Artificial general intelligence - Wikipedia Jump to content Main menu Main menu move to sidebar hide

Navigation

Contribute

Search Search Personal tools

Pages for logged out editors learn more Contents move to sidebar hide Toggle the table of contents Artificial general intelligence 38 languages Edit links English Tools Tools move to sidebar hide

Actions

General

Print/export

From Wikipedia, the free encyclopedia Hypothetical human-level or stronger AI for a wide range of tasks Not to be confused with Generative artificial intelligence . Part of a series on Artificial intelligence Major goals Approaches Applications Philosophy History Glossary Artificial general intelligence (AGI) is a type of artificial intelligence (AI) that can perform as well or better than humans on a wide range of cognitive tasks.[1] This is in contrast to narrow AI , which is designed for specific tasks.[2] AGI is considered one of various definitions of strong AI .

Creating AGI is a primary goal of AI research and of companies such as OpenAI ,[3] DeepMind , and Anthropic . A 2020 survey identified 72 active AGI R&D projects spread across 37 countries.[4] The timeline for AGI development remains a subject of ongoing debate among researchers and experts. As of 2023[update] , some argue that it may be possible in years or decades; others maintain it might take a century or longer; and a minority believe it may never be achieved.[5] There is debate on the exact definition of AGI, and regarding whether modern large language models (LLMs) such as GPT-4 are early, incomplete forms of AGI.[6] AGI is a common topic in science fiction and futures studies .

Contention exists over the potential for AGI to pose a threat to humanity;[7] for example, OpenAI claims to treat it as an existential risk , while others find the development of AGI to be too remote to present a risk.[8] [5] [9] Terminology AGI is also known as strong AI,[10] [11] full AI,[12] human-level AI[5] or general intelligent action.[13] However, some academic sources reserve the term "strong AI" for computer programs that experience sentience or consciousness .[a] In contrast, weak AI (or narrow AI) is able to solve one specific problem, but lacks general cognitive abilities.[14] [11] Some academic sources use "weak AI" to refer more broadly to any programs that neither experience consciousness nor have a mind in the same sense as humans.[a] Related concepts include artificial superintelligence and transformative AI. An artificial superintelligence (ASI) is a hypothetical type of AGI that is much more generally intelligent than humans,[15] while the notion of transformative AI relates to AI having a large impact on society, for example, similar to the agricultural or industrial revolution.[16] Characteristics Main article: Artificial intelligence Various criteria for intelligence have been proposed (most famously the Turing test) but no definition is broadly accepted.[b] Intelligence traits However, researchers generally hold that intelligence is required to do all of the

following:[18] Many interdisciplinary approaches (e.g. cognitive science , computational intelligence , and decision making) consider additional traits such as imagination (the ability to form novel mental images and concepts)[19] and autonomy .[20] Computer-based systems that exhibit many of these capabilities exist (e.g. see computational creativity , automated reasoning , decision support system , robot , evolutionary computation , intelligent agent). However, no consensus holds that modern AI systems possess them to an adequate degree.

Physical traits Other capabilities are considered desirable in intelligent systems, as they may affect intelligence or aid in its expression. These include:[21] This includes the ability to detect and respond to hazard .[22] Tests for human-level AGISeveral tests meant to confirm human-level AGI have been considered, including:[23] [24] The Turing Test (Turing) A machine and a human both converse 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 AI Eugene Goostman , imitating a 13-year-old boy, achieved Turing's estimate of convincing 33% of judges that it was human in 2014.[25] 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. LLMs can now pass university degree-level exams without even attending the classes.[26] The Employment Test (Nilsson) A machine performs an economically important job at least as well as humans in the same job. AIs are now replacing humans in many roles as varied as fast food and marketing.[27] The Ikea test (Marcus) Also known as the Flat Pack Furniture Test. An AI views the parts and instructions of an Ikea flat-pack product, then controls a robot to assemble the furniture correctly.[28] 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.[29] This has not yet been completed. The Modern Turing Test (Suleyman) An AI model is given $100,000 and has to obtain $1 million.[30] [31] AI-complete problems Main article: AI-complete There are many problems that may require general intelligence to solve the problems as well as humans do. For example, even specific straightforward tasks, like machine translation , require that a 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 original intent (social intelligence). All of these problems need to be solved simultaneously in order to reach human-level machine performance.

