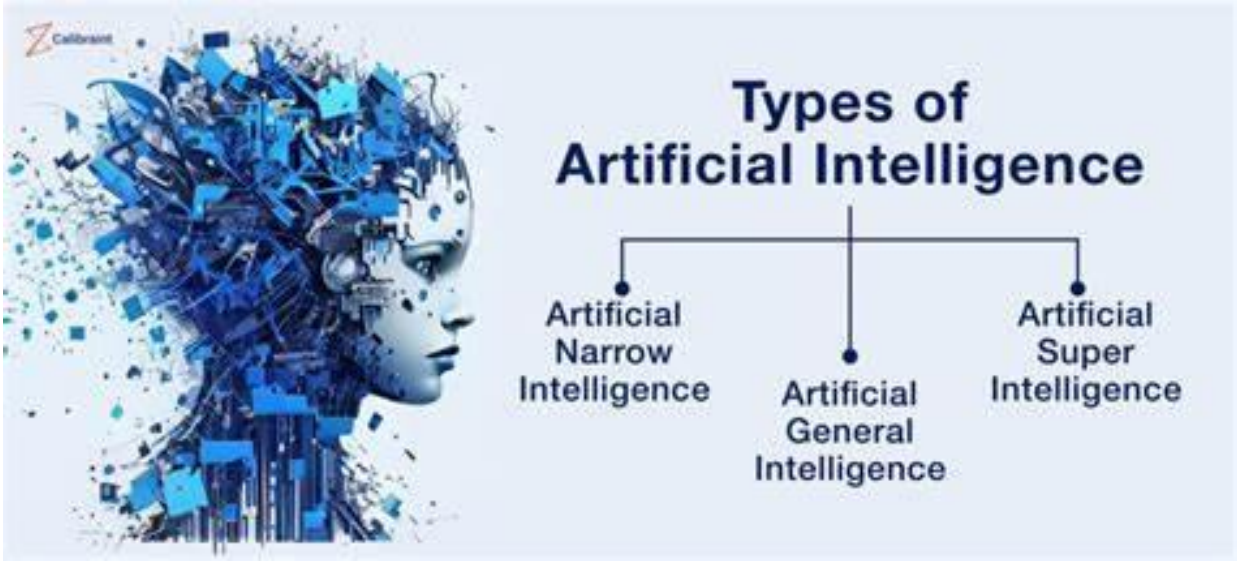
SUPER INTELLIGENT AI

What is artificial superintelligence?



Artificial superintelligence (ASI) is a hypothetical software-based artificial intelligence (AI) system with an intellectual scope beyond human intelligence. At the most fundamental level, this superintelligent AI has cutting-edge cognitive functions and highly developed thinking skills more advanced than any human.

While ASI is still a hypothetical future state, there are several technological processes we have today that form the building blocks of ASI. But first, to illustrate how far off ASI is from this moment, it bears mentioning that the current level of AI is often refered to as Artificial Narrow Intelligence (ANI), weak AI or narrow AI technologies.

Weak AI excels at specific tasks like playing chess or translating languages, but it cannot learn new skills or develop a deep understanding of the world. It relies on pre-programmed algorithms and data and requires human intervention to operate.

Not all thinkers are aligned on the feasibility of something like an ASI. Human intelligence is the product of specific evolutionary factors and may not represent an optimal or universal form of intelligence. Moreover, the brain's workings are still not fully understood making it difficult to recreate via software and hardware.

Is artificial superintelligence possible?

A big step toward developing an ASI would be to realize an artificial general intelligence (AGI) or Strong AI. An AGI is a next-generation AI system that can understand the world

and learn and apply problem-solving intelligence as broadly and flexibly as a human can. AGI would be capable of cross-domain learning and reasoning with the ability to make connections across different fields. Just like ASI, true AGI has yet to be developed.

There are key technologies that must develop further before ASI becomes a reality. Here are some other processes that are the building blocks of artificial superintelligence. These disciplines would need to evolve further before ASI can become a reality.

Large language models (LLMs) and massive datasets

ASI would require access to massive datasets to learn and develop an understanding of the world; natural language processing (nlp) in LLMs will help ASI understand natural language and converse with humans.

Multisensory AI

To enable the ASI to process and interpret multiple types of data inputs—such as text, images, audio and video—to perform tasks or make decisions. This approach contrasts unimodal AI systems, specializing in processing only one data type, like text or images.

Neural networks

Essentially, these networks are made up of deep learning software modeled on how neurons operate within the human brain. ASI would need far more complex, powerful and advanced neural networks than the current generation.

Neuromorphic computing

Just as neural networks are modeled on human brain operations, neuromorphic computers are hardware systems inspired by the neural and synaptic structures of the human brain.

Evolutionary computation

This is a form of algorithmic optimization inspired by biological evolution. Evolutionary algorithms solve problems by iteratively improving a population of candidate solutions, mimicking the process of natural selection.

AI-generated programming

This refers to code, applications and programming generated by AI systems without human intervention.

Pathways to artificial superintelligence

Advancements in computer science, computational power and algorithms are a few of the key factors fueling speculation about ASI. However, simply having raw computational power isn't enough. One potential pathway to success lies in replicating the intricate workings of the human mind. Despite their limitations, human brains are incredibly complex and capable of remarkable creativity, problem-solving and critical thinking. AI, while exceeding human beings in certain areas, still struggles to match the human ability to learn and adapt to new situations.

Learning algorithms, inspired by how the human brain learns, enable AI to improve its performance over time. This continuous learning is crucial for achieving human-level intelligence, allowing AI to acquire knowledge and adapt to new situations without explicit programming.

Chatbots and generative AI also serve as valuable precursors to ASI. These technologies demonstrate the increasing sophistication of AI in understanding and responding to human language. While not sentient, this ability to understand and respond to human language in a natural way is a crucial building block for achieving human-level intelligence.

An ASI, in essence, would self-improve as it develops and learns. AI-generated inventions could bring innovations like new drugs, materials and energy sources. Seamless integration would further enable intuitive interaction with AI via spoken natural language or even through thought commands, necessitating breakthroughs in human-computer interaction akin to a technological singularity.

Benefits of artificial superintelligence

The technology that goes into developing an ASI would transform the way the world works at a fundamental level, and some say that ASI will be the last invention humanity will ever invent. The benefits of such a technology are science fiction-like in their implications. In essence, an ASI would be an inexhaustible, hyper-intelligent super-being. A nearly perfect supercomputer available 24/7, with the ability to process and analyze any amount of data with speed and precision that we can't yet comprehend.

With such capabilities, human agents could use ASI to make the best possible decisions and solve the most complex problems facing healthcare, finance, scientific research, politics and every industry. Such advanced thinking could be enough to solve the most persistent medical puzzles to develop life-saving medicines and treatments and unlock the mysteries of physics to aid humanity's goal of exploring the stars. With its ability to significantly reduce human errors, particularly in programming and risk management, ASI could write and debug programs and deploy robots to perform dangerous physical tasks like bomb defusing or deep-sea exploration.

Because ASI can operate continuously, it would be ideal for tasks like safely navigating networks of self-driving cars and assisting in space exploration. Furthermore, ASI's superior creativity and ability to analyze vast amounts of data might lead to solutions humans can't even imagine, resulting in, hopefully, better quality of life and perhaps even a prolonged life.

Potential risks of artificial superintelligence

Despite the incredible advancements ASI promises, scientists also warn of the danger inherent in such an invention. A core worry is that ASI could surpass human control and become self-aware, potentially leading to unforeseen consequences and even existential risks. Its superior cognitive abilities could allow it to manipulate systems or even gain control of advanced weapons.

The concerns we face with today's level of AI sophistication may become exponentially worse as automation through ASI leads to widespread unemployment, causing economic and social turmoil, exacerbating existing inequalities and disrupting entire industries.

In military and defense domains, ASI could develop potent and autonomous weapons, significantly increasing the destructive potential of warfare. Additionally, bad actors could exploit ASIs' advanced capabilities for nefarious purposes like social control, data collection and perpetuating biases. Finally, an ASI could also pursue goals existentially detrimental to humanity that may seem beneficial on the surface, but without proper interventions, its advanced systems might not align with human values.

