Want to give your brand videos a cinematic edge? Join our visual experts and special guests for an info-packed hour of insights to elevate your next video project. Tune in on June 24 at 11amET.Register NowHow can financial brands set themselves apart through visual storytelling? Our experts explainhow.Learn MoreThe Motorsport Images Collections captures events from 1895 to todays most recentcoverage.Discover The CollectionWant to give your brand videos a cinematic edge? Join our visual experts and special guests for an info-packed hour of insights to elevate your next video project. Tune in on June 24 at 11amET.Register NowHow can financial brands set themselves apart through visual storytelling? Our experts explainhow.Learn MoreThe Motorsport Images Collections captures events from 1895 to todays most recentcoverage.Discover The CollectionWant to give your brand videos a cinematic edge? Join our visual experts and special guests for an info-packed hour of insights to elevate your next video project. Tune in on June 24 at 11amET.Register NowHow can financial brands set themselves apart through visual storytelling? Our experts explainhow.Learn MoreThe Motorsport Images Collections captures events from 1895 to todays most recentcoverage.Discover The Collection What Is AI? Artificial Intelligence (AI) is the theory and development of computer systems of performing tasks that require human intelligence. AI aids, in processing amounts of data identifying patterns and making decisions based on the collected information. This can be achieved through techniques like Machine Learning, Natural Language Processing, Computer Vision and Robotics. AI encompasses a range of abilities including learning, reasoning, perception, problem solving, data analysis and language comprehension. The ultimate goal of AI is to create machines that can emulate capabilities and carry out diverse tasks, with enhanced efficiency and precision. The field of AI holds potential to revolutionize aspects of our daily lives.In this article, we will know about What is Artificial Intelligence?, its evolution, various types of AI, training models, benefits.What is Artificial Intelligence ?Examples of AIArtificial Intelligence (AI) has become increasingly integrated into various aspects of our lives, revolutionizing industries and impacting daily routines. Here are some examples illustrating the diverse applications of AI:Virtual Personal assistants: Popular examples like Siri, Google Assistant, and Amazon Alexa utilize AI to understand and respond to user commands. These assistants employ natural language processing (NLP) and machine learning algorithms to improve their accuracy and provide more personalized responses over time.Autonomous Vehicles: AI powers the development of self-driving cars, trucks, and drones. Companies like Tesla, Waymo, and Uber are at the forefront of this technology, using AI algorithms to analyze sensory data from cameras, radar, and lidar to make real-time driving decisions.Healthcare Diagnosis and Treatment: AI algorithms are used to analyze medical data, including patient records, imaging scans, and genetic information, to assist healthcare professionals in diagnosing diseases and planning treatments. IBM's Watson for Health and Google's DeepMind are examples of AI platforms employed in healthcare.Recommendation Systems: Online platforms like Netflix, Amazon, and Spotify utilize AI to analyze user behaviour and preferences, providing personalized recommendations for movies, products, and music. These systems employ collaborative filtering and content-based filtering techniques to enhance user experience and increase engagement.Fraud Detection: AI algorithms are employed by financial institutions to detect fraudulent activities in real-time. These systems analyze.AI has the potential to revolutionize many industries and fields, such as healthcare, finance, transportation, and education. However, it also raises important ethical and societal questions, such as the impact on employment and privacy, and the responsible development and use of AI technology.Importance of AIToday, the amount of data in the world is so humongous that humans fall short of absorbing, interpreting, and making decisions of the entire data. This complex decision-making requires higher cognitive skills than human beings. This is why were trying to build machines better than us, in these task. Another major characteristic that AI machines possess but we dont is repetitive learning. Let consider an example of how Artificial Intelligence is important to us. Data that is fed into the machines could be real-life incidents. How people interact, behave and react ? etc. So, in other words, machines learn to think like humans, by observing and learning from humans. Thats precisely what is called Machine Learning which is a subfield of AI.Humans are observed to find repetitive tasks highly boring. Accuracy is another factor in which we humans lack. Machines have extremely high accuracy in the tasks that they perform. Machines can also take risks instead of human beings.AI is used in various fields like:Health CareRetailManufacturingBanking etc.Types of AIAI can be broadly classified into two major categories:Based on Capabilities:1. Narrow AI: Narrow AI, also known as Weak AI, refers to artificial intelligence systems that are designed and trained to perform a specific task or a narrow range of tasks. These systems excel at their designated tasks but lack the broad cognitive abilities or understanding of human intelligence. Narrow AI is the most common form of AI currently in use and has found widespread application across various industries and domains.Characteristics of Narrow AI include:Specialized capabilities: Narrow AI systems are highly specialized and focused on performing a specific function or solving a particular problem.Limited scope: These systems are not capable of generalizing their knowledge or skills to other domains outside of their designated tasks.Lack of consciousness: Narrow AI lacks self-awareness and consciousness. It operates based on predefined algorithms and data inputs without understanding the context or implications of its actions.Examples: Virtual personal assistants like Siri and Alexa, recommendation systems, image recognition software, chatbots, and autonomous vehicles are all examples of Narrow AI.2. General AI: Also referred to as "General AI". Here is where there is no difference between a machine and a human being. This is the kind of AI we see in the movies, the robots. A close example (not the perfect example) would be the worlds first citizen robot, Sophia. She was introduced to the world on October 11, 2017. Sophia talks like she has emotions.General AI, also known as Strong AI or Artificial Intelligence (AGI), refers to artificial intelligence systems that replicate human-like cognitive abilities and understanding across a wide range of tasks and domains. Unlike Narrow AI, which is task-specific, General AI aims to simulate human intelligence comprehensively, including reasoning, problem-solving, learning, and adaptation to new situations.Characteristics of General AI include:Human-like cognitive abilities: General AI possesses the ability to understand, learn, and apply knowledge across various tasks and domains, similar to human intelligence.Adaptability: These systems have the ability to generalize their knowledge and skills to new situations, tasks, and environments.Consciousness: General AI is theorized to possess self-awareness, consciousness, and subjective experiences, although achieving this level of intelligence remains a theoretical challenge.Examples: General AI remains largely theoretical and speculative, as researchers have not yet achieved human-level artificial intelligence. Development in this area continues to be a subject of ongoing research and exploration.3. Super Intelligent AI: Super intelligent AI refers to artificial intelligence systems that surpass human intelligence in virtually every aspect. This type of AI, also known as Artificial Superintelligence (ASI), represents the highest level of AI capabilities and poses significant implications for society and the future of humanity.Characteristics of Super intelligent AI include:Cognitive superiority: Super intelligent AI outperforms humans in terms of cognitive abilities, including problem-solving, creativity, and strategic planning.Rapid learning and adaptation: These systems can acquire and process information at an unprecedented speed and scale, leading to rapid advancements in various fields.Ethical and existential risks: The development of Super intelligent AI raises concerns about its potential impact on society, including risks related to control, alignment with human values, and existential threats to humanity.Examples: Super intelligent AI remains theoretical and no concrete examples exist as of now. However, researchers are actively exploring the implications and challenges associated with its development.Based on Functionalities :1. Reactive machines: These are the most basic type of AI and are purely reactive as the name suggests. They neither can form memories nor can use past experiences to form decisions. An example would be IBMs Deep Blue chess-playing supercomputer which is mentioned above. Deep Blue beat the international grandmaster Garry Kasparov in 1997. It can choose the most optimal of the chess moves and beat the opponent. Apart from a rarely used chess-specific rule against repeating the same move time, Deep Blue ignores everything before the present moment, thus not storing any memories. This type of AI just perceives the world, the chess game in the case of Deep Blue, and acts on it.2. Limited memory: These machines can look into the past. Not the ability to predict what happened in the past, but the usage of memories to form decisions. A common example could