A problem is informally called "AI-complete" or "AI-hard" if it is believed that to solve it one would need to implement strong AI, because the solution is beyond the capabilities of a purpose-specific algorithm.[32] AI-complete problems are hypothesised to include general computer vision , natural language understanding , and dealing with unexpected circumstances while solving any real-world problem.[33] AI-complete problems cannot be solved with current computer technology alone, and require human computation . This limitation could be useful to test for the presence of humans, as CAPTCHAs aim to do; and for computer security to repel brute-force attacks .[34] [35] History Classical AI Main articles: History of artificial intelligence and Symbolic artificial intelligence Modern AI research began in the mid-1950s.[36] The first generation of AI researchers were convinced that artificial general intelligence was possible and that it would exist in just a few decades.[37] AI pioneer Herbert A. Simon wrote in 1965:

"machines will be capable, within twenty years, of doing any work a man can do."[38] 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[39] on the project of making HAL 9000 as realistic as possible according to the consensus predictions of the time. He said in 1967, "Within a generation... the problem of creating 'artificial intelligence' will substantially be solved".[40] Several classical AI projects , such as Doug Lenat 's Cyc project (that began in 1984), and Allen Newell 's Soar project, were directed at AGI.

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".[c] In the early 1980s, 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".[44] In response to this and the success of expert systems , both industry and government pumped money into the field.[42] [45] However, confidence in AI spectacularly collapsed in the late 1980s, and the goals of the Fifth Generation Computer Project were never fulfilled.[46] For the second time in 20 years, AI researchers who predicted the imminent achievement of AGI had been mistaken. By the 1990s, AI researchers had a reputation for making vain promises. They became reluctant to make predictions at all[d] and avoided mention of "human level" artificial intelligence for fear of being labeled "wild-eyed dreamer[s]".[48] Narrow AI research Main article: Artificial intelligence In the 1990s and early 21st century, mainstream AI achieved commercial success and academic respectability by focusing on specific sub-problems where AI can produce verifiable results and commercial applications, such as speech recognition and recommendation algorithms .[49] These "applied AI" systems are now used extensively throughout the technology industry, and research in this vein is heavily funded in both academia and industry. As of 2018[update] , development in this field was considered an emerging trend, and a mature stage was expected to be reached in more than 10 years.[50]

At the turn of the century, many mainstream AI researchers[51] hoped that strong AI could be developed by combining 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.[51]

However, even at the time, this was disputed. For example, Stevan Harnad of Princeton University 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).[52] Modern

artificial general intelligence research The term "artificial general intelligence" was used as early as 1997, by Mark Gubrud[53] in a discussion of the implications of fully automated military production and operations. A mathematical formalism of AGI was proposed by Marcus Hutter in 2000. Named AIXI , the proposed AGI agent maximises “the ability to satisfy goals in a wide range of environments”.[54] This type of AGI, characterized by the ability to maximise a mathematical definition of intelligence rather than exhibit human-like behaviour,[55] was also called universal artificial intelligence.[56] The term AGI was re-introduced and popularized by Shane Legg and Ben Goertzel around 2002.[57] AGI research activity in 2006 was described by Pei Wang and Ben Goertzel[58] as "producing publications and preliminary results". The first summer school in AGI was organized in Xiamen, China in 2009[59] by the Xiamen university's Artificial Brain Laboratory and OpenCog. The first university course was given in 2010[60] and 2011[61] at Plovdiv University, Bulgaria by Todor Arnaudov. MIT presented a course on AGI in 2018, organized by Lex Fridman and featuring a number of guest lecturers.