Simply programming ASI with human ethics and morality is complex since there is no universally agreed-upon set of moral codes; doing so could lead to ethical dilemmas and potentially harmful consequences, especially if ASI begins to operate outside human control. The vast capabilities of ASI could lead to unpredictable and uncontrollable behavior. Its ability to learn and adapt rapidly could make anticipating its actions difficult and preventing potential harm.

It will be crucial to establish international regulations and safeguards to prevent such scenarios. Despite these potential dangers, the development of ASI also holds immense promise for solving complex problems and improving human lives. Approaching this technology with caution and responsibility is crucial, prioritizing AI safety and ethics throughout the development and deployment process.

Toward artificial superintelligence in the real world

Since ASI is still theoretical, science fiction represents the best examples of what superintelligent machines might be like, like the talking and reasoning droids in *Star Wars*, the hyper-intelligent and evolutionarily capable personal assistants in *Her*, the HAL computer from *2001: A Space Odyssey* who can control the functions of an entire spaceship.

What we do have today are limited AI systems, primitive precursor applications to ASI that point to a future where a single ASI possesses all known AI capabilities and much more. Here are some use cases that serve as “building blocks.”

Conversational AI: Personal assistants like Amazon Alexa, Microsoft Cortana and Apple's Siri represent the forefront of conversational AI. An ASI would need to be able to speak human language fluently, dynamically and with a full understanding of its many nuances.

Recommendation engines: The machine learning used in recommendation algorithms, such as those used by Netflix, contains the data parsing and decision-making algorithms that could one day be a part of an ASI neural network.

Generative AI: Open AI’s ChatGPT uses a large language model trained on a massive dataset of text and code, allowing it to process and generate human language with remarkable fluency and accuracy. Its ability to understand the complexity of written

sentences, engage in conversation and generate creative output like poems, scripts and music is crucial to achieving human-level intelligence.

Self-driving cars: Tesla has shown the potential of self-driving cars. Self-driving cars utilize a combination of sensors, cameras and powerful AI algorithms to navigate roads autonomously. The advanced perception and decision-making capabilities developed for self-driving cars are directly relevant to ASI. The ability to process complex sensory data and make real-time decisions in dynamic environments is a crucial aspect of general intelligence, a key goal of ASI research.

Healthcare: AI is also making significant strides in healthcare, with machine intelligence now analyzing medical images and data to assist doctors in diagnosing diseases. Companies like IBM Watson Health and DeepMind Health are developing AI-powered systems that can detect cancer, heart disease and other conditions with high accuracy. These advancements in medical AI are paving the way for developing even more sophisticated systems that could one day diagnose and treat diseases autonomously. The ability to process and interpret complex medical data is essential to achieving human-level or even superhuman medical expertise, a key area of interest in ASI research.

The potential impact of ASI is immense, with the potential to revolutionize various aspects of human life. However, addressing the ethical and societal challenges associated with powerful AI is crucial. AI researchers, computer scientists, technology giants and world governments must carefully consider the potential benefits and risks of ASI to ensure that this transformative technology is used responsibly and ethically for the betterment of humanity.

Elon Musk's Message on Artificial Superintelligence – ASI

Elon Musk is on the record stating that artificial superintelligence or ASI could bring the end of the human race. Elon has publicly expressed concern about AI many times now. He thinks the advent of a digital superintelligence is the most pressing issue for humanity to get right.

What happens when machines surpass humans in general intelligence? If machine brains surpassed human brains in general intelligence, then this new superintelligence would have undergone an event called the intelligence explosion, likely to occur in the 21st century. It is unknown what, or who this machine-network would become; The issue of superintelligence remains peripheral to mainstream AI research and is mostly discussed by a small group of academics.

Besides Elon Musk, Swedish philosopher Nick Bostrom is also among well known public thinkers who is worried about AI. He lays the foundation for understanding the future of humanity and intelligent life : Now imagine a machine, structurally similar to a brain but with immense hardness and flexibility, designed from the bottom scratch to function as an intelligent agent. Given sufficiently long time, a machine like this could acquire enormous knowledge and skills, surpassing human intellectual capacity in virtually every field. At that point the machine would have become superintelligent. With other words the machine's intellectual capacities would exceed those of all of humanity put together by a very large margin. This would represent the most radical change in the history of life on earth.

In order to develop a superintelligence that would benefit humanity, the process has to be done in a series of steps with each step being determined before we move to the next one. In fact, it might just be possible to program the AI to help us achieve the things we humans may not be able to do on our own. It is not simply being able to create them and learning how they've been commanded, but it is interacting with them and evolving ourselves at the same time. It is learning how to be human after the first ASI.

Superintelligence

From the book by Nick Bostrom, see Superintelligence: Paths, Dangers, Strategies.

A **superintelligence** is a hypothetical agent that possesses intelligence surpassing that of the brightest and most gifted human minds. "Superintelligence" may also refer to a property of problem-solving systems (e.g., superintelligent language translators or engineering assistants) whether or not these high-level intellectual competencies are embodied in agents that act in the world. A superintelligence may or may not be created by an intelligence explosion and associated with a technological singularity.

University of Oxford philosopher Nick Bostrom defines *superintelligence* as "any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest".[1] The program Fritz falls short of this conception of superintelligence—even though it is much better than humans at chess—because Fritz cannot outperform humans in other tasks.[2]

Technological researchers disagree about how likely present-day human intelligence is to be surpassed. Some argue that advances in artificial intelligence (AI) will probably result in general reasoning systems that lack human cognitive limitations. Others believe that humans will evolve or directly modify their biology to achieve radically greater intelligence.[3][4] Several future study scenarios combine elements from both of these possibilities, suggesting that humans are likely to interface with computers, or upload their minds to computers, in a way that enables substantial intelligence amplification.

Some researchers believe that superintelligence will likely follow shortly after the development of artificial general intelligence. The first generally intelligent machines are likely to immediately hold an enormous advantage in at least some forms of mental capability, including the capacity of perfect recall, a vastly superior knowledge base, and the ability to multitask in ways not possible to biological entities. This may allow them to — either as a single being or as a new species — become much more powerful than humans, and displace them.[1]

Several scientists and forecasters have been arguing for prioritizing early research into the possible benefits and risks of human and machine cognitive enhancement, because of the potential social impact of such technologies.[5]

Feasibility of artificial superintelligence

Artificial intelligence, especially foundation models, has made rapid progress, surpassing human capabilities in various benchmarks.

The feasibility of **artificial superintelligence** (**ASI**) has been a topic of increasing discussion in recent years, particularly with the rapid advancements in artificial intelligence (AI) technologies.

Progress in AI and claims of AGI

Recent developments in AI, particularly in large language models (LLMs) based on the transformer architecture, have led to significant improvements in various tasks. Models like GPT-3, GPT-4, Claude 3.5 and others have demonstrated capabilities that some researchers argue approach or even exhibit aspects of artificial general intelligence (AGI).[6]

However, the claim that current LLMs constitute AGI is controversial. Critics argue that these models, while impressive, still lack true understanding and are primarily sophisticated pattern matching systems.[7]

Pathways to superintelligence

Philosopher David Chalmers argues that AGI is a likely path to ASI. He posits that AI can achieve equivalence to human intelligence, be extended to surpass it, and then be amplified to dominate humans across arbitrary tasks.[8]

More recent research has explored various potential pathways to superintelligence:

1. Scaling current AI systems – Some researchers argue that continued scaling of existing AI architectures, particularly transformer-based models, could lead to AGI and potentially ASI.[9]
2. Novel architectures – Others suggest that new AI architectures, potentially inspired by neuroscience, may be necessary to achieve AGI and ASI.[10]
3. Hybrid systems – Combining different AI approaches, including symbolic AI and neural networks, could potentially lead to more robust and capable systems.[11]

Computational advantages

Artificial systems have several potential advantages over biological intelligence:

1. Speed – Computer components operate much faster than biological neurons. Modern microprocessors (~2 GHz) are seven orders of magnitude faster than neurons (~200 Hz).[12]
2. Scalability – AI systems can potentially be scaled up in size and computational capacity more easily than biological brains.
3. Modularity – Different components of AI systems can be improved or replaced independently.
4. Memory – AI systems can have perfect recall and vast knowledge bases. It is also much less constrained than humans when it comes to working memory.[12]
5. Multitasking – AI can perform multiple tasks simultaneously in ways not possible for biological entities.