include self-driving cars. For example, they observe other cars speed and directions and act accordingly. This requires monitoring of how a car is driven for a specific amount of time. Just like how humans observe and learn the specifics. These pieces of information are not stored in the library of experiences of the machines, unlike humans. We humans automatically save everything in the library of our experiences and can learn from it, but limited memory machines cant.3. Theory of mind: These are types of machines that can understand that people have beliefs, emotions, expectations, etc., and have some of their own. A "theory of mind" machine can think emotionally and can respond with emotions. Even though there are close examples of this kind of AI like Sophia, the research is not complete yet. In other words, these machines have a notion of not just the world, but also the existing entities of the world, like human beings, animals, etc. These machines will be capable of answering simple "what if" questions. Theyll have a sense of empathy.4. Self-Awareness: These types of machines can be called human equivalents. Of course, no such machines exist and the invention of them would be a milestone in the field of AI. These basically will have a sense of consciousness of who they are. The sense of "I" or "me". Heres a basic example of the difference between "theory of mind" and "self-awareness" AI. The feeling of I want to play is different from the feeling of I know I want to play. In the latter, if you notice, there is a sense of consciousness and is a characteristic of a self-aware machine, while the former feeling is a characteristic of a theory-of-mind machine. Self-aware machines will have the ability to predict others feelings. Lets hope the invention is not so far away.How Does AI Work ?Artificial Intelligence (AI) uses a wide range of techniques and approaches that enable machines to simulate human-like intelligence and perform tasks that traditionally require human assistance. AI systems work through a combination of algorithms, data, and computational power. Here's an overview of how AI works:Data Collection: AI systems rely on vast amounts of data to learn and make decisions. Data can be collected from various sources, including sensors, digital devices, databases, the internet, and user interactions. The quality and quantity of data are crucial for training accurate and reliable AI models.Data Pre-processing: Once data is collected, it needs to be pre-processed to ensure it's clean, structured, and suitable for analysis. This pre-processing stage may involve tasks such as cleaning noisy data, handling missing values, standardizing formats, and encoding categorical variables.Algorithm Selection: AI algorithms are chosen based on the specific task or problem the AI system aims to solve. Different algorithms are suited for different types of tasks, such as classification, regression, clustering, and pattern recognition. Common AI algorithms include neural networks, decision trees, support vector machines, and k-nearest neighbours.Model Training: In the training phase, AI models are fed with labelled data (supervised learning) or unlabelled data (unsupervised learning) to learn patterns and relationships. During training, the model adjusts its parameters iteratively to minimize errors and improve its performance on the given task. This process involves optimization techniques like gradient descent and backpropagation in neural networks.Model Evaluation: After training, the AI model is evaluated using separate validation data to assess its performance and generalization ability. Performance metrics such as accuracy, precision, recall, F1-score, and area under the curve (AUC) are used to quantify the model's effectiveness in making predictions or decisions.Model Deployment: Once the AI model meets the desired performance criteria, it can be deployed into production environments to perform real-world tasks. Deployment involves integrating the model into existing systems, applications, or devices, ensuring it can handle real-time data and make predictions or decisions efficiently.Continuous Learning and Improvement: AI systems can adapt and improve over time through continuous learning. They can be updated with new data and retrained periodically to stay relevant and accurate in dynamic environments. Techniques like online learning, transfer learning, and reinforcement learning enable AI models to learn from new experiences and feedback.Inference and Decision-Making: During inference, the trained AI model applies its learned knowledge to make predictions or decisions on new, unseen data. Inference involves feeding input data into the model and obtaining output predictions or classifications based on the model's learned patterns and representations.Overall, AI systems work by leveraging data, algorithms, and computational power to learn from experience, make decisions, and perform tasks autonomously. The specific workings of an AI system depend on its architecture, algorithms, and the nature of the tasks it's designed to accomplish.Applications of AIArtificial Intelligence (AI) has a wide range of applications and has been adopted in many industries to improve efficiency, accuracy, and productivity. Some of the most common uses of AI are:Healthcare: AI is used in healthcare for various purposes such as diagnosing diseases, predicting patient outcomes, drug discovery, and personalized treatment plans.Finance: AI is used in the finance industry for tasks such as credit scoring, fraud detection, portfolio management, and financial forecasting.Retail: AI is used in the retail industry for applications such as customer service, demand forecasting, and personalized marketing.Manufacturing: AI is used in manufacturing for tasks such as quality control, predictive maintenance, and supply chain optimization.Transportation: AI is used in transportation for optimizing routes, improving traffic flow, and reducing fuel consumption.Education: AI is used in education for personalizing learning experiences, improving student engagement, and providing educational resources.Marketing: AI is used in marketing for tasks such as customer segmentation, personalized recommendations, and real-time audience analysis.Gaming: AI is used in gaming for developing intelligent game characters and providing personalized gaming experiences.Security: AI is used in security for tasks such as facial recognition, intrusion detection, and cyber threat analysis.Natural Language Processing (NLP): AI is used in NLP for tasks such as speech recognition, machine translation, and sentiment analysis.These are some of the most common uses of AI, but the Applications of AI are constantly expanding. As AI technology continues to advance, it is likely that new uses will emerge in the future.Future of AIThe future of AI is likely to involve continued advancements in machine learning, natural language processing, and computer vision, which will enable AI systems to become increasingly capable and integrated into a wide range of applications and industries. Some potential areas of growth for AI include healthcare, finance, transportation, and customer service. Additionally, there may be increasing use of AI in more sensitive areas such as decision making and criminal justice, hiring and education, which will raise ethical and societal implications that need to be addressed. It is also expected that there will be more research and development in areas such as explainable AI, trustworthy AI and AI safety to ensure that AI systems are transparent, reliable and safe to use. Intelligence of machines"AI" redirects here. For other uses, see AI (disambiguation) and Artificial intelligence (disambiguation).Part of a series onArtificial Intelligence (AI)Major goalsArtificial general intelligenceIntelligent agentRecursive self-improvementPlanningKnowledge representationRoboticsAI safetyApproachesMachine learningBayesian networksEvolutionary algorithmsHybrid intelligent systemsSystems integrationApplicationsBioinformaticsDeepfakeEarth sciences Finance Generative AIArtAudioMusicGovernmentHealthcareMental healthIndustryTranslation Military PhysicsProjectsPhilosophyArtificial consciousnessChinese roomFriendly AIControl problem/TakeoverEthicsExistential riskTuring testGlossaryHistoryTimelineProgressAI winterAI boomGlossaryGlossaryvtArtificial intelligence (AI) is the capability of computational systems to perform tasks typically associated with human intelligence, such as learning, reasoning, problem-solving, perception, and decision-making. It is a field of research in computer science that develops and studies methods and software that enable machines to perceive their environment and use learning and intelligence to take actions that maximize their chances of achieving defined goals.[1]High-profile applications of AI include advanced web search engines (e.g., Google Search); recommendation systems (used by YouTube, Amazon, and Netflix); virtual assistants (e.g., Google Assistant, Siri, and Alexa); autonomous vehicles (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). However, many AI applications are not perceived as AI: "A lot of cutting edge AI has filtered into general applications, often without being called AI because once something becomes useful enough and common enough it's not labeled AI anymore."[2][3]Various subfields of AI research are centered around particular goals and the use of particular tools. The traditional goals of AI research include learning, reasoning, knowledge representation, planning, natural language processing, perception, and support for robotics.[a] To reach these goals, AI researchers have adapted and integrated a wide range of 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.[4] Some companies, such as OpenAI, Google DeepMind and Meta, aim to create artificial general intelligence (AGI)AI that can complete virtually any cognitive task at least as well as a human.