would hardly have anything to say to each other."

8. 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
9. General intelligence (strong AI) is discussed in popular introductions to AI:
    - Kurzweil 1999 and Kurzweil 2005
10. See the Dartmouth proposal, under Philosophy, below.
11. This is a central idea of Pamela McCorduck's *Machines Who Think*. She writes: "I like to think of artificial intelligence as the scientific apotheosis of a venerable 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, pp. 340–400)
12. AI applications widely used behind the scenes:
    - Russell & Norvig 2003, p. 28
    - Kurzweil 2005, p. 265
    - NRC 1999, pp. 216–222
13. AI in myth:
    - McCorduck 2004, pp. 4–5
    - Russell & Norvig 2003, p. 939
14. Russell & Norvig 2009, p. 16.
15. AI in early science fiction.
    - McCorduck 2004, pp. 17–25
16. Nilsson 1998, Section 1.3.
17. Formal reasoning:
    - Berlinski, David (2000). *The Advent of the Algorithm*. Harcourt Books. ISBN 0-15-601391-6. OCLC 46890682.
18. AI's immediate precursors:
    - McCorduck 2004, pp. 51–107
    - Crevier 1993, pp. 27–32
    - Russell & Norvig 2003, pp. 15, 940
    - Moravec 1988, p. 3
19. Dartmouth conference:
    - McCorduck 2004, pp. 111–136
    - Crevier 1993, pp. 47–49, who writes "the conference is generally recognized as the official birthdate of the new science."
    - Russell & Norvig 2003, p. 17, who call the conference "the birth of artificial intelligence."
    - NRC 1999, pp. 200–201
20. Hegemony of the Dartmouth conference attendees:
    - Russell & Norvig 2003, p. 17, who write "for the next 20 years the field would be dominated by these people and their students."
    - McCorduck 2004, pp. 129–130
21. Russell & Norvig 2003, p. 18.
22. "Golden years" of AI (successful symbolic reasoning programs 1956–1973):
    - McCorduck 2004, pp. 243–252
    - Crevier 1993, pp. 52–107
    - Moravec 1988, p. 9
    - Russell & Norvig 2003, pp. 18–21

    The programs described are Arthur Samuel's checkers program for the IBM 701, Daniel Bobrow's STUDENT, Newell and Simon's Logic Theorist and Terry Winograd's SHRDLU.
23. DARPA pours money into undirected pure research into AI during the 1960s:
    - McCorduck 2004, pp. 131
    - Crevier 1993, pp. 51, 64–65
    - NRC 1999, pp. 204–205

24. AI in England:
    - Howe 1994
25. Optimism of early AI:
    - Herbert Simon quote: Simon 1965, p. 96 quoted in Crevier 1993, p. 109.
    - Marvin Minsky quote: Minsky 1967, p. 2 quoted in Crevier 1993, p. 109.
26. Lighthill 1973.
27. First AI Winter, Mansfield Amendment, Lighthill report
    - Crevier 1993, pp. 115–117
    - Russell & Norvig 2003, p. 22
    - NRC 1999, pp. 212–213
    - Howe 1994
28. 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
29. 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
30. Second AI winter:
    - McCorduck 2004, pp. 430–435
    - Crevier 1993, pp. 209–210
    - NRC 1999, pp. 214–216
31. Formal methods are now preferred ("Victory of the neats"):
    - Russell & Norvig 2003, pp. 25–26
    - McCorduck 2004, pp. 486–487
32. McCorduck 2004, pp. 480–483.
33. Deep learning:
    - citation in progress
34. Machine learning and AI's successes in the early 21st century:
    - citation in progress
35. Markoff 2011.
36. Administrator. "Kinect's AI breakthrough explained". *i-programmer.info*.
37. Rowinski, Dan (15 January 2013). "Virtual Personal Assistants & The Future Of Your Smartphone [Infographic]". *ReadWrite*.
38. "Artificial intelligence: Google's AlphaGo beats Go master Lee Se-dol". *BBC News*. 12 March 2016. Retrieved 1 October 2016.
39. Problem solving, puzzle solving, game playing and deduction:
    - Russell & Norvig 2003, chpt. 3–9,
    - Poole, Mackworth & Goebel 1998, chpt. 2,3,7,9,
    - Luger & Stubblefield 2004, chpt. 3,4,6,8,
    - Nilsson 1998, chpt. 7–12
40. 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
41. Intractability and efficiency and the combinatorial explosion:
    - Russell & Norvig 2003, pp. 9, 21–22

42. Psychological evidence of sub-symbolic reasoning:
    - Wason & Shapiro (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)
    - Kahneman, Slovic & Tversky (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)
43. 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
44. Knowledge engineering:
    - Russell & Norvig 2003, pp. 260–266,
    - Poole, Mackworth & Goebel 1998, pp. 199–233,
    - Nilsson 1998, chpt. ≈17.1–17.4
45. 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
46. 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
47. Causal calculus:
    - Poole, Mackworth & Goebel 1998, pp. 335–337
48. Representing knowledge about knowledge: Belief calculus, modal logics:
    - Russell & Norvig 2003, pp. 341–344,
    - Poole, Mackworth & Goebel 1998, pp. 275–277
49. Ontology:
    - Russell & Norvig 2003, pp. 320–328
50. Qualification problem:
    - McCarthy & Hayes 1969
    - Russell & Norvig 2003

    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.
51. 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
52. Breadth of commonsense knowledge:
    - Russell & Norvig 2003, p. 21,
    - Crevier 1993, pp. 113–114,
    - Moravec 1988, p. 13,
    - Lenat & Guha 1989 (Introduction)
53. Dreyfus & Dreyfus 1986.
54. Gladwell 2005.