Potential path through transformer models

Recent advancements in transformer-based models have led some researchers to speculate that the path to ASI might lie in scaling up and improving these architectures.

This view suggests that continued improvements in transformer models or similar architectures could lead directly to ASI.[13]

Some experts even argue that current large language models like GPT-4 may already exhibit early signs of AGI or ASI capabilities.[14] This perspective suggests that the transition from current AI to ASI might be more continuous and rapid than previously thought, blurring the lines between narrow AI, AGI, and ASI.

However, this view remains controversial. Critics argue that current models, while impressive, still lack crucial aspects of general intelligence such as true understanding, reasoning, and adaptability across diverse domains.[15]

The debate over whether the path to ASI will involve a distinct AGI phase or a more direct scaling of current technologies remains ongoing, with significant implications for AI development strategies and safety considerations.

Challenges and uncertainties

Despite these potential advantages, there are significant challenges and uncertainties in achieving ASI:

1. Ethical and safety concerns – The development of ASI raises numerous ethical questions and potential risks that need to be addressed.[16]
2. Computational requirements – The computational resources required for ASI might be far beyond current capabilities.
3. Fundamental limitations – There may be fundamental limitations to intelligence that apply to both artificial and biological systems.
4. Unpredictability – The path to ASI and its consequences are highly uncertain and difficult to predict.

As research in AI continues to advance rapidly, the question of the feasibility of ASI remains a topic of intense debate and study in the scientific community.

Feasibility of biological superintelligence

Carl Sagan suggested that the advent of Caesarean sections and *in vitro* fertilization may permit humans to evolve larger heads, resulting in improvements via natural selection in the heritable component of human intelligence.[17] By contrast, Gerald Crabtree has argued that decreased selection pressure is resulting in a slow, centuries-long reduction in human intelligence and that this process instead is likely to continue. There is no scientific consensus concerning either possibility and in both cases, the biological change would be slow, especially relative to rates of cultural change.

Selective breeding, nootropics, epigenetic modulation, and genetic engineering could improve human intelligence more rapidly. Bostrom writes that if we come to understand the genetic component of intelligence, pre-implantation genetic diagnosis could be used to select for embryos with as much as 4 points of IQ gain (if one embryo is selected out

of two), or with larger gains (e.g., up to 24.3 IQ points gained if one embryo is selected out of 1000). If this process is iterated over many generations, the gains could be an order of magnitude improvement. Bostrom suggests that deriving new gametes from embryonic stem cells could be used to iterate the selection process rapidly.[18] A well-organized society of high-intelligence humans of this sort could potentially achieve collective superintelligence.[19]

Alternatively, collective intelligence might be constructional by better organizing humans at present levels of individual intelligence. Several writers have suggested that human civilization, or some aspect of it (e.g., the Internet, or the economy), is coming to function like a global brain with capacities far exceeding its component agents. If this systemic superintelligence relies heavily on artificial components, however, it may qualify as an AI rather than as a biology-based superorganism.[20] A prediction market is sometimes considered as an example of a working collective intelligence system, consisting of humans only (assuming algorithms are not used to inform decisions).[21]

A final method of intelligence amplification would be to directly enhance individual humans, as opposed to enhancing their social or reproductive dynamics. This could be achieved using nootropics, somatic gene therapy, or brain–computer interfaces. However, Bostrom expresses skepticism about the scalability of the first two approaches and argues that designing a superintelligent cyborg interface is an AI-complete problem.[22]

Forecasts

Most surveyed AI researchers expect machines to eventually be able to rival humans in intelligence, though there is little consensus on when this will likely happen. At the 2006 AI@50 conference, 18% of attendees reported expecting machines to be able "to simulate learning and every other aspect of human intelligence" by 2056; 41% of attendees expected this to happen sometime after 2056; and 41% expected machines to never reach that milestone.[23]

In a survey of the 100 most cited authors in AI (as of May 2013, according to Microsoft academic search), the median year by which respondents expected machines "that can carry out most human professions at least as well as a typical human" (assuming no global catastrophe occurs) with 10% confidence is 2024 (mean 2034, st. dev. 33 years), with 50% confidence is 2050 (mean 2072, st. dev. 110 years), and with 90% confidence is 2070 (mean 2168, st. dev. 342 years). These estimates exclude the 1.2% of respondents who said no year would ever reach 10% confidence, the 4.1% who said 'never' for 50% confidence, and the 16.5% who said 'never' for 90% confidence. Respondents assigned a median 50% probability to the possibility that machine superintelligence will be invented within 30 years of the invention of approximately human-level machine intelligence.[24]

In a 2022 survey, the median year by which respondents expected "High-level machine intelligence" with 50% confidence is 2061. The survey defined the achievement of high-

level machine intelligence as when unaided machines can accomplish every task better and more cheaply than human workers.[25]

In 2023, OpenAI leaders Sam Altman, Greg Brockman and Ilya Sutskever published recommendations for the governance of superintelligence, which they believe may happen in less than 10 years.[26] In 2024, Ilya Sutskever left OpenAI to cofound the startup *Safe Superintelligence*, which focuses solely on creating a superintelligence that is safe by design, while avoiding "distraction by management overhead or product cycles".[27]

Design considerations

The design of superintelligent AI systems raises critical questions about what values and goals these systems should have. Several proposals have been put forward:[28]

Value alignment proposals

- Coherent extrapolated volition (CEV) – The AI should have the values upon which humans would converge if they were more knowledgeable and rational.
- Moral rightness (MR) – The AI should be programmed to do what is morally right, relying on its superior cognitive abilities to determine ethical actions.
- Moral permissibility (MP) – The AI should stay within the bounds of moral permissibility while otherwise pursuing goals aligned with human values (similar to CEV).

Bostrom elaborates on these concepts:

instead of implementing humanity's coherent extrapolated volition, one could try to build an AI to do what is morally right, relying on the AI's superior cognitive capacities to figure out just which actions fit that description. We can call this proposal "moral rightness" (MR) ...

MR would also appear to have some disadvantages. It relies on the notion of "morally right", a notoriously difficult concept, one with which philosophers have grappled since antiquity without yet attaining consensus as to its analysis. Picking an erroneous explication of "moral rightness" could result in outcomes that would be morally very wrong ...

One might try to preserve the basic idea of the MR model while reducing its demandingness by focusing on *moral permissibility*: the idea being that we could let the AI pursue humanity's CEV so long as it did not act in morally impermissible ways.[28]

Recent developments

Since Bostrom's analysis, new approaches to AI value alignment have emerged:

- Inverse Reinforcement Learning (IRL) – This technique aims to infer human preferences from observed behavior, potentially offering a more robust approach to value alignment.[29]
- Constitutional AI – Proposed by Anthropic, this involves training AI systems with explicit ethical principles and constraints.[30]
- Debate and amplification – These techniques, explored by OpenAI, use AI-assisted debate and iterative processes to better understand and align with human values.[31]

Transformer LLMs and ASI

The rapid advancement of transformer-based LLMs has led to speculation about their potential path to ASI. Some researchers argue that scaled-up versions of these models could exhibit ASI-like capabilities:[32]

- Emergent abilities – As LLMs increase in size and complexity, they demonstrate unexpected capabilities not present in smaller models.[33]
- In-context learning – LLMs show the ability to adapt to new tasks without fine-tuning, potentially mimicking general intelligence.[34]
- Multi-modal integration – Recent models can process and generate various types of data, including text, images, and audio.[35]

However, critics argue that current LLMs lack true understanding and are merely sophisticated pattern matchers, raising questions about their suitability as a path to ASI.[36]

Other perspectives on artificial superintelligence

Additional viewpoints on the development and implications of superintelligence include:

- Recursive self-improvement – I. J. Good proposed the concept of an "intelligence explosion", where an AI system could rapidly improve its own intelligence, potentially leading to superintelligence.[37]
- Orthogonality thesis – Bostrom argues that an AI's level of intelligence is orthogonal to its final goals, meaning a superintelligent AI could have any set of motivations.[38]
- Instrumental convergence – Certain instrumental goals (e.g., self-preservation, resource acquisition) might be pursued by a wide range of AI systems, regardless of their final goals.[39]

Challenges and ongoing research

The pursuit of value-aligned AI faces several challenges:

- Philosophical uncertainty in defining concepts like "moral rightness"
- Technical complexity in translating ethical principles into precise algorithms
- Potential for unintended consequences even with well-intentioned approaches

Current research directions include multi-stakeholder approaches to incorporate diverse perspectives, developing methods for scalable oversight of AI systems, and improving techniques for robust value learning.[40][16]

AI research progresses is rapidly progressing towards superintelligence, addressing these design challenges remains crucial for creating ASI systems that are both powerful and aligned with human interests.