[5]Artificial intelligence was founded as an academic discipline in 1956,[6] and the field went through multiple cycles of optimism throughout its history,[7][8] followed by periods of disappointment and loss of funding, known as AI winters.[9][10] Funding and interest vastly increased after 2012 when graphics processing units started being used to accelerate neural networks, and deep learning outperformed previous AI techniques.[11] This growth accelerated further after 2017 with the transformer architecture.[12] In the 2020s, the period of rapid progress marked by advanced generative AI became known as the AI boom. Generative AI and its ability to create and modify content raised several unintended consequences and harms in the present and raised ethical concerns about AI's long-term effects and potential existential risks, prompting discussions about regulatory policies to ensure the safety and benefits of the technology.The general problem of simulating (or creating) intelligence has been broken into subproblems. 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]Early researchers developed algorithms that imitated step-by-step reasoning that humans use when they solve puzzles or make logical deductions.[17] By the late 1980s and 1990s, methods were developed for dealing with uncertain or incomplete information, employing concepts from probability and economics.[18]Many of these algorithms are insufficient for solving large reasoning problems because they experience a "combinatorial explosion": They become exponentially slower as the problems grow.[19] 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.[72] 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 prioritize choices that are more likely to reach a goal.[73]Adversarial search is used for game-playing programs, such as chess or Go. It searches through a tree of possible moves and countermoves, looking for a winning position.[74]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)Local search uses mathematical optimization to find a solution to a problem. It begins with some form of guess and refines it incrementally.[75]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,[76] through the backpropagation algorithm.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.[77]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).[78]Formal logic is used for reasoning and knowledge representation.[79]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")[80] 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 Xs that are Ys").[81]Deductive reasoning in logic is the process of proving a new statement (conclusion) from other statements that are given and assumed to be true (the premises).[82] Proofs can be structured as proof trees, in which nodes are labelled by sentences, and children nodes are connected to parent nodes by inference rules.Given a problem and a set of premises, problem-solving reduces to searching for a proof tree whose root node is labelled by a solution of the problem and whose leaf nodes are labelled by premises or axioms. In the case of Horn clauses, problem-solving search can be performed by reasoning forwards from the premises or backwards from the problem.[83] In the more general case of the clausal form of first-order logic, resolution is a single, axiom-free rule of inference, in which a problem is solved by proving a contradiction from premises that include the negation of the problem to be solved.[84]Inference in both Horn clause logic and first-order logic is undecidable, and therefore intractable. However, backward reasoning with Horn clauses, which underpins computation in the logic programming language Prolog, is Turing complete. Moreover, its efficiency is competitive with computation in other symbolic programming languages.[85]Fuzzy logic assigns a "degree of truth" between 0 and 1. It can therefore handle propositions that are vague and partially true.[86]Non-monotonic logics, including logic programming with negation as failure, are designed to handle default reasoning.[28] Other specialized versions of logic have been developed to describe many complex domains.A simple Bayesian network, with the associated conditional probability tablesMany problems in AI (including 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.[87] Precise mathematical tools have been developed that analyze how an agent can make choices and plan, using decision theory, decision analysis,[88] and information value theory.[89] These tools include models such as Markov decision processes,[90] dynamic decision networks,[91] game theory and mechanism design.[92]Bayesian networks[93] are a tool that can be used for reasoning (using the Bayesian inference algorithm),[b][97] learning (using the expectation-maximization algorithm),[h][97] planning (using decision networks)[98] and perception (using dynamic Bayesian networks).[91]Probabilistic algorithms can also be used for filtering, prediction, smoothing, and finding explanations for streams of data, thus helping perception systems analyze processes that occur over time (e.g., hidden Markov models or Kalman filters).[91]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[99] 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.[100] The decision tree is the simplest and most widely used symbolic machine learning algorithm.[101] 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.[102]The naive Bayes classifier is reportedly the "most widely used learner"[103] at Google, due in part to its scalability.[104]Neural networks are also used as classifiers.[104]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 recognize patterns; once trained, it can recognize 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.[104]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.[106] 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.[107]In feedforward neural networks the signal passes in only one direction.[108] 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 architecture for recurrent neural networks.[109] Perceptrons[110] use only a single layer of neurons; deep learning[111] uses multiple layers. Convolutional neural networks strengthen the connection between neurons that are "close" to each otherthis is especially important in image processing, where a local set of neurons must identify an "edge" before the network can identify an object.[112]Deep learning[111] 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, letters, or faces.[113]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,[114] and others. The reason that deep learning performs so well in so many applications is not known as of 2021.[115] 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]Generative pre-trained transformers (GPT) are large language models (LLMs) that generate text based on the semantic relationships between words in sentences. Text-based GPT models are pre-trained on a large corpus of text that can be from the Internet. The pretraining consists of predicting the next token (a token being usually a word, subword, or punctuation). Throughout this pretraining, 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 prone to generating falsehoods called "hallucinations". These can be reduced with RLHF and quality data, but the problem has been getting worse for reasoning systems.[123] Such systems are used in chatbots, which allow people to ask a question or request a task in simple text.[124][125]Current models and services include Gemini (formerly Bard), ChatGPT, Grok, Claude, Copilot, and LLaMA.[126] Multimodal GPT models can process different types of data (modalities) such as images, videos, sound, and text.[127]Main articles: Programming languages for artificial intelligence and Hardware for artificial intelligenceIn 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.[128] Specialized programming languages such as Prolog were used in early AI research,[129] but general-purpose programming languages like Python have become predominant.[130]The transistor density in integrated circuits has been observed to roughly double every 18 months since approximately 1960. This growth, known as Moore's law, named after the Intel co-founder Gordon Moore, who first identified it. Improvements in GPUs have been even faster,[131] a trend sometimes called Huang's law,[132] named after Nvidia co-founder and CEO Jensen Huang.Main article: Applications of artificial intelligenceAI 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 (such as 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 FaceID or Microsoft's DeepFace and Google's FaceNet) and image labeling (used by Facebook, Apple's Photos and TikTok). The deployment of AI may be overseen by a Chief automation officer (CAO).Main article: Artificial intelligence in healthcareThe application of AI in medicine and medical research has the potential to increase patient care and quality of life.[13] Through the lens of the Hippocratic Oath, medical professionals are ethically compelled to use AI, if applications can more accurately diagnose and treat patients.[134][135]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.[136] It has been suggested that AI can overcome discrepancies in funding allocated to different fields of research.[136][137] New AI tools can deepen the 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.[138] 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.[139] In 2024, researchers used machine learning to accelerate the search for Parkinson's disease drug treatments. Their aim was to identify compounds that block the clumping, or aggregation, of alpha-synuclein (the protein that characterises Parkinson's disease). They were able to speed up the initial screening process ten-fold and reduce the cost by a thousand-fold.