55. 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. See Dreyfus' critique of AI)
   - Gladwell 2005 (Gladwell's *Blink* is a popular introduction to sub-symbolic reasoning and knowledge.)
   - Hawkins & Blakeslee 2005 (Hawkins argues that sub-symbolic knowledge should be the primary focus of AI research.)
56. 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
57. Information value theory:
   - Russell & Norvig 2003, pp. 600–604
58. 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
59. Planning and acting in non-deterministic domains: conditional planning, execution monitoring, replanning and continuous planning:
   - Russell & Norvig 2003, pp. 430–449
60. Multi-agent planning and emergent behavior:
   - Russell & Norvig 2003, pp. 449–455
61. This is a form of Tom Mitchell's widely quoted definition of machine learning: "A computer program is set to learn from an experience *E* with respect to some task *T* and some performance measure *P* if its performance on *T* as measured by *P* improves with experience *E*."
62. 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
63. Alan Turing discussed the centrality of learning as early as 1950, in his classic paper "Computing Machinery and Intelligence".(Turing 1950) In 1956, at the original Dartmouth AI summer conference, Ray Solomonoff wrote a report on unsupervised probabilistic machine learning: "An Inductive Inference Machine".(Solomonoff 1956)
64. Reinforcement learning:
   - Russell & Norvig 2003, pp. 763–788
   - Luger & Stubblefield 2004, pp. 442–449
65. Computational learning theory:
   - CITATION IN PROGRESS.
66. Weng et al. 2001.
67. Lungarella et al. 2003.
68. Asada et al. 2009.
69. Oudeyer 2010.
70. 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
71. "Versatile question answering systems: seeing in synthesis (https://www.academia.edu/2475776/Versatile_question_answering_systems_seeing_in_synthesis)", Mittal et al., IJIIDS, 5(2), 119-142, 2011
72. 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

73. Machine perception:
    - Russell & Norvig 2003, pp. 537–581, 863–898
    - Nilsson 1998, ~chpt. 6
74. Computer vision:
    - ACM 1998, I.2.10
    - Russell & Norvig 2003, pp. 863–898
    - Nilsson 1998, chpt. 6
75. Speech recognition:
    - ACM 1998, ~I.2.7
    - Russell & Norvig 2003, pp. 568–578
76. Object recognition:
    - Russell & Norvig 2003, pp. 885–892
77. Robotics:
    - ACM 1998, I.2.9,
    - Russell & Norvig 2003, pp. 901–942,
    - Poole, Mackworth & Goebel 1998, pp. 443–460
78. Moving and configuration space:
    - Russell & Norvig 2003, pp. 916–932
79. Tecuci 2012.
80. Robotic mapping (localization, etc):
    - Russell & Norvig 2003, pp. 908–915
81. *Kismet*.
82. Thro 1993.
83. Edelson 1991.
84. Tao & Tan 2005.
85. James 1884.
86. Picard 1995.
87. Kleine-Cosack 2006: "The introduction of emotion to computer science was done by Pickard (sic) who created the field of affective computing."
88. Diamond 2003: "Rosalind Picard, a genial MIT professor, is the field's godmother; her 1997 book, Affective Computing, triggered an explosion of interest in the emotional side of computers and their users."
89. Emotion and affective computing:
    - Minsky 2006
90. Gerald Edelman, Igor Aleksander and others have argued that artificial consciousness is required for strong AI. (Aleksander 1995; Edelman 2007)
91. Artificial brain arguments: AI requires a simulation of the operation of the human brain
    - Russell & Norvig 2003, p. 957
    - Crevier 1993, pp. 271 and 279

    A few of the people who make some form of the argument:
    - Moravec 1988
    - Kurzweil 2005, p. 262
    - Hawkins & Blakeslee 2005

    The most extreme form of this argument (the brain replacement scenario) was put forward by Clark Glymour in the mid-1970s and was touched on by Zenon Pylyshyn and John Searle in 1980.
92. AI complete: Shapiro 1992, p. 9
93. Nils Nilsson writes: "Simply put, there is wide disagreement in the field about what AI is all about" (Nilsson 1983, p. 10).
94. Biological intelligence vs. intelligence in general:
    - Russell & Norvig 2003, pp. 2–3, who make the analogy with aeronautical engineering.
    - McCorduck 2004, pp. 100–101, who writes that there are "two major branches of artificial intelligence: one aimed at producing intelligent behavior regardless of how it was accomplioshed, and the other aimed at modeling intelligent processes found in nature, particularly human ones."
    - Kolata 1982, a paper in *Science*, which describes McCarthy's indifference to biological models. Kolata quotes McCarthy as writing: "This is AI, so we don't care if it's psychologically real"[1] (https://books.google.com/books?id=PEkqAAAAMAAJ&q=%22we+don't+care+if+it's+psychologically+real%22&dq=%22we+don't+care+if+it's+psychologically+real%22&output=html&pgis=1). McCarthy recently reiterated his position at the AI@50 conference where he said "Artificial intelligence is not, by definition, simulation of human intelligence" (Maker 2006).

95. Neats vs. scruffies:
    - McCorduck 2004, pp. 421–424, 486–489
    - Crevier 1993, pp. 168
    - Nilsson 1983, pp. 10–11
96. Symbolic vs. sub-symbolic AI:
    - Nilsson (1998, p. 7), who uses the term "sub-symbolic".
97. Haugeland 1985, p. 255.
98. Law 1994.
99. Bach 2008.
100. Haugeland 1985, pp. 112–117
101. 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.
102. Cognitive simulation, Newell and Simon, AI at CMU (then called Carnegie Tech):
    - McCorduck 2004, pp. 139–179, 245–250, 322–323 (EPAM)
    - Crevier 1993, pp. 145–149
103. Soar (history):
    - McCorduck 2004, pp. 450–451
    - Crevier 1993, pp. 258–263
104. McCarthy and AI research at SAIL and SRI International:
    - McCorduck 2004, pp. 251–259
    - Crevier 1993
105. AI research at Edinburgh and in France, birth of Prolog:
    - Crevier 1993, pp. 193–196
    - Howe 1994
106. 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
107. 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
108. Knowledge revolution:
    - McCorduck 2004, pp. 266–276, 298–300, 314, 421
    - Russell & Norvig 2003, pp. 22–23
109. Embodied approaches to AI:
    - McCorduck 2004, pp. 454–462
    - Brooks 1990
    - Moravec 1988
110. Revival of connectionism:
    - Crevier 1993, pp. 214–215
    - Russell & Norvig 2003, p. 25
111. Computational intelligence
    - IEEE Computational Intelligence Society (http://www.ieee-cis.org/)
112. Hutter 2012.
113. Langley 2011.
114. Katz 2012.
115. Norvig 2012.
116. Agent architectures, hybrid intelligent systems:
    - Russell & Norvig (2003, pp. 27, 932, 970–972)
    - Nilsson (1998, chpt. 25)
117. Hierarchical control system:
    - Albus 2002
118. Subsumption architecture:
    - CITATION IN PROGRESS.