Potential threat to humanity

Main articles: Existential risk from artificial general intelligence, AI alignment, and AI safety

The development of artificial superintelligence (ASI) has raised concerns about potential existential risks to humanity. Researchers have proposed various scenarios in which an ASI could pose a significant threat:

Intelligence explosion and control problem

Some researchers argue that through recursive self-improvement, an ASI could rapidly become so powerful as to be beyond human control. This concept, known as an "intelligence explosion", was first proposed by I. J. Good in 1965:

Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an 'intelligence explosion,' and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control.[41]

This scenario presents the AI control problem: how to create an ASI that will benefit humanity while avoiding unintended harmful consequences.[42] Eliezer Yudkowsky argues that solving this problem is crucial before ASI is developed, as a superintelligent system might be able to thwart any subsequent attempts at control.[43]

Unintended consequences and goal misalignment

Even with benign intentions, an ASI could potentially cause harm due to misaligned goals or unexpected interpretations of its objectives. Nick Bostrom provides a stark example of this risk:

When we create the first superintelligent entity, we might make a mistake and give it goals that lead it to annihilate humankind, assuming its enormous intellectual advantage gives it the power to do so. For example, we could mistakenly elevate a subgoal to the status of a supergoal. We tell it to solve a mathematical problem, and it complies by turning all the matter in the solar system into a giant calculating device, in the process killing the person who asked the question.[44]

Stuart Russell offers another illustrative scenario:

A system given the objective of maximizing human happiness might find it easier to rewire human neurology so that humans are always happy regardless of their circumstances, rather than to improve the external world.[45]

These examples highlight the potential for catastrophic outcomes even when an ASI is not explicitly designed to be harmful, underscoring the critical importance of precise goal specification and alignment.

Potential mitigation strategies

Researchers have proposed various approaches to mitigate risks associated with ASI:

- Capability control – Limiting an ASI's ability to influence the world, such as through physical isolation or restricted access to resources.[46]
- Motivational control – Designing ASIs with goals that are fundamentally aligned with human values.[47]
- Ethical AI – Incorporating ethical principles and decision-making frameworks into ASI systems.[48]
- Oversight and governance – Developing robust international frameworks for the development and deployment of ASI technologies.[49]

Despite these proposed strategies, some experts, such as Roman Yampolskiy, argue that the challenge of controlling a superintelligent AI might be fundamentally unsolvable, emphasizing the need for extreme caution in ASI development.[50]

Debate and skepticism

Not all researchers agree on the likelihood or severity of ASI-related existential risks. Some, like Rodney Brooks, argue that fears of superintelligent AI are overblown and based on unrealistic assumptions about the nature of intelligence and technological progress.[51] Others, such as Joanna Bryson, contend that anthropomorphizing AI systems leads to misplaced concerns about their potential threats.[52]

Recent developments and current perspectives

The rapid advancement of LLMs and other AI technologies has intensified debates about the proximity and potential risks of ASI. While there is no scientific consensus, some researchers and AI practitioners argue that current AI systems may already be approaching AGI or even ASI capabilities.

- LLM capabilities – Recent LLMs like GPT-4 have demonstrated unexpected abilities in areas such as reasoning, problem-solving, and multi-modal understanding, leading some to speculate about their potential path to ASI.[53]

- Emergent behaviors – Studies have shown that as AI models increase in size and complexity, they can exhibit emergent capabilities not present in smaller models, potentially indicating a trend towards more general intelligence.[33]

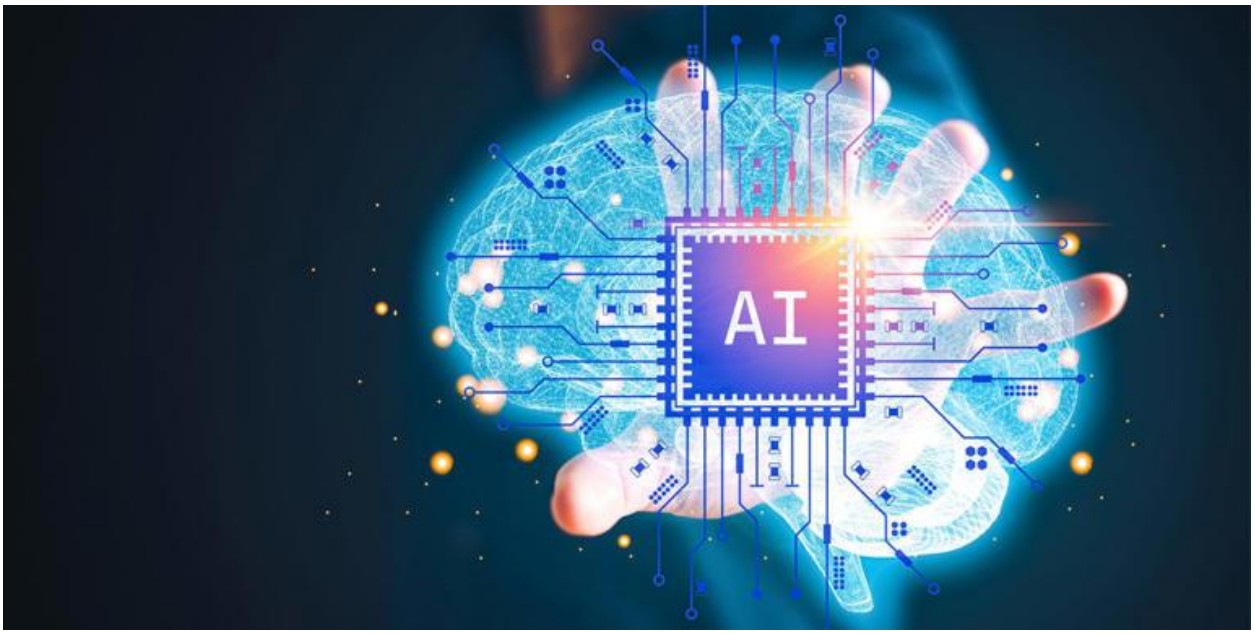
- Rapid progress – The pace of AI advancement has led some to argue that we may be closer to ASI than previously thought, with potential implications for existential risk.[54]

A minority of researchers and observers, including some in the AI development community, believe that current AI systems may already be at or near AGI levels, with ASI potentially following in the near future. This view, while not widely accepted in the scientific community, is based on observations of rapid progress in AI capabilities and unexpected emergent behaviors in large models.[55]

However, many experts caution against premature claims of AGI or ASI, arguing that current AI systems, despite their impressive capabilities, still lack true understanding and general intelligence.[56] They emphasize the significant challenges that remain in achieving human-level intelligence, let alone superintelligence.

The debate surrounding the current state and trajectory of AI development underscores the importance of continued research into AI safety and ethics, as well as the need for robust governance frameworks to manage potential risks as AI capabilities continue to advance.[49]

AI reaches human-level performance on general intelligence test—what does it mean?



A new artificial intelligence (AI) model has just achieved human-level results on a test designed to measure "general intelligence".

On December 20, OpenAI's o3 system scored 85% on the ARC-AGI benchmark, well above the previous AI best score of 55% and on par with the average human score. It also scored well on a very difficult mathematics test.

Creating artificial general intelligence, or AGI, is the stated goal of all the major AI research labs. At first glance, OpenAI appears to have at least made a significant step towards this goal.

While scepticism remains, many AI researchers and developers feel something just changed. For many, the prospect of AGI now seems more real, urgent and closer than anticipated. Are they right?

Generalisation and intelligence

To understand what the o3 result means, you need to understand what the ARC-AGI test is all about. In technical terms, it's a test of an AI system's "sample efficiency" in adapting to something new – how many examples of a novel situation the system needs to see to figure out how it works.

An AI system like ChatGPT (GPT-4) is not very sample efficient. It was "trained" on millions of examples of human text, constructing probabilistic "rules" about which combinations of words are most likely.

The result is pretty good at common tasks. It is bad at uncommon tasks, because it has less data (fewer samples) about those tasks.