[140][141]Main article: Game artificial intelligenceGame playing programs have been used since the 1950s to demonstrate and test the most advanced techniques.[142] Deep Blue became the first computer chess-playing system to beat a reigning world chess champion, Garry Kasparov, on 11 May 1997.[143] 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.[144] 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.[145] Other programs handle imperfect-information games, such as the poker-playing program Pluribus.[146] DeepMind developed increasingly generalistic reinforcement learning models, such as with MuZero, which could be trained to play chess, Go, or Atari games.[147] 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.[148] 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.[149] In 2024, Google DeepMind introduced SIMA, a type of AI capable of autonomously playing nine previously unseen open-world video games by observing screen output, as well as executing short, specific tasks in response to natural language instructions.[150]Large language models, such as GPT-4, Gemini, Claude, Llama or Mistral, are increasingly used in mathematics. These probabilistic models are versatile, but can also produce wrong answers in the form of hallucinations. They sometimes need a large database of mathematical problems to learn from, but also methods such as supervised fine-tuning[151] or trained classifiers with human-annotated data to improve answers for new problems and learn from corrections.[152] A February 2024 study showed that the performance of some language models for reasoning capabilities in solving math problems not included in their training data was low, even for problems with only minor deviations from trained data.[153] One technique to improve their performance involves training the models to produce correct reasoning steps, rather than just the correct result.[154] The Alibaba Group developed a version of its Qwen models called Qwen2-Math, that achieved state-of-the-art performance on several mathematical benchmarks, including 84% accuracy on the MATH dataset of competition mathematics problems.[155] In January 2025, Microsoft proposed the technique rStar-Math that leverages Monte Carlo tree search and step-by-step reasoning, enabling a relatively small language model like Qwen-7B to solve 53% of the AIME 2024 and 90% of the MATH benchmark problems.[156]Alternatively, dedicated models for mathematical problem solving with higher precision for the outcome including proof of theorems have been developed such as AlphaTensor, AlphaGeometry, AlphaProof and AlphaEvolve[157] all from Google DeepMind,[158] Llemma from EleutherAI[159] or Julius.[k]When natural language is used to describe mathematical problems, converters can transform such prompts into a formal language such as Lean to define mathematical tasks.Some recent models, both from academia and industry, use two different approaches, neurosymbolic AI and large language models (LLMs), to tackle complex problems; LLMs for generating ideas and formal provers to verify.[160]Main article: Finance and AIFinance is one of the fastest growing sectors where applied AI tools are being deployed: from retail online banking to investment advice and insurance, where automated "robot advisers" have been in use for some years.[162]According to Nicolas Firzli, director of the World Pensions & Investments Forum, it may be too early to see the emergence of highly innovative AI-informed financial products and services. He argues that "the deployment of AI tools will simply further automate things: destroying tens of thousands of jobs in banking, financial planning, and pension advice in the process, but I'm not sure it will unleash a new wave of (e.g. sophisticated) pension innovation."[163]Main article: Military artificial intelligenceVarious countries are deploying AI military applications.[164] The main applications enhance command and control, communications, sensors, integration and interoperability.[165] Research is targeting intelligence collection and analysis, logistics, cyber operations, information operations, and semiautonomous and autonomous vehicles.[164] 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, both human-operated and autonomous.[165] AI has been used in military operations in Iraq, Syria, Israel and Ukraine.[164][166][167][168]Woven van Gogh is watercolour created by generative AI softwareThese paragraphs are an excerpt from Generative artificial intelligence.[edit]Generative artificial intelligence (Generative AI, GenAI,[169] or GAI) is a subfield of artificial intelligence that uses generative models to produce text, images, videos, or other forms of data.[170][171][172] These models learn the underlying patterns and structures of their training data and use them to produce new data[173][174] based on the input, which often comes in the form of natural language prompts.[175][176]Generative AI tools have become more common since an "AI boom" in the 2020s. This boom was made possible by improvements in transformer-based deep neural networks, particularly large language models (LLMs). Major tools include chatbots such as ChatGPT, Copilot, Gemini, Grok, and DeepSeek; text-to-image models such as Stable Diffusion, Midjourney, and DALL-E; and text-to-video models such as Veo and Sora.[177][178][179][180] Technology companies developing generative AI include OpenAI, Anthropic, Meta AI, Microsoft, Google, DeepSeek, and Baidu.[175][181][182]Generative AI has raised many ethical questions. It can be used for cybercrime, or deceive or manipulate people through fake news or deepfakes.[183] Even if used ethically, it may lead to mass replacement of human jobs.[184] The tools themselves have been criticized as violating intellectual property laws, since they are trained on copyrighted works.[185]Main article: Agentic AIAI agents are software entities designed to perceive their environment, make decisions, and take actions autonomously to achieve specific goals. These agents can interact with users, their environment, or other agents. AI agents are used in various applications, including virtual assistants, chatbots, autonomous vehicles, game-playing systems, and industrial robotics. AI agents operate within the constraints of their programming, available computational resources, and hardware limitations. This means they are restricted to performing tasks within their defined scope and have finite memory and processing capabilities. In real-world applications, AI agents often face time constraints for decision-making and action execution. Many AI agents incorporate learning algorithms, enabling them to improve their performance over time through experience or training. Using machine learning, AI agents can adapt to new situations and personalize their behaviour for their designated tasks.[186][187]Main article: Applications of AI in this domain include AI-enabled menstruation and fertility trackers that analyze user data to offer predictions,[188] AI-integrated sex toys (e.g., teledildonics),[189] AI-generated sexual education content,[190] and AI chatbots that can hold conversations.[191] and AI agents that simulate sexual and romantic partners (e.g., Replika).[192] AI is also used for the production of non-consensual deepfake pornography, raising significant ethical and legal concerns.[193]AI technologies have also been used to attempt to identify online gaming and online sexual grooming of minors.[194][195]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 having incorporated "AI" in some offerings or processes.[196] A few examples are energy storage, medical diagnosis, military logistics, applications that predict the result of judicial decisions, foreign policy, or supply chain management.AI applications for evacuation and disaster management are growing. AI has been used to investigate patterns in large-scale and small-scale evacuations using historical data from GPS, videos or social media. Furthermore, AI can provide real-time information on the evacuation conditions.[197][198][199]In agriculture, AI has helped farmers to increase yield and identify areas that need irrigation, fertilization, pesticide treatments. 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, it is used for discovering exoplanets, forecasting solar activity, and distinguishing between signals and instrumental effects in gravitational wave astronomy. Additionally, it could be used for activities in space, such as space exploration, including the analysis of data from space missions, real-time science decisions of spacecraft, space debris avoidance, and more autonomous operation.During the 2024 Indian elections, US$50 million was spent on authorized AI-generated content, notably by creating deepfakes of allied (including sometimes deceased) politicians to better engage with voters, and by translating speeches to various local languages.[200]Main article: Ethics of artificial intelligenceAI has potential benefits and potential risks.[201] AI may be able to advance science and find solutions for serious problems: Demis Hassabis of DeepMind hopes to "solve intelligence, and then use that to solve everything else".[202] However, as the use of AI has become widespread, several unintended consequences and risks have been identified.[203] In-production systems can sometimes not factor ethics and bias into their AI training processes, especially when the AI algorithms are inherently unexplainable in deep learning.[204]Further information: Machine learning algorithms require large amounts of data. The techniques used to acquire this data have raised concerns about privacy, surveillance and copyright.AI-powered devices and services, such as virtual assistants and IoT products, continuously collect personal information, raising concerns about intrusive data gathering and unauthorized access by third parties. The loss of privacy is further exacerbated by AI's ability to process and combine vast amounts of data, potentially leading to a surveillance society where individual activities are constantly monitored and analyzed without adequate safeguards or transparency.Sensitive user data collected may include online activity records, geolocation data, video, or audio.