

Artificial general intelligence

Artificial general intelligence (**AGI**) is the hypothetical^[1] intelligence of a machine that has the capacity to understand or learn any intellectual task that a human being can. It is a primary goal of some artificial intelligence research and a common topic in science fiction and futures studies. AGI can also be referred to as **strong AI**,^{[2][3][4]} **full AI**,^[5] or **general intelligent action**.^[6] Some academic sources reserve the term "strong AI" for machines that can experience consciousness.^[7] Today's AI is speculated to be decades away from AGI.^{[8][9]}

In contrast to strong AI, weak AI^[10] (also called narrow AI^[3]) is not intended to perform human cognitive abilities, rather, weak AI is limited to the use of software to study or accomplish specific problem solving or reasoning tasks.

As of 2017, over forty organizations are actively researching AGI.^[11]

Contents

Requirements

- Tests for confirming human-level AGI
- Problems requiring AGI to solve

History

- Classical AI
- Narrow AI research
- Modern artificial general intelligence research

Processing power needed to simulate a brain

- Whole brain emulation
- Early estimates
- Modelling the neurons in more detail
- Current research
- Criticisms of simulation-based approaches

Strong AI and consciousness

- Consciousness
- Artificial consciousness research

Possible explanations for the slow progress of AI research

Controversies and dangers

- Feasibility
- Potential threat to human existence

See also

Notes

References

External links

Requirements

Various criteria for intelligence have been proposed (most famously the Turing test) but to date, there is no definition that satisfies everyone.^[12] However, there is wide agreement among artificial intelligence researchers that intelligence is required to do the following:^[13]

- reason, use strategy, solve puzzles, and make judgments under uncertainty;
- represent knowledge, including commonsense knowledge;
- plan;
- learn;
- communicate in natural language;
- and integrate all these skills towards common goals.

Other important capabilities include the ability to sense (e.g. see) and the ability to act (e.g. move and manipulate objects) in the world where intelligent behaviour is to be observed.^[14] This would include an ability to detect and respond to hazard.^[15] Many interdisciplinary approaches to intelligence (e.g. cognitive science, computational intelligence and decision making) tend to emphasise the need to consider additional traits such as imagination (taken as the ability to form mental images and concepts that were not programmed in)^[16] and autonomy.^[17] Computer based systems that exhibit many of these capabilities do exist (e.g. see computational creativity, automated reasoning, decision support system, robot, evolutionary computation, intelligent agent), but not yet at human levels.

Tests for confirming human-level AGI

The following tests to confirm human-level AGI have been considered:^{[18][19]}

The Turing Test (*Turing*)

A machine and a human both converse sight unseen with a second human, who must evaluate which of the two is the machine, which passes the test if it can fool the evaluator a significant fraction of the time. Note: Turing does not prescribe what should qualify as intelligence, only that knowing that it is a machine should disqualify it.

The Coffee Test (*Wozniak*)

A machine is required to enter an average American home and figure out how to make coffee: find the coffee machine, find the coffee, add water, find a mug, and brew the coffee by pushing the proper buttons.

The Robot College Student Test (*Goertzel*)

A machine enrolls in a university, taking and passing the same classes that humans would, and obtaining a degree.

The Employment Test (*Nilsson*)

A machine works an economically important job, performing at least as well as humans in the same job.

Problems requiring AGI to solve

The most difficult problems for computers are informally known as "AI-complete" or "AI-hard", implying that solving them is equivalent to the general aptitude of human intelligence, or strong AI, beyond the capabilities of a purpose-specific algorithm.^[20]

AI-complete problems are hypothesised to include general computer vision, natural language understanding, and dealing with unexpected circumstances while solving any real world problem.^[21]

AI-complete problems cannot be solved with current computer technology alone, and also require human computation. This property could be useful, for example, to test for the presence of humans, as CAPTCHAs aim to do; and for computer security to repel brute-force attacks.^{[22][23]}

History

Classical AI

Modern AI research began in the mid 1950s.^[24] The first generation of AI researchers were convinced that artificial general intelligence was possible and that it would exist in just a few decades. AI pioneer Herbert A. Simon wrote in 1965: "machines will be capable, within twenty years, of doing any work a man can do."^[25] Their predictions were the inspiration for Stanley Kubrick and Arthur C. Clarke's character HAL 9000, who embodied what AI researchers believed they could create by the year 2001. AI pioneer Marvin Minsky was a consultant^[26] on the project of making HAL 9000 as realistic as possible according to the consensus predictions of the time; Crevier quotes him as having said on the subject in 1967, "Within a generation ... the problem of creating 'artificial intelligence' will substantially be solved,"^[27] although Minsky states that he was misquoted.

However, in the early 1970s, it became obvious that researchers had grossly underestimated the difficulty of the project. Funding agencies became skeptical of AGI and put researchers under increasing pressure to produce useful "applied AI".^[28] As the 1980s began, Japan's Fifth Generation Computer Project revived interest in AGI, setting out a ten-year timeline that included AGI goals like "carry on a casual conversation".^[29] In response to this and the success of expert systems, both industry and government pumped money back into the field.^[30] However, confidence in AI spectacularly collapsed in the late 1980s, and the goals of the Fifth Generation Computer Project were never fulfilled.^[31] For the second time in 20 years, AI researchers who had predicted the imminent achievement of AGI had been shown to be fundamentally mistaken. By the 1990s, AI researchers had gained a reputation for making vain promises. They became reluctant to make predictions at all^[32] and to avoid any mention of "human level" artificial intelligence for fear of being labeled "wild-eyed dreamer[s]".^[33]

Narrow AI research

In the 1990s and early 21st century, mainstream AI achieved far greater commercial success and academic respectability by focusing on specific sub-problems where they can produce verifiable results and commercial applications, such as artificial neural networks and statistical machine learning.^[34] These "applied AI" systems are now used extensively throughout the technology industry, and research in this vein is very heavily funded in both academia and industry. Currently, development on this field is considered an emerging trend, and a mature stage is expected to happen in more than 10 years.^[35]

Most mainstream AI researchers hope that strong AI can be developed by combining the programs that solve various sub-problems. Hans Moravec wrote in 1988:

"I am confident that this bottom-up route to artificial intelligence will one day meet the traditional top-down route more than half way, ready to provide the real world competence and the commonsense knowledge that has been so frustratingly elusive in reasoning programs. Fully intelligent machines will result when the metaphorical golden spike is driven uniting the two efforts."^[36]

However, even this fundamental philosophy has been disputed; for example, Stevan Harnad of Princeton concluded his 1990 paper on the Symbol Grounding Hypothesis by stating:

"The expectation has often been voiced that "top-down" (symbolic) approaches to modeling cognition will somehow meet "bottom-up" (sensory) approaches somewhere in between. If the grounding considerations in this paper are valid, then this expectation is hopelessly modular and there is really only one viable route from sense to symbols: from the ground up. A free-floating symbolic level like the software level of a computer will never be reached by this route (or vice versa) – nor is it clear why we should even try to reach such a level, since it looks as if getting there would just amount to uprooting our symbols from their intrinsic meanings (thereby merely reducing ourselves to the functional equivalent of a programmable computer)."^[37]

Modern artificial general intelligence research

The term "artificial general intelligence" was used as early as 1997, by Mark Gubrud^[38] in a discussion of the implications of fully automated military production and operations. The term was re-introduced and popularized by Shane Legg and Ben Goertzel around 2002.^[39] The research objective is much older, for example Doug Lenat's Cyc project (that began in 1984), and Allen Newell's Soar project are regarded as within the scope of AGI. AGI research activity in 2006 was described by Pei Wang and Ben Goertzel^[40] as "producing publications and preliminary results". The first summer school in AGI was organized in Xiamen, China in 2009^[41] by the Xiamen university's Artificial Brain Laboratory and OpenCog. The first university course was given in 2010^[42] and 2011^[43] at Plovdiv University, Bulgaria by Todor Arnaudov. MIT presented a course in AGI in 2018, organized by Lex Fridman and featuring a number of guest lecturers. However, as yet, most AI researchers have devoted little attention to AGI, with some claiming that intelligence is too complex to be completely replicated in the near term. However, a small number of computer scientists are active in AGI research, and many of this group are contributing to a series of AGI conferences. The research is extremely diverse and often pioneering in nature. In the introduction to his book,^[44] Goertzel says that estimates of the time needed before a truly flexible AGI is built vary from 10 years to over a century, but the consensus in the AGI research community seems to be that the timeline discussed by Ray Kurzweil in The Singularity is Near^[45] (i.e. between 2015 and 2045) is plausible.^[46]

However, mainstream AI researchers have given a wide range of opinions on whether progress will be this rapid. A 2012 meta-analysis of 95 such opinions found a bias towards predicting that the onset of AGI would occur within 16–26 years for modern and historical predictions alike. It was later found that the dataset listed some experts as non-experts and vice versa.^[47]