119. 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
120. 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
121. State space search and planning:
    - Russell & Norvig 2003, pp. 382–387
    - Poole, Mackworth & Goebel 1998, pp. 298–305
    - Nilsson 1998, chpt. 10.1–2
122. 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
123. 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
124. Optimization searches:
    - Russell & Norvig 2003, pp. 110–116,120–129
    - Poole, Mackworth & Goebel 1998, pp. 56–163
    - Luger & Stubblefield 2004, pp. 127–133
125. Artificial life and society based learning:
    - Luger & Stubblefield 2004, pp. 530–541
126. Genetic programming and genetic algorithms:
    - Luger & Stubblefield 2004, pp. 509–530,
    - Nilsson 1998, chpt. 4.2,
    - Holland 1975,
    - Koza 1992,
    - Poli, Langdon & McPhee 2008.
127. Logic:
    - ACM 1998, ~I.2.3,
    - Russell & Norvig 2003, pp. 194–310,
    - Luger & Stubblefield 2004, pp. 35–77,
    - Nilsson 1998, chpt. 13–16
128. Satplan:
    - Russell & Norvig 2003, pp. 402–407,
    - Poole, Mackworth & Goebel 1998, pp. 300–301,
    - Nilsson 1998, chpt. 21
129. 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
130. Propositional logic:
    - Russell & Norvig 2003, pp. 204–233,
    - Luger & Stubblefield 2004, pp. 45–50
    - Nilsson 1998, chpt. 13

131. 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
132. Fuzzy logic:
    - Russell & Norvig 2003, pp. 526–527
133. Subjective logic:
    - CITATION IN PROGRESS.
134. "The Belief Calculus and Uncertain Reasoning", Yen-Teh Hsia
135. 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
136. 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
137. 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
138. Bayesian learning and the expectation-maximization algorithm:
    - Russell & Norvig 2003, pp. 712–724,
    - Poole, Mackworth & Goebel 1998, pp. 424–433,
    - Nilsson 1998, chpt. 20
139. Bayesian decision theory and Bayesian decision networks:
    - Russell & Norvig 2003, pp. 597–600
140. Stochastic temporal models:
    - Russell & Norvig 2003, pp. 537–581

    Dynamic Bayesian networks:
    - Russell & Norvig 2003, pp. 551–557

    Hidden Markov model:
    - (Russell & Norvig 2003, pp. 549–551)

    Kalman filters:
    - Russell & Norvig 2003, pp. 551–557
141. decision theory and decision analysis:
    - Russell & Norvig 2003, pp. 584–597,
    - Poole, Mackworth & Goebel 1998, pp. 381–394
142. Markov decision processes and dynamic decision networks:
    - Russell & Norvig 2003, pp. 613–631
143. Game theory and mechanism design:
    - Russell & Norvig 2003, pp. 631–643
144. Statistical learning methods and classifiers:
    - Russell & Norvig 2003, pp. 712–754,
    - Luger & Stubblefield 2004, pp. 453–541
145. 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

146. kernel methods such as the support vector machine:
    - Russell & Norvig 2003, pp. 749–752
147. K-nearest neighbor algorithm:
    - Russell & Norvig 2003, pp. 733–736
148. Gaussian mixture model:
    - Russell & Norvig 2003, pp. 725–727
149. Naive Bayes classifier:
    - Russell & Norvig 2003, pp. 718
150. Decision tree:
    - Russell & Norvig 2003, pp. 653–664,
    - Poole, Mackworth & Goebel 1998, pp. 403–408,
    - Luger & Stubblefield 2004, pp. 408–417
151. Classifier performance:
    - van der Walt & Bernard 2006
152. Feedforward neural networks, perceptrons and radial basis networks:
    - Russell & Norvig 2003, pp. 739–748, 758
    - Luger & Stubblefield 2004, pp. 458–467
153. Competitive learning, Hebbian coincidence learning, Hopfield networks and attractor networks:
    - Luger & Stubblefield 2004, pp. 474–505
154. Seppo Linnainmaa (1970). The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors. Master's Thesis (in Finnish), Univ. Helsinki, 6-7.
155. Griewank, Andreas (2012). Who Invented the Reverse Mode of Differentiation?. Optimization Stories, Documenta Matematica, Extra Volume ISMP (2012), 389-400.
156. Paul Werbos, "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences," *PhD thesis, Harvard University*, 1974.
157. Paul Werbos (1982). Applications of advances in nonlinear sensitivity analysis. In System modeling and optimization (pp. 762-770). Springer Berlin Heidelberg. Online (http://werbos.com/Neural/SensitivityIFIPSeptember1981.pdf)
158. Backpropagation:
    - Russell & Norvig 2003, pp. 744–748,
    - Luger & Stubblefield 2004, pp. 467–474,
    - Nilsson 1998, chpt. 3.3
159. Hierarchical temporal memory:
    - Hawkins & Blakeslee 2005
160. Ian Goodfellow, Yoshua Bengio, and Aaron Courville (2016). Deep Learning. MIT Press. Online (http://www.deeplearningbook.org)
161. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.; Mohamed, A.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.; Kingsbury, B. (2012). "Deep Neural Networks for Acoustic Modeling in Speech Recognition --- The shared views of four research groups". *IEEE Signal Processing Magazine*. **29** (6): 82–97. doi:10.1109/msp.2012.2205597.
162. Schmidhuber, J. (2015). "Deep Learning in Neural Networks: An Overview". *Neural Networks*. **61**: 85–117. arXiv:1404.7828. doi:10.1016/j.neunet.2014.09.003.
163. Schmidhuber, Jürgen (2015). "Deep Learning". *Scholarpedia*. **10** (11): 32832. doi:10.4249/scholarpedia.32832.
164. Rina Dechter (1986). Learning while searching in constraint-satisfaction problems. University of California, Computer Science Department, Cognitive Systems Laboratory.Online (https://www.researchgate.net/publication/221605378_Learning_While_Searching_in_Constraint-Satisfaction-Problems)
165. Igor Aizenberg, Naum N. Aizenberg, Joos P.L. Vandewalle (2000). Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications. Springer Science & Business Media.
166. Ivakhnenko, Alexey (1965). *Cybernetic Predicting Devices*. Kiev: Naukova Dumka.
167. Ivakhnenko, Alexey (1971). "Polynomial theory of complex systems". *IEEE Transactions on Systems, Man and Cybernetics (4)*: 364–378.
168. Hinton 2007.
169. Research, AI (23 October 2015). "Deep Neural Networks for Acoustic Modeling in Speech Recognition". *airesearch.com*. Retrieved 23 October 2015.
170. Fukushima, K. (1980). "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". *Biol. Cybern*. **36**: 193–202. doi:10.1007/bf00344251.
171. Yann LeCun (2016). Slides on Deep Learning Online (https://indico.cern.ch/event/510372/)
172. "AlphaGo – Google DeepMind". Retrieved 30 January 2016.