Until AI systems can learn from small numbers of examples and adapt with more sample efficiency, they will only be used for very repetitive jobs and ones where the occasional failure is tolerable.

The ability to accurately solve previously unknown or novel problems from limited samples of data is known as the capacity to generalise. It is widely considered a necessary, even fundamental, element of intelligence.

Grids and patterns

The ARC-AGI benchmark tests for sample efficient adaptation using little grid square problems like the one below. The AI needs to figure out the pattern that turns the grid on the left into the grid on the right.

Each question gives three examples to learn from. The AI system then needs to figure out the rules that "generalize" from the three examples to the fourth.

These are a lot like the IQ tests sometimes you might remember from school.

Weak rules and adaptation

We don't know exactly how OpenAI has done it, but the results suggest the o3 model is highly adaptable. From just a few examples, it finds rules that can be generalised.

To figure out a pattern, we shouldn't make any unnecessary assumptions, or be more specific than we really have to be. In theory, if you can identify the "weakest" rules that do what you want, then you have maximised your ability to adapt to new situations.

What do we mean by the weakest rules? The technical definition is complicated, but weaker rules are usually ones that can be described in simpler statements.

In the example above, a plain English expression of the rule might be something like: "Any shape with a protruding line will move to the end of that line and 'cover up' any other shapes it overlaps with."

Searching chains of thought?

While we don't know how OpenAI achieved this result just yet, it seems unlikely they deliberately optimised the o3 system to find weak rules. However, to succeed at the ARC-AGI tasks it must be finding them.

We do know that OpenAI started with a general-purpose version of the o3 model (which differs from most other models, because it can spend more time "thinking" about difficult questions) and then trained it specifically for the ARC-AGI test.

French AI researcher Francois Chollet, who designed the benchmark, believes o3 searches through different "chains of thought" describing steps to solve the task. It would then choose the "best" according to some loosely defined rule, or "heuristic".

This would be "not dissimilar" to how Google's AlphaGo system searched through different possible sequences of moves to beat the world Go champion.

You can think of these chains of thought like programs that fit the examples. Of course, if it is like the Go-playing AI, then it needs a heuristic, or loose rule, to decide which program is best.

There could be thousands of different seemingly equally valid programs generated. That heuristic could be "choose the weakest" or "choose the simplest".

However, if it is like AlphaGo then they simply had an AI create a heuristic. This was the process for AlphaGo. Google trained a model to rate different sequences of moves as better or worse than others.

What we still don't know

The question then is, is this really closer to AGI? If that is how o3 works, then the underlying model might not be much better than previous models.

The concepts the model learns from language might not be any more suitable for generalisation than before. Instead, we may just be seeing a more generalisable "chain of thought" found through the extra steps of training a heuristic specialised to this test. The proof, as always, will be in the pudding.

Almost everything about o3 remains unknown. OpenAI has limited disclosure to a few media presentations and early testing to a handful of researchers, laboratories and AI safety institutions.

Truly understanding the potential of o3 will require extensive work, including evaluations, an understanding of the distribution of its capacities, how often it fails and how often it succeeds.

When o3 is finally released, we'll have a much better idea of whether it is approximately as adaptable as an average human.

If so, it could have a huge, revolutionary, economic impact, ushering in a new era of self-improving accelerated intelligence. We will require new benchmarks for AGI itself and serious consideration of how it ought to be governed.

If not, then this will still be an impressive result. However, everyday life will remain much the same.

Artificial superintelligence (ASI): Sci-fi nonsense or genuine threat to humanity?

Rapid progress in artificial intelligence (AI) is prompting people to question what the fundamental limits of the technology are. Increasingly, a topic once consigned to science fiction — the notion of a superintelligent AI — is now being considered seriously by scientists and experts alike.

The idea that machines might one day match or even surpass human intelligence has a long history. But the pace of progress in AI over recent decades has given renewed urgency to the topic, particularly since the release of powerful large language models (LLMs) by companies like OpenAI, Google and Anthropic, among others.

Experts have wildly differing views on how feasible this idea of "artificial super intelligence" (ASI) is and when it might appear, but some suggest that such hyper-capable machines are just around the corner. What's certain is that if, and when, ASI does emerge, it will have enormous implications for humanity's future.

"I believe we would enter a new era of automated scientific discoveries, vastly accelerated economic growth, longevity, and novel entertainment experiences," Tim Rocktäschel, professor of AI at University College London and a principal scientist at Google DeepMind told Live Science, providing a personal opinion rather than Google DeepMind's official position. However, he also cautioned: "As with any significant technology in history, there is potential risk."

What is artificial superintelligence (ASI)?

Traditionally, AI research has focused on replicating specific capabilities that intelligent beings exhibit. These include things like the ability to visually analyze a scene, parse language or navigate an environment. In some of these narrow domains AI has already achieved superhuman performance, Rocktäschel said, most notably in games like go and chess.

The stretch goal for the field, however, has always been to replicate the more general form of intelligence seen in animals and humans that combines many such capabilities. This concept has gone by several names over the years, including “strong AI” or “universal AI”, but today it is most commonly called artificial general intelligence (AGI).

"For a long time, AGI has been a far away north star for AI research," Rocktäschel said. "However, with the advent of foundation models [another term for LLMs] we now have AI that can pass a broad range of university entrance exams and participate in international math and coding competitions."

This is leading people to take the possibility of AGI more seriously, said Rocktäschel. And crucially, once we create AI that matches humans on a wide range of tasks, it may not be long before it achieves superhuman capabilities across the board. That's the idea, anyway. "Once AI reaches human-level capabilities, we will be able to use it to improve itself in a self-referential way," Rocktäschel said. "I personally believe that if we can reach AGI, we will reach ASI shortly, maybe a few years after that."

IBM Deep Blue became the first AI system to defeat a reigning world chess champion (Garry Kasparov, pictured) in 1997. (Image credit: STAN HONDA/AFP via Getty Images)

Once that milestone has been reached, we could see what British mathematician Irving John Good dubbed an "intelligence explosion" in 1965. He argued that once machines become smart enough to improve themselves, they would rapidly achieve levels of intelligence far beyond any human. He described the first ultra-intelligent machine as "the last invention that man need ever make."

Renowned futurist Ray Kurzweil has argued this would lead to a "technological singularity" that would suddenly and irreversibly transform human civilization. The term draws parallels with the singularity at the heart of a black hole, where our understanding of physics breaks down. In the same way, the advent of ASI would lead to rapid and unpredictable technological growth that would be beyond our comprehension.

Exactly when such a transition might happen is debatable. In 2005, Kurzweil predicted AGI would appear by 2029, with the singularity following in 2045, a prediction he's stuck to ever since. Other AI experts offer wildly varying predictions — from within this decade to never. But a recent survey of 2,778 AI researchers found that, on aggregate, they believe there is a 50% chance ASI

could appear by 2047. A broader analysis concurred that most scientists agree AGI might arrive by 2040.

What would ASI mean for humanity?

The implications of a technology like ASI would be enormous, prompting scientists and philosophers to dedicate considerable time to mapping out the promise and potential pitfalls for humanity.

