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]

The creation of artificial superintelligence (ASI) has been a topic of increasing discussion in recent years, particularly with the rapid advancements in artificial intelligence (AI) technologies.[6][7]

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).[8]

However, the claim that current LLMs constitute AGI is controversial. Critics argue that these models, while impressive, still lack true understanding and are prim  
  
  
  
Deep learning is a subset of machine learning that focuses on utilizing neural networks to perform tasks such as classification, regression, and representation learning. The field takes inspiration from biological neuroscience and is centered around stacking artificial neurons into layers and "training" them to process data. The adjective "deep" refers to the use of multiple layers (ranging from three to several hundred or thousands) in the network. Methods used can be either supervised, semi-supervised or unsupervised.[2]

Some common deep learning network architectures include fully connected networks, deep belief networks, recurrent neural networks, convolutional neural networks, generative adversarial networks, transformers, and neural radiance fields. These architectures have been applied to fields including computer vision, speech recognition, natural language processing, machine translation, bioinformatics, drug design, medical image analysis, climate science, material inspection and board game programs, where they have produced results comparable to and in some cases surpassing human expert performance.[3][4][5]

Early forms of neural networks were inspired by information processing and distributed communication nodes in biological systems, particularly the human brain. However, current neural networks do not intend to model the brain function of organisms, and are generally seen as low-quality models for that purpose.[6]

Most modern deep learning models are based on multi-layered neural networks such as convolutional neural networks and transformers, although they can also include propositional formulas or latent variables organized layer-wise in deep generative models such as the nodes in deep belief networks and deep Boltzmann machines.[7]

Fundamentally, deep learning refers to a class of machine learning algorithms in which a hierarchy of layers is used to transform input data into a progressively more abstract and composite representation. For example, in an image recognition model, the raw input may be an image (represented as a tensor of pixels). The first representational layer may attempt to identify basic shapes such as lines and circles, the second layer may compose and encode arrangements of edges, the third layer may encode a nose and eyes, and the fourth layer may recognize that the image contains a face.