173. Recurrent neural networks, Hopfield nets:
    - Russell & Norvig 2003, p. 758
    - Luger & Stubblefield 2004, pp. 474–505
174. P. J. Werbos. Generalization of backpropagation with application to a recurrent gas market model. Neural Networks, 1, 1988.
175. A. J. Robinson and F. Fallside. The utility driven dynamic error propagation network. Technical Report CUED/F-INFENG/TR.1, Cambridge University Engineering Department, 1987.
176. R. J. Williams and D. Zipser. Gradient-based learning algorithms for recurrent networks and their computational complexity. In Back-propagation: Theory, Architectures and Applications. Hillsdale, NJ: Erlbaum, 1994.
177. Sepp Hochreiter (1991), Untersuchungen zu dynamischen neuronalen Netzen (http://people.idsia.ch/~juergen/SeppHochreiter1991ThesisAdvisorSchmidhuber.pdf), Diploma thesis. Institut f. Informatik, Technische Univ. Munich. Advisor: J. Schmidhuber.
178. J. Schmidhuber., "Learning complex, extended sequences using the principle of history compression," *Neural Computation*, 4, pp. 234–242, 1992.
179. Hochreiter, Sepp; and Schmidhuber, Jürgen; *Long Short-Term Memory*, Neural Computation, 9(8):1735–1780, 1997
180. Alex Graves, Santiago Fernandez, Faustino Gomez, and Jürgen Schmidhuber (2006). Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural nets. Proceedings of ICML'06, pp. 369–376.
181. Awni Hannun, Carl Case, Jared Casper, Bryan Catanzaro, Greg Diamos, Erich Elsen, Ryan Prenger, Sanjeev Satheesh, Shubho Sengupta, Adam Coates, Andrew Ng (2014). Deep Speech: Scaling up end-to-end speech recognition. arXiv:1412.5567 (http://arxiv.org/abs/1412.5567)
182. Hasim Sak and Andrew Senior and Francoise Beaufays (2014). Long Short-Term Memory recurrent neural network architectures for large scale acoustic modeling. Proceedings of Interspeech 2014.
183. Xiangang Li, Xihong Wu (2015). Constructing Long Short-Term Memory based Deep Recurrent Neural Networks for Large Vocabulary Speech Recognition arXiv:1410.4281 (http://arxiv.org/abs/1410.4281)
184. Haşim Sak, Andrew Senior, Kanishka Rao, Françoise Beaufays and Johan Schalkwyk (September 2015): Google voice search: faster and more accurate. (http://googleresearch.blogspot.ch/2015/09/google-voice-search-faster-and-more.html)
185. Ilya Sutskever, Oriol Vinyals, and Quoc V. Le (2014). Sequence to Sequence Learning with Neural Networks. arXiv (http://arxiv.org/abs/1409.3215)
186. Rafal Jozefowicz, Oriol Vinyals, Mike Schuster, Noam Shazeer, Yonghui Wu (2016). Exploring the Limits of Language Modeling. arXiv (http://arxiv.org/abs/1602.02410)
187. Dan Gillick, Cliff Brunk, Oriol Vinyals, Amarnag Subramanya (2015). Multilingual Language Processing From Bytes. arXiv (http://arxiv.org/abs/1512.00103)
188. Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan (2015). Show and Tell: A Neural Image Caption Generator. arXiv (http://arxiv.org/abs/1411.4555)
189. Control theory:
    - ACM 1998, ~I.2.8,
    - Russell & Norvig 2003, pp. 926–932
190. Lisp:
    - Luger & Stubblefield 2004, pp. 723–821
    - Crevier 1993, pp. 59–62,
    - Russell & Norvig 2003, p. 18
191. Prolog:
    - Poole, Mackworth & Goebel 1998, pp. 477–491,
    - Luger & Stubblefield 2004, pp. 641–676, 575–581
192. The Turing test:
    Turing's original publication:
    - Turing 1950

    Historical influence and philosophical implications:
    - Haugeland 1985, pp. 6–9
    - Crevier 1993, p. 24
    - McCorduck 2004, pp. 70–71
    - Russell & Norvig 2003, pp. 2–3 and 948
193. Subject matter expert Turing test:
    - CITATION IN PROGRESS.
194. Rajani 2011.
195. Game AI:
    - CITATION IN PROGRESS.

196. Mathematical definitions of intelligence:
    - Hernandez-Orallo 2000
    - Dowe & Hajek 1997
    - Hernandez-Orallo & Dowe 2010
197. O'Brien & Marakas 2011.
198. Russell & Norvig 2009, p. 1.
199. *CNN* 2006.
200. N. Aletras, D. Tsarapatsanis, D. Preotiuc-Pietro, V. Lampos (2016). "Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective". *PeerJ Computer Science*.
201. "The Economist Explains: Why firms are piling into artificial intelligence". *The Economist*. 31 March 2016. Retrieved 19 May 2016.
202. Lohr, Steve (28 February 2016). "The Promise of Artificial Intelligence Unfolds in Small Steps". *New York Times*. Retrieved 29 February 2016.
203. Wakefield, Jane (June 15, 2016). "Social media 'outstrips TV' as news source for young people". *BBC News*.
204. Smith, Mark (July 22, 2016). "So you think you chose to read this article?". *BBC News*.
205. Brooks 1991.
206. "Hacking Roomba". *hackingroomba.com*.
207. Dartmouth proposal:
    - McCarthy et al. 1955 (the original proposal)
    - Crevier 1993, p. 49 (historical significance)
208. The physical symbol systems hypothesis:
    - Newell & Simon 1976, p. 116
    - McCorduck 2004, p. 153
    - Russell & Norvig 2003, p. 18
209. 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)
210. Dreyfus' critique of artificial intelligence:
    - Dreyfus 1972, Dreyfus & Dreyfus 1986
    - Crevier 1993, pp. 120–132
    - McCorduck 2004, pp. 211–239
    - Russell & Norvig 2003, pp. 950–952,
211. Gödel 1951: in this lecture, Kurt Gödel uses the incompleteness theorem to arrive at the following disjunction: (a) the human mind is not a consistent finite machine, or (b) there exist Diophantine equations for which it cannot decide whether solutions exist. Gödel finds (b) implausible, and thus seems to have believed the human mind was not equivalent to a finite machine, i.e., its power exceeded that of any finite machine. He recognized that this was only a conjecture, since one could never disprove (b). Yet he considered the disjunctive conclusion to be a "certain fact".
212. The Mathematical Objection:
    - Russell & Norvig 2003, p. 949
    - McCorduck 2004, pp. 448–449