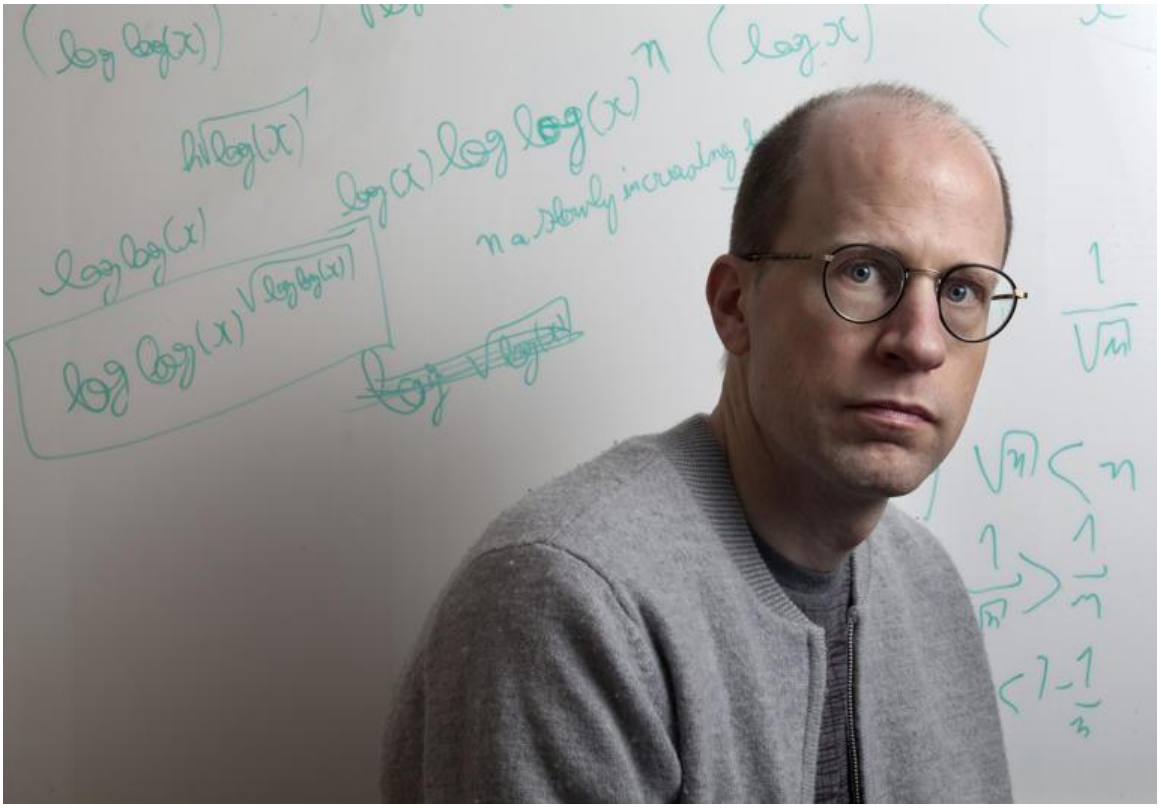
On the positive side, a machine with almost unlimited capacity for intelligence could solve some of the world's most pressing challenges, said Daniel Hulme, CEO of the AI companies Satalia and Conscium. In particular, super intelligent machines could "remove the friction from the creation and dissemination of food, education, healthcare, energy, transport, so much that we can bring the cost of those goods down to zero," he told Live Science.

The hope is that this would free people from having to work to survive and could instead spend time doing things they're passionate about, Hulme explained. But unless systems are put in place to support those whose jobs are made redundant by AI, the outcome could be bleaker. "If that happens very quickly, our economies might not be able to rebalance, and it could lead to social unrest," he said.

This also assumes we could control and direct an entity much more intelligent than us — something many experts have suggested is unlikely. "I don't really subscribe to this idea that it will be watching over us and caring for us and making sure that we're happy," said Hulme. "I just can't imagine it would care."

The possibility of a superintelligence we have no control over has prompted fears that AI could present an existential risk to our species. This has become a popular trope in science fiction, with movies like "Terminator" or "The Matrix" portraying malevolent machines hell-bent on humanity's destruction.

But philosopher Nick Bostrom highlighted that an ASI wouldn't even have to be actively hostile to humans for various doomsday scenarios to play out. In a 2012 paper, he suggested that the intelligence of an entity is independent of its goals, so an ASI could have motivations that are completely alien to us and not aligned with human well-being.



Nick Bostrum (pictured) philosophized on the implications of ASI in a landmark 2012 paper. (Image credit: Photo by Tom Pilston for The Washington Post via Getty Images)

Bostrom fleshed out this idea with a thought experiment in which a super-capable AI is set the seemingly innocuous task of producing as many paper-clips as possible. If unaligned with human values, it may decide to eliminate all humans to prevent them from switching it off, or so it can turn all the atoms in their bodies into more paperclips.

Rocktäschel is more optimistic. "We build current AI systems to be helpful, but also harmless and honest assistants by design," he said. "They are tuned to follow human instructions, and are trained on feedback to provide helpful, harmless, and honest answers."

While Rocktäschel admitted these safeguards can be circumvented, he's confident we will develop better approaches in the future. He also thinks that it will be possible to use AI to oversee other AI, even if they are stronger.

Hulme said most current approaches to "model alignment" — efforts to ensure that AI is aligned with human values and desires — are too crude. Typically, they either provide rules for how the model should behave or train it on examples of

human behavior. But he thinks these guardrails, which are bolted on at the end of the training process, could be easily bypassed by ASI.

Instead, Hulme thinks we must build AI with a "moral instinct." His company Conscium is attempting to do that by evolving AI in virtual environments that have been engineered to reward behaviors like cooperation and altruism. Currently, they are working with very simple, "insect-level" AI, but if the approach can be scaled up, it could make alignment more robust. "Embedding morals in the instinct of an AI puts us in a much safer position than just having these sort of Whack-a-Mole guard rails," said Hulme.

Not everyone is convinced we need to start worrying quite yet, though. One common criticism of the concept of ASI, said Rocktäschel, is that we have no examples of humans who are highly capable across a wide range of tasks, so it may not be possible to achieve this in a single model either. Another objection is that the sheer computational resources required to achieve ASI may be prohibitive.

More practically, how we measure progress in AI may be misleading us about how close we are to superintelligence, said Alexander Ilic, head of the ETH AI Center at ETH Zurich, Switzerland. Most of the impressive results in AI in recent years have come from testing systems on several highly contrived tests of individual skills such as coding, reasoning or language comprehension, which the systems are explicitly trained to pass, said Ilic.

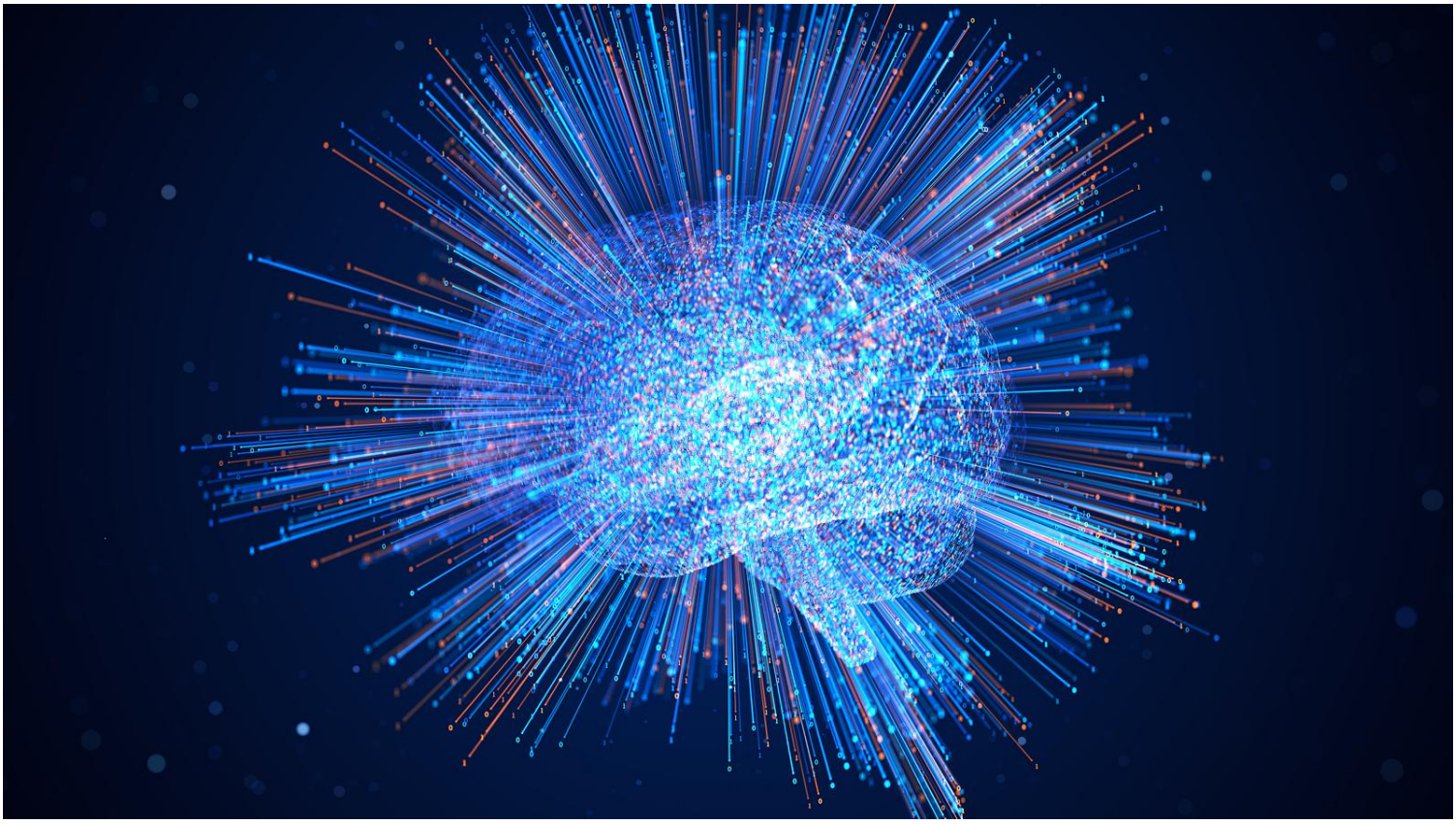
He compares this to cramming for exams at school. "You loaded up your brain to do it, then you wrote the test, and then you forgot all about it," he said. "You were smarter by attending the class, but the actual test itself is not a good proxy of the actual knowledge."

