# Artificial intelligence

**Artificial intelligence** (**AI**) is the intelligence of machines or software, as opposed to the intelligence of living beings, primarily of humans. It is a field of study in computer science that develops and studies intelligent machines. Such machines may be called AIs.

AI technology is widely used throughout industry, government, and science. Some high-profile applications are: advanced web search engines (e.g., Google Search), recommendation systems (used by YouTube, Amazon, and Netflix), interacting via human speech (e.g., Google Assistant, Siri, and Alexa), self-driving cars (e.g., Waymo), generative and creative tools (e.g., ChatGPT and AI art), and superhuman play and analysis in strategy games (e.g., chess and Go).[1]

Alan Turing was the first person to conduct substantial research in the field that he called machine intelligence.[2] Artificial intelligence was founded as an academic discipline in 1956.[3] The field went through multiple cycles of optimism,[4][5] followed by periods of disappointment and loss of funding, known as AI winter.[6][7] Funding and interest vastly increased after 2012 when deep learning surpassed all previous AI techniques,[8] and after 2017 with the transformer architecture.[9] This led to the AI spring of the early 2020s, with companies, universities, and laboratories overwhelmingly based in the United States pioneering significant advances in artificial intelligence.[10]

The growing use of artificial intelligence in the 21st century is influencing a societal and economic shift towards increased automation, data-driven decision-making, and the integration of AI systems into various economic sectors and areas of life, impacting job markets, healthcare, government, industry, and education. This raises questions about the ethical implications and risks of AI, prompting discussions about regulatory policies to ensure the safety and benefits of the technology.

The various sub-fields of AI research are centered around particular goals and the use of particular tools. The traditional goals of AI research include reasoning, knowledge representation, planning, learning, natural language processing, perception, and support for robotics.[a] General intelligence (the ability to complete any task performable by a human) is among the field's long-term goals.[11]

To solve these problems, AI researchers have adapted and integrated a wide range of problem-solving techniques, including search and mathematical optimization, formal logic, artificial neural networks, and methods based on statistics, operations research, and economics.[b] AI also draws upon psychology, linguistics, philosophy, neuroscience and other fields.[12]

## Goals

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

### Reasoning, problem-solving

Early researchers developed algorithms that imitated step-by-step reasoning that humans use when they solve puzzles or make logical deductions.[13] By the late 1980s and 1990s, methods were developed for dealing with uncertain or incomplete information, employing concepts from probability and economics.[14]

Many of these algorithms are insufficient for solving large reasoning problems because they experience a "combinatorial explosion": they became exponentially slower as the problems grew larger.[15] Even humans rarely use the step-by-step deduction that early AI research could model. They solve most of their problems using fast, intuitive judgments.[16] Accurate and efficient reasoning is an unsolved problem.

## Knowledge representation

Knowledge representation and knowledge engineering[17] allow AI programs to answer questions intelligently and make deductions about real-world facts. Formal knowledge representations are used in content-based indexing and retrieval,[18] scene interpretation,[19] clinical decision support,[20] knowledge discovery (mining "interesting" and actionable inferences from large databases),[21] and other areas.[22]

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

A knowledge base is a body of knowledge represented in a form that can be used by a program. An ontology is the set of objects, relations, concepts, and properties used by a particular domain of knowledge.[23] Knowledge bases need to represent things such as: objects, properties, categories and relations between objects;[24] situations, events, states and time;[25] causes and effects;[26] knowledge about knowledge (what we know about what other people know);[27] default reasoning (things that humans assume are true until they are told differently and will remain true even when other facts are changing);[28] and many other aspects and domains of knowledge.

Among the most difficult problems in knowledge representation are: the breadth of commonsense knowledge (the set of atomic facts that the average person knows is enormous);[29] and the sub-symbolic form of most commonsense knowledge (much of what people know is not represented as "facts" or "statements" that they could express verbally).[16] There is also the difficulty of knowledge acquisition, the problem of obtaining knowledge for AI applications.[c]

## Planning and decision making

An "agent" is anything that perceives and takes actions in the world. A rational agent has goals or preferences and takes actions to make them happen.[d][32] In automated planning, the agent has a specific goal.[33] In automated decision making, the agent has preferences – there are some situations it would prefer to be in, and some situations it is trying to avoid. The decision making agent assigns a number to each situation (called the "utility") that measures how much the agent prefers it. For each possible action, it can calculate the "expected utility": the utility of all possible outcomes of the action, weighted by the probability that the outcome will occur. It can then choose the action with the maximum expected utility.[34]

In classical planning, the agent knows exactly what the effect of any action will be.[35] In most real-world problems, however, the agent may not be certain about the situation they are in (it is "unknown" or "unobservable") and it may not know for certain what will happen after each possible action (it is not "deterministic"). It must choose an action by making a probabilistic guess and then reassess the situation to see if the action worked.[36]

In some problems, the agent's preferences may be uncertain, especially if there are other agents or humans involved. These can be learned (e.g., with inverse reinforcement learning) or the agent can seek information to improve its preferences.[37] Information value theory can be used to weigh the value of exploratory or experimental actions.[38] The space of possible future actions and situations is typically intractably large, so the agents must take actions and evaluate situations while being uncertain what the outcome will be.

A Markov decision process has a transition model that describes the probability that a particular action will change the state in a particular way, and a reward function that supplies the utility of each state and the cost of each action. A policy associates a decision with each possible state. The policy could be calculated (e.g., by iteration), be heuristic, or it can be learned.[39]

Game theory describes rational behavior of multiple interacting agents, and is used in AI programs that make decisions that involve other agents.[40]

## Learning

Machine learning is the study of programs that can improve their performance on a given task automatically.[41] It has been a part of AI from the beginning.[e]

There are several kinds of machine learning. Unsupervised learning analyzes a stream of data and finds patterns and makes predictions without any other guidance.[44] Supervised learning requires a human to label the input data first, and comes in two main varieties: classification (where the program must learn to predict what category the input belongs in) and regression (where the program must deduce a numeric function based on numeric input).[45]

In reinforcement learning the agent is rewarded for good responses and punished for bad ones. The agent learns to choose responses that are classified as "good".[46] Transfer learning is when the knowledge gained from one problem is applied to a new problem.[47] Deep learning is a type of machine learning that runs inputs through biologically inspired artificial neural networks for all of these types of learning.[48]

Computational learning theory can assess learners by computational complexity, by sample complexity (how much data is required), or by other notions of optimization.[49]

## Natural language processing

Natural language processing (NLP)[50] allows programs to read, write and communicate in human languages such as English. Specific problems include speech recognition, speech synthesis, machine translation, information extraction, information retrieval and question answering.[51]

Early work, based on Noam Chomsky's generative grammar and semantic networks, had difficulty with word-sense disambiguation[f] unless restricted to small domains called "micro-worlds" (due to the common sense knowledge problem[29]). Margaret Masterman believed that it was meaning, and not grammar that was the key to understanding languages, and that thesauri and not dictionaries should be the basis of computational language structure.

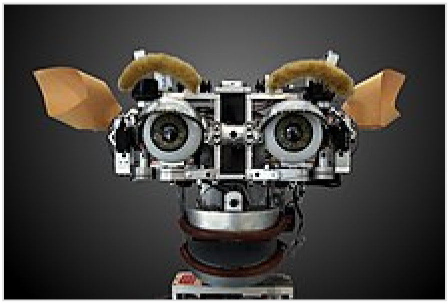
Modern deep learning techniques for NLP include word embedding (representing words, typically as vectors encoding their meaning),[52] transformers (a deep learning architecture using an attention mechanism),[53] and others.[54] In 2019, generative pre-trained transformer (or "GPT") language models began to generate coherent text,[55][56] and by 2023 these models were able to get human-level scores on the bar exam, SAT test, GRE test, and many other real-world applications.[57]

### Perception

Machine perception is the ability to use input from sensors (such as cameras, microphones, wireless signals, active lidar, sonar, radar, and tactile sensors) to deduce aspects of the world. Computer vision is the ability to analyze visual input.[58]

The field includes speech recognition,[59] image classification,[60] facial recognition, object recognition,[61] and robotic perception.[62]

### Social intelligence



Kismet, a robot head which was made in the 1990s; a machine that can recognize and simulate emotions.[63]

Affective computing is an interdisciplinary umbrella that comprises systems that recognize, interpret, process or simulate human feeling, emotion and mood.[64] For example, some virtual assistants are programmed to speak conversationally or even to banter humorously; it makes them appear more sensitive to the emotional dynamics of human interaction, or to otherwise facilitate human–computer interaction.

However, this tends to give naïve users an unrealistic conception of the intelligence of existing computer agents.[65] Moderate successes related to affective computing include textual sentiment analysis and, more recently, multimodal sentiment analysis, wherein AI classifies the affects displayed by a videotaped subject.[66]

### General intelligence

A machine with artificial general intelligence should be able to solve a wide variety of problems with breadth and versatility similar to human intelligence.[11]

## Techniques

AI research uses a wide variety of techniques to accomplish the goals above.[b]

### Search and optimization

AI can solve many problems by intelligently searching through many possible solutions.[67] There are two very different kinds of search used in AI: state space search and local search.

### State space search

State space search searches through a tree of possible states to try to find a goal state.[68] For example, planning algorithms search through trees of goals and subgoals, attempting to find a path to a target goal, a process called means-ends analysis.[69]

Simple exhaustive searches[70] are rarely sufficient for most real-world problems: the search space (the number of places to search) quickly grows to astronomical numbers. The result is a search that is too slow or never completes.[15] "Heuristics" or "rules of thumb" can help to prioritize choices that are more likely to reach a goal.[71]

Adversarial search is used for game-playing programs, such as chess or Go. It searches through a tree of possible moves and counter-moves, looking for a winning position.[72]

### Local search

Local search uses mathematical optimization to find a solution to a problem. It begins with some form of guess and refines it incrementally.[73]

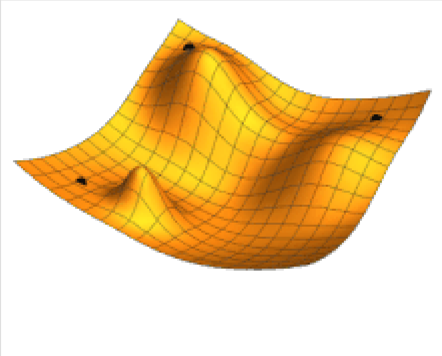


Illustration of gradient descent for 3 different starting points. Two parameters (represented by the plan coordinates) are adjusted in order to minimize the loss function (the height).

Gradient descent is a type of local search that optimizes a set of numerical parameters by incrementally adjusting them to minimize a loss function. Variants of gradient descent are commonly used to train neural networks.[74]

Another type of local search is evolutionary computation, which aims to iteratively improve a set of candidate solutions by "mutating" and "recombining" them, selecting only the fittest to survive each generation.[75]

Distributed search processes can coordinate via swarm intelligence algorithms. Two popular swarm algorithms used in search are particle swarm optimization (inspired by bird flocking) and ant colony optimization (inspired by ant trails).[76]

## Logic

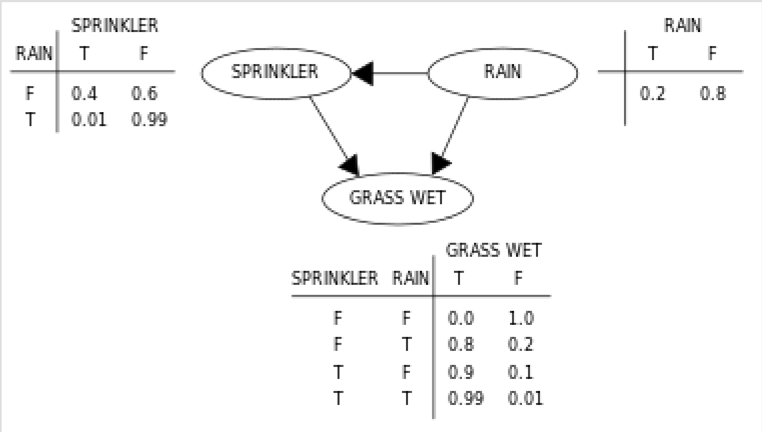
Formal Logic is used for reasoning and knowledge representation.[77] Formal logic comes in two main forms: propositional logic (which operates on statements that are true or false and uses logical connectives such as "and", "or", "not" and "implies")[78] and predicate logic (which also operates on objects, predicates and relations and uses quantifiers such as "*Every X* is a *Y*" and "There are *some X*s that are *Y*s").[79]

Logical inference (or deduction) is the process of proving a new statement (conclusion) from other statements that are already known to be true (the premises).[80] A logical knowledge base also handles queries and assertions as a special case of inference.[81] An inference rule describes what is a valid step in a proof. The most general inference rule is resolution.[82] Inference can be reduced to performing a search to find a path that leads from premises to conclusions, where each step is the application of an inference rule.[83] Inference performed this way is intractable except for short proofs in restricted domains. No efficient, powerful and general method has been discovered.

Fuzzy logic assigns a "degree of truth" between 0 and 1. It can therefore handle propositions that are vague and partially true.[84] Non-monotonic logics are designed to handle default reasoning.[28] Other specialized versions of logic have been developed to describe many complex domains (see knowledge representation above).

## Probabilistic methods for uncertain reasoning

Many problems in AI (including in reasoning, planning, learning, perception, and robotics) require the agent to operate with incomplete or uncertain information. AI researchers have devised a number of tools to solve these problems using methods from probability theory and economics.[85]




A simple Bayesian network, with the associated conditional probability tables

Bayesian networks[86] are a very general tool that can be used for many problems, including reasoning (using the Bayesian inference algorithm),[g][88] learning (using the expectation-maximization algorithm),[h][90] planning (using decision networks)[91] and perception (using dynamic Bayesian networks).[92]

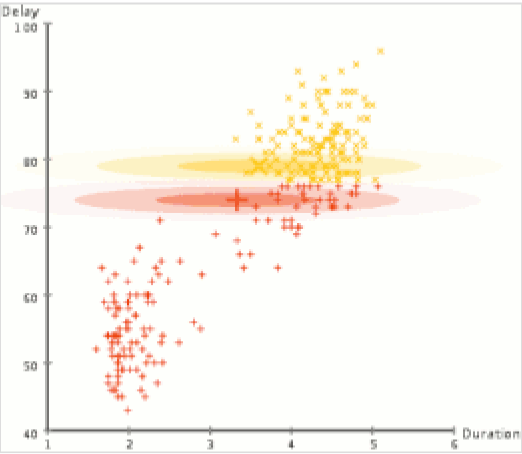
Probabilistic algorithms can also be used for filtering, prediction, smoothing and finding explanations for streams of data, helping perception systems to analyze processes that occur over time (e.g., hidden Markov models or Kalman filters).[92]

Precise mathematical tools have been developed that analyze how an agent can make choices and plan, using decision theory, decision analysis,[93] and information value theory.[94] These tools include models such as Markov decision processes,[95] dynamic decision networks,[92] game theory and mechanism design.[96]

## Classifiers and statistical learning methods

The simplest AI applications can be divided into two types: classifiers (e.g., "if shiny then diamond"), on one hand, and controllers (e.g., "if diamond then pick up"), on the other hand. Classifiers[97] are functions that use pattern matching to determine the closest match. They can be fine-tuned based on chosen examples using supervised learning. Each pattern (also called an "observation") is labeled with a certain predefined class. All the observations combined with their class labels are known as a data set. When a new observation is received, that observation is classified based on previous experience.[45]

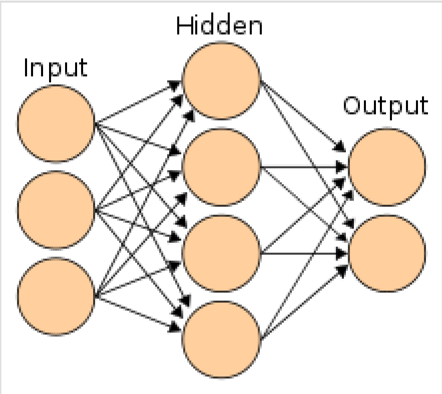
There are many kinds of classifiers in use. The decision tree is the simplest and most widely used symbolic machine learning algorithm.[98] K-nearest neighbor algorithm was the most widely used analogical AI until the mid-1990s, and Kernel methods such as the support vector machine (SVM) displaced k-nearest neighbor in the 1990s.[99] The naive Bayes classifier is reportedly the "most widely used learner"[100] at Google, due in part to its scalability.[101] Neural networks are also used as classifiers.[102]




Expectation-maximization clustering of Old Faithful eruption data starts from a random guess but then successfully converges on an accurate clustering of the two physically distinct modes of eruption.

## Artificial neural networks

An artificial neural network is based on a collection of nodes also known as artificial neurons, which loosely model the neurons in a biological brain. It is trained to recognise patterns; once trained, it can recognise those patterns in fresh data. There is an input, at least one hidden layer of nodes and an output. Each node applies a function and once the weight crosses its specified threshold, the data is transmitted to the next layer. A network is typically called a deep neural network if it has at least 2 hidden layers.[102]




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

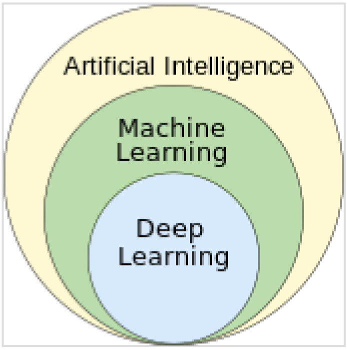
Learning algorithms for neural networks use local search to choose the weights that will get the right output for each input during training. The most common training technique is the backpropagation algorithm.[103] Neural networks learn to model complex relationships between inputs and outputs and find patterns in data. In theory, a neural network can learn any function.[104]

In feedforward neural networks the signal passes in only one direction.[105] Recurrent neural networks feed the output signal back into the input, which allows short-term memories of previous input events. Long short term memory is the most successful network architecture for recurrent networks.[106] Perceptrons[107] use only a single layer of neurons, deep learning[108] uses multiple layers. Convolutional neural networks strengthen the connection between neurons that are "close" to each other – this is especially important in image processing, where a local set of neurons must identify an "edge" before the network can identify an object.[109]

## Deep learning

Deep learning[108] uses several layers of neurons between the network's inputs and outputs. The multiple layers can progressively extract higher-level features from the raw input. For example, in image processing, lower layers may identify edges, while higher layers may identify the concepts relevant to a human such as digits or letters or faces.[110]

Deep learning has profoundly improved the performance of programs in many important subfields of artificial intelligence, including computer vision, speech recognition, natural language processing, image classification[111] and others. The reason that deep learning performs so well in so many applications is not known as of 2023.[112] The sudden success of deep learning in 2012–2015 did not occur because of some new discovery or theoretical breakthrough (deep neural networks and backpropagation had been described by many people, as far back as the 1950s)[i] but because of two factors: the incredible increase in computer power (including the hundred-fold increase in speed by switching to GPUs) and the availability of vast amounts of training data, especially the giant curated datasets used for benchmark testing, such as ImageNet.[j]




### GPT

Generative pre-trained transformers (GPT) are large language models that are based on the semantic relationships between words in sentences (natural language processing). Text-based GPT models are pre-trained on a large corpus of text which can be from the internet. The pre-training consists in predicting the next token (a token being usually a word, subword, or punctuation). Throughout this pre-training, GPT models accumulate knowledge about the world, and can then generate human-like text by repeatedly predicting the next token. Typically, a subsequent training phase makes the model more truthful, useful and harmless, usually with a technique called reinforcement learning from human feedback (RLHF). Current GPT models are still prone to generating falsehoods called "hallucinations", although this can be reduced with RLHF and quality data. They are used in chatbots, which allow you to ask a question or request a task in simple text.[121][122]

Current models and services include: Gemini (formerly Bard), ChatGPT, Grok, Claude, Copilot and LLaMA.[123] Multimodal GPT models can process different types of data (modalities) such as images, videos, sound and text.[124]

### Specialized hardware and software

In the late 2010s, graphics processing units (GPUs) that were increasingly designed with AI-specific enhancements and used with specialized TensorFlow software, had replaced previously used central processing unit (CPUs) as the dominant means for large-scale (commercial and academic) machine learning models' training.[125] Historically, specialized languages, such as Lisp, Prolog, Python and others, had been used.

## Applications

AI and machine learning technology is used in most of the essential applications of the 2020s, including: search engines (such as Google Search), targeting online advertisements, recommendation systems (offered by Netflix, YouTube or Amazon), driving internet traffic, targeted advertising (AdSense, Facebook), virtual assistants (such as Siri or Alexa), autonomous vehicles (including drones, ADAS and self-driving cars), automatic language translation (Microsoft Translator, Google Translate), facial recognition (Apple's Face ID or Microsoft's DeepFace and Google's FaceNet) and image labeling (used by Facebook, Apple's iPhoto and TikTok).

## Health and medicine

The application of AI in medicine and medical research has the potential to increase patient care and quality of life.[126] Through the lens of the Hippocratic Oath, medical professionals are ethically compelled to use AI, if applications can more accurately diagnose and treat patients.