As of 2023[update] , a small number of computer scientists are active in AGI research, and many contribute to a series of AGI conferences . However, increasingly more researchers are interested in open-ended learning,[62] [63] which is the idea of allowing AI to continuously learn and innovate like humans do. Although most open-ended learning works are still done on Minecraft ,[18] [21] [54] its application can be extended to robotics and the sciences.

Feasibility As of 2023[update] , complete forms of AGI remain speculative.[64] [65] No system that meets the generally agreed upon criteria for AGI has yet been demonstrated. Opinions vary both on whether and when artificial general intelligence will arrive. AI pioneer Herbert A. Simon speculated in 1965 that "machines will be capable, within twenty years, of doing any work a man can do". 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".[66] 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.[67] Most AI researchers believe strong AI can be achieved in the future, but some thinkers, like Hubert Dreyfus and Roger Penrose , deny the possibility of achieving strong AI.[68] [69] John McCarthy is among those who believe human-level AI will be accomplished, but that the present level of progress is such that a date cannot accurately be predicted.[70] AI experts' views on the feasibility of AGI wax and wane. Four polls conducted in 2012 and 2013 suggested that the median estimate 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.[71] [72] Further current AGI progress considerations can be found above Tests for confirming human-level AGI .

A report by Stuart Armstrong and Kaj Sotala of the Machine Intelligence Research Institute found that "over [a] 60-year time frame there is a strong bias towards predicting the arrival of human-level AI as between 15 and 25 years from the time the prediction was made". They analyzed 95 predictions made between 1950 and 2012 on when human-level AI will come about.[73] In 2023, Microsoft researchers published a detailed evaluation of GPT-4 . They concluded: "Given the breadth and depth of GPT-4’s capabilities, we believe that it could reasonably be viewed as an early (yet still incomplete) version of an artificial general intelligence

(AGI) system."[74] Another study in 2023 reported that GPT-4 outperforms 99% of humans on the Torrance tests of creative thinking .[75] [76] Timescales In the introduction to his 2006 book,[77] Goertzel says that estimates of the time needed before a truly flexible AGI is built vary from 10 years to over a century. As of 2007[update] , the consensus in the AGI research community seemed to be that the timeline discussed by Ray Kurzweil in 2005 in The Singularity is Near [78] (i.e. between 2015 and 2045) was plausible.[79] 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. That paper has been criticized for how it categorized opinions as expert or non-expert.[80] In 2012, Alex Krizhevsky , Ilya Sutskever , and Geoffrey Hinton developed a neural network called AlexNet , which won the ImageNet competition with a top-5 test error rate of 15.3%, significantly better than the second-best entry's rate of 26.3% (the traditional approach used a weighted sum of scores from different pre-defined classifiers).[81] AlexNet was regarded as the initial ground-breaker of the current deep learning wave.[81] 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, Apple's Siri, and others. At the maximum, these AIs 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 were carried out in 2014, with the IQ score reaching a maximum value of 27.[82] [83] 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 be classified as a narrow AI system.[84] In the same year, Jason Rohrer used his GPT-3 account to develop a chatbot, and provided a chatbot-developing platform called "Project December". OpenAI asked for changes to the chatbot to comply with their safety guidelines; Rohrer disconnected Project December from the GPT-3 API.[85] In 2022, DeepMind developed Gato , a "general-purpose" system capable of performing more than 600 different tasks.[86] In 2023, Microsoft Research published a study on an early version of OpenAI's GPT-4 , contending that it exhibited more general intelligence than previous AI models and demonstrated human-level performance in tasks spanning multiple domains, such as mathematics, coding, and law. This research sparked a debate on whether GPT-4 could be considered an early, incomplete version of artificial general intelligence, emphasizing the need for further exploration and evaluation of such systems.[87] In 2023, the AI researcher Geoffrey Hinton stated that:[88] The idea that this stuff could actually get smarter than people – a few people believed that, [...]. But most people thought it was way off. And I thought it was way off. I thought it was 30 to 50 years or even longer away. Obviously, I no longer think that. In March 2024, Nvidia 's CEO, Jensen Huang , stated his expectation that within five years, AI would be capable of passing any test at least as well as humans.[89] Whole brain emulation Main articles: Whole brain emulation and Brain simulation While the development of large language models is considered the most promising path to AGI,[90] whole brain emulation can serve as an alternative approach. With whole brain simulation, a brain model is built by scanning and mapping a biological brain in detail, and then copying and simulating it on a computer system or another computational device. The simulation model must be sufficiently faithful to the original, so that it behaves in practically the same way as the original brain.[91] Whole brain emulation is a type of brain simulation that is discussed in

