WIKIPEDIA
The Free Encyclopedia

Artificial intelligence

Artificial intelligence (**AI**), in its broadest sense, is intelligence exhibited by machines, particularly computer systems. 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] Such machines may be called AIs.

Some high-profile applications of AI include advanced web search engines (e.g., Google Search); recommendation systems (used by YouTube, Amazon, and Netflix); interacting via human speech (e.g., Google Assistant, Siri, and Alexa); autonomous vehicles (e.g., Waymo); generative and creative tools (e.g., ChatGPT, Apple Intelligence, 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]

Alan Turing was the first person to conduct substantial research in the field that he called "machine intelligence".[4] Artificial intelligence was founded as an academic discipline in 1956,[5] by those now considered the founding fathers of AI: John McCarthy, Marvin Minksy, Nathaniel Rochester, and Claude Shannon.[6][7] The field went through multiple cycles of optimism,[8][9] followed by periods of disappointment and loss of funding, known as AI winter.[10][11] Funding and interest vastly increased after 2012 when deep learning surpassed all previous AI techniques,[12] and after 2017 with the transformer architecture.[13] This led to the AI boom of the early 2020s, with companies, universities, and laboratories overwhelmingly based in the United States pioneering significant advances in artificial intelligence.[14]

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

The various subfields of AI research are centered around particular goals and the use of particular tools. The traditional goals of AI research include reasoning, knowledge representation, planning, learning, natural language processing, perception, and support for robotics.[a] General intelligence—the ability to complete any task performable by a human on an at least equal level—is among the field's long-term goals.[15]

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.[16]

Goals

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]

Reasoning and problem-solving

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.[20] Accurate and efficient reasoning is an unsolved problem.

Knowledge representation

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

Knowledge representation and knowledge engineering[21] allow AI programs to answer questions intelligently and make deductions about real-world facts. Formal knowledge representations are used in content-based indexing and retrieval,[22] scene interpretation,[23] clinical decision support,[24] knowledge discovery (mining "interesting" and actionable inferences from large databases),[25] and other areas.[26]

A knowledge base is a body of knowledge represented in a form that can be used by a program. An ontology is the set of objects, relations, concepts, and properties used by a particular domain of knowledge.[27] Knowledge bases need to represent things such as objects, properties, categories, and relations between objects;[28] situations, events, states, and time;[29] causes and effects;[30] knowledge about knowledge (what we know about what other people know);[31] default reasoning (things that humans assume are true until they are told differently and will remain true even when other facts are changing);[32] and many other aspects and domains of knowledge.

Among the most difficult problems in knowledge representation are the breadth of commonsense knowledge (the set of atomic facts that the average person knows is enormous);[33] and the sub-symbolic form of most commonsense knowledge (much of what people know is not represented as "facts" or "statements" that they could express verbally).[20] There is also the difficulty of knowledge acquisition, the problem of obtaining knowledge for AI applications.[c]

Planning and decision-making

An "agent" is anything that perceives and takes actions in the world. A rational agent has goals or preferences and takes actions to make them happen.[d][36] In automated planning, the agent has a specific goal.[37] In automated decision-making, the agent has preferences—there are some situations it would prefer to be in, and some situations it is trying to avoid. The decision-making agent assigns a number to each situation (called the "utility") that measures how much the agent prefers it. For each possible action, it can calculate the "expected utility": the utility of all possible outcomes of the action, weighted by the probability that the outcome will occur. It can then choose the action with the maximum expected utility.[38]

In classical planning, the agent knows exactly what the effect of any action will be.[39] In most real-world problems, however, the agent may not be certain about the situation they are in (it is "unknown" or "unobservable") and it may not know for certain what will happen after each possible action (it is not "deterministic"). It must choose an action by making a probabilistic guess and then reassess the situation to see if the action worked.[40]

In some problems, the agent's preferences may be uncertain, especially if there are other agents or humans involved. These can be learned (e.g., with inverse reinforcement learning), or the agent can seek information to improve its preferences.[41] Information value theory can be used to weigh the value of exploratory or experimental actions.[42] The space of possible future actions and situations is typically intractably large, so the agents must take actions and evaluate situations while being uncertain of what the outcome will be.

A Markov decision process has a transition model that describes the probability that a particular action will change the state in a particular way and a reward function that supplies the utility of each state and the cost of each action. A policy associates a decision with each possible state. The policy could be calculated (e.g., by iteration), be heuristic, or it can be learned.[43]

Game theory describes the rational behavior of multiple interacting agents and is used in AI programs that make decisions that involve other agents.[44]

Learning

Machine learning is the study of programs that can improve their performance on a given task automatically.[45] It has been a part of AI from the beginning.[e]

There are several kinds of machine learning. Unsupervised learning analyzes a stream of data and finds patterns and makes predictions without any other guidance.[48] Supervised learning requires a human to label the input data first, and comes in two main varieties: classification (where the program must learn to predict what category the input belongs in) and regression (where the program must deduce a numeric function based on numeric input).[49]