Organizations explicitly pursuing AGI include the Swiss AI lab IDSIA,^[48] Nnaisense,^[49] Vicarious, Maluuba,^[11] the OpenCog Foundation, Adaptive AI, LIDA, and Numenta and the associated Redwood Neuroscience Institute.^[50] In addition, organizations such as the Machine Intelligence Research Institute^[51] and OpenAI^[52] have been founded to influence the development path of AGI. Finally, projects such as the Human Brain Project^[53] have the goal of building a functioning simulation of the human brain. A 2017 survey of AGI categorized forty-five known "active R&D projects" that explicitly or implicitly (through published research) research AGI, with the largest three being DeepMind, the Human Brain Project, and OpenAI.^[11]

In 2017, Ben Goertzel founded the AI platform SingularityNET with the aim of facilitating democratic, decentralized control of AGI when it arrives.^[54]

In 2017, researchers Feng Liu, Yong Shi and Ying Liu conducted intelligence tests on publicly available and freely accessible weak AI such as Google AI or Apple's Siri and others. At the maximum, these AI reached an IQ value of about 47, which corresponds approximately to a six-year-old child in first grade. An adult comes to about 100 on average. Similar tests had been carried out in 2014, with the IQ score reaching a maximum value of 27.^{[55][56]}

In 2019, video game programmer and aerospace engineer John Carmack announced plans to research AGI.^[57]

In 2020, OpenAI developed GPT-3, a language model capable of performing many diverse tasks without specific training. According to Gary Grossman in a VentureBeat article, while there is consensus that GPT-3 is not an example of AGI, it is considered by some to be too advanced to classify as a narrow AI system.^[58]

Processing power needed to simulate a brain

Whole brain emulation

A popular discussed approach to achieving general intelligent action is whole brain emulation. A low-level brain model is built by scanning and mapping a biological brain in detail and copying its state into a computer system or another computational device. The computer runs a simulation model so faithful to the original that it will behave in essentially the same way as the original brain, or for all practical purposes, indistinguishably.^[59] Whole brain emulation is discussed in computational neuroscience and neuroinformatics, in the context of brain simulation for medical research purposes. It is discussed in artificial intelligence research^[46] as an approach to strong AI. Neuroimaging technologies that could deliver the necessary detailed understanding are improving rapidly, and futurist Ray Kurzweil in the book *The Singularity Is Near*^[45] predicts that a map of sufficient quality will become available on a similar timescale to the required computing power.

Early estimates

For low-level brain simulation, an extremely powerful computer would be required. The human brain has a huge number of synapses. Each of the 10^{11} (one hundred billion) neurons has on average 7,000 synaptic connections (synapses) to other neurons. It has been estimated that the brain of a three-year-old child has about 10^{15} synapses (1 quadrillion). This number declines with age, stabilizing by adulthood. Estimates vary for an adult, ranging from 10^{14} to 5×10^{14} synapses (100 to 500 trillion).^[61] An estimate of the brain's processing power, based on a simple switch model for neuron activity, is around 10^{14} (100 trillion) synaptic updates per second (SUPS).^[62] In 1997, Kurzweil looked at various estimates for the hardware required to equal the human brain and adopted a figure of 10^{16} computations per second (cps).^[63] (For comparison, if a "computation" was equivalent to one "floating point operation" – a measure used to rate current supercomputers – then 10^{16} "computations" would be equivalent to 10 petaFLOPS, achieved in 2011). He used this figure to predict the necessary hardware would be available sometime between 2015 and 2025, if the exponential growth in computer power at the time of writing continued.

Modelling the neurons in more detail

The artificial neuron model assumed by Kurzweil and used in many current artificial neural network implementations is simple compared with biological neurons. A brain simulation would likely have to capture the detailed cellular behaviour of biological neurons, presently understood only in the broadest of outlines. The overhead introduced by full modeling of the biological, chemical, and physical details of neural behaviour (especially on a molecular scale) would require computational powers several orders of magnitude larger than

Kurzweil's estimate. In addition the estimates do not account for glial cells, which are at least as numerous as neurons, and which may outnumber neurons by as much as 10:1, and are now known to play a role in cognitive processes.^[64]

Current research

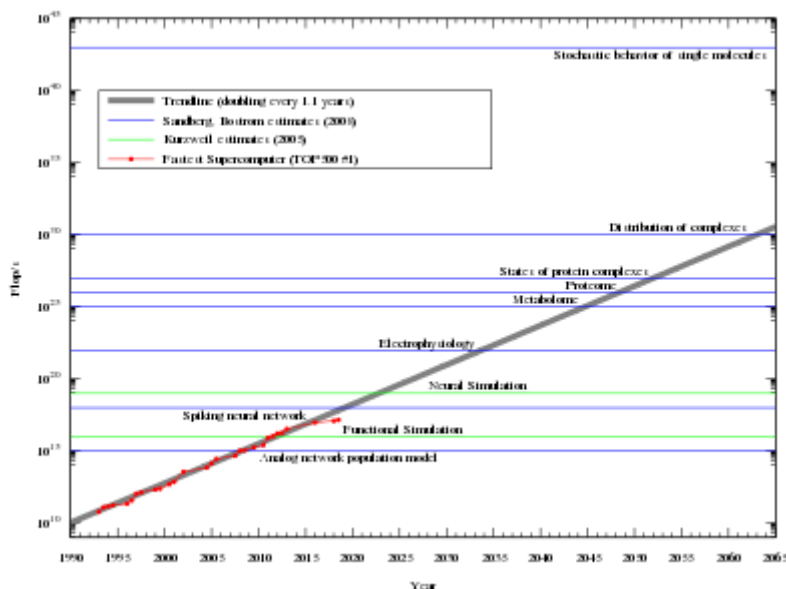
There are some research projects that are investigating brain simulation using more sophisticated neural models, implemented on conventional computing architectures. The Artificial Intelligence System project implemented non-real time simulations of a "brain" (with 10^{11} neurons) in 2005. It took 50 days on a cluster of 27 processors to simulate 1 second of a model.^[65] The Blue Brain project used one of the fastest supercomputer architectures in the world, IBM's Blue Gene platform, to create a real time simulation of a single rat neocortical column consisting of approximately 10,000 neurons and 10^8 synapses in 2006.^[66] A longer-term goal is to build a detailed, functional simulation of the physiological processes in the human brain: "It is not impossible to build a human brain and we can do it in 10 years," Henry Markram, director of the Blue Brain Project said in 2009 at the TED conference in Oxford.^[67] There have also been controversial claims to have simulated a cat brain. Neuro-silicon interfaces have been proposed as an alternative implementation strategy that may scale better.^[68]

Hans Moravec addressed the above arguments ("brains are more complicated", "neurons have to be modeled in more detail") in his 1997 paper "When will computer hardware match the human brain?".^[69] He measured the ability of existing software to simulate the functionality of neural tissue, specifically the retina. His results do not depend on the number of glial cells, nor on what kinds of processing neurons perform where.

The actual complexity of modeling biological neurons has been explored in OpenWorm project that was aimed on complete simulation of a worm that has only 302 neurons in its neural network (among about 1000 cells in total). The animal's neural network has been well documented before the start of the project. However, although the task seemed simple at the beginning, the models based on a generic neural network did not work. Currently, the efforts are focused on precise emulation of biological neurons (partly on the molecular level), but the result cannot be called a total success yet. Even if the number of issues to be solved in a human-brain-scale model is not proportional to the number of neurons, the amount of work along this path is obvious.

Criticisms of simulation-based approaches

A fundamental criticism of the simulated brain approach derives from embodied cognition where human embodiment is taken as an essential aspect of human intelligence. Many researchers believe that embodiment is necessary to ground meaning.^[70] If this view is correct, any fully functional brain model will need to encompass more than just the neurons (i.e., a robotic body). Goertzel^[46] proposes virtual embodiment (like in Second Life), but it is not yet known whether this would be sufficient.



Estimates of how much processing power is needed to emulate a human brain at various levels (from Ray Kurzweil, and Anders Sandberg and Nick Bostrom), along with the fastest supercomputer from TOP500 mapped by year. Note the logarithmic scale and exponential trendline, which assumes the computational capacity doubles every 1.1 years. Kurzweil believes that mind uploading will be possible at neural simulation, while the Sandberg, Bostrom report is less certain about where consciousness arises.^[60]

Desktop computers using microprocessors capable of more than 10^9 cps (Kurzweil's non-standard unit "computations per second", see above) have been available since 2005. According to the brain power estimates used by Kurzweil (and Moravec), this computer should be capable of supporting a simulation of a bee brain, but despite some interest^[71] no such simulation exists. There are at least three reasons for this:

1. The neuron model seems to be oversimplified (see next section).
2. There is insufficient understanding of higher cognitive processes^[72] to establish accurately what the brain's neural activity, observed using techniques such as functional magnetic resonance imaging, correlates with.
3. Even if our understanding of cognition advances sufficiently, early simulation programs are likely to be very inefficient and will, therefore, need considerably more hardware.
4. The brain of an organism, while critical, may not be an appropriate boundary for a cognitive model. To simulate a bee brain, it may be necessary to simulate the body, and the environment. The Extended Mind thesis formalizes the philosophical concept, and research into cephalopods has demonstrated clear examples of a decentralized system.^[73]

In addition, the scale of the human brain is not currently well-constrained. One estimate puts the human brain at about 100 billion neurons and 100 trillion synapses.^{[74][75]} Another estimate is 86 billion neurons of which 16.3 billion are in the cerebral cortex and 69 billion in the cerebellum.^[76] Glial cell synapses are currently unquantified but are known to be extremely numerous.

Strong AI and consciousness

In 1980, philosopher John Searle coined the term "strong AI" as part of his Chinese room argument.^[77] He wanted to distinguish between two different hypotheses about artificial intelligence:^[78]

- An artificial intelligence system can *think* and have a *mind*. (The word "mind" has a specific meaning for philosophers, as used in "the mind body problem" or "the philosophy of mind".)
- An artificial intelligence system can (only) *act like* it thinks and has a mind.

The first one is called "the *strong* AI hypothesis" and the second is "the *weak* AI hypothesis" because the first one makes the *stronger* statement: it assumes something special has happened to the machine that goes beyond all its abilities that we can test. Searle referred to the "strong AI hypothesis" as "strong AI". This usage is also common in academic AI research and textbooks.^[79]

The weak AI hypothesis is equivalent to the hypothesis that artificial general intelligence is possible. According to Russell and Norvig, "Most AI researchers take the weak AI hypothesis for granted, and don't care about the strong AI hypothesis."^[80]

In contrast to Searle, Ray Kurzweil uses the term "strong AI" to describe any artificial intelligence system that acts like it has a mind,^[45] regardless of whether a philosopher would be able to determine if it *actually* has a mind or not. In science fiction, AGI is associated with traits such as consciousness, sentience, sapience, and self-awareness observed in living beings. However, according to Searle, it is an open question whether general intelligence is sufficient for consciousness. "Strong AI" (as defined above by Kurzweil) should not be confused with Searle's "strong AI hypothesis." The strong AI hypothesis is the claim that a computer which behaves as intelligently as a person must also necessarily have a mind and consciousness. AGI refers only to the amount of intelligence that the machine displays, with or without a mind.

Consciousness

There are other aspects of the human mind besides intelligence that are relevant to the concept of strong AI which play a major role in science fiction and the ethics of artificial intelligence:

- consciousness: To have subjective experience and thought.^[81]
- self-awareness: To be aware of oneself as a separate individual, especially to be aware of one's own thoughts.
- sentience: The ability to "feel" perceptions or emotions subjectively.
- sapience: The capacity for wisdom.

These traits have a moral dimension, because a machine with this form of strong AI may have rights, analogous to the rights of non-human animals. As such, preliminary work has been conducted on approaches to integrating full ethical agents with existing legal and social frameworks. These approaches have focused on the legal position and rights of 'strong' AI.^[82]

However, Bill Joy, among others, argues a machine with these traits may be a threat to human life or dignity.^[83] It remains to be shown whether any of these traits are necessary for strong AI. The role of consciousness is not clear, and currently there is no agreed test for its presence. If a machine is built with a device that simulates the neural correlates of consciousness, would it automatically have self-awareness? It is also possible that some of these properties, such as sentience, naturally emerge from a fully intelligent machine, or that it becomes natural to *ascribe* these properties to machines once they begin to act in a way that is clearly intelligent. For example, intelligent action may be sufficient for sentience, rather than the other way around.

Artificial consciousness research

Although the role of consciousness in strong AI/AGI is debatable, many AGI researchers^[84] regard research that investigates possibilities for implementing consciousness as vital. In an early effort Igor Aleksander^[85] argued that the principles for creating a conscious machine already existed but that it would take forty years to train such a machine to understand language.

Possible explanations for the slow progress of AI research

Since the launch of AI research in 1956, the growth of this field has slowed down over time and has stalled the aims of creating machines skilled with intelligent action at the human level.^[86] A possible explanation for this delay is that computers lack a sufficient scope of memory or processing power.^[86] In addition, the level of complexity that connects to the process of AI research may also limit the progress of AI research.^[86]

While most AI researchers believe strong AI can be achieved in the future, there are some individuals like Hubert Dreyfus and Roger Penrose who deny the possibility of achieving strong AI.^[86] John McCarthy was one of various computer scientists who believe human-level AI will be accomplished, but a date cannot accurately be predicted.^[87]

Conceptual limitations are another possible reason for the slowness in AI research.^[86] AI researchers may need to modify the conceptual framework of their discipline in order to provide a stronger base and contribution to the quest of achieving strong AI. As William Clocksin wrote in 2003: "the framework starts from Weizenbaum's observation that intelligence manifests itself only relative to specific social and cultural contexts".^[86]

Furthermore, AI researchers have been able to create computers that can perform jobs that are complicated for people to do, such as mathematics, but conversely they have struggled to develop a computer that is capable of carrying out tasks that are simple for humans to do, such as walking (Moravec's paradox).^[86] A problem

described by David Gelernter is that some people assume thinking and reasoning are equivalent.^[88] However, the idea of whether thoughts and the creator of those thoughts are isolated individually has intrigued AI researchers.^[88]

The problems that have been encountered in AI research over the past decades have further impeded the progress of AI. The failed predictions that have been promised by AI researchers and the lack of a complete understanding of human behaviors have helped diminish the primary idea of human-level AI.^[46] Although the progress of AI research has brought both improvement and disappointment, most investigators have established optimism about potentially achieving the goal of AI in the 21st century.^[46]

Other possible reasons have been proposed for the lengthy research in the progress of strong AI. The intricacy of scientific problems and the need to fully understand the human brain through psychology and neurophysiology have limited many researchers in emulating the function of the human brain in computer hardware.^[89] Many researchers tend to underestimate any doubt that is involved with future predictions of AI, but without taking those issues seriously, people can then overlook solutions to problematic questions.^[46]

Clocks in says that a conceptual limitation that may impede the progress of AI research is that people may be using the wrong techniques for computer programs and implementation of equipment.^[86] When AI researchers first began to aim for the goal of artificial intelligence, a main interest was human reasoning.^[90] Researchers hoped to establish computational models of human knowledge through reasoning and to find out how to design a computer with a specific cognitive task.^[90]

The practice of abstraction, which people tend to redefine when working with a particular context in research, provides researchers with a concentration on just a few concepts.^[90] The most productive use of abstraction in AI research comes from planning and problem solving.^[90] Although the aim is to increase the speed of a computation, the role of abstraction has posed questions about the involvement of abstraction operators.^[91]

A possible reason for the slowness in AI relates to the acknowledgement by many AI researchers that heuristics is a section that contains a significant breach between computer performance and human performance.^[89] The specific functions that are programmed to a computer may be able to account for many of the requirements that allow it to match human intelligence. These explanations are not necessarily guaranteed to be the fundamental causes for the delay in achieving strong AI, but they are widely agreed by numerous researchers.

There have been many AI researchers that debate over the idea whether machines should be created with emotions. There are no emotions in typical models of AI and some researchers say programming emotions into machines allows them to have a mind of their own.^[86] Emotion sums up the experiences of humans because it allows them to remember those experiences.^[88] David Gelernter writes, "No computer will be creative unless it can simulate all the nuances of human emotion."^[88] This concern about emotion has posed problems for AI researchers and it connects to the concept of strong AI as its research progresses into the future.^[92]

Controversies and dangers

Feasibility

As of August 2020, AGI remains speculative^{[8][93]} as no such system has been demonstrated yet. Opinions vary both on *whether* and *when* artificial general intelligence will arrive, at all. At one extreme, AI pioneer Herbert A. Simon speculated in 1965: "machines will be capable, within twenty years, of doing any work a man can do". However, this prediction failed to come true. Microsoft co-founder Paul Allen believed that such intelligence is unlikely in the 21st century because it would require "unforeseeable and fundamentally

unpredictable breakthroughs" and a "scientifically deep understanding of cognition".^[94] Writing in *The Guardian*, roboticist Alan Winfield claimed the gulf between modern computing and human-level artificial intelligence is as wide as the gulf between current space flight and practical faster-than-light spaceflight.^[95]

AI experts' views on the feasibility of AGI wax and wane, and may have seen a resurgence in the 2010s. Four polls conducted in 2012 and 2013 suggested that the median guess among experts for when they would be 50% confident AGI would arrive was 2040 to 2050, depending on the poll, with the mean being 2081. Of the experts, 16.5% answered with "never" when asked the same question but with a 90% confidence instead.^{[96][97]} Further current AGI progress considerations can be found below *Tests for confirming human-level AGI* and *IQ-tests AGI*.

Potential threat to human existence

The thesis that AI poses an existential risk, and that this risk needs much more attention than it currently gets, has been endorsed by many public figures; perhaps the most famous are [Elon Musk](#), [Bill Gates](#), and [Stephen Hawking](#). The most notable AI researcher to endorse the thesis is [Stuart J. Russell](#). Endorsers of the thesis sometimes express bafflement at skeptics: Gates states he does not "understand why some people are not concerned",^[98] and Hawking criticized widespread indifference in his 2014 editorial:

'So, facing possible futures of incalculable benefits and risks, the experts are surely doing everything possible to ensure the best outcome, right? Wrong. If a superior alien civilisation sent us a message saying, 'We'll arrive in a few decades,' would we just reply, 'OK, call us when you get here—we'll leave the lights on?' Probably not—but this is more or less what is happening with AI.'^[99]

Many of the scholars who are concerned about existential risk believe that the best way forward would be to conduct (possibly massive) research into solving the difficult "[control problem](#)" to answer the question: what types of safeguards, algorithms, or architectures can programmers implement to maximize the probability that their recursively-improving AI would continue to behave in a friendly, rather than destructive, manner after it reaches superintelligence?^{[100][101]}

The thesis that AI can pose existential risk also has many strong detractors. Skeptics sometimes charge that the thesis is crypto-religious, with an irrational belief in the possibility of superintelligence replacing an irrational belief in an omnipotent God; at an extreme, [Jaron Lanier](#) argues that the whole concept that current machines are in any way intelligent is "an illusion" and a "stupendous con" by the wealthy.^[102]

Much of existing criticism argues that AGI is unlikely in the short term. Computer scientist [Gordon Bell](#) argues that the human race will already destroy itself before it reaches the [technological singularity](#). [Gordon Moore](#), the original proponent of [Moore's Law](#), declares that "I am a skeptic. I don't believe [a technological singularity] is likely to happen, at least for a long time. And I don't know why I feel that way."^[103] Former Baidu Vice President and Chief Scientist [Andrew Ng](#) states AI existential risk is "like worrying about overpopulation on Mars when we have not even set foot on the planet yet."^[104]

See also

- [Artificial brain](#)
- [AI control problem](#)
- [Automated machine learning](#)
- [BRAIN Initiative](#)

- [China Brain Project](#)
- [Eliezer Yudkowsky](#)
- [Future of Humanity Institute](#)
- [General game playing](#)
- [Human Brain Project](#)
- Intelligence amplification (IA), the use of information technology in augmenting human intelligence instead of creating an external autonomous "AGI"
- [Machine ethics](#)
- [Multi-task learning](#)
- [Nick Bostrom](#)
- [Outline of artificial intelligence](#)
- [Outline of transhumanism](#)
- [Superintelligence](#)
- [Synthetic intelligence](#)
- [Transfer learning](#)

Notes

1. Hodson, Hal (1 March 2019). "DeepMind and Google: the battle to control artificial intelligence" (<https://www.economist.com/1843/2019/03/01/deepmind-and-google-the-battle-to-control-artificial-intelligence>). 1843. Retrieved 7 July 2020. "AGI stands for Artificial General Intelligence, a hypothetical computer program..."
2. Kurzweil, *Singularity* (2005) p. 260
3. Kurzweil, Ray (5 August 2005), "Long Live AI" (https://www.forbes.com/home/free_forbes/2005/0815/030.htmlhttps://www.forbes.com/home/free_forbes/2005/0815/030.html), *Forbes*: Kurzweil describes strong AI as "machine intelligence with the full range of human intelligence."
4. Treder, Mike (10 August 2005), "Advanced Human ntelligence" (https://crnano.typepad.com/crnblog/2005/08/advanced_human_.html), *Responsible Nanotechnology*, archived (https://web.archive.org/web/20191016214415/https://crnano.typepad.com/crnblog/2005/08/advanced_human_.html) from the original on 16 October 2019
5. "The Age of Artificial Intelligence: George John at TEDxLondonBusinessSchool 2013" (<http://tedxtalks.ted.com/video/The-Age-of-Artificial-Intelligence>). Archived (<https://web.archive.org/web/20140226123940/http://tedxtalks.ted.com/video/The-Age-of-Artificial-Intelligence>) from the original on 26 February 2014. Retrieved 22 February 2014.
6. Newell & Simon 1976, This is the term they use for "human-level" intelligence in the [physical symbol system](#) hypothesis.
7. Searle 1980, See below for the origin of the term "strong AI", and see the academic definition of "strong AI" in the article [Chinese room](#).
8. europarl.europa.eu: How artificial intelligence works (<https://www.europarl.europa.eu/at-your-service/files/be-heard/religious-and-non-confessional-dialogue/events/en-20190319-how-artificial-intelligence-works.pdf>), "Concluding remarks: Today's AI is powerful and useful, but remains far from speculated AGI or ASI.", European Parliamentary Research Service, retrieved March 3, 2020
9. Grace, Katja; Salvatier, John; Dafoe, Allan; Zhang, Baobao; Evans, Owain (31 July 2018). "Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts" (<https://doi.org/10.1613%2Fjair.1.11222>). *Journal of Artificial Intelligence Research*. **62**: 729–754. doi:10.1613/jair.1.11222 (<https://doi.org/10.1613%2Fjair.1.11222>). ISSN 1076-9757 (<https://www.worldcat.org/issn/1076-9757>).

10. "The Open University on Strong and Weak AI" (https://web.archive.org/web/20090925043908/http://www.open2.net/nextbigthing/ai/ai_in_depth/in_depth.htm). Archived from the original (http://www.open2.net/nextbigthing/ai/ai_in_depth/in_depth.htm) on 25 September 2009. Retrieved 8 October 2007.
11. Baum, Seth (12 November 2017). "Baum, Seth, A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy (November 12, 2017). Global Catastrophic Risk Institute Working Paper 17-1" (<https://ssrn.com/abstract=3070741>).
12. AI founder John McCarthy writes: "we cannot yet characterize in general what kinds of computational procedures we want to call intelligent." McCarthy, John (2007). "Basic Questions" (<http://www-formal.stanford.edu/jmc/whatisai/node1.html>). Stanford University. Archived (<https://web.archive.org/web/20071026100601/http://www-formal.stanford.edu/jmc/whatisai/node1.html>) from the original on 26 October 2007. Retrieved 6 December 2007. (For a discussion of some definitions of intelligence used by artificial intelligence researchers, see philosophy of artificial intelligence.)
13. This list of intelligent traits is based on the topics covered by major AI textbooks, including: Russell & Norvig 2003, Luger & Stubblefield 2004, Poole, Mackworth & Goebel 1998 and Nilsson 1998.
14. Pfeifer, R. and Bongard J. C., How the body shapes the way we think: a new view of intelligence (The MIT Press, 2007). ISBN 0-262-16239-3
15. White, R. W. (1959). "Motivation reconsidered: The concept of competence". *Psychological Review*. **66** (5): 297–333. doi:10.1037/h0040934 (<https://doi.org/10.1037%2Fh0040934>). PMID 13844397 (<https://pubmed.ncbi.nlm.nih.gov/13844397>).
16. Johnson 1987
17. deCharms, R. (1968). Personal causation. New York: Academic Press.
18. Muehlhauser, Luke (11 August 2013). "What is AGI?" (<http://intelligence.org/2013/08/11/what-is-agi/>). Machine Intelligence Research Institute. Archived (<https://web.archive.org/web/20140425115445/http://intelligence.org/2013/08/11/what-is-agi/>) from the original on 25 April 2014. Retrieved 1 May 2014.
19. "What is Artificial General Intelligence (AGI)? | 4 Tests For Ensuring Artificial General Intelligence" (https://www.talkyblog.com/artificial_general_intelligence_agi/). *Talky Blog*. 13 July 2019. Archived (https://web.archive.org/web/20190717071152/https://www.talkyblog.com/artificial_general_intelligence_agi/) from the original on 17 July 2019. Retrieved 17 July 2019.
20. Shapiro, Stuart C. (1992). Artificial Intelligence (<http://www.cse.buffalo.edu/~shapiro/Papers/ai.pdf>) Archived (<https://web.archive.org/web/20160201014644/http://www.cse.buffalo.edu/~shapiro/Papers/ai.pdf>) 1 February 2016 at the Wayback Machine In Stuart C. Shapiro (Ed.), *Encyclopedia of Artificial Intelligence* (Second Edition, pp. 54–57). New York: John Wiley. (Section 4 is on "AI-Complete Tasks".)
21. Roman V. Yampolskiy. Turing Test as a Defining Feature of AI-Completeness. In Artificial Intelligence, Evolutionary Computation and Metaheuristics (AIECM) --In the footsteps of Alan Turing. Xin-She Yang (Ed.). pp. 3–17. (Chapter 1). Springer, London. 2013. <http://cecs.louisville.edu/ry/TuringTestasaDefiningFeature04270003.pdf> Archived (<https://web.archive.org/web/20130522094547/http://cecs.louisville.edu/ry/TuringTestasaDefiningFeature04270003.pdf>) 22 May 2013 at the Wayback Machine
22. Luis von Ahn, Manuel Blum, Nicholas Hopper, and John Langford. CAPTCHA: Using Hard AI Problems for Security (http://www.captcha.net/captcha_crypt.pdf) Archived (https://web.archive.org/web/20160304001102/http://www.captcha.net/captcha_crypt.pdf) 4 March 2016 at the Wayback Machine. In Proceedings of Eurocrypt, Vol. 2656 (2003), pp. 294–311.
23. Bergmair, Richard (7 January 2006). "Natural Language Steganography and an "AI-complete" Security Primitive". CiteSeerX 10.1.1.105.129 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.105.129>). (unpublished?)
24. Crevier 1993, pp. 48–50
25. Simon 1965, p. 96 quoted in Crevier 1993, p. 109

26. "Scientist on the Set: An Interview with Marvin Minsky" (<http://mitpress.mit.edu/e-books/Hal/chap2/two1.html>). Archived (<https://web.archive.org/web/20120716182537/http://mitpress.mit.edu/e-books/Hal/chap2/two1.html>) from the original on 16 July 2012. Retrieved 5 April 2008.
27. Marvin Minsky to Darrach (1970), quoted in Crevier (1993, p. 109).
28. The Lighthill report specifically criticized AI's "grandiose objectives" and led the dismantling of AI research in England. (Lighthill 1973; Howe 1994) In the U.S., DARPA became determined to fund only "mission-oriented direct research, rather than basic undirected research". See (NRC 1999) under "Shift to Applied Research Increases Investment". See also (Crevier 1993, pp. 115–117) and (Russell & Norvig 2003, pp. 21–22)
29. Crevier 1993, p. 211, Russell & Norvig 2003, p. 24 and see also Feigenbaum & McCorduck 1983
30. Crevier 1993, pp. 161–162, 197–203, 240; Russell & Norvig 2003, p. 25; NRC 1999, under "Shift to Applied Research Increases Investment"
31. Crevier 1993, pp. 209–212
32. As AI founder John McCarthy writes "it would be a great relief to the rest of the workers in AI if the inventors of new general formalisms would express their hopes in a more guarded form than has sometimes been the case." McCarthy, John (2000). "Reply to Lighthill" (<http://www-formal.stanford.edu/jmc/reviews/lighthill/lighthill.html>). Stanford University. Archived (<https://web.archive.org/web/20080930164952/http://www-formal.stanford.edu/jmc/reviews/lighthill/lighthill.html>) from the original on 30 September 2008. Retrieved 29 September 2007.
33. "At its low point, some computer scientists and software engineers avoided the term artificial intelligence for fear of being viewed as wild-eyed dreamers." Markoff, John (14 October 2005). "Behind Artificial Intelligence, a Squadron of Bright Real People" (<https://www.nytimes.com/2005/10/14/technology/14artificial.html?ei=5070&en=11ab55edb7cead5e&ex=1185940800&adxnln=1&adxnnlx=1185805173-o7WsfW7qaP0x5/NU51cQCQ>). *The New York Times*.
34. Russell & Norvig 2003, pp. 25–26
35. "Trends in the Emerging Tech Hype Cycle" (https://blogs.gartner.com/smarterwithgartner/files/2018/08/PR_490866_5_Trends_in_the_Emerging_Tech_Hype_Cycle_2018_Hype_Cycle.png). Gartner Reports. Archived (https://web.archive.org/web/20190522024829/https://blogs.gartner.com/smarterwithgartner/files/2018/08/PR_490866_5_Trends_in_the_Emerging_Tech_Hype_Cycle_2018_Hype_Cycle.png) from the original on 22 May 2019. Retrieved 7 May 2019.
36. Moravec 1988, p. 20
37. Harnad, S (1990). "The Symbol Grounding Problem". *Physica D*. **42** (1–3): 335–346. arXiv:cs/9906002 (<https://arxiv.org/abs/cs/9906002>). Bibcode:1990PhyD...42..335H (<https://ui.adsabs.harvard.edu/abs/1990PhyD...42..335H>). doi:10.1016/0167-2789(90)90087-6 (<https://doi.org/10.1016%2F0167-2789%2890%2990087-6>).
38. Gubrud 1997
39. "Who coined the term "AGI"? » goertzel.org" (<http://goertzel.org/who-coined-the-term-agi/>). Archived (<https://web.archive.org/web/20181228083048/http://goertzel.org/who-coined-the-term-agi/>) from the original on 28 December 2018. Retrieved 28 December 2018., via Life 3.0: 'The term "AGI" was popularized by... Shane Legg, Mark Gubrud and Ben Goertzel'
40. Goertzel & Wang 2006. See also Wang (2006) with an up-to-date summary and lots of links.
41. https://goertzel.org/AGI_Summer_School_2009.htm
42. <http://fmi-plovdiv.org/index.jsp?id=1054&ln=1>
43. <http://fmi.uni-plovdiv.bg/index.jsp?id=1139&ln=1>
44. Goertzel & Pennachin 2006.
45. (Kurzweil 2005, p. 260) or see Advanced Human Intelligence (http://crnano.typepad.com/crnblog/2005/08/advanced_human_.html) Archived (https://web.archive.org/web/20110630032301/http://crnano.typepad.com/crnblog/2005/08/advanced_human_.html) 30 June 2011 at the Wayback Machine where he defines strong AI as "machine intelligence with the full range of human intelligence."

46. Goertzel 2007.
47. "Error in Armstrong and Sotala 2012" (<https://aiimpacts.org/error-in-armstrong-and-sotala-2012/>). AI Impacts. 2016. Retrieved 24 August 2020.
48. "Intelligent machines that learn unaided" (<http://www.hutter1.net/official/aixien.pdf>) (PDF). Ticino Ricerca. 2004. Retrieved 24 August 2020.
49. Markoff, John (27 November 2016). "When A.I. Matures, It May Call Jürgen Schmidhuber 'Dad' " (<https://www.nytimes.com/2016/11/27/technology/artificial-intelligence-pioneer-jurgen-schmidhuber-overlooked.html>). *The New York Times*. Archived (<https://web.archive.org/web/20171226234555/https://www.nytimes.com/2016/11/27/technology/artificial-intelligence-pioneer-jurgen-schmidhuber-overlooked.html>) from the original on 26 December 2017. Retrieved 26 December 2017.
50. James Barrat (2013). "Chapter 11: A Hard Takeoff". *Our Final Invention: Artificial Intelligence and the End of the Human Era* (First ed.). New York: St. Martin's Press. ISBN 9780312622374.
51. "About the Machine Intelligence Research Institute" (<https://intelligence.org/about/>). *Machine Intelligence Research Institute*. Archived (<https://web.archive.org/web/20180121025925/https://intelligence.org/about/>) from the original on 21 January 2018. Retrieved 26 December 2017.
52. "About OpenAI" (<https://openai.com/about/>). *OpenAI*. Archived (<https://web.archive.org/web/20171222181056/https://openai.com/about/>) from the original on 22 December 2017. Retrieved 26 December 2017.
53. Theil, Stefan. "Trouble in Mind" (<https://www.scientificamerican.com/article/why-the-human-brain-project-went-wrong-and-how-to-fix-it/>). *Scientific American*. pp. 36–42. Bibcode:2015SciAm.313d..36T (<https://ui.adsabs.harvard.edu/abs/2015SciAm.313d..36T>). doi:10.1038/scientificamerican1015-36 (<https://doi.org/10.1038%2Fscientificamerican1015-36>). Archived (<https://web.archive.org/web/20171109234151/https://www.scientificamerican.com/article/why-the-human-brain-project-went-wrong-and-how-to-fix-it/>) from the original on 9 November 2017. Retrieved 26 December 2017.
54. Walch, Kathleen (16 October 2020). "Artificial General Intelligence Platform" (<https://www.forbes.com/sites/cognitiveworld/2020/07/14/is-artificial-general-intelligence-agi-on-the-horizon-interview-with-dr-ben-goertzel-ceo--founder-singularitynet-foundation/#1d8239e459d0>). *Forbes*. "Dr. Goertzel is founder of SingularityNET, a decentralized AI platform which lets multiple AI agents cooperate to solve problems in a participatory way without any central controller. ... Dr. Goertzel created SingularityNET for one simple reason: "I intend to create AGI and when I roll out this AGI, I want it to be rolled out in a decentralized and democratically controlled way, rather than in a manner that allows it to be controlled by any one person or corporate or governmental entity.""
55. Liu, Feng; Shi, Yong; Liu, Ying (2017). "Intelligence Quotient and Intelligence Grade of Artificial Intelligence". *Annals of Data Science*. 4 (2): 179–191. arXiv:1709.10242 (<https://arxiv.org/abs/1709.10242>). Bibcode:2017arXiv170910242L (<https://ui.adsabs.harvard.edu/abs/2017arXiv170910242L>). doi:10.1007/s40745-017-0109-0 (<https://doi.org/10.1007%2Fs40745-017-0109-0>). S2CID 37900130 (<https://api.semanticscholar.org/CorpusID:37900130>).
56. "Google-KI doppelt so schlau wie Siri" (<https://t3n.de/news/iq-kind-schlauer-google-ki-siri-864003/>). Archived (<https://web.archive.org/web/20190103055657/https://t3n.de/news/iq-kind-schlauer-google-ki-siri-864003/>) from the original on 3 January 2019. Retrieved 2 January 2019.
57. Lawler, Richard (13 November 2019). "John Carmack takes a step back at Oculus to work on human-like AI" (<https://www.engadget.com/2019-11-13-john-carmack-agi.html>). Engadget. Retrieved 4 April 2020.