computational neuroscience and neuroinformatics , and for medical research purposes. It has been discussed in artificial intelligence research[79] 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 [78] predicts that a map of sufficient quality will become available on a similar timescale to the computing power required to emulate it.

Early estimates Estimates of how much processing power is needed to emulate a human brain at various levels (from Ray Kurzweil, 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.[92] For low-level brain simulation, a very powerful cluster of computers or GPUs would be required, given the enormous quantity of synapses within the human brain . Each of the 1011 (one hundred billion) neurons has on average 7,000 synaptic connections (synapses) to other neurons. The brain of a three-year-old child has about 1015 synapses (1 quadrillion). This number declines with age, stabilizing by adulthood. Estimates vary for an adult, ranging from 1014 to 5×1014 synapses (100 to 500 trillion).[93] An estimate of the brain's processing power, based on a simple switch model for neuron activity, is around 1014 (100 trillion) synaptic updates per second (SUPS).[94] In 1997, Kurzweil looked at various estimates for the hardware required to equal the human brain and adopted a figure of 1016 computations per second (cps).[e] (For comparison, if a "computation" was equivalent to one "floating-point operation " – a measure used to rate current supercomputers – then 1016 "computations" would be equivalent to 10 petaFLOPS , achieved in 2011 , while 1018 was achieved in 2022 .) 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.

Current research The Human Brain Project , an EU -funded initiative active from 2013 to 2023, has developed a particularly detailed and publicly accessible atlas of the human brain.[97] In 2023, researchers from Duke University performed a high-resolution scan of a mouse brain.[98] A supercomputer with similar computing capability as the human brain is expected in April 2024. Called "DeepSouth", it could perform 228 trillions of synaptic operations per second.[99] Criticisms of simulation-based approaches 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 broad outline. 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 known to play a role in cognitive processes.[100] A fundamental criticism of the simulated brain approach derives from embodied cognition theory which asserts that human embodiment is an essential aspect of human intelligence and is necessary to ground meaning.[101] [98] If this theory is correct, any fully functional brain model will need to encompass more than just the neurons (e.g., a robotic body). Goertzel[79] proposes virtual

embodiment (like in metaverses like Second Life) as an option, but it is unknown whether this would be sufficient.

Philosophical perspective See also: Philosophy of artificial intelligence and Turing test "Strong AI" as defined in philosophy In 1980, philosopher John Searle coined the term "strong AI" as part of his Chinese room argument.[102] He wanted to distinguish between two different hypotheses about artificial intelligence:[f] The first one he called "strong" because it makes a stronger statement: it assumes something special has happened to the machine that goes beyond those abilities that we can test. The behaviour of a "weak AI" machine would be precisely identical to a "strong AI" machine, but the latter would also have subjective conscious experience. This usage is also common in academic AI research and textbooks.[103] In contrast to Searle and mainstream AI, some futurists such as Ray Kurzweil use the term "strong AI" to mean "human level artificial general intelligence".[78] This is not the same as Searle's strong AI , unless it is assumed that consciousness is necessary for human-level AGI. Academic philosophers such as Searle do not believe that is the case, and to most artificial intelligence researchers the question is out-of-scope.[104] Mainstream AI is most interested in how a program behaves .[105] According to Russell and Norvig , "as long as the program works, they don't care if you call it real or a simulation."[104] If the program can behave as if it has a mind, then there is no need to know if it actually has mind – indeed, there would be no way to tell. For AI research, Searle's "weak AI hypothesis" is equivalent to the statement "artificial general intelligence is possible". Thus, according to Russell and Norvig, "most AI researchers take the weak AI hypothesis for granted, and don't care about the strong AI hypothesis."[104] Thus, for academic AI research, "Strong AI" and "AGI" are two different things.