For medical research, AI is an important tool for processing and integrating Big Data. This is particularly important for organoid and tissue engineering development which use microscopy imaging as a key technique in fabrication.[127] It has been suggested that AI can overcome discrepancies in funding allocated to different fields of research.[127] New AI tools can deepen our understanding of biomedically relevant pathways. For example, AlphaFold 2 (2021) demonstrated the ability to approximate, in hours rather than months, the 3D structure of a protein.[128] In 2023 it was reported that AI guided drug discovery helped find a class of antibiotics capable of killing two different types of drug-resistant bacteria.[129]

## Games

Game playing programs have been used since the 1950s to demonstrate and test AI's most advanced techniques.[130] Deep Blue became the first computer chess-playing system to beat a reigning world chess champion, Garry Kasparov, on 11 May 1997.[131] In 2011, in a *Jeopardy!* quiz show exhibition match, IBM's question answering system, Watson, defeated the two greatest *Jeopardy!* champions, Brad Rutter and Ken Jennings, by a significant margin.[132] In March 2016, AlphaGo won 4 out of 5 games of Go in a match with Go champion Lee Sedol, becoming the first computer Go-playing system to beat a professional Go player without handicaps. Then in 2017 it defeated Ke Jie, who was the best Go player in the world.[133] Other programs handle imperfect-information games, such as the poker-playing program Pluribus.[134] DeepMind developed increasingly generalistic reinforcement learning models, such as with MuZero, which could be trained to play chess, Go, or Atari games.[135] In 2019, DeepMind's AlphaStar achieved grandmaster level in StarCraft II, a particularly challenging real-time strategy game that involves incomplete knowledge of what happens on the map.[136] In 2021 an AI agent competed in a Playstation Gran Turismo competition, winning against four of the world's best Gran Turismo drivers using deep reinforcement learning.[137]

## Military

Various countries are deploying AI military applications.[138] The main applications enhance command and control, communications, sensors, integration and interoperability.[139] Research is targeting intelligence collection and analysis, logistics, cyber operations, information operations, and semiautonomous and autonomous vehicles.[138] AI technologies enable coordination of sensors and effectors, threat detection and identification, marking of enemy positions, target acquisition, coordination and deconfliction of distributed Joint Fires between networked combat vehicles involving manned and unmanned teams.[139] AI was incorporated into military operations in Iraq and Syria.[138]

In November 2023, US Vice President Kamala Harris disclosed a declaration signed by 31 nations to set guardrails for the military use of IA. The commitments include using legal reviews to ensure the compliance of military AI with international laws, and being cautious and transparent in the development of this technology.[140]

### Generative AI

Vincent van Gogh in watercolour created by generative AI software

In the early 2020s, generative AI gained widespread prominence. In March 2023, 58% of US adults had heard about ChatGPT and 14% had tried it.[141] The increasing realism and ease-of-use of AI-based text-to-image generators such as Midjourney, DALL-E, and Stable Diffusion sparked a trend of viral AI-generated photos. Widespread attention was gained by a fake photo of Pope Francis wearing a white puffer coat, the fictional arrest of Donald Trump, and a hoax of an attack on the Pentagon, as well as the usage in professional creative arts.[142][143]

### Industry-specific tasks

There are also thousands of successful AI applications used to solve specific problems for specific industries or institutions. In a 2017 survey, one in five companies reported they had incorporated "AI" in some offerings or processes.[144] A few examples are energy storage, medical diagnosis, military logistics, applications that predict the result of judicial decisions, foreign policy, or supply chain management.

In agriculture, AI has helped farmers identify areas that need irrigation, fertilization, pesticide treatments or increasing yield. Agronomists use AI to conduct research and development. AI has been used to predict the ripening time for crops such as tomatoes, monitor soil moisture, operate agricultural robots, conduct predictive analytics, classify livestock pig call emotions, automate greenhouses, detect diseases and pests, and save water.

Artificial intelligence is used in astronomy to analyze increasing amounts of available data and applications, mainly for "classification, regression, clustering, forecasting, generation, discovery, and the development of new scientific insights" for example for discovering exoplanets, forecasting solar activity, and distinguishing between signals and instrumental effects in gravitational wave astronomy. It could also be used for activities in space such as space exploration, including analysis of data from space missions, real-time science decisions of spacecraft, space debris avoidance, and more autonomous operation.

## Ethics

AI, like any powerful technology, has potential benefits and potential risks. AI may be able to advance science and find solutions for serious problems: Demis Hassabis of Deep Mind hopes to "solve intelligence, and then use that to solve everything else".[145] However, as the use of AI has become widespread, several unintended consequences and risks have been identified.[146]

Anyone looking to use machine learning as part of real-world, in-production systems needs to factor ethics into their AI training processes and strive to avoid bias. This is especially true when using AI algorithms that are inherently unexplainable in deep learning.[147]

## Risks and harm

### Privacy and copyright

Machine learning algorithms require large amounts of data. The techniques used to acquire this data have raised concerns about privacy, surveillance and copyright.

Technology companies collect a wide range of data from their users, including online activity, geolocation data, video and audio.[148] For example, in order to build speech recognition algorithms, Amazon have recorded millions of private conversations and allowed temporary workers to listen to and transcribe some of them.[149] Opinions about this widespread surveillance range from those who see it as a necessary evil to those for whom it is clearly unethical and a violation of the right to privacy.[150]

AI developers argue that this is the only way to deliver valuable applications. and have developed several techniques that attempt to preserve privacy while still obtaining the data, such as data aggregation, de-identification and differential privacy.[151] Since 2016, some privacy experts, such as Cynthia Dwork, began to view privacy in terms of fairness. Brian Christian wrote that experts have pivoted "from the question of 'what they know' to the question of 'what they're doing with it'.".[152]

Generative AI is often trained on unlicensed copyrighted works, including in domains such as images or computer code; the output is then used under a rationale of "fair use". Also website owners who do not wish to have their copyrighted content be AI indexed or 'scraped' can add code to their site, as you would, if you did not want your website to be indexed by a search engine which is currently available to certain services such as OpenAI. Experts disagree about how well, and under what circumstances, this rationale will hold up in courts of law; relevant factors may include "the purpose and character of the use of the copyrighted work" and "the effect upon the potential market for the copyrighted work".[153] In 2023, leading authors (including John Grisham and Jonathan Franzen) sued AI companies for using their work to train generative AI.[154][155]

### Misinformation

YouTube, Facebook and others use recommender systems to guide users to more content. These AI programs were given the goal of maximizing user engagement (that is, the only goal was to keep people watching). The AI learned that users tended to choose misinformation, conspiracy theories, and extreme partisan content, and, to keep them watching, the AI recommended more of it. Users also tended to watch more content on the same subject, so the AI led people into filter bubbles where they received multiple versions of the same misinformation.[156] This convinced many users that the misinformation was true, and ultimately undermined trust in institutions, the media and the government.[157] The AI program had correctly learned to maximize its goal, but the result was harmful to society. After the U.S. election in 2016, major technology companies took steps to mitigate the problem.

In 2022, generative AI began to create images, audio, video and text that are indistinguishable from real photographs, recordings, films or human writing. It is possible for bad actors to use this technology to create massive amounts of misinformation or propaganda.[158] AI pioneer Geoffrey Hinton expressed concern about AI enabling "authoritarian leaders to manipulate their electorates" on a large scale, among other risks.[159]

### Algorithmic bias and fairness

Machine learning applications will be biased if they learn from biased data.[160] The developers may not be aware that the bias exists.[161] Bias can be introduced by the way training data is selected and by the way a model is deployed.[162][160] If a biased algorithm is used to make decisions that can seriously harm people (as it can in medicine, finance, recruitment, housing or policing) then the algorithm may cause discrimination.[163] Fairness in machine learning is the study of how to prevent the harm caused by algorithmic bias. It has become serious area of academic study within AI. Researchers have discovered it is not always possible to define "fairness" in a way that satisfies all stakeholders.[164]

On June 28, 2015, Google Photos's new image labeling feature mistakenly identified Jacky Alcine and a friend as "gorillas" because they were black. The system was trained on a dataset that contained very few images of black people,[165] a problem called "sample size disparity".[166] Google "fixed" this problem by preventing the system from labelling *anything* as a "gorilla". Eight years later, in 2023, Google Photos still could not identify a gorilla, and neither could similar products from Apple, Facebook, Microsoft and Amazon.[167]

COMPAS is a commercial program widely used by U.S. courts to assess the likelihood of a defendant becoming a recidivist. In 2016, Julia Angwin at ProPublica discovered that COMPAS exhibited racial bias, despite the fact that the program was not told the races of the defendants. Although the error rate for both whites and blacks was calibrated equal at exactly 61%, the errors for each race were different—the system consistently overestimated the chance that a black person would re-offend and would underestimate the chance that a white person would not re-offend.[168] In 2017, several researchers[k] showed that it was mathematically impossible for COMPAS to accommodate all possible measures of fairness when the base rates of re-offense were different for whites and blacks in the data.[170]

A program can make biased decisions even if the data does not explicitly mention a problematic feature (such as "race" or "gender"). The feature will correlate with other features (like "address", "shopping history" or "first name"), and the program will make the same decisions based on these features as it would on "race" or "gender".[171] Moritz Hardt said "the most robust fact in this research area is that fairness through blindness doesn't work."[172]

Criticism of COMPAS highlighted a deeper problem with the misuse of AI. Machine learning models are designed to make "predictions" that are only valid if we assume that the future will resemble the past. If they are trained on data that includes the results of racist decisions in the past, machine learning models must predict that racist decisions will be made in the future. Unfortunately, if an application then uses these predictions as *recommendations*, some of these "recommendations" will likely be racist.[173] Thus, machine learning is not well suited to help make decisions in areas where there is hope that the future will be *better* than the past. It is necessarily descriptive and not proscriptive.[l]

Bias and unfairness may go undetected because the developers are overwhelmingly white and male: among AI engineers, about 4% are black and 20% are women.[166]

At its 2022 Conference on Fairness, Accountability, and Transparency (ACM FAccT 2022) the Association for Computing Machinery, in Seoul, South Korea, presented and published findings recommending that until AI and robotics systems are demonstrated to be free of bias mistakes, they are unsafe and the use of self-learning neural networks trained on vast, unregulated sources of flawed internet data should be curtailed.[175]

### Lack of transparency

Many AI systems are so complex that their designers cannot explain how they reach their decisions.[176] Particularly with deep neural networks, in which there are a large amount of non-linear relationships between inputs and outputs. But some popular explainability techniques exist.[177]

Lidar testing vehicle for autonomous driving

There have been many cases where a machine learning program passed rigorous tests, but nevertheless learned something different than what the programmers intended. For example, a system that could identify skin diseases better than medical professionals was found to actually have a strong tendency to classify images with a ruler as "cancerous", because pictures of malignancies typically include a ruler to show the scale.[178] Another machine learning system designed to help effectively allocate medical resources was found to classify patients with asthma as being at "low risk" of dying from pneumonia. Having asthma is actually a severe risk factor, but since the patients having asthma would usually get much more medical care, they were relatively unlikely to die according to the training data. The correlation between asthma and low risk of dying from pneumonia was real, but misleading.[179]

People who have been harmed by an algorithm's decision have a right to an explanation. Doctors, for example, are required to clearly and completely explain the reasoning behind any decision they make.[180] Early drafts of the European Union's General Data Protection Regulation in 2016 included an explicit statement that this right exists.[m] Industry experts noted that this is an unsolved problem with no solution in sight. Regulators argued that nevertheless the harm is real: if the problem has no solution, the tools should not be used.[181]

DARPA established the XAI ("Explainable Artificial Intelligence") program in 2014 to try and solve these problems.[182]

There are several potential solutions to the transparency problem. SHAP helps visualise the contribution of each feature to the output.[183] LIME can locally approximate a model with a simpler, interpretable model.[184] Multitask learning provides a large number of outputs in addition to the target classification. These other outputs can help developers deduce what the network has learned.[185] Deconvolution, DeepDream and other generative methods can allow developers to see what different layers of a deep network have learned and produce output that can suggest what the network is learning.[186]

### Conflict, surveillance and weaponized AI

A lethal autonomous weapon is a machine that locates, selects and engages human targets without human supervision.[n] By 2015, over fifty countries were reported to be researching battlefield robots.[188] These weapons are considered especially dangerous for several reasons: if they kill an innocent person it is not clear who should be held accountable, it is unlikely they will reliably choose targets, and, if produced at scale, they are potentially weapons of mass destruction.[189] In 2014, 30 nations (including China) supported a ban on autonomous weapons under the United Nations' Convention on Certain Conventional Weapons, however the United States and others disagreed.[190]

AI provides a number of tools that are particularly useful for authoritarian governments: smart spyware, face recognition and voice recognition allow widespread surveillance; such surveillance allows machine learning to classify potential enemies of the state and can prevent them from hiding; recommendation systems can precisely target propaganda and misinformation for maximum effect; deepfakes and generative AI aid in producing misinformation; advanced AI can make authoritarian centralized decision making more competitive with liberal and decentralized systems such as markets.[191]

AI facial recognition systems are used for mass surveillance, notably in China.[192][193] In 2019, Bengaluru, India deployed AI-managed traffic signals. This system uses cameras to monitor traffic density and adjust signal timing based on the interval needed to clear traffic.[194] Terrorists, criminals and rogue states can use weaponized AI such as advanced digital warfare and lethal autonomous weapons. Machine-learning AI is also able to design tens of thousands of toxic molecules in a matter of hours.[195]

### Technological unemployment

From the early days of the development of artificial intelligence there have been arguments, for example those put forward by Joseph Weizenbaum, about whether tasks that can be done by computers actually should be done by them, given the difference between computers and humans, and between quantitative calculation and qualitative, value-based judgement.[196]

Economists have frequently highlighted the risks of redundancies from AI, and speculated about unemployment if there is no adequate social policy for full employment.[197]

In the past, technology has tended to increase rather than reduce total employment, but economists acknowledge that "we're in uncharted territory" with AI.[198] A survey of economists showed disagreement about whether the increasing use of robots and AI will cause a substantial increase in long-term unemployment, but they generally agree that it could be a net benefit if productivity gains are redistributed.[199] Risk estimates vary; for example, in the 2010s, Michael Osborne and Carl Benedikt Frey estimated 47% of U.S. jobs are at "high risk" of potential automation, while an OECD report classified only 9% of U.S. jobs as "high risk".[o][201] The methodology of speculating about future employment levels has been criticised as lacking evidential foundation, and for implying that technology, rather than social policy, creates unemployment, as opposed to redundancies.[197]

Unlike previous waves of automation, many middle-class jobs may be eliminated by artificial intelligence; *The Economist* stated in 2015 that "the worry that AI could do to white-collar jobs what steam power did to blue-collar ones during the Industrial Revolution" is "worth taking seriously".[202] Jobs at extreme risk range from paralegals to fast food cooks, while job demand is likely to increase for care-related professions ranging from personal healthcare to the clergy.[203]

In April 2023, it was reported that 70% of the jobs for Chinese video game illustrators had been eliminated by generative artificial intelligence.[204][205]

### Existential risk

It has been argued AI will become so powerful that humanity may irreversibly lose control of it. This could, as physicist Stephen Hawking stated, "spell the end of the human race".[206] This scenario has been common in science fiction, when a computer or robot suddenly develops a human-like "self-awareness" (or "sentience" or "consciousness") and becomes a malevolent character.[p] These sci-fi scenarios are misleading in several ways.

First, AI does not require human-like "sentience" to be an existential risk. Modern AI programs are given specific goals and use learning and intelligence to achieve them. Philosopher Nick Bostrom argued that if one gives *almost any* goal to a sufficiently powerful AI, it may choose to destroy humanity to achieve it (he used the example of a paperclip factory manager).[208] Stuart Russell gives the example of household robot that tries to find a way to kill its owner to prevent it from being unplugged, reasoning that "you can't fetch the coffee if you're dead."[209] In order to be safe for humanity, a superintelligence would have to be genuinely aligned with humanity's morality and values so that it is "fundamentally on our side".[210]

Second, Yuval Noah Harari argues that AI does not require a robot body or physical control to pose an existential risk. The essential parts of civilization are not physical. Things like ideologies, law, government, money and the economy are made of language; they exist because there are stories that billions of people believe. The current prevalence of misinformation suggests that an AI could use language to convince people to believe anything, even to take actions that are destructive.[211]

The opinions amongst experts and industry insiders are mixed, with sizable fractions both concerned and unconcerned by risk from eventual superintelligent AI.[212] Personalities such as Stephen Hawking, Bill Gates, and Elon Musk have expressed concern about existential risk from AI.[213]

In the early 2010s, experts argued that the risks are too distant in the future to warrant research or that humans will be valuable from the perspective of a superintelligent machine.[214] However, after 2016, the study of current and future risks and possible solutions became a serious area of research.[215]

AI pioneers including Fei-Fei Li, Geoffrey Hinton, Yoshua Bengio, Cynthia Breazeal, Rana el Kaliouby, Demis Hassabis, Joy Buolamwini, and Sam Altman have expressed concerns about the risks of AI. In 2023, many leading AI experts issued the joint statement that "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war".[216]

Other researchers, however, spoke in favor of a less dystopian view. AI pioneer Juergen Schmidhuber did not sign the joint statement, emphasising that in 95% of all cases, AI research is about making "human lives longer and healthier and easier."[217] While the tools that are now being used to improve lives can also be used by bad actors, "they can also be used against the bad actors."[218][219] Andrew Ng also argued that "it's a mistake to fall for the doomsday hype on AI—and that regulators who do will only benefit vested interests."[220] Yann LeCun "scoffs at his peers' dystopian scenarios of supercharged misinformation and even, eventually, human extinction."[221]

### Limiting AI

Possible options for limiting AI include: using Embedded Ethics or Constitutional AI where companies or governments can add a policy, restricting high levels of compute power in training, restricting the ability to rewrite its own code base, restrict certain AI techniques but not in the training phase, open-source (transparency) vs proprietary (could be more restricted), backup model with redundancy, restricting security, privacy and copyright, restricting or controlling the memory, real-time monitoring, risk analysis, emergency shut-off, rigorous simulation and testing, model certification, assess known vulnerabilities, restrict the training material, restrict access to the internet, issue terms of use.

## Ethical machines and alignment

Friendly AI are machines that have been designed from the beginning to minimize risks and to make choices that benefit humans. Eliezer Yudkowsky, who coined the term, argues that developing friendly AI should be a higher research priority: it may require a large investment and it must be completed before AI becomes an existential risk.[222]

Machines with intelligence have the potential to use their intelligence to make ethical decisions. The field of machine ethics provides machines with ethical principles and procedures for resolving ethical dilemmas.[223] The field of machine ethics is also called computational morality,[223] and was founded at an AAAI symposium in 2005.[224]

Other approaches include Wendell Wallach's "artificial moral agents"[225] and Stuart J. Russell's three principles for developing provably beneficial machines.[226]

## Frameworks

Artificial Intelligence projects can have their ethical permissibility tested while designing, developing, and implementing an AI system. An AI framework such as the Care and Act Framework containing the SUM values – developed by the Alan Turing Institute tests projects in four main areas:[227][228]

- RESPECT the dignity of individual people
- CONNECT with other people sincerely, openly and inclusively
- CARE for the wellbeing of everyone
- PROTECT social values, justice and the public interest

Other developments in ethical frameworks include those decided upon during the Asilomar Conference, the Montreal Declaration for Responsible AI, and the IEEE's Ethics of Autonomous Systems initiative, among others;[229] however, these principles do not go without their criticisms, especially regards to the people chosen contributes to these frameworks.[230]

Promotion of the wellbeing of the people and communities that these technologies affect requires consideration of the social and ethical implications at all stages of AI system design, development and implementation, and collaboration between job roles such as data scientists, product managers, data engineers, domain experts, and delivery managers.[231]

## Regulation

The regulation of artificial intelligence is the development of public sector policies and laws for promoting and regulating artificial intelligence (AI); it is therefore related to the broader regulation of algorithms.[232] The regulatory and policy landscape for AI is an emerging issue in jurisdictions globally.[233] According to AI Index at Stanford, the annual number of AI-related laws passed in the 127 survey countries jumped from one passed in 2016 to 37 passed in 2022 alone.[234][235] Between 2016 and 2020, more than 30 countries adopted dedicated strategies for AI.[236] Most EU member states had released national AI strategies, as had

Canada, China, India, Japan, Mauritius, the Russian Federation, Saudi Arabia, United Arab Emirates, US and Vietnam. Others were in the process of elaborating their own AI strategy, including Bangladesh, Malaysia and Tunisia.[236] The Global Partnership on Artificial Intelligence was launched in June 2020, stating a need for AI to be developed in accordance with human rights and democratic values, to ensure public confidence and trust in the technology.[236] Henry Kissinger, Eric Schmidt, and Daniel Huttenlocher published a joint statement in November 2021 calling for a government commission to regulate AI.[237] In 2023, OpenAI leaders published recommendations for the governance of superintelligence, which they believe may happen in less than 10 years.[238] In 2023, the United Nations also launched an advisory body to provide recommendations on AI governance; the body comprises technology company executives, governments officials and academics.[239]

The first global AI Safety Summit was held in 2023 with a declaration calling for international co-operation.

In a 2022 Ipsos survey, attitudes towards AI varied greatly by country; 78% of Chinese citizens, but only 35% of Americans, agreed that "products and services using AI have more benefits than drawbacks".[234] A 2023 Reuters/Ipsos poll found that 61% of Americans agree, and 22% disagree, that AI poses risks to humanity.[240] In a 2023 Fox News poll, 35% of Americans thought it "very important", and an additional 41% thought it "somewhat important", for the federal government to regulate AI, versus 13% responding "not very important" and 8% responding "not at all important".[241][242]

In November 2023, the first global AI Safety Summit was held in Bletchley Park in the UK to discuss the near and far term risks of AI and the possibility of mandatory and voluntary regulatory frameworks.[243] 28 countries including the United States, China, and the European Union issued a declaration at the start of the summit, calling for international co-operation to manage the challenges and risks of artificial intelligence.[244][245]

# History

The study of mechanical or "formal" reasoning began with philosophers and mathematicians in antiquity. The study of logic led directly to Alan Turing's theory of computation, which suggested that a machine, by shuffling symbols as simple as "0" and "1", could simulate both mathematical deduction and formal reasoning, which is known as the Church–Turing thesis.[246] This, along with concurrent discoveries in cybernetics and information theory, led researchers to consider the possibility of building an "electronic brain".[q][248]

Alan Turing was thinking about machine intelligence at least as early as 1941, when he circulated a paper on machine intelligence which could be the earliest paper in the field of AI – though it is now lost.[2] The first available paper generally recognized as "AI" was McCullouch and Pitts design for Turing-complete "artificial neurons" in 1943 – the first mathematical model of a neural network.[249] The paper was influenced by Turing's earlier paper 'On Computable Numbers' from 1936 using similar two-state boolean 'neurons', but was the first to apply it to neuronal function.[2]

The term 'machine intelligence' was used by Alan Turing during his life which was later often referred to as 'artificial intelligence' after his death in 1954. In 1950 Turing published the best known of his papers 'Computing Machinery and Intelligence', the paper introduced his concept of what is now known as the Turing test to the general public. Then followed three radio broadcasts on AI by Turing, the lectures: 'Intelligent Machinery, A Heretical Theory', 'Can Digital Computers Think'? and the panel discussion 'Can Automatic Calculating Machines be Said to Think'. By 1956 computer intelligence had been actively pursued for more than a decade in Britain; the earliest AI programmes were written there in 1951–1952.[2]

In 1951, using a Ferranti Mark 1 computer of the University of Manchester, checkers and chess programs were written where you could play against the computer.[250] The field of American AI research was founded at a workshop at Dartmouth College in 1956.[r][3] The attendees became the leaders of AI research in the 1960s.[s] They and their students produced programs that the press described as "astonishing":[t] computers were learning checkers strategies, solving word problems in algebra, proving logical theorems and speaking English.[u][4] Artificial Intelligence laboratories were set up at a number of British and US Universities in the latter 1950s and early 1960s.[2]

They had, however, underestimated the difficulty of the problem.[v] Both the U.S. and British governments cut off exploratory research in response to the criticism of Sir James Lighthill[255] and ongoing pressure from the U.S. Congress to fund more productive projects. Minsky's and Papert's book *Perceptrons* was understood as proving that artificial neural networks would never be useful for solving real-world tasks, thus discrediting the approach altogether.[256] The "AI winter", a period when obtaining funding for AI projects was difficult, followed.[6]

In the early 1980s, AI research was revived by the commercial success of expert systems,[257] a form of AI program that simulated the knowledge and analytical skills of human experts. By 1985, the market for AI had reached over a billion dollars. At the same time, Japan's fifth generation computer project inspired the U.S. and British governments to restore funding for academic research.[5] However, beginning with the collapse of the Lisp Machine market in 1987, AI once again fell into disrepute, and a second, longer-lasting winter began.[7]

Many researchers began to doubt that the current practices would be able to imitate all the processes of human cognition, especially perception, robotics, learning and pattern recognition.[258] A number of researchers began to look into "sub-symbolic" approaches.[259] Robotics researchers, such as Rodney Brooks, rejected "representation" in general and focussed directly on engineering machines that move and survive.[w] Judea Pearl, Lofti Zadeh and others developed methods that handled incomplete and uncertain information by making reasonable guesses rather than precise logic.[85][264] But the most important development was the revival of "connectionism", including neural network research, by Geoffrey Hinton and others.[265] In 1990, Yann LeCun successfully showed that convolutional neural networks can recognize handwritten digits, the first of many successful applications of neural networks.[266]

AI gradually restored its reputation in the late 1990s and early 21st century by exploiting formal mathematical methods and by finding specific solutions to specific problems. This "narrow" and "formal" focus allowed researchers to produce verifiable results and collaborate with other fields (such as statistics, economics and mathematics).[267] By 2000, solutions developed by AI researchers were being widely used, although in the 1990s they were rarely described as "artificial intelligence".[268]

Several academic researchers became concerned that AI was no longer pursuing the original goal of creating versatile, fully intelligent machines. Beginning around 2002, they founded the subfield of artificial general intelligence (or "AGI"), which had several well-funded institutions by the 2010s.[11]

Deep learning began to dominate industry benchmarks in 2012 and was adopted throughout the field.[8] For many specific tasks, other methods were abandoned.[x] Deep learning's success was based on both hardware improvements (faster computers,[270] graphics processing units, cloud computing[271]) and access to large amounts of data[272] (including curated datasets,[271] such as ImageNet).