58. Grossman, Gary (3 September 2020). "We're entering the AI twilight zone between narrow and general AI" (<https://venturebeat.com/2020/09/03/were-entering-the-ai-twilight-zone-between-narrow-and-general-ai/>). VentureBeat. Retrieved 5 September 2020. "Certainly, too, there are those who claim we are already seeing an early example of an AGI system in the recently announced GPT-3 natural language processing (NLP) neural network. ... So is GPT-3 the first example of an AGI system? This is debatable, but the consensus is that it is not AGI. ... If nothing else, GPT-3 tells us there is a middle ground between narrow and general AI."
59. Sandberg & Boström 2008. "The basic idea is to take a particular brain, scan its structure in detail, and construct a software model of it that is so faithful to the original that, when run on appropriate hardware, it will behave in essentially the same way as the original brain."
60. Sandberg & Boström 2008.
61. Drachman 2005.
62. Russell & Norvig 2003.
63. In "Mind Children" Moravec 1988, p. 61 10^{15} cps is used. More recently, in 1997, <"Archived copy" (<https://web.archive.org/web/20060615031852/http://transhumanist.com/volume1/moravec.htm>). Archived from the original (<https://www.transhumanist.com/volume1/moravec.htm>) on 15 June 2006. Retrieved 23 June 2006.> Moravec argued for 10^8 MIPS which would roughly correspond to 10^{14} cps. Moravec talks in terms of MIPS, not "cps", which is a non-standard term Kurzweil introduced.
64. Swaminathan, Nikhil (January–February 2011). "Glia—the other brain cells" (<http://discovermagazine.com/2011/jan-feb/62>). *Discover*. Archived (<https://web.archive.org/web/20140208071350/http://discovermagazine.com/2011/jan-feb/62>) from the original on 8 February 2014. Retrieved 24 January 2014.
65. Izhikevich, Eugene M.; Edelman, Gerald M. (4 March 2008). "Large-scale model of mammalian thalamocortical systems" (https://web.archive.org/web/20090612095651/http://vesicle.nsi.edu/users/izhikevich/publications/large-scale_model_of_human_brain.pdf) (PDF). *PNAS*. **105** (9): 3593–3598. Bibcode:2008PNAS..105.3593I (<https://ui.adsabs.harvard.edu/abs/2008PNAS..105.3593I>). doi:10.1073/pnas.0712231105 (<https://doi.org/10.1073%2Fpnas.0712231105>). PMC 2265160 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2265160>). PMID 18292226 (<http://pubmed.ncbi.nlm.nih.gov/18292226>). Archived from the original (http://vesicle.nsi.edu/users/izhikevich/publications/large-scale_model_of_human_brain.pdf) (PDF) on 12 June 2009. Retrieved 23 June 2015.
66. "Project Milestones" (<http://bluebrain.epfl.ch/Jahia/site/bluebrain/op/edit/pid/19085>). *Blue Brain*. Retrieved 11 August 2008.
67. "Artificial brain '10 years away' 2009 BBC news" (<http://news.bbc.co.uk/1/hi/technology/8164060.stm>). 22 July 2009. Archived (<https://web.archive.org/web/20170726040959/http://news.bbc.co.uk/1/hi/technology/8164060.stm>) from the original on 26 July 2017. Retrieved 25 July 2009.
68. University of Calgary news (<http://gauntlet.ucalgary.ca/story/10343>) Archived (<https://web.archive.org/web/20090818081044/http://gauntlet.ucalgary.ca/story/10343>) 18 August 2009 at the Wayback Machine, NBC News news (<http://www.nbcnews.com/id/12037941>) Archived (<https://web.archive.org/web/20170704063922/http://www.nbcnews.com/id/12037941>) 4 July 2017 at the Wayback Machine
69. "Archived copy" (<https://web.archive.org/web/20060615031852/http://transhumanist.com/volume1/moravec.htm>). Archived from the original (<https://www.transhumanist.com/volume1/moravec.htm>) on 15 June 2006. Retrieved 23 June 2006.
70. de Vega, Glenberg & Graesser 2008. A wide range of views in current research, all of which require grounding to some degree
71. "some links to bee brain studies" (https://web.archive.org/web/20111005162232/http://www.setiai.com/archives/cat_honey_bee_brain.html). Archived from the original (http://www.setiai.com/archives/cat_honey_bee_brain.html) on 5 October 2011. Retrieved 30 March 2010.

72. In Goertzel's AGI book ([Yudkowsky 2006](#)), Yudkowsky proposes 5 levels of organisation that must be understood – code/data, sensory modality, concept & category, thought, and deliberation (consciousness) – in order to use the available hardware
73. Yekutieli, Y; Sagiv-Zohar, R; Aharonov, R; Engel, Y; Hochner, B; Flash, T (August 2005). "Dynamic model of the octopus arm. I. Biomechanics of the octopus reaching movement". *J. Neurophysiol.* **94** (2): 1443–58. doi:10.1152/jn.00684.2004 (<https://doi.org/10.1152%2Fjn.00684.2004>). PMID 15829594 (<https://pubmed.ncbi.nlm.nih.gov/15829594>).
74. [Williams & Herrup 1988](#)
75. "nervous system, human." (<http://search.eb.com/eb/article-75525>) *Encyclopædia Britannica*. 9 January 2007
76. [Azevedo et al. 2009](#).
77. [Searle 1980](#)
78. As defined in a standard AI textbook: "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." ([Russell & Norvig 2003](#))
79. For example:
 - [Russell & Norvig 2003](#),
 - [Oxford University Press Dictionary of Psychology](#) (<http://www.encyclopedia.com/doc/1O87-strongAI.html>) Archived (<https://web.archive.org/web/20071203103022/http://www.encyclopedia.com/doc/1O87-strongAI.html>) 3 December 2007 at the [Wayback Machine](#) (quoted in "High Beam Encyclopedia"),
 - [MIT Encyclopedia of Cognitive Science](#) (<http://www.aaai.org/AITopics/html/phil.html>) Archived (<https://web.archive.org/web/20080719074502/http://www.aaai.org/AITopics/html/phil.html>) 19 July 2008 at the [Wayback Machine](#) (quoted in "AITopics")
 - [Planet Math](#) (<https://planetmath.org/encyclopedia/StrongAIThesis.html>) Archived (<https://web.archive.org/web/20070919012830/http://planetmath.org/encyclopedia/StrongAIThesis.html>) 19 September 2007 at the [Wayback Machine](#)
 - [Will Biological Computers Enable Artificially Intelligent Machines to Become Persons?](#) (http://www.cbhd.org/resources/biotech/tongen_2003-11-07.htm) Archived (https://web.archive.org/web/20080513031753/http://www.cbhd.org/resources/biotech/tongen_2003-11-07.htm) 13 May 2008 at the [Wayback Machine](#) Anthony Tongen
80. [Russell & Norvig 2003](#), p. 947.
81. Note that [consciousness](#) is difficult to define. A popular definition, due to [Thomas Nagel](#), is that it "feels like" something to be conscious. If we are not conscious, then it doesn't feel like anything. Nagel uses the example of a bat: we can sensibly ask "what does it feel like to be a bat?" However, we are unlikely to ask "what does it feel like to be a toaster?" Nagel concludes that a bat appears to be conscious (i.e. has consciousness) but a toaster does not. See ([Nagel 1974](#))
82. Sotala, Kaj; Yampolskiy, Roman V (19 December 2014). "Responses to catastrophic AGI risk: a survey" (<https://doi.org/10.1088%2F0031-8949%2F90%2F1%2F018001>). *Physica Scripta*. **90** (1): 8. doi:10.1088/0031-8949/90/1/018001 (<https://doi.org/10.1088%2F0031-8949%2F90%2F1%2F018001>). ISSN 0031-8949 (<https://www.worldcat.org/issn/0031-8949>).
83. Joy, Bill (April 2000). "Why the future doesn't need us". *Wired*.
84. [Yudkowsky 2006](#).
85. [Aleksander 1996](#).
86. [Clocksin 2003](#).
87. [McCarthy 2003](#).
88. [Gelernter 2010](#).
89. [McCarthy 2007](#).