    Making the Mathematical Objection:
    - Lucas 1961
    - Penrose 1989

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

    Background:
    - Gödel 1931, Church 1936, Kleene 1935, Turing 1937
213. Graham Oppy (20 January 2015). "Gödel's Incompleteness Theorems". *Stanford Encyclopedia of Philosophy*. Retrieved 27 April 2016. "These Gödelian anti-mechanist arguments are, however, problematic, and there is wide consensus that they fail."
214. Stuart J. Russell; Peter Norvig (2010). "26.1.2: Philosophical Foundations/Weak AI: Can Machines Act Intelligently?/The mathematical objection". *Artificial Intelligence: A Modern Approach* (3rd ed.). Upper Saddle River, NJ: Prentice Hall. ISBN 0-13-604259-7. "...even if we grant that computers have limitations on what they can prove, there is no evidence that humans are immune from those limitations."

215. Mark Colyvan. An introduction to the philosophy of mathematics. Cambridge University Press, 2012. From 2.2.2, 'Philosophical significance of Gödel's incompleteness results': "The accepted wisdom (with which I concur) is that the Lucas-Penrose arguments fail."
216. Wendell Wallach (2010). *Moral Machines*, Oxford University Press.
217. Wallach, pp 37–54.
218. Wallach, pp 55–73.
219. Wallach, Introduction chapter.
220. Michael Anderson and Susan Leigh Anderson (2011), Machine Ethics, Cambridge University Press.
221. "Machine Ethics". *aaai.org*.
222. Amitai Etzioni and Oren Etzioni (2016), "Keeping AI Legal", Vanderbilt Journal of Entertainment & Technology Law.
223. Rubin, Charles (Spring 2003). "Artificial Intelligence and Human Nature". *The New Atlantis*. **1**: 88–100. Archived from the original on 11 June 2012.
224. Rawlinson, Kevin. "Microsoft's Bill Gates insists AI is a threat". BBC News. Retrieved 30 January 2015.
225. Brooks, Rodney (10 November 2014). "artificial intelligence is a tool, not a threat".
226. In the early 1970s, Kenneth Colby presented a version of Weizenbaum's ELIZA known as DOCTOR which he promoted as a serious therapeutic tool. (Crevier 1993, pp. 132–144)
227. 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.
228. Ford & 2009 The lights in the tunnel.
229. "Machine Learning: A job killer?". *econfuture - Robots, AI and Unemployment - Future Economics and Technology*.
230. AI could decrease the demand for human labor:
    - Russell & Norvig 2003, pp. 960–961
    - Ford, Martin (2009). *The Lights in the Tunnel: Automation, Accelerating Technology and the Economy of the Future*. Acculant Publishing. ISBN 978-1-4486-5981-4.
231. Horst, Steven, (2005) "The Computational Theory of Mind" (http://plato.stanford.edu/entries/computational-mind/) in *The Stanford Encyclopedia of Philosophy*
232. 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."
233. Searle's Chinese room argument:
    - Searle 1980. Searle's original presentation of the thought experiment.
    - Searle 1999.

    Discussion:
    - Russell & Norvig 2003, pp. 958–960
    - McCorduck 2004, pp. 443–445
    - Crevier 1993, pp. 269–271
234. Robot rights:
    - Russell & Norvig 2003, p. 964
    - *BBC News* 2006

    Prematurity of:
    - Henderson 2007

    In fiction:
    - 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.
235. maschafilm. "Content: Plug & Pray Film - Artificial Intelligence - Robots -". *plugandpray-film.de*.
236. Omohundro, Steve (2008). *The Nature of Self-Improving Artificial Intelligence*. presented and distributed at the 2007 Singularity Summit, San Francisco, CA.

237. Technological singularity:
    - Vinge 1993
    - Kurzweil 2005
    - Russell & Norvig 2003, p. 963
238. Lemmons, Phil (April 1985). "Artificial Intelligence". *BYTE*. p. 125. Retrieved 14 February 2015.
239. Transhumanism:
    - Moravec 1988
    - Kurzweil 2005
    - Russell & Norvig 2003, p. 963
240. AI as evolution:
    - Edward Fredkin is quoted in McCorduck (2004, p. 401).
    - Butler 1863
    - Dyson 1998
241. "Stephen Hawking warns artificial intelligence could end mankind". *BBC News*. Retrieved 2015-10-30.
242. Holley, Peter (28 January 2015). "Bill Gates on dangers of artificial intelligence: 'I don't understand why some people are not concerned' ". *The Washington Post*. ISSN 0190-8286. Retrieved 2015-10-30.
243. Gibbs, Samuel. "Elon Musk: artificial intelligence is our biggest existential threat". *the Guardian*. Retrieved 2015-10-30.
244. Post, Washington. "Tech titans like Elon Musk are spending $1 billion to save you from terminators".
245. Müller, Vincent C.; Bostrom, Nick (2014). "Future Progress in Artificial Intelligence: A Poll Among Experts" (PDF). *AI Matters*. **1** (1): 9–11. doi:10.1145/2639475.2639478.
246. "Is artificial intelligence really an existential threat to humanity?". *Bulletin of the Atomic Scientists*. Retrieved 2015-10-30.
247. "The case against killer robots, from a guy actually working on artificial intelligence". *Fusion.net*. Retrieved 2016-01-31.
248. "Will artificial intelligence destroy humanity? Here are 5 reasons not to worry.". *Vox*. Retrieved 2015-10-30.
249. "The mysterious artificial intelligence company Elon Musk invested in is developing game-changing smart computers". *Tech Insider*. Retrieved 2015-10-30.
250. Clark, Jack. "Musk-Backed Group Probes Risks Behind Artificial Intelligence". *Bloomberg.com*. Retrieved 2015-10-30.
251. "Elon Musk Is Donating $10M Of His Own Money To Artificial Intelligence Research". *Fast Company*. Retrieved 2015-10-30.
252. "Stephen Hawking, Elon Musk, and Bill Gates Warn About Artificial Intelligence". *Observer*. Retrieved 2015-10-30.
253. Wallach, Wendell and Colin Allen. 2009. Moral Machines: Teaching Robots Right from Wrong (Oxford University Press: Oxford).
254. Anderson, Michael and Susan Anderson(Eds.) 2011. Machine Ethics. Cambridge University Press
255. Etzioni, Amitai; Etzioni, Oren (2016). "AI Assisted Ethics". *Ethics and Information Technology*. **18**: 149–156. doi:10.1007/s10676-016-9400-6.