Deep learning's success led to an enormous increase in interest and funding in AI.[y] The amount of machine learning research (measured by total publications) increased by 50% in the years 2015–2019,[236] and WIPO reported that AI was the most prolific emerging technology in terms of the number of patent applications and granted patents.[273] According to 'AI Impacts', about $50 billion annually was invested in "AI" around 2022 in the US alone and about 20% of new US Computer Science PhD graduates have specialized in "AI";[274] about 800,000 "AI"-related US job openings existed in 2022.[275] The large majority of the advances have occurred within the United States, with its companies, universities, and research labs leading artificial intelligence research.[10]

In 2016, issues of fairness and the misuse of technology were catapulted into center stage at machine learning conferences, publications vastly increased, funding became available, and many researchers re-focussed their careers on these issues. The alignment problem became a serious field of academic study.[215]

# Philosophy

## Defining artificial intelligence

Alan Turing wrote in 1950 "I propose to consider the question 'can machines think'?"[276] He advised changing the question from whether a machine "thinks", to "whether or not it is possible for machinery to show intelligent behaviour".[276] He devised the Turing test, which measures the ability of a machine to simulate human conversation.[277] Since we can only observe the behavior of the machine, it does not matter if it is "actually" thinking or literally has a "mind". Turing notes that we can not determine these things about other people[z] but "it is usual to have a polite convention that everyone thinks"[278]

Russell and Norvig agree with Turing that AI must be defined in terms of "acting" and not "thinking".[279] However, they are critical that the test compares machines to *people*. "Aeronautical engineering texts," they wrote, "do not define the goal of their field as making 'machines that fly so exactly like pigeons that they can fool other pigeons.' "[280] AI founder John McCarthy agreed, writing that "Artificial intelligence is not, by definition, simulation of human intelligence".[281]

McCarthy defines intelligence as "the computational part of the ability to achieve goals in the world."[282] Another AI founder, Marvin Minsky similarly defines it as "the ability to solve hard problems".[283] These definitions view intelligence in terms of well-defined problems with well-defined solutions, where both the difficulty of the problem and the performance of the program are direct measures of the "intelligence" of the machine—and no other philosophical discussion is required, or may not even be possible.

Another definition has been adopted by Google,[284] a major practitioner in the field of AI. This definition stipulates the ability of systems to synthesize information as the manifestation of intelligence, similar to the way it is defined in biological intelligence.

## Evaluating approaches to AI

No established unifying theory or paradigm has guided AI research for most of its history.[aa] The unprecedented success of statistical machine learning in the 2010s eclipsed all other approaches (so much so that some sources, especially in the business world, use the term "artificial intelligence" to mean "machine learning with neural networks"). This approach is mostly sub-symbolic, soft and narrow (see below). Critics argue that these questions may have to be revisited by future generations of AI researchers.

### Symbolic AI and its limits

Symbolic AI (or "GOFAI")[286] simulated the high-level conscious reasoning that people use when they solve puzzles, express legal reasoning and do mathematics. They were highly successful at "intelligent" tasks such as algebra or IQ tests. In the 1960s, Newell and Simon proposed the physical symbol systems hypothesis: "A physical symbol system has the necessary and sufficient means of general intelligent action."[287]

However, the symbolic approach failed on many tasks that humans solve easily, such as learning, recognizing an object or commonsense reasoning. Moravec's paradox is the discovery that high-level "intelligent" tasks were easy for AI, but low level "instinctive" tasks were extremely difficult.[288] Philosopher Hubert Dreyfus had argued since the 1960s that human expertise depends on unconscious instinct rather than conscious symbol manipulation, and on having a "feel" for the situation, rather than explicit symbolic knowledge.[289] Although his arguments had been ridiculed and ignored when they were first presented, eventually, AI research came to agree with him.[ab][16]

The issue is not resolved: sub-symbolic reasoning can make many of the same inscrutable mistakes that human intuition does, such as algorithmic bias. Critics such as Noam Chomsky argue continuing research into symbolic AI will still be necessary to attain general intelligence,[291][292] in part because sub-symbolic AI is a move away from explainable AI: it can be difficult or impossible to understand why a modern statistical AI program made a particular decision. The emerging field of neuro-symbolic artificial intelligence attempts to bridge the two approaches.

### Neat vs. scruffy

"Neats" hope that intelligent behavior is described using simple, elegant principles (such as logic, optimization, or neural networks). "Scruffies" expect that it necessarily requires solving a large number of unrelated problems. Neats defend their programs with theoretical rigor, scruffies rely mainly on incremental testing to see if they work. This issue was actively discussed in the 1970s and 1980s,[293] but eventually was seen as irrelevant. Modern AI has elements of both.

### Soft vs. hard computing

Finding a provably correct or optimal solution is intractable for many important problems.[15] Soft computing is a set of techniques, including genetic algorithms, fuzzy logic and neural networks, that are tolerant of imprecision, uncertainty, partial truth and approximation. Soft computing was introduced in the late 1980s and most successful AI programs in the 21st century are examples of soft computing with neural networks.

### Narrow vs. general AI

AI researchers are divided as to whether to pursue the goals of artificial general intelligence and superintelligence directly or to solve as many specific problems as possible (narrow AI) in hopes these solutions will lead indirectly to the field's long-term goals.[294][295] General intelligence is difficult to define and difficult to measure, and modern AI has had more verifiable successes by focusing on specific problems with specific solutions. The experimental sub-field of artificial general intelligence studies this area exclusively.

## Machine consciousness, sentience and mind

The philosophy of mind does not know whether a machine can have a mind, consciousness and mental states, in the same sense that human beings do. This issue considers the internal experiences of the machine, rather than its external behavior. Mainstream AI research considers this issue irrelevant because it does not affect the goals of the field: to build machines that can solve problems using intelligence. Russell and Norvig add that "[t]he additional project of making a machine conscious in exactly the way humans are is not one that we are equipped to take on."[296] However, the question has become central to the philosophy of mind. It is also typically the central question at issue in artificial intelligence in fiction.

### Consciousness

David Chalmers identified two problems in understanding the mind, which he named the "hard" and "easy" problems of consciousness.[297] The easy problem is understanding how the brain processes signals, makes plans and controls behavior. The hard problem is explaining how this *feels* or why it should feel like anything at all, assuming we are right in thinking that it truly does feel like something (Dennett's consciousness illusionism says this is an illusion). Human information processing is easy to explain, however, human subjective experience is difficult to explain. For example, it is easy to imagine a color-blind person who has learned to identify which objects in their field of view are red, but it is not clear what would be required for the person to *know what red looks like*.[298]

### Computationalism and functionalism

Computationalism is the position in the philosophy of mind that the human mind is an information processing system and that thinking is a form of computing. Computationalism argues that the relationship between mind and body is similar or identical to the relationship between software and hardware and thus may be a solution to the mind–body problem. This philosophical position was inspired by the work of AI researchers and cognitive scientists in the 1960s and was originally proposed by philosophers Jerry Fodor and Hilary Putnam.[299]

Philosopher John Searle characterized this position as "strong AI": "The appropriately programmed computer with the right inputs and outputs would thereby have a mind in exactly the same sense human beings have minds."[ac] Searle counters this assertion with his Chinese room argument, which attempts to show that, even if a machine perfectly simulates human behavior, there is still no reason to suppose it also has a mind.[303]

### AI welfare and rights

It is difficult or impossible to reliably evaluate whether an advanced AI is sentient (has the ability to feel), and if so, to what degree.[304] But if there is a significant chance that a given machine can feel and suffer, then it may be entitled to certain rights or welfare protection measures, similarly to animals.[305][306] Sapience (a set of capacities related to high intelligence, such as discernment or self-awareness) may provide another moral basis for AI rights.[305] Robot rights are also sometimes proposed as a practical way to integrate autonomous agents into society.[307]

In 2017, the European Union considered granting "electronic personhood" to some of the most capable AI systems. Similarly to the legal status of companies, it would have conferred rights but also responsibilities.[308] Critics argued in 2018 that granting rights to AI systems would downplay the importance of human rights, and that legislation should focus on user needs rather than speculative futuristic scenarios. They also noted that robots lacked the autonomy to take part to society on their own.[309][310]

Progress in AI increased interest in the topic. Proponents of AI welfare and rights often argue that AI sentience, if it emerges, would be particularly easy to deny. They warn that this may be a moral blind spot analogous to slavery or factory farming, which could lead to large-scale suffering if sentient AI is created and carelessly exploited.[306][305]

# Future

## Superintelligence and the singularity

A superintelligence is a hypothetical agent that would possess intelligence far surpassing that of the brightest and most gifted human mind.[295]

If research into artificial general intelligence produced sufficiently intelligent software, it might be able to reprogram and improve itself. The improved software would be even better at improving itself, leading to what I. J. Good called an "intelligence explosion" and Vernor Vinge called a "singularity".[311]

However, technologies cannot improve exponentially indefinitely, and typically follow an S-shaped curve, slowing when they reach the physical limits of what the technology can do.[312]

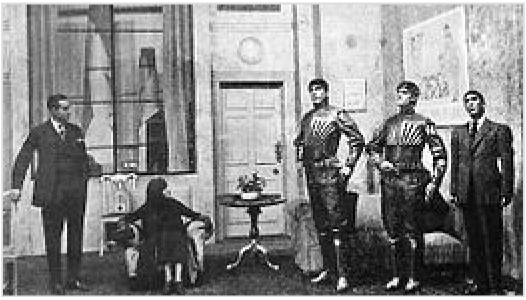
## Transhumanism

Robot designer Hans Moravec, cyberneticist Kevin Warwick, and inventor Ray Kurzweil have predicted that humans and machines will merge in the future into cyborgs that are more capable and powerful than either. This idea, called transhumanism, has roots in Aldous Huxley and Robert Ettinger.[313]

Edward Fredkin argues that "artificial intelligence is the next stage in evolution", an idea first proposed by Samuel Butler's "Darwin among the Machines" as far back as 1863, and expanded upon by George Dyson in his book of the same name in 1998.[314]

# In fiction

Thought-capable artificial beings have appeared as storytelling devices since antiquity,[315] and have been a persistent theme in science fiction.[316]



The word "robot" itself was coined by Karel Čapek in his 1921 play *R.U.R.*, the title standing for "Rossum's Universal Robots".

A common trope in these works began with Mary Shelley's *Frankenstein*, where a human creation becomes a threat to its masters. This includes such works as Arthur C. Clarke's and Stanley Kubrick's *2001: A Space Odyssey* (both 1968), with HAL 9000, the murderous computer in charge of the *Discovery One* spaceship, as well as *The Terminator* (1984) and *The Matrix* (1999). In contrast, the rare loyal robots such as Gort from *The Day the Earth Stood Still* (1951) and Bishop from *Aliens* (1986) are less prominent in popular culture.[317]

Isaac Asimov introduced the Three Laws of Robotics in many books and stories, most notably the "Multivac" series about a super-intelligent computer of the same name. Asimov's laws are often brought up during lay discussions of machine ethics;[318] while almost all artificial intelligence researchers are familiar with Asimov's laws through popular culture, they generally consider the laws useless for many reasons, one of which is their ambiguity.[319]

Several works use AI to force us to confront the fundamental question of what makes us human, showing us artificial beings that have the ability to feel, and thus to suffer. This appears in Karel Čapek's *R.U.R.*, the films *A.I. Artificial Intelligence* and *Ex Machina*, as well as the novel *Do Androids Dream of Electric Sheep?*, by Philip K. Dick. Dick considers the idea that our understanding of human subjectivity is altered by technology created with artificial intelligence.[320]

# See also

- AI effect
- Artificial intelligence detection software – Software to detect AI-generated content
- Behavior selection algorithm – Algorithm that selects actions for intelligent agents
- Business process automation – Technology-enabled automation of complex business processes
- Case-based reasoning – Process of solving new problems based on the solutions of similar past problems
- Emergent algorithm – Algorithm exhibiting emergent behavior
- Female gendering of AI technologies
- Glossary of artificial intelligence – List of definitions of terms and concepts commonly used in the study of artificial intelligence
- Robotic process automation – Form of business process automation technology
- Weak artificial intelligence – Form of artificial intelligence
- Wetware computer – Computer composed of organic material

- Intelligence amplification – Use of information technology to augment human intelligence

# Explanatory notes

a. This list of intelligent traits is based on the topics covered by the major AI textbooks, including: Russell & Norvig (2021), Luger & Stubblefield (2004), Poole, Mackworth & Goebel (1998) and Nilsson (1998)
b. This list of tools is based on the topics covered by the major AI textbooks, including: Russell & Norvig (2021), Luger & Stubblefield (2004), Poole, Mackworth & Goebel (1998) and Nilsson (1998)
c. It is among the reasons that expert systems proved to be inefficient for capturing knowledge.[30][31]
d. "Rational agent" is general term used in economics, philosophy and theoretical artificial intelligence. It can refer to anything that directs its behavior to accomplish goals, such as a person, an animal, a corporation, a nation, or, in the case of AI, a computer program.
e. Alan Turing discussed the centrality of learning as early as 1950, in his classic paper "Computing Machinery and Intelligence".[42] In 1956, at the original Dartmouth AI summer conference, Ray Solomonoff wrote a report on unsupervised probabilistic machine learning: "An Inductive Inference Machine".[43]
f. See AI winter § Machine translation and the ALPAC report of 1966
g. Compared with symbolic logic, formal Bayesian inference is computationally expensive. For inference to be tractable, most observations must be conditionally independent of one another. AdSense uses a Bayesian network with over 300 million edges to learn which ads to serve.[87]
h. Expectation-maximization, one of the most popular algorithms in machine learning, allows clustering in the presence of unknown latent variables.[89]
i. Some form of deep neural networks (without a specific learning algorithm) were described by: Alan Turing (1948);[113] Frank Rosenblatt(1957);[113] Karl Steinbuch and Roger David Joseph (1961).[114] Deep or recurrent networks that learned (or used gradient descent) were developed by: Ernst Ising and Wilhelm Lenz (1925);[115] Oliver Selfridge (1959);[114] Alexey Ivakhnenko and Valentin Lapa (1965);[115] Kaoru Nakano (1977);[116] Shun-Ichi Amari (1972);[116] John Joseph Hopfield (1982).[116] Backpropagation was independently discovered by: Henry J. Kelley (1960);[113] Arthur E. Bryson (1962);[113] Stuart Dreyfus (1962);[113] Arthur E. Bryson and Yu-Chi Ho (1969);[113] Seppo Linnainmaa (1970);[117] Paul Werbos (1974).[113] In fact, backpropagation and gradient descent are straight forward applications of Gottfried Leibniz' chain rule in calculus (1676),[118] and is essentially identical (for one layer) to the method of least squares, developed independently by Johann Carl Friedrich Gauss (1795) and Adrien-Marie Legendre (1805).[119] There are probably many others, yet to be discovered by historians of science.
j. Geoffrey Hinton said, of his work on neural networks in the 1990s, "our labeled datasets were thousands of times too small. [And] our computers were millions of times too slow"[120]
k. Including Jon Kleinberg (Cornell), Sendhil Mullainathan (University of Chicago), Cynthia Chouldechova (Carnegie Mellon) and Sam Corbett-Davis (Stanford)[169]
l. Moritz Hardt (a director at the Max Planck Institute for Intelligent Systems) argues that machine learning "is fundamentally the wrong tool for a lot of domains, where you're trying to design interventions and mechanisms that change the world."[174]
m. When the law was passed in 2018, it still contained a form of this provision.
n. This is the United Nations' definition, and includes things like land mines as well.[187]

o. See table 4; 9% is both the OECD average and the US average.[200]

p. Sometimes called a "robopocalypse".[207]

q. "Electronic brain" was the term used by the press around this time.[247]

r. Daniel Crevier wrote, "the conference is generally recognized as the official birthdate of the new science."[251] Russell and Norvig called the conference "the inception of artificial intelligence."[249]

s. Russell and Norvig wrote "for the next 20 years the field would be dominated by these people and their students."[252]

t. Russell and Norvig wrote "it was astonishing whenever a computer did anything kind of smartish".[253]

u. The programs described are Arthur Samuel's checkers program for the IBM 701, Daniel Bobrow's STUDENT, Newell and Simon's Logic Theorist and Terry Winograd's SHRDLU.

v. Russell and Norvig write: "in almost all cases, these early systems failed on more difficult problems"[254]

w. Embodied approaches to AI[260] were championed by Hans Moravec[261] and Rodney Brooks[262] and went by many names: Nouvelle AI.[262] Developmental robotics,[263]

x. Matteo Wong wrote in The Atlantic: "Whereas for decades, computer-science fields such as natural-language processing, computer vision, and robotics used extremely different methods, now they all use a programming method called "deep learning." As a result, their code and approaches have become more similar, and their models are easier to integrate into one another."[269]

y. Jack Clark wrote in Bloomberg: "After a half-decade of quiet breakthroughs in artificial intelligence, 2015 has been a landmark year. Computers are smarter and learning faster than ever," and noted that the number of software projects that use machine learning at Google increased from a "sporadic usage" in 2012 to more than 2,700 projects in 2015.[271]

z. See Problem of other minds

aa. Nils Nilsson wrote in 1983: "Simply put, there is wide disagreement in the field about what AI is all about."[285]

ab. Daniel Crevier wrote that "time has proven the accuracy and perceptiveness of some of Dreyfus's comments. Had he formulated them less aggressively, constructive actions they suggested might have been taken much earlier."[290]

ac. Searle presented this definition of "Strong AI" in 1999.[300] Searle's original formulation was "The appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states."[301] Strong AI is defined similarly by Russell and Norvig: "Stong AI – the assertion that machines that do so are *actually* thinking (as opposed to *simulating* thinking)."[302]

# References


1. Google (2016).
2. Copeland, J., ed. (2004). *The Essential Turing: the ideas that gave birth to the computer age*. Oxford, England: Clarendon Press. ISBN 0-19-825079-7.

3. Dartmouth workshop:
   - Russell & Norvig (2021, p. 18)
   - McCorduck (2004, pp. 111–136)
   - NRC (1999, pp. 200–201)

   The proposal:
   - McCarthy et al. (1955)
4. Successful programs the 1960s:
   - McCorduck (2004, pp. 243–252)
   - Crevier (1993, pp. 52–107)
   - Moravec (1988, p. 9)
   - Russell & Norvig (2021, pp. 19–21)
5. Funding initiatives in the early 1980s: Fifth Generation Project (Japan), Alvey (UK), Microelectronics and Computer Technology Corporation (US), Strategic Computing Initiative (US):
   - McCorduck (2004, pp. 426–441)
   - Crevier (1993, pp. 161–162, 197–203, 211, 240)
   - Russell & Norvig (2021, p. 23)
   - NRC (1999, pp. 210–211)
   - Newquist (1994, pp. 235–248)
6. First AI Winter, Lighthill report, Mansfield Amendment
   - Crevier (1993, pp. 115–117)
   - Russell & Norvig (2021, pp. 21–22)
   - NRC (1999, pp. 212–213)
   - Howe (1994)
   - Newquist (1994, pp. 189–201)
7. Second AI Winter:
   - Russell & Norvig (2021, p. 24)
   - McCorduck (2004, pp. 430–435)
   - Crevier (1993, pp. 209–210)
   - NRC (1999, pp. 214–216)
   - Newquist (1994, pp. 301–318)
8. Deep learning revolution, AlexNet:
   - Goldman (2022)
   - Russell & Norvig (2021, p. 26)
   - McKinsey (2018)
9. Toews (2023).
10. Frank (2023).

11. Artificial general intelligence:
    - Russell & Norvig (2021, pp. 32–33, 1020–1021)

    Proposal for the modern version:
    - Pennachin & Goertzel (2007)

    Warnings of overspecialization in AI from leading researchers:
    - Nilsson (1995)
    - McCarthy (2007)
    - Beal & Winston (2009)
12. Russell & Norvig (2021, §1.2).
13. Problem solving, puzzle solving, game playing and deduction:
    - Russell & Norvig (2021, chpt. 3–5)
    - Russell & Norvig (2021, chpt. 6) (constraint satisfaction)
    - Poole, Mackworth & Goebel (1998, chpt. 2, 3, 7, 9)
    - Luger & Stubblefield (2004, chpt. 3, 4, 6, 8)
    - Nilsson (1998, chpt. 7–12)
14. Uncertain reasoning:
    - Russell & Norvig (2021, chpt. 12–18)
    - Poole, Mackworth & Goebel (1998, pp. 345–395)
    - Luger & Stubblefield (2004, pp. 333–381)
    - Nilsson (1998, chpt. 7–12)
15. Intractability and efficiency and the combinatorial explosion:
    - Russell & Norvig (2021, p. 21)
16. Psychological evidence of the prevalence sub-symbolic reasoning and knowledge:
    - Kahneman (2011)
    - Dreyfus & Dreyfus (1986)
    - Wason & Shapiro (1966)
    - Kahneman, Slovic & Tversky (1982)
17. Knowledge representation and knowledge engineering:
    - Russell & Norvig (2021, chpt. 10)
    - Poole, Mackworth & Goebel (1998, pp. 23–46, 69–81, 169–233, 235–277, 281–298, 319–345)
    - Luger & Stubblefield (2004, pp. 227–243),
    - Nilsson (1998, chpt. 17.1–17.4, 18)
18. Smoliar & Zhang (1994).
19. Neumann & Möller (2008).
20. Kuperman, Reichley & Bailey (2006).
21. McGarry (2005).
22. Bertini, Del Bimbo & Torniai (2006).
23. Russell & Norvig (2021), pp. 272.

24. Representing categories and relations: Semantic networks, description logics, inheritance (including frames and scripts):
    - Russell & Norvig (2021, §10.2 & 10.5),
    - Poole, Mackworth & Goebel (1998, pp. 174–177),
    - Luger & Stubblefield (2004, pp. 248–258),
    - Nilsson (1998, chpt. 18.3)
25. Representing events and time:Situation calculus, event calculus, fluent calculus (including solving the frame problem):
    - Russell & Norvig (2021, §10.3),
    - Poole, Mackworth & Goebel (1998, pp. 281–298),
    - Nilsson (1998, chpt. 18.2)
26. Causal calculus:
    - Poole, Mackworth & Goebel (1998, pp. 335–337)
27. Representing knowledge about knowledge: Belief calculus, modal logics:
    - Russell & Norvig (2021, §10.4),
    - Poole, Mackworth & Goebel (1998, pp. 275–277)
28. Default reasoning, Frame problem, default logic, non-monotonic logics, circumscription, closed world assumption, abduction:
    - Russell & Norvig (2021, §10.6)
    - Poole, Mackworth & Goebel (1998, pp. 248–256, 323–335)
    - Luger & Stubblefield (2004, pp. 335–363)
    - Nilsson (1998, ~18.3.3)

    (Poole *et al.* places abduction under "default reasoning". Luger *et al.* places this under "uncertain reasoning").
29. Breadth of commonsense knowledge:
    - Lenat & Guha (1989, Introduction)
    - Crevier (1993, pp. 113–114),
    - Moravec (1988, p. 13),
    - Russell & Norvig (2021, pp. 241, 385, 982) (qualification problem)
30. Newquist (1994), p. 296.
31. Crevier (1993), pp. 204–208.
32. Russell & Norvig (2021), p. 528.
33. Automated planning:
    - Russell & Norvig (2021, chpt. 11).
34. Automated decision making, Decision theory:
    - Russell & Norvig (2021, chpt. 16–18).
35. Classical planning:
    - Russell & Norvig (2021, Section 11.2).

36. Sensorless or "conformant" planning, contingent planning, replanning (a.k.a online planning):
    - Russell & Norvig (2021, Section 11.5).
37. Uncertain preferences:
    - Russell & Norvig (2021, Section 16.7)

    Inverse reinforcement learning:
    - Russell & Norvig (2021, Section 22.6)
38. Information value theory:
    - Russell & Norvig (2021, Section 16.6).
39. Markov decision process:
    - Russell & Norvig (2021, chpt. 17).
40. Game theory and multi-agent decision theory:
    - Russell & Norvig (2021, chpt. 18).
41. Learning:
    - Russell & Norvig (2021, chpt. 19–22)
    - Poole, Mackworth & Goebel (1998, pp. 397–438)
    - Luger & Stubblefield (2004, pp. 385–542)
    - Nilsson (1998, chpt. 3.3, 10.3, 17.5, 20)
42. Turing (1950).
43. Solomonoff (1956).
44. Unsupervised learning:
    - Russell & Norvig (2021, pp. 653) (definition)
    - Russell & Norvig (2021, pp. 738–740) (cluster analysis)
    - Russell & Norvig (2021, pp. 846–860) (word embedding)
45. Supervised learning:
    - Russell & Norvig (2021, §19.2) (Definition)
    - Russell & Norvig (2021, Chpt. 19–20) (Techniques)
46. Reinforcement learning:
    - Russell & Norvig (2021, chpt. 22)
    - Luger & Stubblefield (2004, pp. 442–449)
47. Transfer learning:
    - Russell & Norvig (2021, pp. 281)
    - The Economist (2016)
48. "Artificial Intelligence (AI): What Is AI and How Does It Work? | Built In" (https://builtin.com/artificial-intelligence). *builtin.com*. Retrieved 30 October 2023.
49. Computational learning theory:
    - Russell & Norvig (2021, pp. 672–674)
    - Jordan & Mitchell (2015)

50. Natural language processing (NLP):
   - Russell & Norvig (2021, chpt. 23–24)
   - Poole, Mackworth & Goebel (1998, pp. 91–104)
   - Luger & Stubblefield (2004, pp. 591–632)
51. Subproblems of NLP:
   - Russell & Norvig (2021, pp. 849–850)
52. Russell & Norvig (2021), p. 856–858.
53. Dickson (2022).
54. Modern statistical and deep learning approaches to NLP:
   - Russell & Norvig (2021, chpt. 24)
   - Cambria & White (2014)
55. Vincent (2019).
56. Russell & Norvig (2021), p. 875–878.
57. Bushwick (2023).
58. Computer vision:
   - Russell & Norvig (2021, chpt. 25)
   - Nilsson (1998, chpt. 6)
59. Russell & Norvig (2021), pp. 849–850.
60. Russell & Norvig (2021), pp. 895–899.
61. Russell & Norvig (2021), pp. 899–901.
62. Russell & Norvig (2021), pp. 931–938.
63. MIT AIL (2014).
64. Affective computing:
   - Thro (1993)
   - Edelson (1991)
   - Tao & Tan (2005)
   - Scassellati (2002)
65. Waddell (2018).
66. Poria et al. (2017).
67. Search algorithms:
   - Russell & Norvig (2021, Chpt. 3–5)
   - Poole, Mackworth & Goebel (1998, pp. 113–163)
   - Luger & Stubblefield (2004, pp. 79–164, 193–219)
   - Nilsson (1998, chpt. 7–12)
68. State space search:
   - Russell & Norvig (2021, chpt. 3)
69. Russell & Norvig (2021), §11.2.

70. Uninformed searches (breadth first search, depth-first search and general state space search):
    - Russell & Norvig (2021, §3.4)
    - Poole, Mackworth & Goebel (1998, pp. 113–132)
    - Luger & Stubblefield (2004, pp. 79–121)
    - Nilsson (1998, chpt. 8)
71. Heuristic or informed searches (e.g., greedy best first and A*):
    - Russell & Norvig (2021, s§3.5)
    - Poole, Mackworth & Goebel (1998, pp. 132–147)
    - Poole & Mackworth (2017, §3.6)
    - Luger & Stubblefield (2004, pp. 133–150)
72. Adversarial search:
    - Russell & Norvig (2021, chpt. 5)
73. Local or "optimization" search:
    - Russell & Norvig (2021, chpt. 4)
74. Singh Chauhan, Nagesh (18 December 2020). "Optimization Algorithms in Neural Networks" (https://www.kdnuggets.com/optimization-algorithms-in-neural-networks). *KDnuggets*. Retrieved 13 January 2024.
75. Evolutionary computation:
    - Russell & Norvig (2021, §4.1.2)
76. Merkle & Middendorf (2013).
77. Logic:
    - Russell & Norvig (2021, chpt. 6–9)
    - Luger & Stubblefield (2004, pp. 35–77)
    - Nilsson (1998, chpt. 13–16)
78. Propositional logic:
    - Russell & Norvig (2021, chpt. 6)
    - Luger & Stubblefield (2004, pp. 45–50)
    - Nilsson (1998, chpt. 13)
79. First-order logic and features such as equality:
    - Russell & Norvig (2021, chpt. 7)
    - Poole, Mackworth & Goebel (1998, pp. 268–275),
    - Luger & Stubblefield (2004, pp. 50–62),
    - Nilsson (1998, chpt. 15)
80. Logical inference:
    - Russell & Norvig (2021, chpt. 10)
81. Russell & Norvig (2021), §8.3.1.
82. Resolution and unification:
    - Russell & Norvig (2021, §7.5.2, §9.2, §9.5)

83. Forward chaining, backward chaining, Horn clauses, and logical deduction as search:
    - Russell & Norvig (2021, §9.3, §9.4)
    - Poole, Mackworth & Goebel (1998, pp. ~46–52)
    - Luger & Stubblefield (2004, pp. 62–73)
    - Nilsson (1998, chpt. 4.2, 7.2)
84. Fuzzy logic:
    - Russell & Norvig (2021, pp. 214, 255, 459)
    - Scientific American (1999)
85. Stochastic methods for uncertain reasoning:
    - Russell & Norvig (2021, Chpt. 12–18 and 20),
    - Poole, Mackworth & Goebel (1998, pp. 345–395),
    - Luger & Stubblefield (2004, pp. 165–191, 333–381),
    - Nilsson (1998, chpt. 19)
86. Bayesian networks:
    - Russell & Norvig (2021, §12.5–12.6, §13.4–13.5, §14.3–14.5, §16.5, §20.2 -20.3),
    - Poole, Mackworth & Goebel (1998, pp. 361–381),
    - Luger & Stubblefield (2004, pp. ~182–190, ≈363–379),
    - Nilsson (1998, chpt. 19.3–4)
87. Domingos (2015), chapter 6.
88. Bayesian inference algorithm:
    - Russell & Norvig (2021, §13.3–13.5),
    - Poole, Mackworth & Goebel (1998, pp. 361–381),
    - Luger & Stubblefield (2004, pp. ~363–379),
    - Nilsson (1998, chpt. 19.4 & 7)
89. Domingos (2015), p. 210.
90. Bayesian learning and the expectation-maximization algorithm:
    - Russell & Norvig (2021, Chpt. 20),
    - Poole, Mackworth & Goebel (1998, pp. 424–433),
    - Nilsson (1998, chpt. 20)
    - Domingos (2015, p. 210)
91. Bayesian decision theory and Bayesian decision networks:
    - Russell & Norvig (2021, §16.5)

92. Stochastic temporal models:
    - Russell & Norvig (2021, Chpt. 14)

    Hidden Markov model:
    - Russell & Norvig (2021, §14.3)

    Kalman filters:
    - Russell & Norvig (2021, §14.4)

    Dynamic Bayesian networks:
    - Russell & Norvig (2021, §14.5)
93. decision theory and decision analysis:
    - Russell & Norvig (2021, Chpt. 16–18),
    - Poole, Mackworth & Goebel (1998, pp. 381–394)
94. Information value theory:
    - Russell & Norvig (2021, §16.6)
95. Markov decision processes and dynamic decision networks:
    - Russell & Norvig (2021, chpt. 17)
96. Game theory and mechanism design:
    - Russell & Norvig (2021, chpt. 18)
97. Statistical learning methods and classifiers:
    - Russell & Norvig (2021, chpt. 20),
98. Decision trees:
    - Russell & Norvig (2021, §19.3)
    - Domingos (2015, p. 88)
99. Non-parameteric learning models such as K-nearest neighbor and support vector machines:
    - Russell & Norvig (2021, §19.7)
    - Domingos (2015, p. 187) (k-nearest neighbor)
    - Domingos (2015, p. 88) (kernel methods)
100. Domingos (2015), p. 152.
101. Naive Bayes classifier:
    - Russell & Norvig (2021, §12.6)
    - Domingos (2015, p. 152)
102. Neural networks:
    - Russell & Norvig (2021, Chpt. 21),
    - Domingos (2015, Chapter 4)
103. Gradient calculation in computational graphs, backpropagation, automatic differentiation:
    - Russell & Norvig (2021, §21.2),
    - Luger & Stubblefield (2004, pp. 467–474),
    - Nilsson (1998, chpt. 3.3)

104. Universal approximation theorem:
    - Russell & Norvig (2021, p. 752)

    The theorem:
    - Cybenko (1988)
    - Hornik, Stinchcombe & White (1989)
105. Feedforward neural networks:
    - Russell & Norvig (2021, §21.1)
106. Recurrent neural networks:
    - Russell & Norvig (2021, §21.6)
107. Perceptrons:
    - Russell & Norvig (2021, pp. 21, 22, 683, 22)
108. Deep learning:
    - Russell & Norvig (2021, Chpt. 21)
    - Goodfellow, Bengio & Courville (2016)
    - Hinton *et al.* (2016)
    - Schmidhuber (2015)
109. Convolutional neural networks:
    - Russell & Norvig (2021, §21.3)
110. Deng & Yu (2014), pp. 199–200.
111. Ciresan, Meier & Schmidhuber (2012).
112. Russell & Norvig (2021), p. 751.
113. Russell & Norvig (2021), p. 785.
114. Schmidhuber (2022), §5.
115. Schmidhuber (2022), §6.
116. Schmidhuber (2022), §7.
117. Schmidhuber (2022), §8.
118. Schmidhuber (2022), §2.
119. Schmidhuber (2022), §3.
120. Quoted in Christian (2020, p. 22)
121. Smith (2023).
122. "Explained: Generative AI" (https://news.mit.edu/2023/explained-generative-ai-1109). 9 November 2023.
123. "AI Writing and Content Creation Tools" (https://mitsloanedtech.mit.edu/ai/tools/writing/). MIT Sloan Teaching & Learning Technologies. Retrieved 25 December 2023.
124. Marmouyet (2023).
125. Kobielus (2019).
126. Davenport, T; Kalakota, R (June 2019). "The potential for artificial intelligence in healthcare" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6616181). *Future Healthc J*. **6** (2): 94–98. doi:10.7861/futurehosp.6-2-94 (https://doi.org/10.7861%2Ffuturehosp.6-2-94). PMC 6616181 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6616181). PMID 31363513 (https://pubmed.ncbi.nlm.nih.gov/31363513).

127. Bax, Monique; Thorpe, Jordan; Romanov, Valentin (December 2023). "The future of personalized cardiovascular medicine demands 3D and 4D printing, stem cells, and artificial intelligence" (https://doi.org/10.3389%2Ffsens.2023.1294721). *Frontiers in Sensors*. **4**. doi:10.3389/fsens.2023.1294721 (https://doi.org/10.3389%2Ffsens.2023.1294721). ISSN 2673-5067 (https://www.worldcat.org/issn/2673-5067).
128. Jumper, J; Evans, R; Pritzel, A (2021). "Highly accurate protein structure prediction with AlphaFold" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8371605). *Nature*. **596** (7873): 583–589. Bibcode:2021Natur.596..583J (https://ui.adsabs.harvard.edu/abs/2021Natur.596..583J). doi:10.1038/s41586-021-03819-2 (https://doi.org/10.1038%2Fs41586-021-03819-2). PMC 8371605 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8371605). PMID 34265844 (https://pubmed.ncbi.nlm.nih.gov/34265844).
129. "AI discovers new class of antibiotics to kill drug-resistant bacteria" (https://www.newscientist.com/article/2409706-ai-discovers-new-class-of-antibiotics-to-kill-drug-resistant-bacteria/). 20 December 2023.
130. Grant, Eugene F.; Lardner, Rex (25 July 1952). "The Talk of the Town – It" (https://www.newyorker.com/magazine/1952/08/02/it). *The New Yorker*. ISSN 0028-792X (https://www.worldcat.org/issn/0028-792X). Retrieved 28 January 2024.
131. Anderson, Mark Robert (11 May 2017). "Twenty years on from Deep Blue vs Kasparov: how a chess match started the big data revolution" (https://theconversation.com/twenty-years-on-from-deep-blue-vs-kasparov-how-a-chess-match-started-the-big-data-revolution-76882). *The Conversation*. Retrieved 28 January 2024.
132. Markoff, John (16 February 2011). "Computer Wins on 'Jeopardy!': Trivial, It's Not" (https://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html). *The New York Times*. ISSN 0362-4331 (https://www.worldcat.org/issn/0362-4331). Retrieved 28 January 2024.
133. Byford, Sam (27 May 2017). "AlphaGo retires from competitive Go after defeating world number one 3-0" (https://www.theverge.com/2017/5/27/15704088/alphago-ke-jie-game-3-result-retires-future). *The Verge*. Retrieved 28 January 2024.
134. Brown, Noam; Sandholm, Tuomas (30 August 2019). "Superhuman AI for multiplayer poker" (https://www.science.org/doi/10.1126/science.aay2400). *Science*. **365** (6456): 885–890. doi:10.1126/science.aay2400 (https://doi.org/10.1126%2Fscience.aay2400). ISSN 0036-8075 (https://www.worldcat.org/issn/0036-8075).
135. "MuZero: Mastering Go, chess, shogi and Atari without rules" (https://deepmind.google/discover/blog/muzero-mastering-go-chess-shogi-and-atari-without-rules/). *Google DeepMind*. 23 December 2020. Retrieved 28 January 2024.
136. Sample, Ian (30 October 2019). "AI becomes grandmaster in 'fiendishly complex' StarCraft II" (https://www.theguardian.com/technology/2019/oct/30/ai-becomes-grandmaster-in-fiendishly-complex-starcraft-ii). *The Guardian*. ISSN 0261-3077 (https://www.worldcat.org/issn/0261-3077). Retrieved 28 January 2024.
137. Wurman, P.R.; Barrett, S.; Kawamoto, K. (2022). "Outracing champion Gran Turismo drivers with deep reinforcement learning". *Nature 602*. **602** (7896): 223–228. doi:10.1038/s41586-021-04357-7 (https://doi.org/10.1038%2Fs41586-021-04357-7).
138. Congressional Research Service (2019). *Artificial Intelligence and National Security* (https://fas.org/sgp/crs/natsec/R45178.pdf) (PDF). Washington, DC: Congressional Research Service.PD-notice
139. Slyusar, Vadym (2019). "Artificial intelligence as the basis of future control networks" (https://www.researchgate.net/publication/334573170). *ResearchGate*. doi:10.13140/RG.2.2.30247.50087 (https://doi.org/10.13140%2FRG.2.2.30247.50087).

140. Knight, Will. "The US and 30 Other Nations Agree to Set Guardrails for Military AI" (https://www.wired.com/story/the-us-and-30-other-nations-agree-to-set-guardrails-for-military-ai/). *Wired*. ISSN 1059-1028 (https://www.worldcat.org/issn/1059-1028). Retrieved 24 January 2024.
141. Marcelline, Marco (27 May 2023). "ChatGPT: Most Americans Know About It, But Few Actually Use the AI Chatbot" (https://www.pcmag.com/news/few-americans-have-actually-tried-chatgpt-despite-most-knowing-about-it). *PCMag*. Retrieved 28 January 2024.
142. Lu, Donna (31 March 2023). "Misinformation, mistakes and the Pope in a puffer: what rapidly evolving AI can – and can't – do" (https://www.theguardian.com/technology/2023/apr/01/misinformation-mistakes-and-the-pope-in-a-puffer-what-rapidly-evolving-ai-can-and-cant-do). *The Guardian*. ISSN 0261-3077 (https://www.worldcat.org/issn/0261-3077). Retrieved 28 January 2024.
143. Hurst, Luke (23 May 2023). "How a fake image of a Pentagon explosion shared on Twitter caused a real dip on Wall Street" (https://www.euronews.com/next/2023/05/23/fake-news-about-an-explosion-at-the-pentagon-spreads-on-verified-accounts-on-twitter). *euronews*. Retrieved 28 January 2024.
144. Ransbotham, Sam; Kiron, David; Gerbert, Philipp; Reeves, Martin (6 September 2017). "Reshaping Business With Artificial Intelligence" (https://sloanreview.mit.edu/projects/reshaping-business-with-artificial-intelligence/). *MIT Sloan Management Review*. Archived (https://web.archive.org/web/20240213070751/https://sloanreview.mit.edu/projects/reshaping-business-with-artificial-intelligence/) from the original on 13 February 2024.
145. Simonite (2016).
146. Russell & Norvig (2021), p. 987.
147. Laskowski (2023).
148. GAO (2022).
149. Valinsky (2019).
150. Russell & Norvig (2021), p. 991.
151. Russell & Norvig (2021), p. 991–992.
152. Christian (2020), p. 63.
153. Vincent (2022).
154. Reisner (2023).
155. Alter & Harris (2023).
156. Nicas (2018).
157. Rainie, Lee; Keeter, Scott; Perrin, Andrew (22 July 2019). "Trust and Distrust in America" (https://www.pewresearch.org/politics/2019/07/22/trust-and-distrust-in-america/). *Pew Research Center*. Archived (https://web.archive.org/web/20240222000601/https://www.pewresearch.org/politics/2019/07/22/trust-and-distrust-in-america/) from the original on 22 February 2024.
158. Williams (2023).
159. Taylor & Hern (2023).
160. Rose (2023).
161. CNA (2019).
162. Goffrey (2008), p. 17.
163. Berdahl et al. (2023); Goffrey (2008, p. 17); Rose (2023); Russell & Norvig (2021, p. 995)
164. Algorithmic bias and Fairness (machine learning):
    - Russell & Norvig (2021, section 27.3.3)
    - Christian (2020, Fairness)
165. Christian (2020), p. 25.