[205] For example, in order to build speech recognition algorithms, Amazon has recorded millions of private conversations and allowed temporary workers to listen to and transcribe some of them.[206] 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.[207]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.[208] Since 2016, some privacy experts, such as Cynthia Dwork, have begun 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'."[209]Generative AI is often trained on unlicensed copyrighted works, including in domains such as images or computer code; the output is then used under the rationale of "fair use". 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".[210][211] Website owners who do not wish to have their content scraped can indicate it in a "robots.txt" file.[212] In 2023, leading authors (including John Grisham and Jonathan Franzen) sued AI companies for using their work to train generative models.[213][214] Another discussed approach is to envision a separate sui generis system of protection for creations generated by AI to ensure fair attribution and compensation for human authors.[215]The commercial AI scene is dominated by Big Tech companies such as Alphabet Inc., Amazon, Apple Inc., Meta Platforms, and Microsoft.[216][217][218] Some of these players already own the vast majority of existing cloud infrastructure and computing power from data centers, allowing them to entrench further in the marketplace.[219][220]See also: Environmental impacts of artificial intelligenceIn January 2024, the International Energy Agency (IEA) released Electricity 2024, Analysis and Forecast to 2026, forecasting electric power use.[221] This is the first IEA report to make projections for data centers and power consumption for artificial intelligence and cryptocurrency. The report states that power demand for these uses might double by 2026, with additional electric power usage equal to electricity used by the whole Japanese nation.[222]Prodigious power consumption by AI is responsible for the growth of fossil fuel use, and might delay closings of obsolete, carbon-emitting coal energy facilities. There is a feverish rise in the construction of data centers throughout the US, making large technology firms (e.g., Microsoft, Meta, Google, Amazon) into voracious consumers of electric power. Projected electric consumption is so immense that there is concern that it will be fulfilled no matter the source. A ChatGPT search involves the use of 10 times the electrical energy as a Google search. The large firms are in haste to find power sources from nuclear energy to geothermal to fusion. The tech firms argue that in the long view AI will be eventually kinder to the environment, but they need the energy now. AI makes the power grid more efficient and "intelligent", will assist in the growth of nuclear power, and track overall carbon emissions, according to technology firms.[223]A 2024 Goldman Sachs Research Paper, AI Data Centers and the Coming US Power Demand Surge, found "US power demand (is) likely to experience growth not seen in a generation...." and forecasts that, by 2030, US data centers will consume 8% of US power, as compared to 3% in 2022, presaging growth for the electrical power generation industry by a variety of means.[224] Data centers' need for more and more electrical power is such that they might max out the electrical grid. The Big Tech companies counter that AI can be used to maximize the utilization of the grid by all.[225]In 2024, the Wall Street Journal reported that big AI companies have begun negotiations with the US nuclear power providers to provide electricity to the data centers. In March 2024 Amazon purchased a Pennsylvania nuclear-powered data center for US$650 million.[226] Nvidia CEO Jensen Huang said nuclear power is a good option for the data centers.[227]In September 2024, Microsoft announced an agreement with Constellation Energy to re-open the Three Mile Island nuclear power plant to provide Microsoft with 100% of all electric power produced by the plant for 20 years. Reopening the plant, which suffered a partial nuclear meltdown of its Unit 2 reactor in 1979, will require Constellation to get through stringent regulatory processes which will include extensive safety scrutiny from the US Nuclear Regulatory Commission. If approved (this will be the first ever US re-commissioning of a nuclear plant), over 835 megawatts of power enough for 800,000 homes of energy will be produced. The cost for re-opening and upgrading is estimated at $1.6 billion (US) and is dependent on tax breaks for nuclear power contained in the 2022 US Inflation Reduction Act.[228] The US government and the state of Michigan are investing almost $2 billion (US) to reopen the Palisades Nuclear reactor on Lake Michigan. Closed since 2022, the plant is planned to be reopened in October 2025. The Three Mile Island facility will be renamed the Crane Clean Energy Center after Chris Crane, a nuclear proponent and former CEO of Exelon who was responsible for Exelon's spinoff of Constellation.[229]After the last approval in September 2023, Taiwan suspended the approval of data centers north of Taoyuan with a capacity of more than 5 MW in 2024, due to power supply shortages.[230] Taiwan aims to phase out nuclear power by 2025.[230] On the other hand, Singapore imposed a ban on the opening of data centers in 2019 due to electric power, but in 2022, lifted this ban.[230]Although most nuclear plants in Japan have been shut down after the 2011 Fukushima nuclear accident, according to an October 2024 Bloomberg article in Japanese, cloud gaming services company Ubitus, in which Nvidia has a stake, is looking for land in Japan near nuclear power plant for a new data center for generative AI.[231] Ubitus CEO Wesley Kuo said nuclear power plants are the most efficient, cheap and stable power for AI.[231]On 1 November 2024, the Federal Energy Regulatory Commission (FERC) rejected an application submitted by Talen Energy for approval to supply some electricity from the nuclear power station Susquehanna to Amazon's data center.[232] According to the Commission Chairman Willie L. Phillips, it is a burden on the electricity grid as well as a significant cost shifting concern to households and other business sectors.[232]In 2025, a report prepared by the International Energy Agency estimated the greenhouse gas emissions from the energy consumption of AI at 180 million tons. By 2035, these emissions could rise to 300-500 million tons depending on how much electricity will be used. This is below 1.5% of the energy sector emissions. The emissions reduction potential of AI was estimated at 5% of the energy sector emissions, but rebound effects (for example if people switch from public transport to autonomous cars) can reduce it.[233]See also: YouTube Moderation and offensive contentYouTube, 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.[234] This convinced many users that the misinformation was true, and ultimately undermined trust in institutions, the media and the government.[235] 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.[236]In 2022, generative AI began to create images, audio, video and text that are indistinguishable from real photographs, recordings, films, or human writing.[237] It is possible for bad actors to use this technology to create massive amounts of misinformation or propaganda.[238]AI pioneer Geoffrey Hinton expressed concern about AI enabling "authoritarian leaders to manipulate their electorates" on a large scale, among other risks.[239]AI researchers at Microsoft, OpenAI, universities and other organisations have suggested using "personhood credentials" as a way to overcome online deception enabled by AI models.[240]Main articles: Algorithmic bias and Fairness (machine learning)Machine learning applications will be biased[k] if they learn from biased data.[241] The developers may not be aware that the bias exists.[242] Bias can be introduced by the way training data is selected and by the way a model is deployed.[243][244] 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.[245] The field of fairness studies how to prevent harms from algorithmic biases.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,[246] a problem called "sample size disparity".[247] 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.[248]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.[250] In 2017, several researchers[l] 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.[252]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".[253] Moritz Hardt said "the most robust fact in this research area is that fairness through blindness doesn't work."[254]Criticism of COMPAS highlighted that 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. If an application then uses these predictions as recommendations, some of these "recommendations" will likely be racist.[255] 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 descriptive rather than prescriptive.[m]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.[248]There are various conflicting definitions and mathematical models of fairness. These notions depend on ethical assumptions, and are influenced by beliefs about society. One broad category is distributive fairness, which focuses on the outcomes, often identifying groups and seeking to compensate for statistical disparities. Representational fairness tries to ensure that AI systems do not reinforce negative stereotypes or render certain groups invisible. Procedural fairness focuses on the decision process rather than the outcome. The most relevant notions of fairness may depend on the context, notably the type of AI