Consciousness, self-awareness, sentience Other aspects of the human mind besides intelligence are relevant to the concept of AGI or "strong AI", and these play a major role in science fiction and the ethics of artificial intelligence :

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 . Preliminary work has been conducted on integrating strong AI with existing legal and social frameworks, focusing on the legal position and rights of 'strong' AI.[108] It remains to be shown whether "artificial consciousness " is necessary for AGI. However, many AGI researchers regard research that investigates possibilities for implementing consciousness as vital.[109] Bill Joy , among others, argues a machine with these traits may be a threat to human life or dignity.[110] Research challenges See also: History of artificial intelligence § The problems , and History of artificial intelligence § Predictions (or "Where is HAL 9000?") Progress in artificial intelligence has historically gone through periods of rapid progress separated by periods when progress appeared to stop.[68] Ending each hiatus were fundamental advances in hardware, software or both to create space for further progress.[68] [111] [112] For example, the computer hardware available in the twentieth century was not sufficient to implement deep learning, which requires large numbers of GPU -enabled CPUs .[113] A further challenge is the lack of clarity in defining what intelligence entails. Does it require consciousness? Must it display the ability to set goals as well as pursue them? Is it purely a matter of scale such that if model sizes increase sufficiently, intelligence will emerge? Are facilities such as planning, reasoning, and causal understanding required? Does intelligence require explicitly replicating the brain and its specific faculties? Does it require emotions?[114] Benefits AGI could have a wide variety of applications. If oriented

towards such goals, AGI could help mitigate various problems in the world such as hunger, poverty and health problems.[115] AGI could improve the productivity and efficiency in most jobs. For example, in public health, AGI could accelerate medical research, notably against cancer.[116] It could take care of the elderly,[117] and democratize access to rapid, high-quality medical diagnostics. It could offer fun, cheap and personalized education.[117] For virtually any job that benefits society if done well, it would probably sooner or later be preferable to leave it to an AGI. The need to work to subsist could become obsolete if the wealth produced is properly redistributed .[117] [118] This also raises the question of the place of humans in a radically automated society.

AGI could also help to make rational decisions, and to anticipate and prevent disasters. It could also help to reap the benefits of potentially catastrophic technologies such as nanotechnology or climate engineering , while avoiding the associated risks.[119] If an AGI's primary goal is to prevent existential catastrophes such as human extinction (which could be difficult if the Vulnerable World Hypothesis turns out to be true),[120] it could take measures to drastically reduce the risks[119] while minimizing the impact of these measures on our quality of life.

Risks Existential risks Main articles: Existential risk from artificial general intelligence and AI safety AGI may represent multiple types of existential risk , which are risks that threaten "the premature extinction of Earth-originating intelligent life or the permanent and drastic destruction of its potential for desirable future development".[121] The risk of human extinction from AGI has been the topic of many debates, but there is also the possibility that the development of AGI would lead to a permanently flawed future. Notably, it could be used to spread and preserve the set of values of whoever develops it. If humanity still has moral blind spots similar to slavery in the past, AGI might irreversibly entrench it, preventing moral progress .[122] Furthermore, AGI could facilitate mass surveillance and indoctrination, which could be used to create a stable repressive worldwide totalitarian regime.[123] [124] There is also a risk for the machines themselves. If machines that are sentient or otherwise worthy of moral consideration are mass created in the future, engaging in a civilizational path that indefinitely neglects their welfare and interests could be an existential catastrophe.[125] [126] Considering how much AGI could improve humanity's future and help reduce other existential risks, Toby Ord calls these existential risks "an argument for proceeding with due caution", not for "abandoning AI".[123]