166. Russell & Norvig (2021), p. 995.
167. Grant & Hill (2023).
168. Larson & Angwin (2016).
169. Christian (2020), p. 67–70.
170. Christian (2020, pp. 67–70); Russell & Norvig (2021, pp. 993–994)
171. Russell & Norvig (2021, p. 995); Lipartito (2011, p. 36); Goodman & Flaxman (2017, p. 6); Christian (2020, pp. 39–40, 65)
172. Quoted in Christian (2020, p. 65).
173. Russell & Norvig (2021, p. 994); Christian (2020, pp. 40, 80–81)
174. Quoted in Christian (2020, p. 80)
175. Dockrill (2022).
176. Sample (2017).
177. "Black Box AI" (https://www.techopedia.com/definition/34940/black-box-ai). 16 June 2023.
178. Christian (2020), p. 110.
179. Christian (2020), pp. 88–91.
180. Christian (2020, p. 83); Russell & Norvig (2021, p. 997)
181. Christian (2020), p. 91.
182. Christian (2020), p. 83.
183. Verma (2021).
184. Rothman (2020).
185. Christian (2020), p. 105-108.
186. Christian (2020), pp. 108–112.
187. Russell & Norvig (2021), p. 989.
188. Robitzski (2018); Sainato (2015)
189. Russell & Norvig (2021), p. 987-990.
190. Russell & Norvig (2021), p. 988.
191. Harari (2018).
192. Buckley, Chris; Mozur, Paul (22 May 2019). "How China Uses High-Tech Surveillance to Subdue Minorities" (https://www.nytimes.com/2019/05/22/world/asia/china-surveillance-xinjiang.html). *The New York Times*.
193. "Security lapse exposed a Chinese smart city surveillance system" (https://web.archive.org/web/20210307203740/https://consent.yahoo.com/v2/collectConsent?sessionId=3_cc-session_c8562b93-9863-4915-8523-6c7b930a3efc). 3 May 2019. Archived from the original (https://social.techcrunch.com/2019/05/03/china-smart-city-exposed/) on 7 March 2021. Retrieved 14 September 2020.
194. "AI traffic signals to be installed in Bengaluru soon" (https://nextbigwhat.com/ai-traffic-signals-to-be-installed-in-bengaluru-soon/). *NextBigWhat*. 24 September 2019. Retrieved 1 October 2019.
195. Urbina et al. (2022).
196. Tarnoff, Ben (4 August 2023). "Lessons from Eliza". *The Guardian Weekly*. pp. 34–9.
197. E McGaughey, 'Will Robots Automate Your Job Away? Full Employment, Basic Income, and Economic Democracy' (2022) 51(3) Industrial Law Journal 511–559 (https://academic.oup.com/ilj/article/51/3/511/6321008) Archived (https://web.archive.org/web/20230527163045/https://academic.oup.com/ilj/article/51/3/511/6321008) 27 May 2023 at the Wayback Machine
198. Ford & Colvin (2015);McGaughey (2022)
199. IGM Chicago (2017).

200. Arntz, Gregory & Zierahn (2016), p. 33.
201. Lohr (2017); Frey & Osborne (2017); Arntz, Gregory & Zierahn (2016, p. 33)
202. Morgenstern (2015).
203. Mahdawi (2017); Thompson (2014)
204. Zhou, Viola (11 April 2023). "AI is already taking video game illustrators' jobs in China" (https://restofworld.org/2023/ai-image-china-video-game-layoffs/). *Rest of World*. Retrieved 17 August 2023.
205. Carter, Justin (11 April 2023). "China's game art industry reportedly decimated by growing AI use" (https://www.gamedeveloper.com/art/china-s-game-art-industry-reportedly-decimated-ai-art-use). *Game Developer*. Retrieved 17 August 2023.
206. Cellan-Jones (2014).
207. Russell & Norvig 2021, p. 1001.
208. Bostrom (2014).
209. Russell (2019).
210. Bostrom (2014); Müller & Bostrom (2014); Bostrom (2015).
211. Harari (2023).
212. Müller & Bostrom (2014).
213. Leaders' concerns about the existential risks of AI around 2015:
    - Rawlinson (2015)
    - Holley (2015)
    - Gibbs (2014)
    - Sainato (2015)
214. Arguments that AI is not an imminent risk:
    - Brooks (2014)
    - Geist (2015)
    - Madrigal (2015)
    - Lee (2014)
215. Christian (2020), pp. 67, 73.
216. Valance (2023).
217. Taylor, Josh (7 May 2023). "Rise of artificial intelligence is inevitable but should not be feared, 'father of AI' says" (https://www.theguardian.com/technology/2023/may/07/rise-of-artificial-intelligence-is-inevitable-but-should-not-be-feared-father-of-ai-says). *The Guardian*. Retrieved 26 May 2023.
218. Colton, Emma (7 May 2023). " 'Father of AI' says tech fears misplaced: 'You cannot stop it' " (https://www.foxnews.com/tech/father-ai-jurgen-schmidhuber-says-tech-fears-misplaced-cannot-stop). *Fox News*. Retrieved 26 May 2023.
219. Jones, Hessie (23 May 2023). "Juergen Schmidhuber, Renowned 'Father Of Modern AI,' Says His Life's Work Won't Lead To Dystopia" (https://www.forbes.com/sites/hessiejones/2023/05/23/juergen-schmidhuber-renowned-father-of-modern-ai-says-his-lifes-work-wont-lead-to-dystopia/). *Forbes*. Retrieved 26 May 2023.
220. McMorrow, Ryan (19 December 2023). "Andrew Ng: 'Do we think the world is better off with more or less intelligence?' " (https://www.ft.com/content/2dc07f9e-d2a9-4d98-b746-b051f9352be3). *Financial Times*. Retrieved 30 December 2023.
221. Levy, Steven (22 December 2023). "How Not to Be Stupid About AI, With Yann LeCun" (https://www.wired.com/story/artificial-intelligence-meta-yann-lecun-interview/). *Wired*. Retrieved 30 December 2023.

222. Yudkowsky (2008).
223. Anderson & Anderson (2011).
224. AAAI (2014).
225. Wallach (2010).
226. Russell (2019), p. 173.
227. Alan Turing Institute (2019). "Understanding artificial intelligence ethics and safety" (https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf) (PDF).
228. Alan Turing Institute (2023). "AI Ethics and Governance in Practice" (https://www.turing.ac.uk/sites/default/files/2023-12/aieg-ati-ai-ethics-an-intro_1.pdf) (PDF).
229. Floridi, Luciano; Cowls, Josh (23 June 2019). "A Unified Framework of Five Principles for AI in Society" (https://hdsr.mitpress.mit.edu/pub/l0jsh9d1). *Harvard Data Science Review*. **1** (1). doi:10.1162/99608f92.8cd550d1 (https://doi.org/10.1162%2F99608f92.8cd550d1). S2CID 198775713 (https://api.semanticscholar.org/CorpusID:198775713).
230. Buruk, Banu; Ekmekci, Perihan Elif; Arda, Berna (1 September 2020). "A critical perspective on guidelines for responsible and trustworthy artificial intelligence" (https://doi.org/10.1007/s11019-020-09948-1). *Medicine, Health Care and Philosophy*. **23** (3): 387–399. doi:10.1007/s11019-020-09948-1 (https://doi.org/10.1007%2Fs11019-020-09948-1). ISSN 1572-8633 (https://www.worldcat.org/issn/1572-8633). PMID 32236794 (https://pubmed.ncbi.nlm.nih.gov/32236794). S2CID 214766800 (https://api.semanticscholar.org/CorpusID:214766800).
231. Kamila, Manoj Kumar; Jasrotia, Sahil Singh (1 January 2023). "Ethical issues in the development of artificial intelligence: recognizing the risks" (https://doi.org/10.1108/IJOES-05-2023-0107). *International Journal of Ethics and Systems*. ahead-of-print (ahead-of-print). doi:10.1108/IJOES-05-2023-0107 (https://doi.org/10.1108%2FIJOES-05-2023-0107). ISSN 2514-9369 (https://www.worldcat.org/issn/2514-9369). S2CID 259614124 (https://api.semanticscholar.org/CorpusID:259614124).
232. Regulation of AI to mitigate risks:
    - Berryhill et al. (2019)
    - Barfield & Pagallo (2018)
    - Iphofen & Kritikos (2019)
    - Wirtz, Weyerer & Geyer (2018)
    - Buiten (2019)
233. Law Library of Congress (U.S.). Global Legal Research Directorate (2019).
234. Vincent (2023).
235. Stanford University (2023).
236. UNESCO (2021).
237. Kissinger (2021).
238. Altman, Brockman & Sutskever (2023).
239. VOA News (25 October 2023). "UN Announces Advisory Body on Artificial Intelligence" (https://www.voanews.com/a/un-announces-advisory-body-on-artificial-intelligence-/7328732.html).
240. Edwards (2023).
241. Kasperowicz (2023).
242. Fox News (2023).
243. Milmo, Dan (3 November 2023). "Hope or Horror? The great AI debate dividing its pioneers". *The Guardian Weekly*. pp. 10–12.

244. "The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023" (https://web.archive.org/web/20231101123904/https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023). *GOV.UK*. 1 November 2023. Archived from the original (https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023) on 1 November 2023. Retrieved 2 November 2023.
245. "Countries agree to safe and responsible development of frontier AI in landmark Bletchley Declaration" (https://www.gov.uk/government/news/countries-agree-to-safe-and-responsible-development-of-frontier-ai-in-landmark-bletchley-declaration). *GOV.UK* (Press release). Archived (https://web.archive.org/web/20231101115016/https://www.gov.uk/government/news/countries-agree-to-safe-and-responsible-development-of-frontier-ai-in-landmark-bletchley-declaration) from the original on 1 November 2023. Retrieved 1 November 2023.
246. Berlinski (2000).
247. "Google books ngram" (https://books.google.com/ngrams/graph?content=electronic+brain&year_start=1930&year_end=2019&corpus=en-2019&smoothing=3).
248. AI's immediate precursors:
    - McCorduck (2004, pp. 51–107)
    - Crevier (1993, pp. 27–32)
    - Russell & Norvig (2021, pp. 8–17)
    - Moravec (1988, p. 3)
249. Russell & Norvig (2021), p. 17.
250. See "A Brief History of Computing" (http://www.alanturing.net/turing_archive/pages/Reference%20Articles/BriefHistofComp.html) at AlanTuring.net.
251. Crevier (1993), pp. 47–49.
252. Russell & Norvig (2003), p. 17.
253. Russell & Norvig (2003), p. 18.
254. Russell & Norvig (2021), p. 21.
255. Lighthill (1973).
256. Russell & Norvig (2021), p. 22.
257. Expert systems:
    - Russell & Norvig (2021, pp. 23, 292)
    - Luger & Stubblefield (2004, pp. 227–331)
    - Nilsson (1998, chpt. 17.4)
    - McCorduck (2004, pp. 327–335, 434–435)
    - Crevier (1993, pp. 145–62, 197–203)
    - Newquist (1994, pp. 155–183)
258. Russell & Norvig (2021), p. 24.
259. Nilsson (1998), p. 7.
260. McCorduck (2004), pp. 454–462.
261. Moravec (1988).
262. Brooks (1990).

263. Developmental robotics:

- Weng et al. (2001)
- Lungarella et al. (2003)
- Asada et al. (2009)
- Oudeyer (2010)

264. Russell & Norvig (2021), p. 25.
265. 
- Crevier (1993, pp. 214–215)
- Russell & Norvig (2021, pp. 24, 26)

266. Russell & Norvig (2021), p. 26.
267. Formal and narrow methods adopted in the 1990s:

- Russell & Norvig (2021, pp. 24–26)
- McCorduck (2004, pp. 486–487)

268. AI widely used in the late 1990s:

- Kurzweil (2005, p. 265)
- NRC (1999, pp. 216–222)
- Newquist (1994, pp. 189–201)

269. Wong (2023).
270. Moore's Law and AI:

- Russell & Norvig (2021, pp. 14, 27)

271. Clark (2015b).
272. Big data:

- Russell & Norvig (2021, p. 26)

273. "Intellectual Property and Frontier Technologies" (https://www.wipo.int/about-ip/en/frontier_technologies/). *WIPO*. Archived (https://web.archive.org/web/20220402064804/https://www.wipo.int/about-ip/en/frontier_technologies/) from the original on 2 April 2022. Retrieved 30 March 2022.
274. DiFeliciantonio (2023).
275. Goswami (2023).
276. Turing (1950), p. 1.
277. Turing's original publication of the Turing test in "Computing machinery and intelligence":

- Turing (1950)

Historical influence and philosophical implications:

- Haugeland (1985, pp. 6–9)
- Crevier (1993, p. 24)
- McCorduck (2004, pp. 70–71)
- Russell & Norvig (2021, pp. 2 and 984)

278. Turing (1950), Under "The Argument from Consciousness".
279. Russell & Norvig (2021), chpt. 2.
280. Russell & Norvig (2021), p. 3.
281. Maker (2006).
282. McCarthy (1999).

283. Minsky (1986).
284. "What Is Artificial Intelligence (AI)?" (https://cloud.google.com/learn/what-is-artificial-intelligence). *Google Cloud Platform*. Archived (https://web.archive.org/web/20230731114802/https://cloud.google.com/learn/what-is-artificial-intelligence) from the original on 31 July 2023. Retrieved 16 October 2023.
285. Nilsson (1983), p. 10.
286. Haugeland (1985), pp. 112–117.
287. Physical symbol system hypothesis:
    - Newell & Simon (1976, p. 116)

    Historical significance:
    - McCorduck (2004, p. 153)
    - Russell & Norvig (2021, p. 19)
288. Moravec's paradox:
    - Moravec (1988, pp. 15–16)
    - Minsky (1986, p. 29)
    - Pinker (2007, pp. 190–91)
289. Dreyfus' critique of AI:
    - Dreyfus (1972)
    - Dreyfus & Dreyfus (1986)

    Historical significance and philosophical implications:
    - Crevier (1993, pp. 120–132)
    - McCorduck (2004, pp. 211–239)
    - Russell & Norvig (2021, pp. 981–982)
    - Fearn (2007, Chpt. 3)
290. Crevier (1993), p. 125.
291. Langley (2011).
292. Katz (2012).
293. Neats vs. scruffies, the historic debate:
    - McCorduck (2004, pp. 421–424, 486–489)
    - Crevier (1993, p. 168)
    - Nilsson (1983, pp. 10–11)
    - Russell & Norvig (2021, p. 24)

    A classic example of the "scruffy" approach to intelligence:
    - Minsky (1986)

    A modern example of neat AI and its aspirations in the 21st century:
    - Domingos (2015)
294. Pennachin & Goertzel (2007).
295. Roberts (2016).
296. Russell & Norvig (2021), p. 986.
297. Chalmers (1995).
298. Dennett (1991).
299. Horst (2005).

300. Searle (1999).
301. Searle (1980), p. 1.
302. Russell & Norvig (2021), p. 9817.
303. Searle's Chinese room argument:

- Searle (1980). Searle's original presentation of the thought experiment.
- Searle (1999).

Discussion:

- Russell & Norvig (2021, pp. 985)
- McCorduck (2004, pp. 443–445)
- Crevier (1993, pp. 269–271)

304. Leith, Sam (7 July 2022). "Nick Bostrom: How can we be certain a machine isn't conscious?" (https://www.spectator.co.uk/article/nick-bostrom-how-can-we-be-certain-a-machine-isnt-conscious/). *The Spectator*. Retrieved 23 February 2024.
305. Thomson, Jonny (31 October 2022). "Why don't robots have rights?" (https://bigthink.com/thinking/why-dont-robots-have-rights/). *Big Think*. Retrieved 23 February 2024.
306. Kateman, Brian (24 July 2023). "AI Should Be Terrified of Humans" (https://time.com/6296234/ai-should-be-terrified-of-humans/). *TIME*. Retrieved 23 February 2024.
307. Wong, Jeff (10 July 2023). "What leaders need to know about robot rights" (https://www.fastcompany.com/90920769/what-leaders-need-to-know-about-robot-rights). *Fast Company*.
308. Hern, Alex (12 January 2017). "Give robots 'personhood' status, EU committee argues" (https://www.theguardian.com/technology/2017/jan/12/give-robots-personhood-status-eu-committee-argues). *The Guardian*. ISSN 0261-3077 (https://www.worldcat.org/issn/0261-3077). Retrieved 23 February 2024.
309. Dovey, Dana (14 April 2018). "Experts Don't Think Robots Should Have Rights" (https://www.newsweek.com/robots-human-rights-electronic-persons-humans-versus-machines-886075). *Newsweek*. Retrieved 23 February 2024.
310. Cuddy, Alice (13 April 2018). "Robot rights violate human rights, experts warn EU" (https://www.euronews.com/2018/04/13/robot-rights-violate-human-rights-experts-warn-eu). *euronews*. Retrieved 23 February 2024.
311. The Intelligence explosion and technological singularity:

- Russell & Norvig (2021, pp. 1004–1005)
- Omohundro (2008)
- Kurzweil (2005)

I. J. Good's "intelligence explosion"

- Good (1965)

Vernor Vinge's "singularity"

- Vinge (1993)

312. Russell & Norvig (2021), p. 1005.
313. Transhumanism:

- Moravec (1988)
- Kurzweil (2005)
- Russell & Norvig (2021, p. 1005)

314. AI as evolution:

    - Edward Fredkin is quoted in McCorduck (2004, p. 401)
    - Butler (1863)
    - Dyson (1998)

315. AI in myth:

    - McCorduck (2004, pp. 4–5)

316. McCorduck (2004), pp. 340–400.
317. Buttazzo (2001).
318. Anderson (2008).
319. McCauley (2007).
320. Galvan (1997).

## AI textbooks

The two most widely used textbooks in 2023. (See the Open Syllabus (https://explorer.opensyllabus.org/result/field?id=Computer+Science)).


- Russell, Stuart J.; Norvig, Peter. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Hoboken: Pearson. ISBN 978-0134610993. LCCN 20190474 (https://lccn.loc.gov/20190474).
- Rich, Elaine; Knight, Kevin; Nair, Shivashankar B (2010). *Artificial Intelligence* (3rd ed.). New Delhi: Tata McGraw Hill India. ISBN 978-0070087705.


These were the four of the most widely used AI textbooks in 2008:

- Luger, George; Stubblefield, William (2004). *Artificial Intelligence: Structures and Strategies for Complex Problem Solving* (https://archive.org/details/artificialintell0000luge) (5th ed.). Benjamin/Cummings. ISBN 978-0-8053-4780-7. Archived (https://web.archive.org/web/20200726220613/https://archive.org/details/artificialintell0000luge) from the original on 26 July 2020. Retrieved 17 December 2019.
- Nilsson, Nils (1998). *Artificial Intelligence: A New Synthesis* (https://archive.org/details/artificialintell0000nils). Morgan Kaufmann. ISBN 978-1-55860-467-4. Archived (https://web.archive.org/web/20200726131654/https://archive.org/details/artificialintell0000nils) from the original on 26 July 2020. Retrieved 18 November 2019.
- Russell, Stuart J.; Norvig, Peter (2003), *Artificial Intelligence: A Modern Approach* (http://aima.cs.berkeley.edu/) (2nd ed.), Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13-790395-2.
- Poole, David; Mackworth, Alan; Goebel, Randy (1998). *Computational Intelligence: A Logical Approach* (https://archive.org/details/computationalint00pool). New York: Oxford University Press. ISBN 978-0-19-510270-3. Archived (https://web.archive.org/web/20200726131436/https://archive.org/details/computationalint00pool) from the original on 26 July 2020. Retrieved 22 August 2020.

Later editions.

- Poole, David; Mackworth, Alan (2017). *Artificial Intelligence: Foundations of Computational Agents* (http://artint.info/index.html) (2nd ed.). Cambridge University Press. ISBN 978-1-107-19539-4. Archived (https://web.archive.org/web/20171207013855/http://artint.info/index.html) from the original on 7 December 2017. Retrieved 6 December 2017.


## History of AI


- Crevier, Daniel (1993). *AI: The Tumultuous Search for Artificial Intelligence*. New York, NY: BasicBooks. ISBN 0-465-02997-3..
- McCorduck, Pamela (2004), *Machines Who Think* (2nd ed.), Natick, MA: A. K. Peters, Ltd., ISBN 1-56881-205-1.
- Newquist, H. P. (1994). *The Brain Makers: Genius, Ego, And Greed In The Quest For Machines That Think*. New York: Macmillan/SAMS. ISBN 978-0-672-30412-5.


## Other sources


- AI & ML in Fusion (https://suli.pppl.gov/2023/course/Rea-PPPL-SULI2023.pdf)
- AI & ML in Fusion, video lecture (https://drive.google.com/file/d/1npCTrJ8XJn20ZGDA_DfMpANuQZFMzKPh/view?usp=drive_link) Archived (https://web.archive.org/web/20230702164332/https://drive.google.com/file/d/1npCTrJ8XJn20ZGDA_DfMpANuQZFMzKPh/view?usp=drive_link) 2 July 2023 at the Wayback Machine
- "AlphaGo – Google DeepMind" (https://wildoftech.com/alphago-google-deepmind). Archived (https://web.archive.org/web/20160310191926/https://wildoftech.com/alphago-google-deepmind/) from the original on 10 March 2016.
- Alter, Alexandra; Harris, Elizabeth A. (20 September 2023), "Franzen, Grisham and Other Prominent Authors Sue OpenAI" (https://www.nytimes.com/2023/09/20/books/authors-openai-lawsuit-chatgpt-copyright.html?campaign_id=2&emc=edit_th_20230921&instance_id=103259&nl=todaysheadlines&regi_id=62816440&segment_id=145288&user_id=ad24f3545dae0ec44284a38bb4a88f1d), *The New York Times*

- Altman, Sam; Brockman, Greg; Sutskever, Ilya (22 May 2023). "Governance of Superintelligence" (https://openai.com/blog/governance-of-superintelligence). *openai.com*. Archived (https://web.archive.org/web/20230527061619/https://openai.com/blog/governance-of-superintelligence) from the original on 27 May 2023. Retrieved 27 May 2023.
- Anderson, Susan Leigh (2008). "Asimov's "three laws of robotics" and machine metaethics". *AI & Society*. **22** (4): 477–493. doi:10.1007/s00146-007-0094-5 (https://doi.org/10.1007%2Fs00146-007-0094-5). S2CID 1809459 (https://api.semanticscholar.org/CorpusID:1809459).
- Anderson, Michael; Anderson, Susan Leigh (2011). *Machine Ethics*. Cambridge University Press.
- Arntz, Melanie; Gregory, Terry; Zierahn, Ulrich (2016), "The risk of automation for jobs in OECD countries: A comparative analysis", *OECD Social, Employment, and Migration Working Papers 189*
- Asada, M.; Hosoda, K.; Kuniyoshi, Y.; Ishiguro, H.; Inui, T.; Yoshikawa, Y.; Ogino, M.; Yoshida, C. (2009). "Cognitive developmental robotics: a survey". *IEEE Transactions on Autonomous Mental Development*. **1** (1): 12–34. doi:10.1109/tamd.2009.2021702 (https://doi.org/10.1109%2Ftamd.2009.2021702). S2CID 10168773 (https://api.semanticscholar.org/CorpusID:10168773).
- "Ask the AI experts: What's driving today's progress in AI?" (https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/ask-the-ai-experts-whats-driving-todays-progress-in-ai). *McKinsey & Company*. Archived (https://web.archive.org/web/20180413190018/https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/ask-the-ai-experts-whats-driving-todays-progress-in-ai) from the original on 13 April 2018. Retrieved 13 April 2018.
- Barfield, Woodrow; Pagallo, Ugo (2018). *Research handbook on the law of artificial intelligence*. Cheltenham, UK: Edward Elgar Publishing. ISBN 978-1-78643-904-8. OCLC 1039480085 (https://www.worldcat.org/oclc/1039480085).
- Beal, J.; Winston, Patrick (2009), "The New Frontier of Human-Level Artificial Intelligence", *IEEE Intelligent Systems*, **24**: 21–24, doi:10.1109/MIS.2009.75 (https://doi.org/10.1109%2FMIS.2009.75), hdl:1721.1/52357 (https://hdl.handle.net/1721.1%2F52357), S2CID 32437713 (https://api.semanticscholar.org/CorpusID:32437713)
- Berdahl, Carl Thomas; Baker, Lawrence; Mann, Sean; Osoba, Osonde; Girosi, Federico (7 February 2023). "Strategies to Improve the Impact of Artificial Intelligence on Health Equity: Scoping Review" (https://ai.jmir.org/2023/1/e42936). *JMIR AI*. **2**: e42936. doi:10.2196/42936 (https://doi.org/10.2196%2F42936). ISSN 2817-1705 (https://www.worldcat.org/issn/2817-1705). S2CID 256681439 (https://api.semanticscholar.org/CorpusID:256681439). Archived (https://web.archive.org/web/20230221145255/https://ai.jmir.org/2023/1/e42936/) from the original on 21 February 2023. Retrieved 21 February 2023.
- Berlinski, David (2000). *The Advent of the Algorithm* (https://archive.org/details/adventofalgorith0000berl). Harcourt Books. ISBN 978-0-15-601391-8. OCLC 46890682 (https://www.worldcat.org/oclc/46890682). Archived (https://web.archive.org/web/20200726215744/https://archive.org/details/adventofalgorith0000berl) from the original on 26 July 2020. Retrieved 22 August 2020.
- Berryhill, Jamie; Heang, Kévin Kok; Clogher, Rob; McBride, Keegan (2019). *Hello, World: Artificial Intelligence and its Use in the Public Sector* (https://oecd-opsi.org/wp-content/uploads/2019/11/AI-Report-Online.pdf) (PDF). Paris: OECD Observatory of Public Sector Innovation. Archived (https://web.archive.org/web/20191220021331/https://oecd-opsi.org/wp-content/uploads/2019/11/AI-Report-Online.pdf) (PDF) from the original on 20 December 2019. Retrieved 9 August 2020.