90. Holte & Choueiry 2003.
91. Zucker 2003.
92. Kaplan, Andreas; Haenlein, Michael (2019). "Kaplan Andreas and Haelein Michael (2019) Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence". *Business Horizons*. **62**: 15–25.
doi:10.1016/j.bushor.2018.08.004 (<https://doi.org/10.1016%2Fj.bushor.2018.08.004>).
93. itu.int: Beyond Mad?: The Race For Artificial General Intelligence (<https://www.itu.int/en/journal/001/Documents/itu2018-9.pdf>), "AGI represents a level of power that remains firmly in the realm of speculative fiction as on date." February 2, 2018, retrieved March 3, 2020
94. Allen, Paul. "The Singularity Isn't Near" (<http://www.technologyreview.com/view/425733/paul-allen-the-singularity-isnt-near/>). *MIT Technology Review*. Retrieved 17 September 2014.
95. Winfield, Alan. "Artificial intelligence will not turn into a Frankenstein's monster" (<https://www.theguardian.com/technology/2014/aug/10/artificial-intelligence-will-not-become-a-frankensteins-monster-ian-winfield>). *The Guardian*. Archived (<https://web.archive.org/web/20140917135230/http://www.theguardian.com/technology/2014/aug/10/artificial-intelligence-will-not-become-a-frankensteins-monster-ian-winfield>) from the original on 17 September 2014. Retrieved 17 September 2014.
96. Raffi Khatchadourian (23 November 2015). "The Doomsday Invention: Will artificial intelligence bring us utopia or destruction?" (<http://www.newyorker.com/magazine/2015/11/23/doomsday-invention-artificial-intelligence-nick-bostrom>). *The New Yorker*. Archived (<https://web.archive.org/web/20160128105955/http://www.newyorker.com/magazine/2015/11/23/doomsday-invention-artificial-intelligence-nick-bostrom>) from the original on 28 January 2016. Retrieved 7 February 2016.
97. Müller, V. C., & Bostrom, N. (2016). Future progress in artificial intelligence: A survey of expert opinion. In *Fundamental issues of artificial intelligence* (pp. 555–572). Springer, Cham.
98. Rawlinson, Kevin. "Microsoft's Bill Gates insists AI is a threat" (<https://www.bbc.co.uk/news/31047780>). *BBC News*. Retrieved 30 January 2015.
99. "Stephen Hawking: 'Transcendence looks at the implications of artificial intelligence – but are we taking AI seriously enough?'" (<https://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence--but-are-we-taking-ai-seriously-enough-9313474.html>). *The Independent (UK)*. Retrieved 3 December 2014.
00. Bostrom, Nick (2014). *Superintelligence: Paths, Dangers, Strategies* (First ed.). ISBN 978-0199678112.
01. Kaj Sotala; Roman Yampolskiy (19 December 2014). "Responses to catastrophic AGI risk: a survey". *Physica Scripta*. **90** (1).
02. "But What Would the End of Humanity Mean for Me?" (<https://www.theatlantic.com/health/archive/2014/05/but-what-does-the-end-of-humanity-mean-for-me/361931/>). *The Atlantic*. 9 May 2014. Retrieved 12 December 2015.
03. "Tech Luminaries Address Singularity" (<https://spectrum.ieee.org/computing/hardware/tech-luminaries-address-singularity>). *IEEE Spectrum: Technology, Engineering, and Science News* (SPECIAL REPORT: THE SINGULARITY). 1 June 2008. Retrieved 8 April 2020.
04. Shermer, Michael (1 March 2017). "Apocalypse AI" (<https://www.scientificamerican.com/article/artificial-intelligence-is-not-a-threat-mdash-yet/>). *Scientific American*. p. 77.
Bibcode:2017SciAm.316c..77S (<https://ui.adsabs.harvard.edu/abs/2017SciAm.316c..77S>).
doi:10.1038/scientificamerican0317-77 (<https://doi.org/10.1038%2Fscientificamerican0317-77>). Retrieved 27 November 2017.

References

- Halal, William E. "TechCast Article Series: The Automation of Thought" (https://web.archive.org/web/20130606101835/http://www.techcast.org/Upload/PDFs/633615794236495345_TCThe

AutomationofThought.pdf) (PDF). Archived from the original (http://www.techcast.org/Upload/PDFs/633615794236495345_TCTheAutomationofThought.pdf) (PDF) on 6 June 2013.

- Aleksander, Igor (1996), *Impossible Minds* (<https://archive.org/details/impossiblemindsm0000alek>), World Scientific Publishing Company, ISBN 978-1-86094-036-1
- Omohundro, Steve (2008), *The Nature of Self-Improving Artificial Intelligence*, presented and distributed at the 2007 Singularity Summit, San Francisco, CA.
- Sandberg, Anders; Boström, Nick (2008), *Whole Brain Emulation: A Roadmap* (<http://www.fhi.ox.ac.uk/Reports/2008-3.pdf>) (PDF), Technical Report #2008-3, Future of Humanity Institute, Oxford University, retrieved 5 April 2009
- Azevedo FA, Carvalho LR, Grinberg LT, et al. (April 2009), "Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain" (<https://www.researchgate.net/publication/24024444>), *The Journal of Comparative Neurology*, **513** (5): 532–41, doi:10.1002/cne.21974 (<https://doi.org/10.1002%2Fcne.21974>), PMID 19226510 (<http://pubmed.ncbi.nlm.nih.gov/19226510/>), retrieved 4 September 2013
- Berglas, Anthony (2008), *Artificial Intelligence will Kill our Grandchildren* (<http://berglas.org/Articles/AIKillGrandchildren/AIKillGrandchildren.html>)
- Chalmers, David (1996), *The Conscious Mind*, Oxford University Press.
- Clocksin, William (August 2003), "Artificial intelligence and the future", *Philosophical Transactions of the Royal Society A*, **361** (1809): 1721–1748, Bibcode:2003RSPTA.361.1721C (<https://ui.adsabs.harvard.edu/abs/2003RSPTA.361.1721C>), doi:10.1098/rsta.2003.1232 (<https://doi.org/10.1098%2Frsta.2003.1232>), PMID 12952683 (<https://pubmed.ncbi.nlm.nih.gov/12952683/>), S2CID 31032007 (<https://api.semanticscholar.org/CorpusID:31032007>).
- Crevier, Daniel (1993), *AI: The Tumultuous Search for Artificial Intelligence*, New York, NY: BasicBooks, ISBN 0-465-02997-3
- Darrach, Brad (20 November 1970), "Meet Shakey, the First Electronic Person", *Life Magazine*, pp. 58–68.
- Drachman, D (2005), "Do we have brain to spare?", *Neurology*, **64** (12): 2004–5, doi:10.1212/01.WNL.0000166914.38327.BB (<https://doi.org/10.1212%2F01.WNL.0000166914.38327.BB>), PMID 15985565 (<https://pubmed.ncbi.nlm.nih.gov/15985565/>), S2CID 38482114 (<https://api.semanticscholar.org/CorpusID:38482114>).
- Feigenbaum, Edward A.; McCorduck, Pamela (1983), *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World*, Michael Joseph, ISBN 978-0-7181-2401-4
- Gelernter, David (2010), *Dream-logic, the Internet and Artificial Thought* (http://www.edge.org/3rd_culture/gelernter10.1/gelernter10.1_index.html), retrieved 25 July 2010
- Goertzel, Ben; Pennachin, Cassio, eds. (2006), *Artificial General Intelligence* (<https://web.archive.org/web/20130320184603/http://people.inf.elte.hu/csizsekp/ai/books/artificial-general-intelligence-cognitive-technologies.9783540237334.27156.pdf>) (PDF), Springer, ISBN 978-3-540-23733-4, archived from the original (<http://people.inf.elte.hu/csizsekp/ai/books/artificial-general-intelligence-cognitive-technologies.9783540237334.27156.pdf>) (PDF) on 20 March 2013
- Goertzel, Ben; Wang, Pei (2006), *Introduction: Aspects of Artificial General Intelligence* (https://sites.google.com/site/narswang/publications/wang-goertzel.AGI_Aspects.pdf?attredirects=1) (PDF)
- Goertzel, Ben (December 2007), "Human-level artificial general intelligence and the possibility of a technological singularity: a reaction to Ray Kurzweil's The Singularity Is Near, and McDermott's critique of Kurzweil" (<https://scholar.google.com/scholar?hl=sv&lr=&cluster=15189798216526465792>), *Artificial Intelligence*, **171** (18, Special Review Issue): 1161–1173, doi:10.1016/j.artint.2007.10.011 (<https://doi.org/10.1016%2Fj.artint.2007.10.011>), retrieved 1 April 2009.
- Gubrud, Mark (November 1997), "Nanotechnology and International Security" (<http://www.foresight.org/Conferences/MNT05/Papers/Gubrud/>), *Fifth Foresight Conference on Molecular Nanotechnology*, retrieved 7 May 2011