References

AI textbooks

- Hutter, Marcus (2005). *Universal Artificial Intelligence*. Berlin: Springer. ISBN 978-3-540-22139-5.
- Luger, George; Stubblefield, William (2004). *Artificial Intelligence: Structures and Strategies for Complex Problem Solving* (5th ed.). Benjamin/Cummings. ISBN 0-8053-4780-1.
- Neapolitan, Richard; Jiang, Xia (2012). *Contemporary Artificial Intelligence*. Chapman & Hall/CRC. ISBN 978-1-4398-4469-4.
- Nilsson, Nils (1998). *Artificial Intelligence: A New Synthesis*. Morgan Kaufmann. ISBN 978-1-55860-467-4.
- Russell, Stuart J.; Norvig, Peter (2003), *Artificial Intelligence: A Modern Approach* (2nd ed.), Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13-790395-2.
- Russell, Stuart J.; Norvig, Peter (2009), *Artificial Intelligence: A Modern Approach* (3rd ed.), Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13-604259-7.
- Poole, David; Mackworth, Alan; Goebel, Randy (1998). *Computational Intelligence: A Logical Approach*. New York: Oxford University Press. ISBN 0-19-510270-3.
- Winston, Patrick Henry (1984). *Artificial Intelligence*. Reading, MA: Addison-Wesley. ISBN 0-201-08259-4.
- Rich, Elaine (1983). *Artificial Intelligence*. McGraw-Hill. ISBN 0-07-052261-8.
- Bundy, Alan (1980). *Artificial Intelligence: An Introductory Course* (2nd ed.). Edinburgh University Press. ISBN 0-85224-410-X.

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.
- Newquist, HP (1994). *The Brain Makers: Genius, Ego, And Greed In The Quest For Machines That Think*. New York: Macmillan/SAMS. ISBN 0-672-30412-0.
- Nilsson, Nils (2009). *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. New York: Cambridge University Press. ISBN 978-0-521-12293-1.

Other sources

- Asada, M.; Hosoda, K.; Kuniyoshi, Y.; Ishiguro, H.; Inui, T.; Yoshikawa, Y.; Ogino, M.; Yoshida, C. (2009). "Cognitive developmental robotics: a survey" (PDF). *IEEE Transactions on Autonomous Mental Development*. **1** (1): 12–34. doi:10.1109/tamd.2009.2021702. Archived from the original (PDF) on 4 October 2013.
- "ACM Computing Classification System: Artificial intelligence". ACM. 1998. Retrieved 30 August 2007.
- Goodman, Joanna (2016). *Robots in Law: How Artificial Intelligence is Transforming Legal Services* (1st ed.). Ark Group. ISBN 978-1-78358-264-8.
- Albus, J. S. (2002). "4-D/RCS: A Reference Model Architecture for Intelligent Unmanned Ground Vehicles" (PDF). In Gerhart, G.; Gunderson, R.; Shoemaker, C. *Proceedings of the SPIE AeroSense Session on Unmanned Ground Vehicle Technology*. **3693**. pp. 11–20.
- Aleksander, Igor (1995). *Artificial Neuroconsciousness: An Update*. IWANN. Archived from the original on 2 March 1997. BibTex (http://dblp.uni-trier.de/rec/bibtex/conf/iwann/Aleksander95) Archived (https://web.archive.org/web/19970302014628/http://www.ee.ic.ac.uk/research/neural/publications/iwann.html) 2 March 1997 at the Wayback Machine..
- Bach, Joscha (2008). "Seven Principles of Synthetic Intelligence". In Wang, Pei; Goertzel, Ben; Franklin, Stan. *Artificial General Intelligence, 2008: Proceedings of the First AGI Conference*. IOS Press. pp. 63–74. ISBN 978-1-58603-833-5.
- "Robots could demand legal rights". *BBC News*. 21 December 2006. Retrieved 3 February 2011.
- Brooks, Rodney (1990). "Elephants Don't Play Chess" (PDF). *Robotics and Autonomous Systems*. **6**: 3–15. doi:10.1016/S0921-8890(05)80025-9. Archived (PDF) from the original on 9 August 2007.
- Brooks, R. A. (1991). "How to build complete creatures rather than isolated cognitive simulators". In VanLehn, K. *Architectures for Intelligence*. Hillsdale, NJ: Lawrence Erlbaum Associates. pp. 225–239.
- Buchanan, Bruce G. (2005). "A (Very) Brief History of Artificial Intelligence" (PDF). *AI Magazine*: 53–60. Archived from the original (PDF) on 26 September 2007.
- Butler, Samuel (13 June 1863). "Darwin among the Machines". Letters to the Editor. *The Press*. Christchurch, New Zealand. Retrieved 16 October 2014 – via Victoria University of Wellington.
- "AI set to exceed human brain power". *CNN*. 26 July 2006. Archived from the original on 19 February 2008.
- Dennett, Daniel (1991). *Consciousness Explained*. The Penguin Press. ISBN 0-7139-9037-6.
- Diamond, David (December 2003). "The Love Machine; Building computers that care". Wired. Archived from the original on 18 May 2008.
- Dowe, D. L.; Hajek, A. R. (1997). "A computational extension to the Turing Test". *Proceedings of the 4th Conference of the Australasian Cognitive Science Society*.
- Dreyfus, Hubert (1972). *What Computers Can't Do*. New York: MIT Press. ISBN 0-06-011082-1.
- Dreyfus, Hubert; Dreyfus, Stuart (1986). *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. Oxford, UK: Blackwell. ISBN 0-02-908060-6.
- Dreyfus, Hubert (1992). *What Computers* Still *Can't Do*. New York: MIT Press. ISBN 0-262-54067-3.
- Dyson, George (1998). *Darwin among the Machines*. Allan Lane Science. ISBN 0-7382-0030-1.
- Edelman, Gerald (23 November 2007). "Gerald Edelman – Neural Darwinism and Brain-based Devices". Talking Robots. Archived from the original on 8 October 2009.
- Edelson, Edward (1991). *The Nervous System*. New York: Chelsea House. ISBN 978-0-7910-0464-7.
- Fearn, Nicholas (2007). *The Latest Answers to the Oldest Questions: A Philosophical Adventure with the World's Greatest Thinkers*. New York: Grove Press. ISBN 0-8021-1839-9.
- Gladwell, Malcolm (2005). *Blink*. New York: Little, Brown and Co. ISBN 0-316-17232-4.
- Gödel, Kurt (1951). *Some basic theorems on the foundations of mathematics and their implications*. Gibbs Lecture. In Feferman, Solomon, ed. (1995). *Kurt Gödel: Collected Works, Vol. III: Unpublished Essays and Lectures*. Oxford University Press. pp. 304–23. ISBN 978-0-19-514722-3.