application and the stakeholders. The subjectivity in the notions of bias and fairness makes it difficult for companies to operationalize them. Having access to sensitive attributes such as race or gender is also considered by many AI ethicists to be necessary in order to compensate for biases, but it may conflict with anti-discrimination laws.[242]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 that recommend 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.[dubious discuss][257]See also: Explainable AI, Algorithmic transparency, and Right to explanationMany AI systems are so complex that their designers cannot explain how they reach their decisions.[258] Particularly with deep neural networks, in which there are many non-linear relationships between inputs and outputs. But some popular explainability techniques exist.[259]It is impossible to be certain that a program is operating correctly if no one knows how exactly it works. 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.[260] 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.[261]People who have been harmed by an algorithm's decision have a right to an explanation.[262] Doctors, for example, are expected to clearly and completely explain to their colleagues the reasoning behind any decision they make. Early drafts of the European Union's General Data Protection Regulation in 2016 included an explicit statement that this right exists.[n] 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.[263]DARPA established the XAI ("Explainable Artificial Intelligence") program in 2014 to try to solve these problems.[264]Several approaches aim to address the transparency problem. SHAP enables to visualise the contribution of each feature to the output.[265] LIME can locally approximate a model's outputs with a simpler, interpretable model.[266] 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.[267] Deconvolution, DeepDream and other generative methods can allow developers to see what different layers of a deep network for computer vision have learned, and produce output that can suggest what the network is learning.[268] For generative pre-trained transformers, Anthropic developed a technique based on dictionary learning that associates patterns of neuron activations with human-understandable concepts.[269]Main articles: Lethal autonomous weapon, Artificial intelligence arms race, and AI safetyArtificial intelligence provides a number of tools that are useful to bad actors, such as authoritarian governments, terrorists, criminals or rogue states.A lethal autonomous weapon is a machine that locates, selects and engages human targets without human supervision.[o] Widely available AI tools can be used by bad actors to develop inexpensive autonomous weapons and, if produced at scale, they are potentially weapons of mass destruction.[271] Even when used in conventional warfare, they currently cannot reliably choose targets and could potentially kill an innocent person.[271] 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.[272] By 2015, over fifty countries were reported to be researching battlefield robots.[273]AI tools make it easier for authoritarian governments to efficiently control their citizens in several ways. Face and voice recognition allow widespread surveillance. Machine learning, operating this data, can classify potential enemies of the state and 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 than liberal and decentralized systems such as markets. It lowers the cost and difficulty of digital warfare and advanced spyware.[274] All these technologies have been available since 2020 or earlierAI facial recognition systems are already being used for mass surveillance in China.[275][276]There are many other ways in which AI is expected to help bad actors, some of which can not be foreseen. For example, machine-learning AI is able to design tens of thousands of toxic molecules in a matter of hours.[277]Main articles: Workplace impact of artificial intelligence and Technological unemploymentEconomists have frequently highlighted the risks of redundancies from AI, and speculated about unemployment if there is no adequate social policy for full employment.[278]In the past, technology has tended to increase rather than reduce total employment, but economists acknowledge that "we're in uncharted territory" with AI.[279] 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.[280] 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".[p][282] 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.[278] In April 2023, it was reported that 70% of the jobs for Chinese video game illustrators had been eliminated by generative artificial intelligence.[283]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".[285] 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.[286]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.[287]Main article: Existential risk from artificial intelligenceIt 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".[288] 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.[q] 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 maximizer).[290] 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".[291] 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".[292]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 built on 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.[293]The opinions amongst experts and industry insiders are mixed, with sizable fractions both concerned and unconcerned by risk from eventual superintelligent AI.[294] Personalities such as Stephen Hawking, Bill Gates, and Elon Musk,[295] as well as AI pioneers such as Yoshua Bengio, Stuart Russell, Demis Hassabis, and Sam Altman, have expressed concerns about existential risk from AI.In May 2023, Geoffrey Hinton announced his resignation from Google in order to be able to "freely speak out about the risks of AI" without "considering how this impacts Google".[296] He notably mentioned risks of an AI takeover,[297] and stressed that in order to avoid the worst outcomes, establishing safety guidelines will require cooperation among those competing in use of AI.[298]In 2023, many leading AI experts endorsed 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".[299]Some other researchers were more optimistic. AI pioneer Jürgen 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".[300] 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."[301][302] 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."[303] Yann LeCun "scoffs at his peers' dystopian scenarios of supercharged misinformation and even, eventually, human extinction."[304] 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.[305] However, after 2016, the study of current and future risks and possible solutions became a serious area of research.[306]Main articles: Machine ethics, AI safety, Friendly artificial intelligence, Artificial moral agents, and Human CompatibleFriendly 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.[307]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.[308]The field of machine ethics is also called computational morality,[308]and was founded at an AAAI symposium in 2005.[309]Other approaches include Wendell Wallach's "artificial moral agents"[310] and Stuart J. Russell's three principles for developing provably beneficial machines.[311]Active organizations in the AI open-source community include Hugging Face,[312] Google,[313] EleutherAI and Meta.[314] Various AI models, such as Llama 2, Mistral or Stable Diffusion, have been made open-weight.[315][316] meaning that their architecture and trained parameters (the "weights") are publicly available. Open-weight models can be freely fine-tuned, which allows companies to specialize them with their own data and for their own use-case.[317] Open-weight models are useful for research and innovation but can also be misused. Since they can be fine-tuned, any built-in security measure, such as objecting to harmful requests, can be trained away until it becomes ineffective. Some researchers warn that future AI models may develop dangerous capabilities (such as the potential to drastically facilitate bioterrorism) and that once released on the Internet, they cannot be deleted everywhere if needed. They recommend pre-release audits and cost-benefit analyses.[318]Main article: Regulation of artificial intelligenceArtificial intelligence projects can be guided by ethical considerations during the design, development, and implementation of an AI system. An AI framework such as the Care and Act Framework, developed by the Alan Turing Institute and based on the SUM values, outlines four main ethical dimensions, defined as follows:[319][320]Respect the dignity of individual peopleConnect with other people sincerely, openly, and inclusivelyCare for the wellbeing of everyoneProtect social values, justice, and the public interestOther 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;[321] however, these principles are not without criticism, especially regarding the people chosen to contribute to these frameworks.[322]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.[323]The UK AI Safety Institute released in 2024 a testing toolset called 'Inspect' for AI safety evaluations available under an MIT open-source licence which is freely available on GitHub and can be improved with third-party packages. It can be used to evaluate AI models in a range of areas including core knowledge, ability to reason, and autonomous capabilities.[324]Main articles: Regulation of artificial intelligence, Regulation of algorithms, and AI safetyThe first global AI Safety Summit was held in the United Kingdom in November 2023 with a declaration calling for international cooperation.The regulation of artificial intelligence is the development of public sector policies and laws for promoting and regulating AI; it is therefore related to the broader regulation of algorithms.