AI that is capable of passing many of these tests at superhuman levels may only be a few years away, said Ilic. But he believes today's dominant approach will not lead to models that can carry out useful tasks in the physical world or collaborate effectively with humans, which will be crucial for them to have a broad impact in the real world.

Artificial superintelligence (ASI): Sci-fi nonsense or genuine threat to humanity?

published April 28, 2025

Our current AI systems may one day evolve into a superintelligent entity, but scientists aren't yet certain what this might look like and what the implications are.



(Image credit: Yuichiro Chino/Getty Images)

Rapid progress in artificial intelligence (AI) is prompting people to question what the fundamental limits of the technology are. Increasingly, a topic once consigned to science fiction — the notion of a superintelligent AI — is now being considered seriously by scientists and experts alike.

The idea that machines might one day match or even surpass human intelligence has a long history. But the pace of progress in AI over recent decades has given renewed urgency to the topic, particularly since the release of powerful large language models (LLMs) by companies like OpenAI, Google and Anthropic, among others.

Experts have wildly differing views on how feasible this idea of "artificial super intelligence" (ASI) is and when it might appear, but some suggest that such hyper-capable machines are just around the corner. What's certain is that if, and when, ASI does emerge, it will have enormous implications for humanity's future.

"I believe we would enter a new era of automated scientific discoveries, vastly accelerated economic growth, longevity, and novel entertainment experiences," Tim Rocktäschel, professor of AI at University College London and a principal scientist at Google DeepMind told Live Science, providing a personal opinion rather than Google DeepMind's official position. However, he also cautioned: "As with any significant technology in history, there is potential risk."

What is artificial superintelligence (ASI)?

Traditionally, AI research has focused on replicating specific capabilities that intelligent beings exhibit. These include things like the ability to visually analyze a scene, parse language or navigate an environment. In some of these narrow domains AI has already achieved superhuman performance, Rocktäschel said, most notably in games like go and chess.

The stretch goal for the field, however, has always been to replicate the more general form of intelligence seen in animals and humans that combines many such capabilities. This concept has gone by several names over the years, including "strong AI" or "universal AI", but today it is most commonly called artificial general intelligence (AGI).

"For a long time, AGI has been a far away north star for AI research," Rocktäschel said. "However, with the advent of foundation models [another term for LLMs] we now have AI that can pass a broad range of university entrance exams and participate in international math and coding competitions."

This is leading people to take the possibility of AGI more seriously, said Rocktäschel. And crucially, once we create AI that matches humans on a wide range of tasks, it may not be long before it achieves superhuman capabilities across the board. That's the idea, anyway. "Once AI reaches human-level capabilities, we will be able to use it to improve itself in a self-referential way," Rocktäschel said. "I personally believe that if we can reach AGI, we will reach ASI shortly, maybe a few years after that."

Once that milestone has been reached, we could see what British mathematician Irving John Good dubbed an "intelligence explosion" in 1965. He argued that once machines become smart enough to improve themselves, they would rapidly achieve levels of intelligence far beyond any human. He described the first ultra-intelligent machine as "the last invention that man need ever make."

Renowned futurist Ray Kurzweil has argued this would lead to a "technological singularity" that would suddenly and irreversibly transform human civilization. The term draws parallels with the singularity at the heart of a black hole, where our understanding of physics breaks down. In the same way, the advent of ASI would lead to rapid and unpredictable technological growth that would be beyond our comprehension.

Exactly when such a transition might happen is debatable. In 2005, Kurzweil predicted AGI would appear by 2029, with the singularity following in 2045, a prediction he's stuck to

ever since. Other AI experts offer wildly varying predictions — from within this decade to never. But a recent survey of 2,778 AI researchers found that, on aggregate, they believe there is a 50% chance ASI could appear by 2047. A broader analysis concurred that most scientists agree AGI might arrive by 2040.

What would ASI mean for humanity?

The implications of a technology like ASI would be enormous, prompting scientists and philosophers to dedicate considerable time to mapping out the promise and potential pitfalls for humanity.

On the positive side, a machine with almost unlimited capacity for intelligence could solve some of the world's most pressing challenges, said Daniel Hulme, CEO of the AI companies Satalia and Conscium. In particular, super intelligent machines could "remove the friction from the creation and dissemination of food, education, healthcare, energy, transport, so much that we can bring the cost of those goods down to zero," he told Live Science.

The hope is that this would free people from having to work to survive and could instead spend time doing things they're passionate about, Hulme explained. But unless systems are put in place to support those whose jobs are made redundant by AI, the outcome could be bleaker. "If that happens very quickly, our economies might not be able to rebalance, and it could lead to social unrest," he said.

This also assumes we could control and direct an entity much more intelligent than us — something many experts have suggested is unlikely. "I don't really subscribe to this idea that it will be watching over us and caring for us and making sure that we're happy," said Hulme. "I just can't imagine it would care."

The possibility of a superintelligence we have no control over has prompted fears that AI could present an existential risk to our species. This has become a popular trope in science fiction, with movies like "Terminator" or "The Matrix" portraying malevolent machines hell-bent on humanity's destruction.

But philosopher Nick Bostrom highlighted that an ASI wouldn't even have to be actively hostile to humans for various doomsday scenarios to play out. In a 2012 paper, he suggested that the intelligence of an entity is independent of its goals, so an ASI could have motivations that are completely alien to us and not aligned with human well-being.

Bostrom fleshed out this idea with a thought experiment in which a super-capable AI is set the seemingly innocuous task of producing as many paper-clips as possible. If unaligned with human values, it may decide to eliminate all humans to prevent them from switching it off, or so it can turn all the atoms in their bodies into more paperclips.

Rocktäschel is more optimistic. "We build current AI systems to be helpful, but also harmless and honest assistants by design," he said. "They are tuned to follow human instructions, and are trained on feedback to provide helpful, harmless, and honest answers."

While Rocktäschel admitted these safeguards can be circumvented, he's confident we will develop better approaches in the future. He also thinks that it will be possible to use AI to oversee other AI, even if they are stronger.