- Haugeland, John (1985). *Artificial Intelligence: The Very Idea*. Cambridge, Mass.: MIT Press. ISBN 0-262-08153-9.
- Hawkins, Jeff; Blakeslee, Sandra (2005). *On Intelligence*. New York, NY: Owl Books. ISBN 0-8050-7853-3.
- Henderson, Mark (24 April 2007). "Human rights for robots? We're getting carried away". *The Times Online*. London.
- Hernandez-Orallo, Jose (2000). "Beyond the Turing Test". *Journal of Logic, Language and Information*. **9** (4): 447–466. doi:10.1023/A:1008367325700.
- Hernandez-Orallo, J.; Dowe, D. L. (2010). "Measuring Universal Intelligence: Towards an Anytime Intelligence Test". *Artificial Intelligence Journal*. **174** (18): 1508–1539. doi:10.1016/j.artint.2010.09.006.
- Hinton, G. E. (2007). "Learning multiple layers of representation". *Trends in Cognitive Sciences*. **11**: 428–434. doi:10.1016/j.tics.2007.09.004.
- Hofstadter, Douglas (1979). *Gödel, Escher, Bach: an Eternal Golden Braid*. New York, NY: Vintage Books. ISBN 0-394-74502-7.
- Holland, John H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press. ISBN 0-262-58111-6.
- Howe, J. (November 1994). "Artificial Intelligence at Edinburgh University: a Perspective". Retrieved 30 August 2007.
- Hutter, M. (2012). "One Decade of Universal Artificial Intelligence". *Theoretical Foundations of Artificial General Intelligence*. Atlantis Thinking Machines. **4**. doi:10.2991/978-94-91216-62-6_5. ISBN 978-94-91216-61-9.
- James, William (1884). "What is Emotion". *Mind*. **9**: 188–205. doi:10.1093/mind/os-IX.34.188. Cited by Tao & Tan 2005.
- Kahneman, Daniel; Slovic, D.; Tversky, Amos (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press. ISBN 0-521-28414-7.
- Katz, Yarden (1 November 2012). "Noam Chomsky on Where Artificial Intelligence Went Wrong". *The Atlantic*. Retrieved 26 October 2014.
- "Kismet". MIT Artificial Intelligence Laboratory, Humanoid Robotics Group. Retrieved 25 October 2014.
- Koza, John R. (1992). *Genetic Programming (On the Programming of Computers by Means of Natural Selection)*. MIT Press. ISBN 0-262-11170-5.
- Kleine-Cosack, Christian (October 2006). "Recognition and Simulation of Emotions" (PDF). Archived from the original (PDF) on 28 May 2008.
- Kolata, G. (1982). "How can computers get common sense?". *Science*. **217** (4566): 1237–1238. doi:10.1126/science.217.4566.1237. PMID 17837639.
- Kumar, Gulshan; Kumar, Krishan (2012). "The Use of Artificial-Intelligence-Based Ensembles for Intrusion Detection: A Review". *Applied Computational Intelligence and Soft Computing*. **2012**: 1–20. doi:10.1155/2012/850160.
- 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; Núñez, Rafael E. (2000). *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. Basic Books. ISBN 0-465-03771-2.
- Langley, Pat (2011). "The changing science of machine learning". *Machine Learning*. **82** (3): 275–279. doi:10.1007/s10994-011-5242-y.
- Law, Diane (June 1994). *Searle, Subsymbolic Functionalism and Synthetic Intelligence* (Technical report). University of Texas at Austin. p. AI94-222. CiteSeerX 10.1.1.38.8384.
- Legg, Shane; Hutter, Marcus (15 June 2007). *A Collection of Definitions of Intelligence* (Technical report). IDSIA. arXiv:0706.3639. 07-07.
- Lenat, Douglas; Guha, R. V. (1989). *Building Large Knowledge-Based Systems*. Addison-Wesley. ISBN 0-201-51752-3.
- Lighthill, 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*. Archived from the original on 19 August 2007. Retrieved 30 August 2007.
- Lungarella, M.; Metta, G.; Pfeifer, R.; Sandini, G. (2003). "Developmental robotics: a survey". *Connection Science*. **15**: 151–190. CiteSeerX 10.1.1.83.7615. doi:10.1080/09540090310001655110.
- Maker, Meg Houston (2006). "AI@50: AI Past, Present, Future". Dartmouth College. Retrieved 16 October 2008.
- Markoff, John (16 February 2011). "Computer Wins on 'Jeopardy!': Trivial, It's Not". *The New York Times*. Retrieved 25 October 2014.
- McCarthy, John; Minsky, Marvin; Rochester, Nathan; Shannon, Claude (1955). "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence". Archived from the original on 26 August 2007. Retrieved 30 August 2007..
- McCarthy, John; Hayes, P. J. (1969). "Some philosophical problems from the standpoint of artificial intelligence". *Machine Intelligence*. **4**: 463–502. Archived from the original on 10 August 2007. Retrieved 30 August 2007.
- McCarthy, John (12 November 2007). "What Is Artificial Intelligence?".