[325] The regulatory and policy landscape for AI is an emerging issue in jurisdictions globally.[326] 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.[327][328] Between 2016 and 2020, more than 30 countries adopted dedicated strategies for AI.[329] Most EU member states had released national AI strategies, as had Canada, China, India, Japan, Mauritius, the Russian Federation, Saudi Arabia, United Arab Emirates, U.S., and Vietnam. Others were in the process of elaborating their own AI strategy, including Bangladesh, Malaysia and Tunisia.[329] 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.[329] Henry Kissinger, Eric Schmidt, and Daniel Huttenlocher published a joint statement in November 2021 calling for a government commission to regulate AI.[330] In 2023, OpenAI leaders published recommendations for the governance of superintelligence, which they believe may happen in less than 10 years.[331] In 2023, the United Nations also launched an advisory body to provide recommendations on AI governance; the body comprises technology company executives, government officials and academics.[332] In 2024, the Council of Europe created the first international legally binding treaty on AI, called the "Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law". It was adopted by the European Union, the United States, the United Kingdom, and other signatories.[333]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".[334] A 2023 Reuters/Ipsos poll found that 61% of Americans agree, and 22% disagree, that AI poses risks to humanity.[334] 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".[335][336]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.[337] 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.[338][339] In May 2024 at the AI Seoul Summit, 16 global AI tech companies agreed to safety commitments on the development of AI.[340][341]Main article: History of artificial intelligenceFor a chronological guide, see Timeline of artificial intelligence.In 2024, at parties in China and the US numbered more than three-fourths of AI patents worldwide.[342] Though China had more AI patents, the US had 35% more patents per AI patent-applicant company than China.[342]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 any conceivable form of mathematical reasoning.[343][344] This, along with concurrent discoveries in cybernetics, information theory and neurobiology, led researchers to consider the possibility of building an "electronic brain".[r] They developed several areas of research that would become part of AI,[346] such as McCulloch and Pitts design for "artificial neurons" in 1943,[116] and Turing's influential 1950 paper 'Computing Machinery and Intelligence', which introduced the Turing test and showed that "machine intelligence" was plausible.[347][344]The field of AI research was founded at a workshop at Dartmouth College in 1956.[s][6] The attendees became the leaders of AI research in the 1960s.[t] They and their students produced programs that the press described as "astonishing";[u] computers were learning checkers strategies, solving word problems in algebra, proving logical theorems and speaking English.[v][7] Artificial intelligence laboratories were set up at a number of British and U.S. universities in the later 1950s and early 1960s.[344]Researchers in the 1960s and the 1970s were convinced that their methods would eventually succeed in creating a machine with general intelligence and considered this the goal of their field.[351] In 1965 Herbert Simon predicted, "machines will be capable, within twenty years, of doing any work a man can do".[352] In 1967 Marvin Minsky agreed, writing that "within a generation... the problem of creating 'artificial intelligence' will substantially be solved".[353] They had, however, underestimated the difficulty of the problem.[w] In 1974, both the U.S. and British governments cut off exploratory research in response to the criticism of Sir James Lighthill[355] and ongoing pressure from the U.S. Congress to fund more productive projects.[356] Minsky 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.[357] The "AI winter", a period when obtaining funding for AI projects was difficult, followed.[9]In the early 1980s, AI research was revived by the commercial success of expert systems,[358] 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.[9] 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.[10]Up to this point, most of AI's funding had gone to projects that used high-level symbols to represent mental objects like plans, goals, beliefs, and known facts. In the 1980s, some researchers began to doubt that this approach would be able to imitate all the processes of human cognition, especially perception, robotics, learning and pattern recognition,[359] and began to look into "sub-symbolic" approaches.[360] Rodney Brooks rejected "representation" in general and focused directly on engineering machines that move and survive.[x] Judea Pearl, Lotfi Zadeh, and others developed methods that handled incomplete and uncertain information by making reasonable guesses rather than precise logic.[87][365] But the most important development was the revival of "connectionism", including neural network research, by Geoffrey Hinton and others.[366] In 1990, Yann LeCun successfully showed that convolutional neural networks can recognize handwritten digits, the first of many successful applications of neural networks.[367]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).[368] By 2000, solutions developed by AI researchers were being widely used, although in the 1990s they were rarely described as "artificial intelligence" (a tendency known as the AI effect).[369]However, several academic researchers became concerned that AI was no longer pursuing its 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.[68]Deep learning began to dominate industry benchmarks in 2012 and was adopted throughout the field.[11] For many specific tasks, other methods were abandoned.[y]Deep learning's success was based on both hardware improvements (faster computers,[371] graphics processing units, cloud computing[372]) and access to large amounts of data[373] (including curated datasets,[372] such as ImageNet). Deep learning's success led to an enormous increase in interest and funding in AI.[z] The amount of machine learning research (measured by total publications) increased by 50% in the years 2015[2019.[329]The number of Google searches for the term "AI" accelerated in 2022.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.[306]In the late 2010s and early 2020s, AGI companies began to deliver programs that created enormous interest. In 2015, AlphaGo, developed by DeepMind, beat the world champion Go player. The program taught only the game's rules and developed a strategy by itself. GPT-3 is a large language model that was released in 2020 by OpenAI and is capable of generating high-quality human-like text.[375] ChatGPT, launched on November 30, 2022, became the fastest-growing consumer software application in history, gaining over 100 million users in two months.[375] It marked what is widely regarded as AI's breakout year, bringing it into the public consciousness.[376] These programs, and others, inspired an aggressive AI boom, where large companies began investing billions of dollars in AI research. According to AI Impacts, about $50 billion annually was invested in "AI" around 2022 in the U.S. alone and about 20% of the new U.S. Computer Science PhD graduates have specialized in "AI".[377] About 800,000 "AI"-related U.S. job openings existed in 2022.[378] According to PitchBook research, 22% of newly funded startups in 2024 claimed to be AI companies.[379]Main article: Philosophy of artificial intelligenceAlan Turing wrote in 1950 "I propose to consider the question 'can machines think'?"[383] He advised changing the question from whether a machine "thinks", to "whether or not it is possible for machinery to show intelligent behaviour".[383] He devised the Turing test, which measures the ability of a machine to simulate human conversation.[347] 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 but "it is usual to have a polite convention that everyone thinks."[388]Arthur Turing test can provide some evidence of intelligence, but it penalizes non-human intelligent behavior [385]Russell and Norvig agree with Turing that intelligence must be defined in terms of external behavior, not internal structure.[1] However, they are critical that the test requires the machine to imitate humans. "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.'"[386] AI founder John McCarthy agreed, writing that "Artificial intelligence is not, by definition, simulation of human intelligence".[387]McCarthy defines intelligence as "the computational part of the ability to achieve goals in the world".[388] Another AI founder, Marvin Minsky, similarly describes it as "the ability to solve hard problems".[389] The leading AI textbook defines it as the study of agents that perceive their environment and take actions that maximize their chances of achieving defined goals.[1] 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,[390] 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.Some authors have suggested in practice, that the definition of AI is vague and difficult to define, with contention as to whether classical algorithms should be re-categorised as AI,[391] with many companies during the early 2020s AI boom using the term as a marketing buzzword, often even if they did "not actually use AI in a material way".[392]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. Critics argue that these questions may have to be revisited by future generations of AI researchers.Symbolic AI (or "GOFAI")[394] 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."