- Holte, RC; Choueiry, BY (2003), "Abstraction and reformulation in artificial intelligence", *Philosophical Transactions of the Royal Society B*, **358** (1435): 1197–1204, doi:10.1098/rstb.2003.1317 (<https://doi.org/10.1098%2Frstb.2003.1317>), PMC 1693218 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1693218>), PMID 12903653 (<https://pubmed.ncbi.nlm.nih.gov/12903653>).
- Howe, J. (November 1994), *Artificial Intelligence at Edinburgh University : a Perspective* (http://www.dai.ed.ac.uk/AI_at_Edinburgh_perspective.html), retrieved 30 August 2007
- Johnson, Mark (1987), *The body in the mind*, Chicago, ISBN 978-0-226-40317-5
- Kurzweil, Ray (2005), *The Singularity is Near*, Viking Press
- Lighthill, Professor Sir James (1973), "Artificial Intelligence: A General Survey", *Artificial Intelligence: a paper symposium*, Science Research Council
- Luger, George; Stubblefield, William (2004), *Artificial Intelligence: Structures and Strategies for Complex Problem Solving* (<https://archive.org/details/artificialintell0000luge/page/720>) (5th ed.), The Benjamin/Cummings Publishing Company, Inc., p. 720 (<https://archive.org/details/artificialintell0000luge/page/720>), ISBN 978-0-8053-4780-7
- McCarthy, John (October 2007), "From here to human-level AI", *Artificial Intelligence*, **171** (18): 1174–1182, doi:10.1016/j.artint.2007.10.009 (<https://doi.org/10.1016%2Fj.artint.2007.10.009>).
- McCorduck, Pamela (2004), *Machines Who Think* (http://www.pamelamc.com/html/machines_who_think.html) (2nd ed.), Natick, MA: A. K. Peters, Ltd., ISBN 1-56881-205-1
- Moravec, Hans (1976), *The Role of Raw Power in Intelligence* (<https://web.archive.org/web/20160303232511/http://www.frc.ri.cmu.edu/users/hpm/project.archive/general.articles/1975/Raw.Power.html>), archived from the original (<http://www.frc.ri.cmu.edu/users/hpm/project.archive/general.articles/1975/Raw.Power.html>) on 3 March 2016, retrieved 29 September 2007
- Moravec, Hans (1988), *Mind Children*, Harvard University Press
- Nagel (1974), "What Is it Like to Be a Bat" (http://organizations.utep.edu/Portals/1475/nagel_bat.pdf) (PDF), *Philosophical Review*, **83** (4): 435–50, doi:10.2307/2183914 (<https://doi.org/10.2307%2F2183914>), JSTOR 2183914 (<https://www.jstor.org/stable/2183914>).
- Newell, Allen; Simon, H. A. (1963), "GPS: A Program that Simulates Human Thought", in Feigenbaum, E.A.; Feldman, J. (eds.), *Computers and Thought*, New York: McGraw-Hill
- Newell, Allen; Simon, H. A. (1976). "Computer Science as Empirical Inquiry: Symbols and Search" (<https://doi.org/10.1145%2F360018.360022>). *Communications of the ACM*. **19** (3): 113–126. doi:10.1145/360018.360022 (<https://doi.org/10.1145%2F360018.360022>).
- Nilsson, Nils (1998), *Artificial Intelligence: A New Synthesis*, Morgan Kaufmann Publishers, ISBN 978-1-55860-467-4
- Russell, Stuart J.; Norvig, Peter (2003), *Artificial Intelligence: A Modern Approach* (<http://aima.cs.berkeley.edu/>) (2nd ed.), Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13-790395-2
- NRC (1999), "Developments in Artificial Intelligence" (<http://www.nap.edu/readingroom/books/far/ch9.html>), *Funding a Revolution: Government Support for Computing Research*, National Academy Press
- Poole, David; Mackworth, Alan; Goebel, Randy (1998), *Computational Intelligence: A Logical Approach* (<http://www.cs.ubc.ca/spider/poole/ci.html>), New York: Oxford University Press
- Searle, John (1980), "Minds, Brains and Programs" (<http://cogprints.org/7150/1/10.1.1.83.5248.pdf>) (PDF), *Behavioral and Brain Sciences*, **3** (3): 417–457, doi:10.1017/S0140525X00005756 (<https://doi.org/10.1017%2FS0140525X00005756>)
- Simon, H. A. (1965), *The Shape of Automation for Men and Management*, New York: Harper & Row
- Sutherland, J.G. (1990), "Holographic Model of Memory, Learning, and Expression", *International Journal of Neural Systems*, **1–3**: 256–267.
- Williams RW, Herrup K (1988), "The control of neuron number", *Annual Review of Neuroscience*, **11**: 423–53, doi:10.1146/annurev.ne.11.030188.002231 (<https://doi.org/10.1146/annurev.ne.11.030188.002231>)

6%2Fannurev.ne.11.030188.002231), PMID 3284447 (<https://pubmed.ncbi.nlm.nih.gov/3284447>).

- de Vega, Manuel; Glenberg, Arthur; Graesser, Arthur, eds. (2008), *Symbols and Embodiment: Debates on meaning and cognition*, Oxford University Press, ISBN 978-0-19-921727-4
- Yudkowsky, Eliezer (2006), Goertzel, Ben; Pennachin, Cassio (eds.), "Artificial General Intelligence" (<https://web.archive.org/web/20090411050423/http://www.singinst.org/upload/LOGI/LOGI.pdf>) (PDF), *Annual Review of Psychology*, Springer, **49**: 585–612, doi:10.1146/annurev.psych.49.1.585 (<https://doi.org/10.1146%2Fannurev.psych.49.1.585>), ISBN 978-3-540-23733-4, PMID 9496632 (<https://pubmed.ncbi.nlm.nih.gov/9496632>), archived from the original (<http://www.singinst.org/upload/LOGI/LOGI.pdf>) (PDF) on 11 April 2009
- Zucker, Jean-Daniel (July 2003), "A grounded theory of abstraction in artificial intelligence", *Philosophical Transactions of the Royal Society B*, **358** (1435): 1293–1309, doi:10.1098/rstb.2003.1308 (<https://doi.org/10.1098%2Frstb.2003.1308>), PMC 1693211 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1693211>), PMID 12903672 (<https://pubmed.ncbi.nlm.nih.gov/12903672>).
- Yudkowsky, Eliezer (2008), "Artificial Intelligence as a Positive and Negative Factor in Global Risk", *Global Catastrophic Risks*, Bibcode:2008gcr..book..303Y (<https://ui.adsabs.harvard.edu/abs/2008gcr..book..303Y>).

External links

- The AGI portal maintained by Pei Wang (<https://cis.temple.edu/~pwang/AGI-Intro.html>)
 - The Genesis Group at MIT's CSAIL (<https://web.archive.org/web/20050405071221/http://genesiss.csail.mit.edu/index.html>) – Modern research on the computations that underlay human intelligence
 - OpenCog – open source project to develop a human-level AI (<http://www.opencog.org/>)
 - Simulating logical human thought (http://academia.wikia.com/wiki/A_Method_for_Simulating_the_Process_of_Logical_Human_Thought)
 - What Do We Know about AI Timelines? (<http://www.openphilanthropy.org/focus/global-catastrophic-risks/potential-risks-advanced-artificial-intelligence/ai-timelines>) – Literature review
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Artificial_general_intelligence&oldid=984693782"

This page was last edited on 21 October 2020, at 15:02 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.