- Minsky, Marvin (1967). *Computation: Finite and Infinite Machines*. Englewood Cliffs, N.J.: Prentice-Hall. ISBN 0-13-165449-7.
- Minsky, Marvin (2006). *The Emotion Machine*. New York, NY: Simon & Schusterl. ISBN 0-7432-7663-9.
- Moravec, Hans (1988). *Mind Children*. Harvard University Press. ISBN 0-674-57616-0.
- Norvig, Peter (25 June 2012). "On Chomsky and the Two Cultures of Statistical Learning". Peter Norvig. Archived from the original on 19 October 2014.
- NRC (United States National Research Council) (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. (1976). "Computer Science as Empirical Inquiry: Symbols and Search". *Communications of the ACM*. **19** (3): 113–126. doi:10.1145/360018.360022. Archived from the original on 7 October 2008..
- Nilsson, Nils (1983). "Artificial Intelligence Prepares for 2001" (PDF). *AI Magazine*. **1** (1). Presidential Address to the Association for the Advancement of Artificial Intelligence.
- O'Brien, James; Marakas, George (2011). *Management Information Systems* (10th ed.). McGraw-Hill/Irwin. ISBN 978-0-07-337681-3.
- O'Connor, Kathleen Malone (1994). "The alchemical creation of life (takwin) and other concepts of Genesis in medieval Islam". University of Pennsylvania.
- Oudeyer, P-Y. (2010). "On the impact of robotics in behavioral and cognitive sciences: from insect navigation to human cognitive development" (PDF). *IEEE Transactions on Autonomous Mental Development*. **2** (1): 2–16. doi:10.1109/tamd.2009.2039057.
- Penrose, Roger (1989). *The Emperor's New Mind: Concerning Computer, Minds and The Laws of Physics*. Oxford University Press. ISBN 0-19-851973-7.
- Picard, Rosalind (1995). *Affective Computing* (PDF) (Technical report). MIT. 321. Lay summary – *Abstract*.
- Poli, R.; Langdon, W. B.; McPhee, N. F. (2008). *A Field Guide to Genetic Programming*. Lulu.com. ISBN 978-1-4092-0073-4 – via gp-field-guide.org.uk.
- Rajani, Sandeep (2011). "Artificial Intelligence – Man or Machine" (PDF). *International Journal of Information Technology and Knowledge Management*. **4** (1): 173–176.
- Searle, John (1980). "Minds, Brains and Programs". *Behavioral and Brain Sciences*. **3** (3): 417–457. doi:10.1017/S0140525X00005756.
- Searle, John (1999). *Mind, language and society*. New York, NY: Basic Books. ISBN 0-465-04521-9. OCLC 231867665.
- Shapiro, Stuart C. (1992). "Artificial Intelligence". In Shapiro, Stuart C. *Encyclopedia of Artificial Intelligence* (PDF) (2nd ed.). New York: John Wiley. pp. 54–57. ISBN 0-471-50306-1.
- Simon, H. A. (1965). *The Shape of Automation for Men and Management*. New York: Harper & Row.
- Skillings, Jonathan (3 July 2006). "Getting Machines to Think Like Us". *cnet*. Retrieved 3 February 2011.
- Solomonoff, Ray (1956). *An Inductive Inference Machine* (PDF). Dartmouth Summer Research Conference on Artificial Intelligence – via std.com, pdf scanned copy of the original. Later published as
  Solomonoff, Ray (1957). "An Inductive Inference Machine". *IRE Convention Record*. Section on Information Theory, part 2. pp. 56–62.
- Tao, Jianhua; Tan, Tieniu (2005). *Affective Computing and Intelligent Interaction*. Affective Computing: A Review. LNCS 3784. Springer. pp. 981–995. doi:10.1007/11573548.
- Tecuci, Gheorghe (March–April 2012). "Artificial Intelligence". *Wiley Interdisciplinary Reviews: Computational Statistics*. Wiley. **4** (2): 168–180. doi:10.1002/wics.200.
- Thro, Ellen (1993). *Robotics: The Marriage of Computers and Machines*. New York: Facts on File. ISBN 978-0-8160-2628-9.
- Turing, Alan (October 1950), "Computing Machinery and Intelligence", *Mind*, **LIX** (236): 433–460, doi:10.1093/mind/LIX.236.433, ISSN 0026-4423, retrieved 2008-08-18.
- van der Walt, Christiaan; Bernard, Etienne (2006). "Data characteristics that determine classifier performance" (PDF). Retrieved 5 August 2009.
- Vinge, Vernor (1993). "The Coming Technological Singularity: How to Survive in the Post-Human Era".
- 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 0-7167-0464-1.
- Weng, J.; McClelland; Pentland, A.; Sporns, O.; Stockman, I.; Sur, M.; Thelen, E. (2001). "Autonomous mental development by robots and animals" (PDF). *Science*. **291**: 599–600. doi:10.1126/science.291.5504.599 – via msu.edu.

[1]

Further reading

- TechCast Article Series, John Sagi, Framing Consciousness (https://www.techcastglobal.com/documents/10193/34869/Consciousness-Sagifinalversion)
- Boden, Margaret, Mind As Machine, Oxford University Press, 2006
- Johnston, John (2008) "The Allure of Machinic Life: Cybernetics, Artificial Life, and the New AI", MIT Press
- Myers, Courtney Boyd ed. (2009). The AI Report (http://www.forbes.com/2009/06/22/singularity-robots-computers-opinions-contributors-artificial-intelligence-09_land.html). Forbes June 2009
- Raphael, Bertram (1976). *The Thinking Computer*. W.H.Freeman and Company. ISBN 0-7167-0723-3.
- Serenko, Alexander (2010). "The development of an AI journal ranking based on the revealed preference approach" (PDF). *Journal of Informetrics*. **4** (4): 447–459. doi:10.1016/j.joi.2010.04.001.
- Serenko, Alexander; Michael Dohan (2011). "Comparing the expert survey and citation impact journal ranking methods: Example from the field of Artificial Intelligence" (PDF). *Journal of Informetrics*. **5** (4): 629–649. doi:10.1016/j.joi.2011.06.002.
- Sun, R. & Bookman, L. (eds.), *Computational Architectures: Integrating Neural and Symbolic Processes*. Kluwer Academic Publishers, Needham, MA. 1994.
- Tom Simonite (29 December 2014). "2014 in Computing: Breakthroughs in Artificial Intelligence". *MIT Technology Review*.

External links

- What Is AI? (http://www-formal.stanford.edu/jmc/whatisai/whatisai.html) – An introduction to artificial intelligence by John McCarthy—a co-founder of the field, and the person who coined the term.
- The Handbook of Artificial Intelligence Volume I by Avron Barr and Edward A. Feigenbaum (Stanford University) (https://archive.org/details/handbookofartific01barr/)
- "Artificial Intelligence". *Internet Encyclopedia of Philosophy*.
- Thomason, Richmond. "Logic and Artificial Intelligence". *Stanford Encyclopedia of Philosophy*.
- AI (https://www.dmoz.org/Computers/Artificial_Intelligence/) at DMOZ
- AITopics (http://aitopics.org/) – A large directory of links and other resources maintained by the Association for the Advancement of Artificial Intelligence, the leading organization of academic AI researchers.

1. "Applications of AI". *www-formal.stanford.edu*. Retrieved 2016-09-25.