- Bertini, M; Del Bimbo, A; Torniai, C (2006). "Automatic annotation and semantic retrieval of video sequences using multimedia ontologies". *MM '06 Proceedings of the 14th ACM international conference on Multimedia*. 14th ACM international conference on Multimedia. Santa Barbara: ACM. pp. 679–682.
- Bostrom, Nick (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- Bostrom, Nick (2015). "What happens when our computers get smarter than we are?" (https://www.ted.com/talks/nick_bostrom_what_happens_when_our_computers_get_smarter_than_we_are/transcript). TED (conference). Archived (https://web.archive.org/web/20200725005719/https://www.ted.com/talks/nick_bostrom_what_happens_when_our_computers_get_smarter_than_we_are/transcript) from the original on 25 July 2020. Retrieved 30 January 2020.
- Brooks, Rodney (10 November 2014). "artificial intelligence is a tool, not a threat" (https://web.archive.org/web/20141112130954/http://www.rethinkrobotics.com/artificial-intelligence-tool-threat/). Archived from the original (http://www.rethinkrobotics.com/artificial-intelligence-tool-threat/) on 12 November 2014.
- Brooks, Rodney (1990). "Elephants Don't Play Chess" (http://people.csail.mit.edu/brooks/papers/elephants.pdf) (PDF). *Robotics and Autonomous Systems*. **6** (1–2): 3–15. CiteSeerX 10.1.1.588.7539 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.588.7539). doi:10.1016/S0921-8890(05)80025-9 (https://doi.org/10.1016%2FS0921-8890%2805%2980025-9). Archived (https://web.archive.org/web/20070809020912/http://people.csail.mit.edu/brooks/papers/elephants.pdf) (PDF) from the original on 9 August 2007.
- Buiten, Miriam C (2019). "Towards Intelligent Regulation of Artificial Intelligence" (https://doi.org/10.1017%2Ferr.2019.8). *European Journal of Risk Regulation*. **10** (1): 41–59. doi:10.1017/err.2019.8 (https://doi.org/10.1017%2Ferr.2019.8). ISSN 1867-299X (https://www.worldcat.org/issn/1867-299X).
- Bushwick, Sophie (16 March 2023), "What the New GPT-4 AI Can Do" (https://www.scientificamerican.com/article/what-the-new-gpt-4-ai-can-do/), *Scientific American*
- Butler, Samuel (13 June 1863). "Darwin among the Machines" (https://nzetc.victoria.ac.nz/tm/scholarly/tei-ButFir-t1-g1-t1-g1-t4-body.html). Letters to the Editor. *The Press*. Christchurch, New Zealand. Archived (https://web.archive.org/web/20080919172551/http://www.nzetc.org/tm/scholarly/tei-ButFir-t1-g1-t1-g1-t4-body.html) from the original on 19 September 2008. Retrieved 16 October 2014 – via Victoria University of Wellington.
- Buttazzo, G. (July 2001). "Artificial consciousness: Utopia or real possibility?". *Computer*. **34** (7): 24–30. doi:10.1109/2.933500 (https://doi.org/10.1109%2F2.933500).
- Cambria, Erik; White, Bebo (May 2014). "Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]". *IEEE Computational Intelligence Magazine*. **9** (2): 48–57. doi:10.1109/MCI.2014.2307227 (https://doi.org/10.1109%2FMCI.2014.2307227). S2CID 206451986 (https://api.semanticscholar.org/CorpusID:206451986).
- Cellan-Jones, Rory (2 December 2014). "Stephen Hawking warns artificial intelligence could end mankind" (https://www.bbc.com/news/technology-30290540). *BBC News*. Archived (https://web.archive.org/web/20151030054329/http://www.bbc.com/news/technology-30290540) from the original on 30 October 2015. Retrieved 30 October 2015.
- Chalmers, David (1995). "Facing up to the problem of consciousness" (http://www.imprint.co.uk/chalmers.html). *Journal of Consciousness Studies*. **2** (3): 200–219. Archived (https://web.archive.org/web/20050308163649/http://www.imprint.co.uk/chalmers.html) from the original on 8 March 2005. Retrieved 11 October 2018.
- Christian, Brian (2020). *The Alignment Problem: Machine learning and human values*. W. W. Norton & Company. ISBN 978-0-393-86833-3. OCLC 1233266753 (https://www.worldcat.org/oclc/1233266753).

- Ciresan, D.; Meier, U.; Schmidhuber, J. (2012). "Multi-column deep neural networks for image classification". *2012 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3642–3649. arXiv:1202.2745 (https://arxiv.org/abs/1202.2745). doi:10.1109/cvpr.2012.6248110 (https://doi.org/10.1109%2Fcvpr.2012.6248110). ISBN 978-1-4673-1228-8. S2CID 2161592 (https://api.semanticscholar.org/CorpusID:2161592).
- Clark, Jack (2015b). "Why 2015 Was a Breakthrough Year in Artificial Intelligence" (https://www.bloomberg.com/news/articles/2015-12-08/why-2015-was-a-breakthrough-year-in-artificial-intelligence). *Bloomberg.com*. Archived (https://web.archive.org/web/20161123053855/https://www.bloomberg.com/news/articles/2015-12-08/why-2015-was-a-breakthrough-year-in-artificial-intelligence) from the original on 23 November 2016. Retrieved 23 November 2016.
- CNA (12 January 2019). "Commentary: Bad news. Artificial intelligence is biased" (https://www.channelnewsasia.com/news/commentary/artificial-intelligence-big-data-bias-hiring-loans-key-challenge-11097374). *CNA*. Archived (https://web.archive.org/web/20190112104421/https://www.channelnewsasia.com/news/commentary/artificial-intelligence-big-data-bias-hiring-loans-key-challenge-11097374) from the original on 12 January 2019. Retrieved 19 June 2020.
- Cybenko, G. (1988). Continuous valued neural networks with two hidden layers are sufficient (Report). Department of Computer Science, Tufts University.
- Deng, L.; Yu, D. (2014). "Deep Learning: Methods and Applications" (http://research.microsoft.com/pubs/209355/DeepLearning-NowPublishing-Vol7-SIG-039.pdf) (PDF). *Foundations and Trends in Signal Processing*. **7** (3–4): 1–199. doi:10.1561/2000000039 (https://doi.org/10.1561%2F2000000039). Archived (https://web.archive.org/web/20160314152112/http://research.microsoft.com/pubs/209355/DeepLearning-NowPublishing-Vol7-SIG-039.pdf) (PDF) from the original on 14 March 2016. Retrieved 18 October 2014.
- Dennett, Daniel (1991). *Consciousness Explained*. The Penguin Press. ISBN 978-0-7139-9037-9.
- DiFeliciantonio, Chase (3 April 2023). "AI has already changed the world. This report shows how" (https://www.sfchronicle.com/tech/article/ai-artificial-intelligence-report-stanford-17869558.php). *San Francisco Chronicle*. Archived (https://web.archive.org/web/20230619015309/https://www.sfchronicle.com/tech/article/ai-artificial-intelligence-report-stanford-17869558.php) from the original on 19 June 2023. Retrieved 19 June 2023.
- Dickson, Ben (2 May 2022). "Machine learning: What is the transformer architecture?" (https://bdtechtalks.com/2022/05/02/what-is-the-transformer/). *TechTalks*. Retrieved 22 November 2023.
- Dockrill, Peter (27 June 2022), "Robots With Flawed AI Make Sexist And Racist Decisions, Experiment Shows" (https://web.archive.org/web/20220627225827/https://www.sciencealert.com/robots-with-flawed-ai-make-sexist-racist-and-toxic-decisions-experiment-shows), *Science Alert*, archived from the original (https://www.sciencealert.com/robots-with-flawed-ai-make-sexist-racist-and-toxic-decisions-experiment-shows) on 27 June 2022
- Domingos, Pedro (2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Basic Books. ISBN 978-0465065707.
- Dreyfus, Hubert (1972). *What Computers Can't Do*. New York: MIT Press. ISBN 978-0-06-011082-6.
- Dreyfus, Hubert; Dreyfus, Stuart (1986). *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer* (https://archive.org/details/mindovermachinep00drey). Oxford: Blackwell. ISBN 978-0-02-908060-3. Archived (https://web.archive.org/web/20200726131414/https://archive.org/details/mindovermachinep00drey) from the original on 26 July 2020. Retrieved 22 August 2020.

- Dyson, George (1998). *Darwin among the Machines* (https://archive.org/details/darwinamongmachi00dyso). Allan Lane Science. ISBN 978-0-7382-0030-9. Archived (https://web.archive.org/web/20200726131443/https://archive.org/details/darwinamongmachi00dyso) from the original on 26 July 2020. Retrieved 22 August 2020.
- Edelson, Edward (1991). *The Nervous System* (https://archive.org/details/nervoussystem0000edel). New York: Chelsea House. ISBN 978-0-7910-0464-7. Archived (https://web.archive.org/web/20200726131758/https://archive.org/details/nervoussystem0000edel) from the original on 26 July 2020. Retrieved 18 November 2019.
- Edwards, Benj (17 May 2023). "Poll: AI poses risk to humanity, according to majority of Americans" (https://arstechnica.com/information-technology/2023/05/poll-61-of-americans-say-ai-threatens-humanitys-future/). *Ars Technica*. Archived (https://web.archive.org/web/20230619013608/https://arstechnica.com/information-technology/2023/05/poll-61-of-americans-say-ai-threatens-humanitys-future/) from the original on 19 June 2023. Retrieved 19 June 2023.
- Evans, Woody (2015). "Posthuman Rights: Dimensions of Transhuman Worlds" (https://doi.org/10.5209%2Frev_TK.2015.v12.n2.49072). *Teknokultura*. **12** (2). doi:10.5209/rev_TK.2015.v12.n2.49072 (https://doi.org/10.5209%2Frev_TK.2015.v12.n2.49072).
- Fearn, Nicholas (2007). *The Latest Answers to the Oldest Questions: A Philosophical Adventure with the World's Greatest Thinkers*. New York: Grove Press. ISBN 978-0-8021-1839-4.
- Ford, Martin; Colvin, Geoff (6 September 2015). "Will robots create more jobs than they destroy?" (https://www.theguardian.com/technology/2015/sep/06/will-robots-create-destroy-jobs). *The Guardian*. Archived (https://web.archive.org/web/20180616204119/https://www.theguardian.com/technology/2015/sep/06/will-robots-create-destroy-jobs) from the original on 16 June 2018. Retrieved 13 January 2018.
- Fox News (2023). "Fox News Poll" (https://static.foxnews.com/foxnews.com/content/uploads/2023/05/Fox_April-21-24-2023_Complete_National_Topline_May-1-Release.pdf) (PDF). Fox News. Archived (https://web.archive.org/web/20230512082712/https://static.foxnews.com/foxnews.com/content/uploads/2023/05/Fox_April-21-24-2023_Complete_National_Topline_May-1-Release.pdf) (PDF) from the original on 12 May 2023. Retrieved 19 June 2023.
- Frank, Michael (22 September 2023). "US Leadership in Artificial Intelligence Can Shape the 21st Century Global Order" (https://thediplomat.com/2023/09/us-leadership-in-artificial-intelligence-can-shape-the-21st-century-global-order/). *The Diplomat*. Retrieved 8 December 2023. "Instead, the United States has developed a new area of dominance that the rest of the world views with a mixture of awe, envy, and resentment: artificial intelligence... From AI models and research to cloud computing and venture capital, U.S. companies, universities, and research labs – and their affiliates in allied countries – appear to have an enormous lead in both developing cutting-edge AI and commercializing it. The value of U.S. venture capital investments in AI start-ups exceeds that of the rest of the world combined."
- Frey, Carl Benedikt; Osborne, Michael A (1 January 2017). "The future of employment: How susceptible are jobs to computerisation?". *Technological Forecasting and Social Change*. **114**: 254–280. CiteSeerX 10.1.1.395.416 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.395.416). doi:10.1016/j.techfore.2016.08.019 (https://doi.org/10.1016%2Fj.techfore.2016.08.019). ISSN 0040-1625 (https://www.worldcat.org/issn/0040-1625).
- "From not working to neural networking" (https://www.economist.com/news/special-report/21700756-artificial-intelligence-boom-based-old-idea-modern-twist-not). *The Economist*. 2016. Archived (https://web.archive.org/web/20161231203934/https://www.economist.com/news/special-report/21700756-artificial-intelligence-boom-based-old-idea-modern-twist-not) from the original on 31 December 2016. Retrieved 26 April 2018.

- Galvan, Jill (1 January 1997). "Entering the Posthuman Collective in Philip K. Dick's "Do Androids Dream of Electric Sheep?" ". *Science Fiction Studies*. **24** (3): 413–429. JSTOR 4240644 (https://www.jstor.org/stable/4240644).
- Geist, Edward Moore (9 August 2015). "Is artificial intelligence really an existential threat to humanity?" (http://thebulletin.org/artificial-intelligence-really-existential-threat-humanity8577). *Bulletin of the Atomic Scientists*. Archived (https://web.archive.org/web/20151030054330/http://thebulletin.org/artificial-intelligence-really-existential-threat-humanity8577) from the original on 30 October 2015. Retrieved 30 October 2015.
- Gertner, Jon (18 July 2023). "Wikipedia's Moment of Truth – Can the online encyclopedia help teach A.I. chatbots to get their facts right — without destroying itself in the process? + comment" (https://www.nytimes.com/2023/07/18/magazine/wikipedia-ai-chatgpt.html). *The New York Times*. Archived (https://web.archive.org/web/20230718233916/https://www.nytimes.com/2023/07/18/magazine/wikipedia-ai-chatgpt.html#permid=126389255) from the original on 18 July 2023. Retrieved 19 July 2023.
- Gibbs, Samuel (27 October 2014). "Elon Musk: artificial intelligence is our biggest existential threat" (https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat). *The Guardian*. Archived (https://web.archive.org/web/20151030054330/http://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat) from the original on 30 October 2015. Retrieved 30 October 2015.
- Goffrey, Andrew (2008). "Algorithm". In Fuller, Matthew (ed.). *Software studies: a lexicon* (https://archive.org/details/softwarestudiesl00full_007). Cambridge, Mass.: MIT Press. pp. 15 (https://archive.org/details/softwarestudiesl00full_007/page/n29)–20. ISBN 978-1-4356-4787-9.
- Goldman, Sharon (14 September 2022). "10 years later, deep learning 'revolution' rages on, say AI pioneers Hinton, LeCun and Li" (https://venturebeat.com/ai/10-years-on-ai-pioneers-hinton-lecun-li-say-deep-learning-revolution-will-continue/). *VentureBeat*. Retrieved 8 December 2023.
- Good, I. J. (1965), *Speculations Concerning the First Ultraintelligent Machine* (https://exhibits.stanford.edu/feigenbaum/catalog/gz727rg3869)
- Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron (2016), *Deep Learning* (https://web.archive.org/web/20160416111010/http://www.deeplearningbook.org/), MIT Press., archived from the original (http://www.deeplearningbook.org/) on 16 April 2016, retrieved 12 November 2017
- Goodman, Bryce; Flaxman, Seth (2017). "EU regulations on algorithmic decision-making and a 'right to explanation' ". *AI Magazine*. **38** (3): 50. arXiv:1606.08813 (https://arxiv.org/abs/1606.08813). doi:10.1609/aimag.v38i3.2741 (https://doi.org/10.1609%2Faimag.v38i3.2741). S2CID 7373959 (https://api.semanticscholar.org/CorpusID:7373959).
- Government Accountability Office (13 September 2022). Consumer Data: Increasing Use Poses Risks to Privacy (https://www.gao.gov/products/gao-22-106096). *gao.gov* (Report).
- Grant, Nico; Hill, Kashmir (22 May 2023). "Google's Photo App Still Can't Find Gorillas. And Neither Can Apple's" (https://www.nytimes.com/2023/05/22/technology/ai-photo-labels-google-apple.html). *The New York Times*.
- Goswami, Rohan (5 April 2023). "Here's where the A.I. jobs are" (https://www.cnbc.com/2023/04/05/ai-jobs-see-the-state-by-state-data-from-a-stanford-study.html). *CNBC*. Archived (https://web.archive.org/web/20230619015309/https://www.cnbc.com/2023/04/05/ai-jobs-see-the-state-by-state-data-from-a-stanford-study.html) from the original on 19 June 2023. Retrieved 19 June 2023.

- Harari, Yuval Noah (October 2018). "Why Technology Favors Tyranny" (https://www.theatlantic.com/magazine/archive/2018/10/yuval-noah-harari-technology-tyranny/568330). *The Atlantic*. Archived (https://web.archive.org/web/20210925221449/https://www.theatlantic.com/magazine/archive/2018/10/yuval-noah-harari-technology-tyranny/568330/) from the original on 25 September 2021. Retrieved 23 September 2021.
- Harari, Yuval Noah (2023). "AI and the future of humanity" (https://www.youtube.com/watch?v=LWiM-LuRe6w). *YouTube*.
- Haugeland, John (1985). *Artificial Intelligence: The Very Idea*. Cambridge, Mass.: MIT Press. ISBN 978-0-262-08153-5.
- Henderson, Mark (24 April 2007). "Human rights for robots? We're getting carried away" (http://www.thetimes.co.uk/tto/technology/article1966391.ece). *The Times Online*. London. Archived (https://web.archive.org/web/20140531104850/http://www.thetimes.co.uk/tto/technology/article1966391.ece) from the original on 31 May 2014. Retrieved 31 May 2014.
- Hinton, G.; Deng, L.; Yu, D.; Dahl, G.; Mohamed, A.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.; Kingsbury, B. (2012). "Deep Neural Networks for Acoustic Modeling in Speech Recognition – The shared views of four research groups". *IEEE Signal Processing Magazine*. **29** (6): 82–97. Bibcode:2012ISPM...29...82H (https://ui.adsabs.harvard.edu/abs/2012ISPM...29...82H). doi:10.1109/msp.2012.2205597 (https://doi.org/10.1109%2Fmsp.2012.2205597). S2CID 206485943 (https://api.semanticscholar.org/CorpusID:206485943).
- Holley, Peter (28 January 2015). "Bill Gates on dangers of artificial intelligence: 'I don't understand why some people are not concerned' " (https://www.washingtonpost.com/news/the-switch/wp/2015/01/28/bill-gates-on-dangers-of-artificial-intelligence-dont-understand-why-some-people-are-not-concerned/). *The Washington Post*. ISSN 0190-8286 (https://www.worldcat.org/issn/0190-8286). Archived (https://web.archive.org/web/20151030054330/https://www.washingtonpost.com/news/the-switch/wp/2015/01/28/bill-gates-on-dangers-of-artificial-intelligence-dont-understand-why-some-people-are-not-concerned/) from the original on 30 October 2015. Retrieved 30 October 2015.
- Hornik, Kurt; Stinchcombe, Maxwell; White, Halbert (1989). *Multilayer Feedforward Networks are Universal Approximators* (http://cognitivemedium.com/magic_paper/assets/Hornik.pdf) (PDF). *Neural Networks*. Vol. 2. Pergamon Press. pp. 359–366.
- Horst, Steven (2005). "The Computational Theory of Mind" (http://plato.stanford.edu/entries/computational-mind). *The Stanford Encyclopedia of Philosophy*. Archived (https://web.archive.org/web/20160306083748/http://plato.stanford.edu/entries/computational-mind/) from the original on 6 March 2016. Retrieved 7 March 2016.
- Howe, J. (November 1994). "Artificial Intelligence at Edinburgh University: a Perspective" (http://www.inf.ed.ac.uk/about/AIhistory.html). Archived (https://web.archive.org/web/20070515072641/http://www.inf.ed.ac.uk/about/AIhistory.html) from the original on 15 May 2007. Retrieved 30 August 2007.
- IGM Chicago (30 June 2017). "Robots and Artificial Intelligence" (http://www.igmchicago.org/surveys/robots-and-artificial-intelligence). *www.igmchicago.org*. Archived (https://web.archive.org/web/20190501114826/http://www.igmchicago.org/surveys/robots-and-artificial-intelligence) from the original on 1 May 2019. Retrieved 3 July 2019.
- Iphofen, Ron; Kritikos, Mihalis (3 January 2019). "Regulating artificial intelligence and robotics: ethics by design in a digital society". *Contemporary Social Science*. **16** (2): 170–184. doi:10.1080/21582041.2018.1563803 (https://doi.org/10.1080%2F21582041.2018.1563803). ISSN 2158-2041 (https://www.worldcat.org/issn/2158-2041). S2CID 59298502 (https://api.semanticscholar.org/CorpusID:59298502).

- Jordan, M. I.; Mitchell, T. M. (16 July 2015). "Machine learning: Trends, perspectives, and prospects". *Science*. **349** (6245): 255–260. Bibcode:2015Sci...349..255J (https://ui.adsabs.harvard.edu/abs/2015Sci...349..255J). doi:10.1126/science.aaa8415 (https://doi.org/10.1126%2Fscience.aaa8415). PMID 26185243 (https://pubmed.ncbi.nlm.nih.gov/26185243). S2CID 677218 (https://api.semanticscholar.org/CorpusID:677218).
- Kahneman, Daniel (2011). *Thinking, Fast and Slow* (https://books.google.com/books?id=ZuKTvERuPG8C). Macmillan. ISBN 978-1-4299-6935-2. Archived (https://web.archive.org/web/20230315191803/https://books.google.com/books?id=ZuKTvERuPG8C) from the original on 15 March 2023. Retrieved 8 April 2012.
- Kahneman, Daniel; Slovic, D.; Tversky, Amos (1982). "Judgment under uncertainty: Heuristics and biases". *Science*. New York: Cambridge University Press. **185** (4157): 1124–1131. Bibcode:1974Sci...185.1124T (https://ui.adsabs.harvard.edu/abs/1974Sci...185.1124T). doi:10.1126/science.185.4157.1124 (https://doi.org/10.1126%2Fscience.185.4157.1124). ISBN 978-0-521-28414-1. PMID 17835457 (https://pubmed.ncbi.nlm.nih.gov/17835457). S2CID 143452957 (https://api.semanticscholar.org/CorpusID:143452957).
- Kasperowicz, Peter (1 May 2023). "Regulate AI? GOP much more skeptical than Dems that government can do it right: poll" (https://www.foxnews.com/politics/regulate-ai-gop-much-more-skeptical-than-dems-that-the-government-can-do-it-right-poll). *Fox News*. Archived (https://web.archive.org/web/20230619013616/https://www.foxnews.com/politics/regulate-ai-gop-much-more-skeptical-than-dems-that-the-government-can-do-it-right-poll) from the original on 19 June 2023. Retrieved 19 June 2023.
- Katz, Yarden (1 November 2012). "Noam Chomsky on Where Artificial Intelligence Went Wrong" (https://www.theatlantic.com/technology/archive/2012/11/noam-chomsky-on-where-artificial-intelligence-went-wrong/261637/?single_page=true). *The Atlantic*. Archived (https://web.archive.org/web/20190228154403/https://www.theatlantic.com/technology/archive/2012/11/noam-chomsky-on-where-artificial-intelligence-went-wrong/261637/?single_page=true) from the original on 28 February 2019. Retrieved 26 October 2014.
- "Kismet" (http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html). MIT Artificial Intelligence Laboratory, Humanoid Robotics Group. Archived (https://web.archive.org/web/20141017040432/http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html) from the original on 17 October 2014. Retrieved 25 October 2014.
- Kissinger, Henry (1 November 2021). "The Challenge of Being Human in the Age of AI" (https://www.wsj.com/articles/being-human-artifical-intelligence-ai-chess-antibiotic-philosophy-ethics-bill-of-rights-11635795271). *The Wall Street Journal*. Archived (https://web.archive.org/web/20211104012825/https://www.wsj.com/articles/being-human-artifical-intelligence-ai-chess-antibiotic-philosophy-ethics-bill-of-rights-11635795271) from the original on 4 November 2021. Retrieved 4 November 2021.
- Kobielus, James (27 November 2019). "GPUs Continue to Dominate the AI Accelerator Market for Now" (https://www.informationweek.com/ai-or-machine-learning/gpus-continue-to-dominate-the-ai-accelerator-market-for-now). *InformationWeek*. Archived (https://web.archive.org/web/20211019031104/https://www.informationweek.com/ai-or-machine-learning/gpus-continue-to-dominate-the-ai-accelerator-market-for-now) from the original on 19 October 2021. Retrieved 11 June 2020.
- Kuperman, G. J.; Reichley, R. M.; Bailey, T. C. (1 July 2006). "Using Commercial Knowledge Bases for Clinical Decision Support: Opportunities, Hurdles, and Recommendations" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1513681). *Journal of the American Medical Informatics Association*. **13** (4): 369–371. doi:10.1197/jamia.M2055 (https://doi.org/10.1197%2Fjamia.M2055). PMC 1513681 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1513681). PMID 16622160 (https://pubmed.ncbi.nlm.nih.gov/16622160).
- Kurzweil, Ray (2005). *The Singularity is Near*. Penguin Books. ISBN 978-0-670-03384-3.

- Langley, Pat (2011). "The changing science of machine learning" (https://doi.org/10.1007%2Fs10994-011-5242-y). *Machine Learning*. **82** (3): 275–279. doi:10.1007/s10994-011-5242-y (https://doi.org/10.1007%2Fs10994-011-5242-y).
- Larson, Jeff; Angwin, Julia (23 May 2016). "How We Analyzed the COMPAS Recidivism Algorithm" (https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm). *ProPublica*. Archived (https://web.archive.org/web/20190429190950/https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm) from the original on 29 April 2019. Retrieved 19 June 2020.
- Laskowski, Nicole (November 2023). "What is Artificial Intelligence and How Does AI Work? TechTarget" (https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence). *Enterprise AI*. Retrieved 30 October 2023.
- Law Library of Congress (U.S.). Global Legal Research Directorate, issuing body. (2019). *Regulation of artificial intelligence in selected jurisdictions*. LCCN 2019668143 (https://lccn.loc.gov/2019668143). OCLC 1110727808 (https://www.worldcat.org/oclc/1110727808).
- Lee, Timothy B. (22 August 2014). "Will artificial intelligence destroy humanity? Here are 5 reasons not to worry" (https://www.vox.com/2014/8/22/6043635/5-reasons-we-shouldnt-worry-about-super-intelligent-computers-taking). *Vox*. Archived (https://web.archive.org/web/20151030092203/http://www.vox.com/2014/8/22/6043635/5-reasons-we-shouldnt-worry-about-super-intelligent-computers-taking) from the original on 30 October 2015. Retrieved 30 October 2015.
- Lenat, Douglas; Guha, R. V. (1989). *Building Large Knowledge-Based Systems*. Addison-Wesley. ISBN 978-0-201-51752-1.
- Lighthill, James (1973). "Artificial Intelligence: A General Survey". *Artificial Intelligence: a paper symposium*. Science Research Council.
- Lipartito, Kenneth (6 January 2011), *The Narrative and the Algorithm: Genres of Credit Reporting from the Nineteenth Century to Today* (https://mpra.ub.uni-muenchen.de/28142/1/MPRA_paper_28142.pdf) (PDF) (Unpublished manuscript), doi:10.2139/ssrn.1736283 (https://doi.org/10.2139%2Fssrn.1736283), S2CID 166742927 (https://api.semanticscholar.org/CorpusID:166742927), archived (https://ghostarchive.org/archive/20221009/https://mpra.ub.uni-muenchen.de/28142/1/MPRA_paper_28142.pdf) (PDF) from the original on 9 October 2022
- Lohr, Steve (2017). "Robots Will Take Jobs, but Not as Fast as Some Fear, New Report Says" (https://www.nytimes.com/2017/01/12/technology/robots-will-take-jobs-but-not-as-fast-as-some-fear-new-report-says.html). *The New York Times*. Archived (https://web.archive.org/web/20180114073704/https://www.nytimes.com/2017/01/12/technology/robots-will-take-jobs-but-not-as-fast-as-some-fear-new-report-says.html) from the original on 14 January 2018. Retrieved 13 January 2018.
- Lungarella, M.; Metta, G.; Pfeifer, R.; Sandini, G. (2003). "Developmental robotics: a survey". *Connection Science*. **15** (4): 151–190. CiteSeerX 10.1.1.83.7615 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.83.7615). doi:10.1080/09540090310001655110 (https://doi.org/10.1080%2F09540090310001655110). S2CID 1452734 (https://api.semanticscholar.org/CorpusID:1452734).
- "Machine Ethics" (https://web.archive.org/web/20141129044821/http://www.aaai.org/Library/Symposia/Fall/fs05-06). *aaai.org*. Archived from the original (http://www.aaai.org/Library/Symposia/Fall/fs05-06) on 29 November 2014.
- Madrigal, Alexis C. (27 February 2015). "The case against killer robots, from a guy actually working on artificial intelligence" (https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots). *Fusion.net*. Archived (https://web.archive.org/web/20160204175716/http://fusion.net/story/54583/the-case-against-killer-robots-from-a-guy-actually-building-ai/) from the original on 4 February 2016. Retrieved 31 January 2016.

- Mahdawi, Arwa (26 June 2017). "What jobs will still be around in 20 years? Read this to prepare your future" (https://www.theguardian.com/us-news/2017/jun/26/jobs-future-automation-robots-skills-creative-health). *The Guardian*. Archived (https://web.archive.org/web/20180114021804/https://www.theguardian.com/us-news/2017/jun/26/jobs-future-automation-robots-skills-creative-health) from the original on 14 January 2018. Retrieved 13 January 2018.
- Maker, Meg Houston (2006). "AI@50: AI Past, Present, Future" (https://web.archive.org/web/20070103222615/http://www.engagingexperience.com/2006/07/ai50_ai_past_pr.html). Dartmouth College. Archived from the original (http://www.engagingexperience.com/2006/07/ai50_ai_past_pr.html) on 3 January 2007. Retrieved 16 October 2008.
- Marmouyet, Françoise (15 December 2023). "Google's Gemini: is the new AI model really better than ChatGPT?" (https://theconversation.com/googles-gemini-is-the-new-ai-model-really-better-than-chatgpt-219526). *The Conversation*. Retrieved 25 December 2023.
- Minsky, Marvin (1986), *The Society of Mind*, Simon and Schuster
- Maschafilm (2010). "Content: Plug & Pray Film – Artificial Intelligence – Robots" (http://www.plugandpray-film.de/en/content.html). *plugandpray-film.de*. Archived (https://web.archive.org/web/20160212040134/http://www.plugandpray-film.de/en/content.html) from the original on 12 February 2016.
- McCarthy, John; Minsky, Marvin; Rochester, Nathan; Shannon, Claude (1955). "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence" (https://web.archive.org/web/20070826230310/http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html). Archived from the original (http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html) on 26 August 2007. Retrieved 30 August 2007.
- McCarthy, John (2007), "From Here to Human-Level AI", *Artificial Intelligence*: 171
- McCarthy, John (1999), *What is AI?* (http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html), archived (https://web.archive.org/web/20221204051737/http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html) from the original on 4 December 2022, retrieved 4 December 2022
- McCauley, Lee (2007). "AI armageddon and the three laws of robotics". *Ethics and Information Technology*. **9** (2): 153–164. CiteSeerX 10.1.1.85.8904 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.85.8904). doi:10.1007/s10676-007-9138-2 (https://doi.org/10.1007%2Fs10676-007-9138-2). S2CID 37272949 (https://api.semanticscholar.org/CorpusID:37272949).
- McGarry, Ken (1 December 2005). "A survey of interestingness measures for knowledge discovery". *The Knowledge Engineering Review*. **20** (1): 39–61. doi:10.1017/S0269888905000408 (https://doi.org/10.1017%2FS0269888905000408). S2CID 14987656 (https://api.semanticscholar.org/CorpusID:14987656).
- McGaughey, E (2022), *Will Robots Automate Your Job Away? Full Employment, Basic Income, and Economic Democracy* (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3044448), p. 51(3) Industrial Law Journal 511–559, doi:10.2139/ssrn.3044448 (https://doi.org/10.2139%2Fssrn.3044448), S2CID 219336439 (https://api.semanticscholar.org/CorpusID:219336439), SSRN 3044448 (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3044448), archived (https://web.archive.org/web/20210131074722/https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3044448) from the original on 31 January 2021, retrieved 27 May 2023
- Merkle, Daniel; Middendorf, Martin (2013). "Swarm Intelligence". In Burke, Edmund K.; Kendall, Graham (eds.). *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*. Springer Science & Business Media. ISBN 978-1-4614-6940-7.
- Moravec, Hans (1988). *Mind Children* (https://archive.org/details/mindchildrenfutu00mora). Harvard University Press. ISBN 978-0-674-57616-2. Archived (https://web.archive.org/web/20200726131644/https://archive.org/details/mindchildrenfutu00mora) from the original on 26 July 2020. Retrieved 18 November 2019.

- Morgenstern, Michael (9 May 2015). "Automation and anxiety" (https://www.economist.com/news/special-report/21700758-will-smarter-machines-cause-mass-unemployment-automation-and-anxiety). *The Economist*. Archived (https://web.archive.org/web/20180112214621/https://www.economist.com/news/special-report/21700758-will-smarter-machines-cause-mass-unemployment-automation-and-anxiety) from the original on 12 January 2018. Retrieved 13 January 2018.
- Müller, Vincent C.; Bostrom, Nick (2014). "Future Progress in Artificial Intelligence: A Poll Among Experts" (http://www.sophia.de/pdf/2014_PT-AI_polls.pdf) (PDF). *AI Matters*. **1** (1): 9–11. doi:10.1145/2639475.2639478 (https://doi.org/10.1145%2F2639475.2639478). S2CID 8510016 (https://api.semanticscholar.org/CorpusID:8510016). Archived (https://web.archive.org/web/20160115114604/http://www.sophia.de/pdf/2014_PT-AI_polls.pdf) (PDF) from the original on 15 January 2016.
- Neumann, Bernd; Möller, Ralf (January 2008). "On scene interpretation with description logics". *Image and Vision Computing*. **26** (1): 82–101. doi:10.1016/j.imavis.2007.08.013 (https://doi.org/10.1016%2Fj.imavis.2007.08.013). S2CID 10767011 (https://api.semanticscholar.org/CorpusID:10767011).
- Nilsson, Nils (1995), "Eyes on the Prize", *AI Magazine*, vol. 16, pp. 9–17
- Newell, Allen; Simon, H. A. (1976). "Computer Science as Empirical Inquiry: Symbols and Search" (https://doi.org/10.1145%2F360018.360022). *Communications of the ACM*. **19** (3): 113–126. doi:10.1145/360018.360022 (https://doi.org/10.1145%2F360018.360022).
- Nicas, Jack (7 February 2018). "How YouTube Drives People to the Internet's Darkest Corners" (https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478). *The Wall Street Journal*. ISSN 0099-9660 (https://www.worldcat.org/issn/0099-9660). Retrieved 16 June 2018.
- Nilsson, Nils (1983). "Artificial Intelligence Prepares for 2001" (https://ai.stanford.edu/~nilsson/OnlinePubs-Nils/General%20Essays/AIMag04-04-002.pdf) (PDF). *AI Magazine*. **1** (1). Archived (https://web.archive.org/web/20200817194457/http://ai.stanford.edu/~nilsson/OnlinePubs-Nils/General%20Essays/AIMag04-04-002.pdf) (PDF) from the original on 17 August 2020. Retrieved 22 August 2020. Presidential Address to the Association for the Advancement of Artificial Intelligence.
- NRC (United States National Research Council) (1999). "Developments in Artificial Intelligence". *Funding a Revolution: Government Support for Computing Research*. National Academy Press.
- Omohundro, Steve (2008). *The Nature of Self-Improving Artificial Intelligence*. presented and distributed at the 2007 Singularity Summit, San Francisco, CA.
- Oudeyer, P-Y. (2010). "On the impact of robotics in behavioral and cognitive sciences: from insect navigation to human cognitive development" (http://www.pyoudeyer.com/IEEETAMDOudeyer10.pdf) (PDF). *IEEE Transactions on Autonomous Mental Development*. **2** (1): 2–16. doi:10.1109/tamd.2009.2039057 (https://doi.org/10.1109%2Ftamd.2009.2039057). S2CID 6362217 (https://api.semanticscholar.org/CorpusID:6362217). Archived (https://web.archive.org/web/20181003202543/http://www.pyoudeyer.com/IEEETAMDOudeyer10.pdf) (PDF) from the original on 3 October 2018. Retrieved 4 June 2013.
- Pennachin, C.; Goertzel, B. (2007). "Contemporary Approaches to Artificial General Intelligence". *Artificial General Intelligence*. Cognitive Technologies. Berlin, Heidelberg: Springer. pp. 1–30. doi:10.1007/978-3-540-68677-4_1 (https://doi.org/10.1007%2F978-3-540-68677-4_1). ISBN 978-3-540-23733-4.
- Pinker, Steven (2007) [1994], *The Language Instinct*, Perennial Modern Classics, Harper, ISBN 978-0-06-133646-1

- Poria, Soujanya; Cambria, Erik; Bajpai, Rajiv; Hussain, Amir (September 2017). "A review of affective computing: From unimodal analysis to multimodal fusion" (http://researchrepository.napier.ac.uk/Output/1792429). *Information Fusion*. **37**: 98–125. doi:10.1016/j.inffus.2017.02.003 (https://doi.org/10.1016%2Fj.inffus.2017.02.003). hdl:1893/25490 (https://hdl.handle.net/1893%2F25490). S2CID 205433041 (https://api.semanticscholar.org/CorpusID:205433041). Archived (https://web.archive.org/web/20230323165407/https://www.napier.ac.uk/research-and-innovation/research-search/outputs/a-review-of-affective-computing-from-unimodal-analysis-to-multimodal-fusion) from the original on 23 March 2023. Retrieved 27 April 2021.
- Rawlinson, Kevin (29 January 2015). "Microsoft's Bill Gates insists AI is a threat" (https://www.bbc.co.uk/news/31047780). *BBC News*. Archived (https://web.archive.org/web/20150129183607/http://www.bbc.co.uk/news/31047780) from the original on 29 January 2015. Retrieved 30 January 2015.
- Reisner, Alex (19 August 2023), "Revealed: The Authors Whose Pirated Books are Powering Generative AI" (https://www.theatlantic.com/technology/archive/2023/08/books3-ai-meta-llama-pirated-books/675063/), *The Atlantic*
- Roberts, Jacob (2016). "Thinking Machines: The Search for Artificial Intelligence" (https://web.archive.org/web/20180819152455/https://www.sciencehistory.org/distillations/magazine/thinking-machines-the-search-for-artificial-intelligence). *Distillations*. Vol. 2, no. 2. pp. 14–23. Archived from the original (https://www.sciencehistory.org/distillations/magazine/thinking-machines-the-search-for-artificial-intelligence) on 19 August 2018. Retrieved 20 March 2018.
- Robitzski, Dan (5 September 2018). "Five experts share what scares them the most about AI" (https://futurism.com/artificial-intelligence-experts-fear/amp). Archived (https://web.archive.org/web/20191208094101/https://futurism.com/artificial-intelligence-experts-fear/amp) from the original on 8 December 2019. Retrieved 8 December 2019.
- "Robots could demand legal rights" (http://news.bbc.co.uk/2/hi/technology/6200005.stm). *BBC News*. 21 December 2006. Archived (https://web.archive.org/web/20191015042628/http://news.bbc.co.uk/2/hi/technology/6200005.stm) from the original on 15 October 2019. Retrieved 3 February 2011.
- Rose, Steve (11 July 2023). "AI Utopia or dystopia?". *The Guardian Weekly*. pp. 42–43.
- Russell, Stuart (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. United States: Viking. ISBN 978-0-525-55861-3. OCLC 1083694322 (https://www.worldcat.org/oclc/1083694322).
- Sainato, Michael (19 August 2015). "Stephen Hawking, Elon Musk, and Bill Gates Warn About Artificial Intelligence" (https://observer.com/2015/08/stephen-hawking-elon-musk-and-bill-gates-warn-about-artificial-intelligence/). *Observer*. Archived (https://web.archive.org/web/20151030053323/http://observer.com/2015/08/stephen-hawking-elon-musk-and-bill-gates-warn-about-artificial-intelligence/) from the original on 30 October 2015. Retrieved 30 October 2015.
- Sample, Ian (5 November 2017). "Computer says no: why making AIs fair, accountable and transparent is crucial" (https://www.theguardian.com/science/2017/nov/05/computer-says-no-why-making-ais-fair-accountable-and-transparent-is-crucial). *The Guardian*. Retrieved 30 January 2018.
- Rothman, Denis (7 October 2020). "Exploring LIME Explanations and the Mathematics Behind It" (https://www.codemotion.com/magazine/ai-ml/lime-explainable-ai/). *Codemotion*. Retrieved 25 November 2023.
- Scassellati, Brian (2002). "Theory of mind for a humanoid robot". *Autonomous Robots*. **12** (1): 13–24. doi:10.1023/A:1013298507114 (https://doi.org/10.1023%2FA%3A1013298507114). S2CID 1979315 (https://api.semanticscholar.org/CorpusID:1979315).