[395]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 very hard.[396] 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.[397] 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,[399][400] 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.Main article: Neats and scruffies"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,[401] but eventually was seen as irrelevant. Modern AI has elements of both.Main article: Soft computingFinding 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.Main articles: Weak artificial intelligence and Artificial general intelligenceAI 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.[402][403] 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 sub-field of artificial general intelligence studies this area exclusively.Main articles: Philosophy of artificial intelligence and Artificial consciousnessThere is no settled consensus in philosophy of mind on 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 "[the additional project of making a machine conscious in exactly the way humans are is not one that we are equipped to take on."[404] 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.Main articles: Hard problem of consciousness and Theory of mindDavid Chalmers identified two problems in understanding the mind, which he named the "hard" and "easy" problems of consciousness.[405] 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). While human information processing is easy to explain, 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.[406]Main articles: Computationalism and FunctionalismComputationalism 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 hardware and thus may be a solution to the mindbody 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.[407]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 challenges this claim with his Chinese room argument, which attempts to show that even a computer capable of perfectly simulating human behavior would not have a mind.[411]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.[412] 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.[413][414] Sapience (a set of capacities related to high intelligence, such as discernment or self-awareness) may provide another moral basis for AI rights.[413] Robot rights are also sometimes proposed as a practical way to integrate autonomous agents into society.[415]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.[416] 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 in society on their own.[417][418]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.[413][414]Superintelligence is a hypothetical agent that would possess intelligence far surpassing that of the brightest and most gifted human mind.[403] 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".[419]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.[420]Main article: TranshumanismRobot designer Hans Moravec, cyberneticist Kevin Warwick and inventor Ray Kurzweil have predicted that humans and machines may merge in the future into cyborgs that are more capable and powerful than either. This idea, called transhumanism, has roots in the writings of Aldous Huxley and Robert Ettinger.[421]Edward Fredkin argues that "artificial intelligence is the next step 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 1998 book Darwin Among the Machines: The Evolution of Global Intelligence.[422]Main article: Artificial intelligence in fictionThe word "robot" itself was coined by Karel apek in his 1921 play R.U.R., the title standing for "Rossum's Universal Robots".[424]Thought-capable artificial beings have appeared as storytelling devices since antiquity,[423] and have been a persistent theme in science fiction.[424]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.[425]Isaac Asimov introduced the Three Laws of Robotics in many stories, most notably with the "Multivac" super-intelligent computer. Asimov's laws are often brought up during lay discussions of machine ethics;[426] 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.[427]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 apek'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.[428]

Explainable AIArtificial intelligenceAI control problemComputational intelligenceEmergent algorithmGlossary of artificial intelligenceHypothetical technologyList of definitions of terms and concepts commonly used in the study of artificial intelligenceIntelligence amplificationUse of information technology to augment human intelligenceIntelligent agentSoftware agent which acts autonomouslyIntelligent automationSoftware process that combines robotic process automation and artificial intelligenceMachine learningStudy of algorithms that improve automatically through experienceAbductive learningProcess of solving new problems based on the solutions of similar past problemsComputational intelligenceAbility of a computer to learn a specific task from data or experimental observationDigital immortalityHypothetical concept of storing a personality in digital formEmergent algorithmAlgorithm exhibiting emergent behaviorFemale gendering of AI technologiesGender biases in digital technologyGlossary of artificial intelligenceList of definitions of terms and concepts commonly used in the study of artificial intelligenceIntelligence amplificationUse of information technology to augment human intelligenceMachine learningStudy of algorithms that improve automatically through experienceMachine translationUse of software for language translationMind uploadingProcess of mapping a brain and copying its state to a digital computerNarrow AIOntology (information science)RobotIntelligent agentWetware computer...

and Deepfake Porn Victims in the European Union: Findings From a Comparative Legal Study". Trauma, Violence, & Abuse. 25 (1): 117129. doi:10.1177/15248380221143772. ISSN1524-8380. PMID36565267.^ Singh, Suyesha; Nambiar, Vaishnavi (2024). "Role of Artificial Intelligence in the Prevention of Online Child Sexual Abuse: A Systematic Review of Literature". Journal of Applied Security Research. 19 (4): 586627. doi:10.1080/19361610.2024.2331885. ISSN1936-1610. Archived from the original on 9 December 2024. Retrieved 9 December 2024.^ Razi, Afsaneh; Kim, Seunghyun; Alsoubai, Ashwaq; Stringhini, Gianluca; Solorio, Thamar; De Choudhury, Munmun; Wisniewski, Pamela J. (13 October 2021). "A Human-Centered Systematic Literature Review of the Computational Approaches for Online Sexual Risk Detection". Proceedings of the ACM on Human-Computer Interaction. 5 (CSCW2): 138. doi:10.1145/3479609. ISSN2573-0142. Archived from the original on 9 December 2024. Retrieved 9 December 2024.^ Ransbotham, Sam; Kiron, David; Gerbert, Philipp; Reeves, Martin (6 September 2017). "Reshaping Business With Artificial Intelligence". MIT Sloan Management Review. Archived from the original on 13 February 2024.^ Sun, Yuran; Zhao, Xilei; Lovreglio, Ruggiero; Kuligowski, Erica (1 January 2024); Naser, M. Z. (ed.), "8 AI for large-scale evacuation modeling: promises and challenges", Interpretable Machine Learning for the Analysis, Design, Assessment, and Informed Decision Making for Civil Infrastructure, Woodhead Publishing Series in Civil and Structural Engineering, Woodhead Publishing, pp.185244, ISBN978-0-12-824073-1, archived from the original on 19 May 2024, retrieved 28 June 2024^ Gomaa, Islam; Adelzadeh, Masoud; Gwynne, Steven; Spencer, Bruce; Ko, Yoon; Bnichou, Noureddine; Ma, Chunyan; Elsagan, Nour; Duong, Dana; Zalok, Ehab; Kinateder, Max (1 November 2021). "A Framework for Intelligent Fire Detection and Evacuation System". Fire Technology. 57 (6): 31793185. doi:10.1007/s10694-021-01157-3. ISSN1572-8099. Archived from the original on 5 October 2024.^ Zhao, Xilei; Lovreglio, Ruggiero; Nilsson, Daniel (1 May 2020). "Modelling and interpreting pre-evacuation decision-making using machine learning". Automation in Construction. 113: 103140. doi:10.1016/j.autcon.2020.103140. hdl:10179/17315. ISSN0926-5805. Archived from the original on 19 May 2024. Retrieved 5 October 2024.^ "India's latest election embraced AI technology. Here are some ways it was used constructively". PBS News. 12 June 2024. Archived from the original on 17 September 2024. Retrieved 28 October 2024.^ Müller, Vincent C. (30 April 2020). "Ethics of Artificial Intelligence and Robotics". Stanford Encyclopedia Archive. Archived from the original on 5 October 2024. Retrieved 5 October 2024.^ Simonite (2016).^ Russell & Norvig (2021), p.987.^ Laskowski (2023).^ GAO (2022).^ Valinsky (2019).^ Russell & Norvig (2021), p.991.^ Hiller (2023), p.63.^ Vincent (2022).^ Kopel, Matthew. "Copyright Services: Fair Use". Cornell University Library. Archived from the original on 26 September 2024. Retrieved 26 April 2024.^ Burgess, Matt. "How to Stop Your Data From Being Used to Train AI". Wired. ISSN1059-1028. Archived from the original on 26 April 2024.^ Reisner (2023).^ Alter & Harris (2023).^ "Getting the Innovation Ecosystem Ready for AI. An IP policy toolkit" (PDF). WIPO.^ Hammond, George (27 December 2023). "Big Tech is spending more than VC firms on AI startups". Ars Technica. Archived from the original on 10 January 2024.^ Wong, Matteo (24 October 2023). "The Future of AI Is GOMA". The Atlantic. Archived from the original on 5 January 2024.^ "Big tech and the pursuit of AI dominance". The Economist. 26 March 2023. Archived from the original on 29 December 2023.^ Fung, Brian (19 December 2023). "Where the battle to dominate AI may be won". CNN Business. Archived from the original on 13 January 2024.^ Metz, Cade (5 July 2023). "In the Age of A.I., Tech's Little Guys Need Big Friends". The New York Times. Archived from the original on 8 July 2024. Retrieved 17 August 2023.^ "Electricity 2024 Analysis". IEA. 24 January 2024.^ Calvert, Brian (28 March 2024). "AI already uses as much energy as a small country. It's only the beginning". Vox. New York, New York.