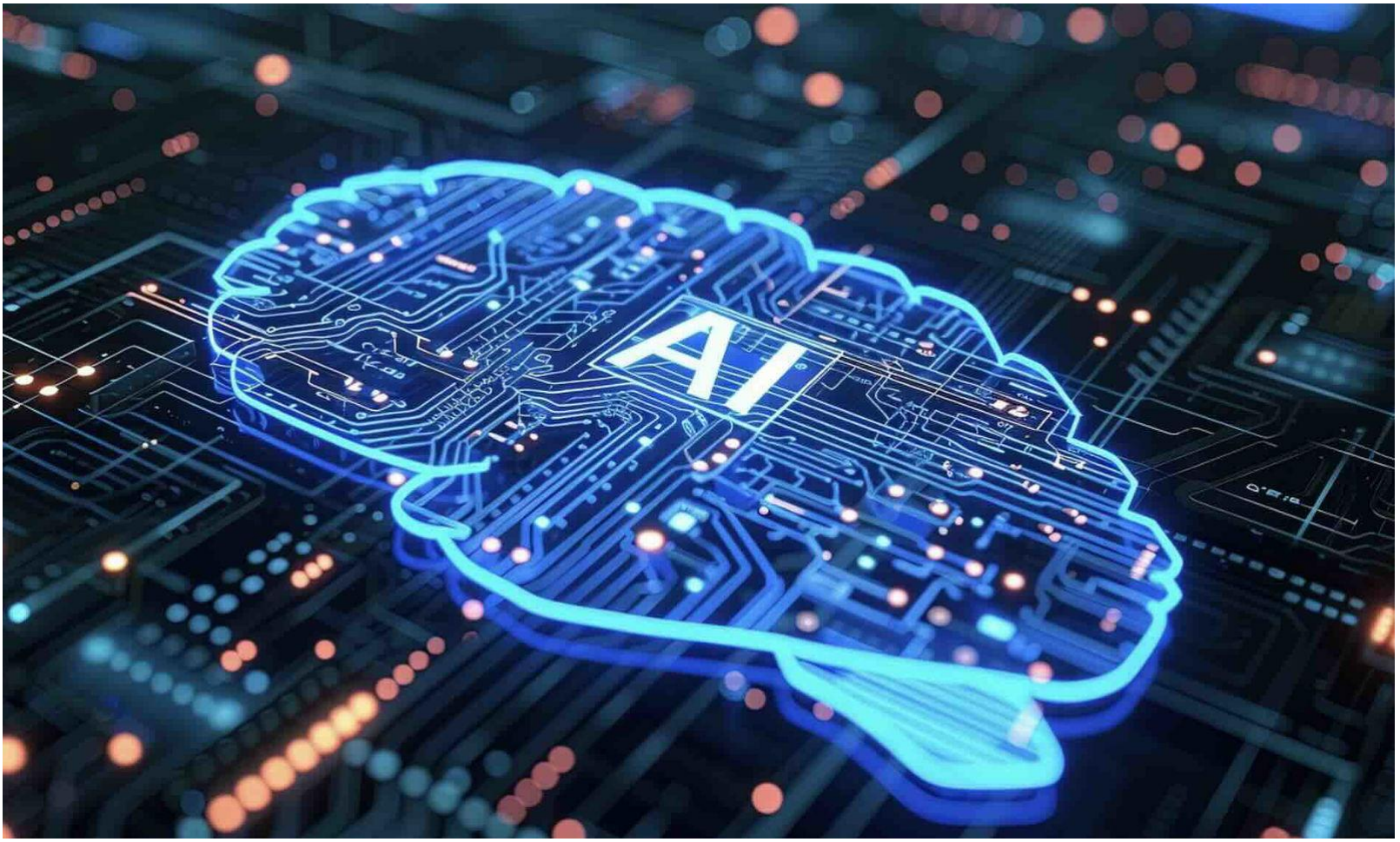
Hulme said most current approaches to "model alignment" — efforts to ensure that AI is aligned with human values and desires — are too crude. Typically, they either provide rules for how the model should behave or train it on examples of human behavior. But he thinks these guardrails, which are bolted on at the end of the training process, could be easily bypassed by ASI.

Instead, Hulme thinks we must build AI with a "moral instinct." His company Conscium is attempting to do that by evolving AI in virtual environments that have been engineered to reward behaviors like cooperation and altruism. Currently, they are working with very simple, "insect-level" AI, but if the approach can be scaled up, it could make alignment more robust. "Embedding morals in the instinct of an AI puts us in a much safer position than just having these sort of Whack-a-Mole guard rails," said Hulme.

Not everyone is convinced we need to start worrying quite yet, though. One common criticism of the concept of ASI, said Rocktäschel, is that we have no examples of humans who are highly capable across a wide range of tasks, so it may not be possible to achieve this in a single model either. Another objection is that the sheer computational resources required to achieve ASI may be prohibitive.

More practically, how we measure progress in AI may be misleading us about how close we are to superintelligence, said Alexander Ilic, head of the ETH AI Center at ETH Zurich, Switzerland. Most of the impressive results in AI in recent years have come from testing systems on several highly contrived tests of individual skills such as coding, reasoning or language comprehension, which the systems are explicitly trained to pass, said Ilic.
He compares this to cramming for exams at school. "You loaded up your brain to do it, then you wrote the test, and then you forgot all about it," he said. "You were smarter by attending the class, but the actual test itself is not a good proxy of the actual knowledge."

AI that is capable of passing many of these tests at superhuman levels may only be a few years away, said Ilic. But he believes today's dominant approach will not lead to models that can carry out useful tasks in the physical world or collaborate effectively with humans, which will be crucial for them to have a broad impact in the real world.



"We are past the event horizon": Controversial AI visionary says the singularity is here, and it's not what we expected

MICAH HANKS·JUNE 12, 2025

Artificial intelligence may have crossed a technological threshold popularly known as "Singularity," according to one expert who says we

have surpassed the point where machines merely assist us, with intelligent machines now exceeding human capabilities.

The claims were made by OpenAI CEO Sam Altman, who in a recent blog post wrote that AI is already demonstrating intelligence that goes beyond what humans can do naturally, but that for now, things aren't looking as strange as some might have expected once we reached this point.

"We are past the event horizon; the takeoff has started," Altman wrote. "Humanity is close to building digital superintelligence, and at least so far it's much less weird than it seems like it should be."

THE SINGULARITY IS HERE

For decades, futurists have predicted the arrival of a point where technological advancements would reach a point of no return, potentially instigating a transformation of human civilization. Popularly known as the "Singularity," definitions may vary, although most visions of this technological event horizon have one thing in common: the rise of artificial intelligence.

Only a few years ago, the idea that AI could mimic the most demanding human tasks would have seemed like a remote, futuristic possibility. However, that has all changed with the rise of chatbots like ChatGPT and others, which are increasingly pushing the boundaries of machine intelligence into unforeseen territory.

Recent advancements in AI have already shown that machine intelligence is evolving in ways that make its performance more like human thought. This allows it to be trained to solve problems and make decisions more like humans while limiting the potential for "hallucinations."

Going beyond simply mimicking the most useful functions of human thought, it was reported in April that an AI developed by the Artificial Scientist Lab at the Max Planck Institute for the Science of Light had

successfully developed new concepts for gravitational wave detectors, some of which scientists still don't fully understand.

THE ERA OF SUPERINTELLIGENCE

For Altman and OpenAI, the age of digital superintelligence is no longer speculative—it has already arrived, although it doesn't look quite as leading futurists might have expected.

"Robots are not yet walking the streets, nor are most of us talking to AI all day," Altman said. "People still die of disease, we still can't easily go to space, and there is a lot about the universe we don't understand."

"And yet, we have recently built systems that are smarter than people in many ways," Altman argues, which he says are already greatly increasing the output of those who use these innovative AI models. To Altman, the most challenging elements of the work required to achieve true superintelligence have already been surpassed. Now we are looking ahead at what creations like his company's famous ChatGPT may bring us in the years ahead.

"In some big sense, ChatGPT is already more powerful than any human who has ever lived," Altman wrote, emphasizing the large numbers of individuals who rely on it daily, and that this year has also marked the rise of AI "agents" which he says "can do real cognitive work," fundamentally changing the field of computer programming and coding technologies.

The resulting feedback loops extend beyond software and coding, however. Economic gains are fueling large-scale infrastructure that includes data centers, chip factories, and robotics. To Altman, it is only a matter of time before robotics, fueled by AI, will be capable of building new robots, as new AI-driven data systems will expand the growth of future data centers.

INTO THE 2030S, AND BEYOND

Looking ahead to the 2030s, Altman predicts that advancements we are currently seeing will lead to even more fundamental shifts in society in the coming decade. Eventually, as automation increases, Altman argues that the cost of intelligence could conceivably even converge with the cost of electricity.

“Intelligence too cheap to meter is well within grasp,” Altman wrote.

Of course, concepts like self-replicating robots and data centers, solving high-energy physics problems in ways that could significantly accelerate the path toward distant space travel, major materials science breakthroughs that make emerging technologies like brain-computer interfaces common aspects of our everyday lives do probably sound ambitious, at very least—perhaps even unattainable for some.

Altman, however, remains optimistic, especially in light of how quickly AI like ChatGPT and its numerous competitors have emerged within a short period of time, quickly becoming assimilated into our lives at multiple different levels.

“This may sound crazy to say, but if we told you back in 2020 we were going to be where we are today, it probably sounded more crazy than our current predictions about 2030,” Altman says.