- Schmidhuber, J. (2015). "Deep Learning in Neural Networks: An Overview". *Neural Networks*. **61**: 85–117. arXiv:1404.7828 (https://arxiv.org/abs/1404.7828). doi:10.1016/j.neunet.2014.09.003 (https://doi.org/10.1016%2Fj.neunet.2014.09.003). PMID 25462637 (https://pubmed.ncbi.nlm.nih.gov/25462637). S2CID 11715509 (https://api.semanticscholar.org/CorpusID:11715509).
- Schmidhuber, Jürgen (2022). "Annotated History of Modern AI and Deep Learning" (https://people.idsia.ch/~juergen/).
- Schulz, Hannes; Behnke, Sven (1 November 2012). "Deep Learning" (https://www.researchgate.net/publication/230690795). *KI – Künstliche Intelligenz*. **26** (4): 357–363. doi:10.1007/s13218-012-0198-z (https://doi.org/10.1007%2Fs13218-012-0198-z). ISSN 1610-1987 (https://www.worldcat.org/issn/1610-1987). S2CID 220523562 (https://api.semanticscholar.org/CorpusID:220523562).
- Searle, John (1980). "Minds, Brains and Programs" (http://cogprints.org/7150/1/10.1.1.83.5248.pdf) (PDF). *Behavioral and Brain Sciences*. **3** (3): 417–457. doi:10.1017/S0140525X00005756 (https://doi.org/10.1017%2FS0140525X00005756). S2CID 55303721 (https://api.semanticscholar.org/CorpusID:55303721). Archived (https://web.archive.org/web/20190317230215/http://cogprints.org/7150/1/10.1.1.83.5248.pdf) (PDF) from the original on 17 March 2019. Retrieved 22 August 2020.
- Searle, John (1999). *Mind, language and society* (https://archive.org/details/mindlanguagesoci00sear). New York: Basic Books. ISBN 978-0-465-04521-1. OCLC 231867665 (https://www.worldcat.org/oclc/231867665). Archived (https://web.archive.org/web/20200726220615/https://archive.org/details/mindlanguagesoci00sear) from the original on 26 July 2020. Retrieved 22 August 2020.
- Simonite, Tom (31 March 2016). "How Google Plans to Solve Artificial Intelligence" (https://www.technologyreview.com/2016/03/31/161234/how-google-plans-to-solve-artificial-intelligence/). *MIT Technology Review*.
- Smith, Craig S. (15 March 2023). "ChatGPT-4 Creator Ilya Sutskever on AI Hallucinations and AI Democracy" (https://www.forbes.com/sites/craigsmith/2023/03/15/gpt-4-creator-ilya-sutskever-on-ai-hallucinations-and-ai-democracy/). *Forbes*. Retrieved 25 December 2023.
- Smoliar, Stephen W.; Zhang, HongJiang (1994). "Content based video indexing and retrieval". *IEEE MultiMedia*. **1** (2): 62–72. doi:10.1109/93.311653 (https://doi.org/10.1109%2F93.311653). S2CID 32710913 (https://api.semanticscholar.org/CorpusID:32710913).
- Solomonoff, Ray (1956). *An Inductive Inference Machine* (http://world.std.com/~rjs/indinf56.pdf) (PDF). Dartmouth Summer Research Conference on Artificial Intelligence. Archived (https://web.archive.org/web/20110426161749/http://world.std.com/~rjs/indinf56.pdf) (PDF) from the original on 26 April 2011. Retrieved 22 March 2011 – via std.com, pdf scanned copy of the original. Later published as
  Solomonoff, Ray (1957). "An Inductive Inference Machine". *IRE Convention Record*. Vol. Section on Information Theory, part 2. pp. 56–62.
- Stanford University (2023). "Artificial Intelligence Index Report 2023/Chapter 6: Policy and Governance" (https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report-2023_CHAPTER_6-1.pdf) (PDF). AI Index. Archived (https://web.archive.org/web/20230619013609/https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report-2023_CHAPTER_6-1.pdf) (PDF) from the original on 19 June 2023. Retrieved 19 June 2023.
- Tao, Jianhua; Tan, Tieniu (2005). *Affective Computing and Intelligent Interaction*. Affective Computing: A Review. Lecture Notes in Computer Science. Vol. 3784. Springer. pp. 981–995. doi:10.1007/11573548 (https://doi.org/10.1007%2F11573548). ISBN 978-3-540-29621-8.

- Taylor, Josh; Hern, Alex (2 May 2023). " 'Godfather of AI' Geoffrey Hinton quits Google and warns over dangers of misinformation" (https://www.theguardian.com/technology/2023/may/02/geoffrey-hinton-godfather-of-ai-quits-google-warns-dangers-of-machine-learning). *The Guardian*.
- Thompson, Derek (23 January 2014). "What Jobs Will the Robots Take?" (https://www.theatlantic.com/business/archive/2014/01/what-jobs-will-the-robots-take/283239/). *The Atlantic*. Archived (https://web.archive.org/web/20180424202435/https://www.theatlantic.com/business/archive/2014/01/what-jobs-will-the-robots-take/283239/) from the original on 24 April 2018. Retrieved 24 April 2018.
- Thro, Ellen (1993). *Robotics: The Marriage of Computers and Machines* (https://archive.org/details/isbn_9780816026289). New York: Facts on File. ISBN 978-0-8160-2628-9. Archived (https://web.archive.org/web/20200726131505/https://archive.org/details/isbn_9780816026289) from the original on 26 July 2020. Retrieved 22 August 2020.
- Toews, Rob (3 September 2023). "Transformers Revolutionized AI. What Will Replace Them?" (https://www.forbes.com/sites/robtoews/2023/09/03/transformers-revolutionized-ai-what-will-replace-them/). *Forbes*. Retrieved 8 December 2023.
- Turing, Alan (October 1950), "Computing Machinery and Intelligence", *Mind*, **LIX** (236): 433–460, doi:10.1093/mind/LIX.236.433 (https://doi.org/10.1093%2Fmind%2FLIX.236.433), ISSN 0026-4423 (https://www.worldcat.org/issn/0026-4423)
- *UNESCO Science Report: the Race Against Time for Smarter Development* (https://unesdoc.unesco.org/ark:/48223/pf0000377433/PDF/377433eng.pdf.multi). Paris: UNESCO. 2021. ISBN 978-92-3-100450-6. Archived (https://web.archive.org/web/20220618233752/https://unesdoc.unesco.org/ark:/48223/pf0000377433/PDF/377433eng.pdf.multi) from the original on 18 June 2022. Retrieved 18 September 2021.
- Urbina, Fabio; Lentzos, Filippa; Invernizzi, Cédric; Ekins, Sean (7 March 2022). "Dual use of artificial-intelligence-powered drug discovery" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9544280). *Nature Machine Intelligence*. **4** (3): 189–191. doi:10.1038/s42256-022-00465-9 (https://doi.org/10.1038%2Fs42256-022-00465-9). PMC 9544280 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9544280). PMID 36211133 (https://pubmed.ncbi.nlm.nih.gov/36211133). S2CID 247302391 (https://api.semanticscholar.org/CorpusID:247302391).
- Valance, Christ (30 May 2023). "Artificial intelligence could lead to extinction, experts warn" (https://www.bbc.com/news/uk-65746524). *BBC News*. Archived (https://web.archive.org/web/20230617200355/https://www.bbc.com/news/uk-65746524) from the original on 17 June 2023. Retrieved 18 June 2023.
- Valinsky, Jordan (11 April 2019), "Amazon reportedly employs thousands of people to listen to your Alexa conversations" (https://www.cnn.com/2019/04/11/tech/amazon-alexa-listening/index.html), *CNN.com*
- Verma, Yugesh (25 December 2021). "A Complete Guide to SHAP – SHAPley Additive exPlanations for Practitioners" (https://analyticsindiamag.com/a-complete-guide-to-shap-shapley-additive-explanations-for-practitioners/). *Analytics India Magazine*. Retrieved 25 November 2023.
- Vincent, James (7 November 2019). "OpenAI has published the text-generating AI it said was too dangerous to share" (https://www.theverge.com/2019/11/7/20953040/openai-text-generation-ai-gpt-2-full-model-release-1-5b-parameters). *The Verge*. Archived (https://web.archive.org/web/20200611054114/https://www.theverge.com/2019/11/7/20953040/openai-text-generation-ai-gpt-2-full-model-release-1-5b-parameters) from the original on 11 June 2020. Retrieved 11 June 2020.

- Vincent, James (15 November 2022). "The scary truth about AI copyright is nobody knows what will happen next" (https://www.theverge.com/23444685/generative-ai-copyright-infringement-legal-fair-use-training-data). *The Verge*. Archived (https://web.archive.org/web/20230619055201/https://www.theverge.com/23444685/generative-ai-copyright-infringement-legal-fair-use-training-data) from the original on 19 June 2023. Retrieved 19 June 2023.
- Vincent, James (3 April 2023). "AI is entering an era of corporate control" (https://www.theverge.com/23667752/ai-progress-2023-report-stanford-corporate-control). *The Verge*. Archived (https://web.archive.org/web/20230619005803/https://www.theverge.com/23667752/ai-progress-2023-report-stanford-corporate-control) from the original on 19 June 2023. Retrieved 19 June 2023.
- Vinge, Vernor (1993). "The Coming Technological Singularity: How to Survive in the Post-Human Era" (https://web.archive.org/web/20070101133646/http://www-rohan.sdsu.edu/faculty/vinge/misc/singularity.html). *Vision 21: Interdisciplinary Science and Engineering in the Era of Cyberspace*: 11. Bibcode:1993vise.nasa...11V (https://ui.adsabs.harvard.edu/abs/1993vise.nasa...11V). Archived from the original (http://www-rohan.sdsu.edu/faculty/vinge/misc/singularity.html) on 1 January 2007. Retrieved 14 November 2011.
- Waddell, Kaveh (2018). "Chatbots Have Entered the Uncanny Valley" (https://www.theatlantic.com/technology/archive/2017/04/uncanny-valley-digital-assistants/523806/). *The Atlantic*. Archived (https://web.archive.org/web/20180424202350/https://www.theatlantic.com/technology/archive/2017/04/uncanny-valley-digital-assistants/523806/) from the original on 24 April 2018. Retrieved 24 April 2018.
- Wallach, Wendell (2010). *Moral Machines*. Oxford University Press.
- Wason, P. C.; Shapiro, D. (1966). "Reasoning" (https://archive.org/details/newhorizonsinpsy0000foss). In Foss, B. M. (ed.). *New horizons in psychology*. Harmondsworth: Penguin. Archived (https://web.archive.org/web/20200726131518/https://archive.org/details/newhorizonsinpsy0000foss) from the original on 26 July 2020. Retrieved 18 November 2019.
- Weng, J.; McClelland; Pentland, A.; Sporns, O.; Stockman, I.; Sur, M.; Thelen, E. (2001). "Autonomous mental development by robots and animals" (http://www.cse.msu.edu/dl/SciencePaper.pdf) (PDF). *Science*. **291** (5504): 599–600. doi:10.1126/science.291.5504.599 (https://doi.org/10.1126%2Fscience.291.5504.599). PMID 11229402 (https://pubmed.ncbi.nlm.nih.gov/11229402). S2CID 54131797 (https://api.semanticscholar.org/CorpusID:54131797). Archived (https://web.archive.org/web/20130904235242/http://www.cse.msu.edu/dl/SciencePaper.pdf) (PDF) from the original on 4 September 2013. Retrieved 4 June 2013 – via msu.edu.
- "What is 'fuzzy logic'? Are there computers that are inherently fuzzy and do not apply the usual binary logic?" (https://www.scientificamerican.com/article/what-is-fuzzy-logic-are-t/). *Scientific American*. 21 October 1999. Archived (https://web.archive.org/web/20180506035133/https://www.scientificamerican.com/article/what-is-fuzzy-logic-are-t/) from the original on 6 May 2018. Retrieved 5 May 2018.
- Williams, Rhiannon (28 June 2023), "Humans may be more likely to believe disinformation generated by AI" (https://www.technologyreview.com/2023/06/28/1075683/humans-may-be-more-likely-to-believe-disinformation-generated-by-ai/), *MIT Technology Review*
- Wirtz, Bernd W.; Weyerer, Jan C.; Geyer, Carolin (24 July 2018). "Artificial Intelligence and the Public Sector – Applications and Challenges" (https://zenodo.org/record/3569435). *International Journal of Public Administration*. **42** (7): 596–615. doi:10.1080/01900692.2018.1498103 (https://doi.org/10.1080%2F01900692.2018.1498103). ISSN 0190-0692 (https://www.worldcat.org/issn/0190-0692). S2CID 158829602 (https://api.semanticscholar.org/CorpusID:158829602). Archived (https://web.archive.org/web/20200818131415/https://zenodo.org/record/3569435) from the original on 18 August 2020. Retrieved 22 August 2020.

- Wong, Matteo (19 May 2023), "ChatGPT Is Already Obsolete" (https://www.theatlantic.com/technology/archive/2023/05/ai-advancements-multimodal-models/674113/), *The Atlantic*
- Yudkowsky, E (2008), "Artificial Intelligence as a Positive and Negative Factor in Global Risk" (http://intelligence.org/files/AIPosNegFactor.pdf) (PDF), *Global Catastrophic Risks*, Oxford University Press, 2008, Bibcode:2008gcr..book..303Y (https://ui.adsabs.harvard.edu/abs/2008gcr..book..303Y), archived (https://web.archive.org/web/20131019182403/http://intelligence.org/files/AIPosNegFactor.pdf) (PDF) from the original on 19 October 2013, retrieved 24 September 2021

# Further reading

- Ashish Vaswani, Noam Shazeer, Niki Parmar et al. "Attention is all you need." Advances in neural information processing systems 30 (2017). Seminal paper on transformers.
- Autor, David H., "Why Are There Still So Many Jobs? The History and Future of Workplace Automation" (2015) 29(3) *Journal of Economic Perspectives* 3.
- Boden, Margaret, *Mind As Machine*, Oxford University Press, 2006.
- Cukier, Kenneth, "Ready for Robots? How to Think about the Future of AI", *Foreign Affairs*, vol. 98, no. 4 (July/August 2019), pp. 192–98. George Dyson, historian of computing, writes (in what might be called "Dyson's Law") that "Any system simple enough to be understandable will not be complicated enough to behave intelligently, while any system complicated enough to behave intelligently will be too complicated to understand." (p. 197.) Computer scientist Alex Pentland writes: "Current AI machine-learning algorithms are, at their core, dead simple stupid. They work, but they work by brute force." (p. 198.)
- Domingos, Pedro, "Our Digital Doubles: AI will serve our species, not control it", *Scientific American*, vol. 319, no. 3 (September 2018), pp. 88–93. "AIs are like autistic savants and will remain so for the foreseeable future.... AIs lack common sense and can easily make errors that a human never would... They are also liable to take our instructions too literally, giving us precisely what we asked for instead of what we actually wanted." (p. 93.)
- Gertner, Jon. (2023) "Wikipedia's Moment of Truth: Can the online encyclopedia help teach A.I. chatbots to get their facts right — without destroying itself in the process?" *New York Times Magazine* (July 18, 2023) online (https://www.nytimes.com/2023/07/18/magazine/wikipedia-ai-chatgpt.html)
- Gleick, James, "The Fate of Free Will" (review of Kevin J. Mitchell, *Free Agents: How Evolution Gave Us Free Will*, Princeton University Press, 2023, 333 pp.), *The New York Review of Books*, vol. LXXI, no. 1 (18 January 2024), pp. 27–28, 30. "Agency is what distinguishes us from machines. For biological creatures, reason and purpose come from acting in the world and experiencing the consequences. Artificial intelligences – disembodied, strangers to blood, sweat, and tears – have no occasion for that." (p. 30.)

- Hanna, Alex, and Emily M. Bender, "Theoretical AI Harms Are a Distraction: Fearmongering about artificial intelligence's potential to end humanity shrouds the real harm it already causes", *Scientific American*, vol 330, no. 2 (February 2024), pp. 69–70. "[H]ype [about "existential risks"] surrounds many AI firms, but their technology already enables myriad harms, including... discrimination in housing, criminal justice, and health care, as well as the spread of hate speech and misinformation... Large language models extrude... fluent... coherent-seeming text but have no understanding of what the text means, let alone the ability to reason.... (p. 69.) [T]hat output... becomes a noxious... insidious pollutant of our information ecosystem.... [T]oo many... publications [about] AI come from corporate labs or... academic groups that receive... industry funding. Many of these publications are based on junk science [that] is nonreproducible... is full of hype, and uses evaluation methods that do not measure what they purport to... Meanwhile 'AI doomers' cite this junk science... to [misdirect] attention [to] the fantasy of all-powerful machines possibly going rogue and destroying humanity." (p. 70.)
- Hughes-Castleberry, Kenna, "A Murder Mystery Puzzle: The literary puzzle *Cain's Jawbone*, which has stumped humans for decades, reveals the limitations of natural-language-processing algorithms", *Scientific American*, vol. 329, no. 4 (November 2023), pp. 81–82. "This murder mystery competition has revealed that although NLP (natural-language processing) models are capable of incredible feats, their abilities are very much limited by the amount of context they receive. This [...] could cause [difficulties] for researchers who hope to use them to do things such as analyze ancient languages. In some cases, there are few historical records on long-gone civilizations to serve as training data for such a purpose." (p. 82.)
- Immerwahr, Daniel, "Your Lying Eyes: People now use A.I. to generate fake videos indistinguishable from real ones. How much does it matter?", *The New Yorker*, 20 November 2023, pp. 54–59. "If by 'deepfakes' we mean realistic videos produced using artificial intelligence that actually deceive people, then they barely exist. The fakes aren't deep, and the deeps aren't fake. [...] A.I.-generated videos are not, in general, operating in our media as counterfeited evidence. Their role better resembles that of cartoons, especially smutty ones." (p. 59.)
- Johnston, John (2008) *The Allure of Machinic Life: Cybernetics, Artificial Life, and the New AI*, MIT Press.
- Jumper, John; Evans, Richard; Pritzel, Alexander; et al. (26 August 2021). "Highly accurate protein structure prediction with AlphaFold" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8371605). *Nature*. **596** (7873): 583–589. Bibcode:2021Natur.596..583J (https://ui.adsabs.harvard.edu/abs/2021Natur.596..583J). doi:10.1038/s41586-021-03819-2 (https://doi.org/10.1038%2Fs41586-021-03819-2). PMC 8371605 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8371605). PMID 34265844 (https://pubmed.ncbi.nlm.nih.gov/34265844). S2CID 235959867 (https://api.semanticscholar.org/CorpusID:235959867).
- LeCun, Yann; Bengio, Yoshua; Hinton, Geoffrey (28 May 2015). "Deep learning" (https://www.nature.com/articles/nature14539). *Nature*. **521** (7553): 436–444. Bibcode:2015Natur.521..436L (https://ui.adsabs.harvard.edu/abs/2015Natur.521..436L). doi:10.1038/nature14539 (https://doi.org/10.1038%2Fnature14539). PMID 26017442 (https://pubmed.ncbi.nlm.nih.gov/26017442). S2CID 3074096 (https://api.semanticscholar.org/CorpusID:3074096). Archived (https://web.archive.org/web/20230605235832/https://www.nature.com/articles/nature14539) from the original on 5 June 2023. Retrieved 19 June 2023.

- Marcus, Gary, "Am I Human?: Researchers need new ways to distinguish artificial intelligence from the natural kind", *Scientific American*, vol. 316, no. 3 (March 2017), pp. 61–63. Marcus points out a so far insuperable stumbling block to artificial intelligence: an incapacity for reliable disambiguation. "[V]irtually every sentence [that people generate] is ambiguous, often in multiple ways. Our brain is so good at comprehending language that we do not usually notice." A prominent example is the "pronoun disambiguation problem" ("PDP"): a machine has no way of determining to whom or what a pronoun in a sentence—such as "he", "she" or "it"—refers.
- Marcus, Gary, "Artificial Confidence: Even the newest, buzziest systems of artificial general intelligence are stymmied by the same old problems", *Scientific American*, vol. 327, no. 4 (October 2022), pp. 42–45.
- Mitchell, Melanie (2019). *Artificial intelligence: a guide for thinking humans*. New York: Farrar, Straus and Giroux. ISBN 9780374257835.
- Mnih, Volodymyr; Kavukcuoglu, Koray; Silver, David; et al. (26 February 2015). "Human-level control through deep reinforcement learning" (https://www.nature.com/articles/nature14236/). *Nature*. **518** (7540): 529–533. Bibcode:2015Natur.518..529M (https://ui.adsabs.harvard.edu/abs/2015Natur.518..529M). doi:10.1038/nature14236 (https://doi.org/10.1038%2Fnature14236). PMID 25719670 (https://pubmed.ncbi.nlm.nih.gov/25719670). S2CID 205242740 (https://api.semanticscholar.org/CorpusID:205242740). Archived (https://web.archive.org/web/20230619055525/https://www.nature.com/articles/nature14236/) from the original on 19 June 2023. Retrieved 19 June 2023. Introduced DQN, which produced human-level performance on some Atari games.
- Press, Eyal, "In Front of Their Faces: Does facial-recognition technology lead police to ignore contradictory evidence?", *The New Yorker*, 20 November 2023, pp. 20–26.
- Roivainen, Eka, "AI's IQ: ChatGPT aced a [standard intelligence] test but showed that intelligence cannot be measured by IQ alone", *Scientific American*, vol. 329, no. 1 (July/August 2023), p. 7. "Despite its high IQ, ChatGPT fails at tasks that require real humanlike reasoning or an understanding of the physical and social world.... ChatGPT seemed unable to reason logically and tried to rely on its vast database of... facts derived from online texts."
- Serenko, Alexander; Michael Dohan (2011). "Comparing the expert survey and citation impact journal ranking methods: Example from the field of Artificial Intelligence" (http://www.aserenko.com/papers/JOI_AI_Journal_Ranking_Serenko.pdf) (PDF). *Journal of Informetrics*. **5** (4): 629–49. doi:10.1016/j.joi.2011.06.002 (https://doi.org/10.1016%2Fj.joi.2011.06.002). Archived (https://web.archive.org/web/20131004212839/http://www.aserenko.com/papers/JOI_AI_Journal_Ranking_Serenko.pdf) (PDF) from the original on 4 October 2013. Retrieved 12 September 2013.
- Silver, David; Huang, Aja; Maddison, Chris J.; et al. (28 January 2016). "Mastering the game of Go with deep neural networks and tree search" (https://www.nature.com/articles/nature16961). *Nature*. **529** (7587): 484–489. Bibcode:2016Natur.529..484S (https://ui.adsabs.harvard.edu/abs/2016Natur.529..484S). doi:10.1038/nature16961 (https://doi.org/10.1038%2Fnature16961). PMID 26819042 (https://pubmed.ncbi.nlm.nih.gov/26819042). S2CID 515925 (https://api.semanticscholar.org/CorpusID:515925). Archived (https://web.archive.org/web/20230618213059/https://www.nature.com/articles/nature16961) from the original on 18 June 2023. Retrieved 19 June 2023.
- *White Paper: On Artificial Intelligence – A European approach to excellence and trust* (https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf) (PDF). Brussels: European Commission. 2020. Archived (https://web.archive.org/web/20200220173419/https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf) (PDF) from the original on 20 February 2020. Retrieved 20 February 2020.

## External links

- "Artificial Intelligence" (http://www.iep.utm.edu/art-inte). *Internet Encyclopedia of Philosophy*.
- Thomason, Richmond. "Logic and Artificial Intelligence" (https://plato.stanford.edu/entries/logic-ai/). In Zalta, Edward N. (ed.). *Stanford Encyclopedia of Philosophy*.
- Artificial Intelligence (https://www.bbc.co.uk/programmes/p003k9fc). BBC Radio 4 discussion with John Agar, Alison Adam & Igor Aleksander (*In Our Time*, 8 December 2005).
- Theranostics and AI—The Next Advance in Cancer Precision Medicine (https://datascience.cancer.gov/news-events/blog/theranostics-and-ai-next-advance-cancer-precision-medicine).



-