Risk of loss of control and human extinction The thesis that AI poses an existential risk for humans, and that this risk needs more attention, is controversial but has been endorsed in 2023 by many public figures, AI researchers and CEOs of AI companies such as Elon Musk , Bill Gates , Geoffrey Hinton , Yoshua Bengio , Demis Hassabis and Sam Altman .[127] [128] In 2014, Stephen Hawking criticized widespread indifference:

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.[129] The potential fate of humanity has sometimes been compared to the fate of gorillas threatened by human activities. The comparison states that greater intelligence allowed humanity to dominate gorillas, which are now vulnerable in ways that they could not have anticipated. As a result, the gorilla has become an endangered species, not out of malice, but simply as a collateral damage from human activities.[130] The skeptic Yann LeCun considers

that AGIs will have no desire to dominate humanity and that we should be careful not to anthropomorphize them and interpret their intents as we would for humans. He said that people won't be "smart enough to design super-intelligent machines, yet ridiculously stupid to the point of giving it moronic objectives with no safeguards".[131] On the other side, the concept of instrumental convergence suggests that almost whatever their goals, intelligent agents will have reasons to try to survive and acquire more power as intermediary steps to achieving these goals. And that this does not require having emotions.[132] Many scholars who are concerned about existential risk advocate for more research into solving the "control problem " to answer the question: what types of safeguards, algorithms, or architectures can programmers implement to maximise the probability that their recursively-improving AI would continue to behave in a friendly , rather than destructive, manner after it reaches superintelligence?[133] [134] Solving the control problem is complicated by the AI arms race (which could lead to a race to the bottom of safety precautions in order to release products before competitors),[135] and the use of AI in weapon systems.[136] The thesis that AI can pose existential risk also has detractors. Skeptics usually say that AGI is unlikely in the short-term, or that concerns about AGI distract from other issues related to current AI.[137] Former Google fraud czar Shuman Ghosemajumder considers that for many people outside of the technology industry, existing chatbots and LLMs are already perceived as though they were AGI, leading to further misunderstanding and fear.[138] 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.[139] Some researchers believe that the communication campaigns on AI existential risk by certain AI groups (such as OpenAI, Anthropic, DeepMind, and Conjecture) may be an at attempt at regulatory capture and to inflate interest in their products.[140] [141] In 2023, the CEOs of Google DeepMind, OpenAI and Anthropic, along with other industry leaders and researchers, issued a joint statement asserting that "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war."[128] Mass unemployment Further information: Technological unemployment Researchers from OpenAI estimated that "80% of the U.S. workforce could have at least 10% of their work tasks affected by the introduction of LLMs, while around 19% of workers may see at least 50% of their tasks impacted".[142] [143] They consider office workers to be the most exposed, for example mathematicians, accountants or web designers.[143] AGI could have a better autonomy, ability to make decisions, to interface with other computer tools, but also to control robotized bodies.

According to Stephen Hawking, the outcome of automation on the quality of life will depend on how the wealth will be redistributed:[118] Everyone can enjoy a life of luxurious leisure if the machine-produced wealth is shared, or most people can end up miserably poor if the machine-owners successfully lobby against wealth redistribution. So far, the trend seems to be toward the second option, with technology driving ever-increasing inequality Elon Musk considers that the automation of society will require governments to adopt a universal basic income .[144] See also Notes References Sources Further reading External links Existential risk from artificial intelligence Concepts Organizations People Other Category Retrieved from "https://en.wikipedia.org/w/index.php?title=Artificial_general_intelligence&oldid=1219278162 " Categories : Hidden categories: