**Machine learning** (**ML**) is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve performance on some set of tasks.[[1]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-1) It is seen as a part of [artificial intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence). Machine learning algorithms build a model based on sample data, known as [training data](https://en.wikipedia.org/wiki/Training_data), in order to make predictions or decisions without being explicitly programmed to do so.[[2]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-2) Machine learning algorithms are used in a wide variety of applications, such as in medicine, [email filtering](https://en.wikipedia.org/wiki/Email_filtering), [speech recognition](https://en.wikipedia.org/wiki/Speech_recognition), and [computer vision](https://en.wikipedia.org/wiki/Computer_vision), where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.[[3]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-tvt-3)

A subset of machine learning is closely related to [computational statistics](https://en.wikipedia.org/wiki/Computational_statistics), which focuses on making predictions using computers, but not all machine learning is statistical learning. The study of [mathematical optimization](https://en.wikipedia.org/wiki/Mathematical_optimization) delivers methods, theory and application domains to the field of machine learning. [Data mining](https://en.wikipedia.org/wiki/Data_mining) is a related field of study, focusing on [exploratory data analysis](https://en.wikipedia.org/wiki/Exploratory_data_analysis) through [unsupervised learning](https://en.wikipedia.org/wiki/Unsupervised_learning).[[5]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-5)[[6]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-6) Some implementations of machine learning use data and [neural networks](https://en.wikipedia.org/wiki/Neural_networks) in a way that mimics the working of a biological brain.[[7]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-7)[[8]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-8) In its application across business problems, machine learning is also referred to as [predictive analytics](https://en.wikipedia.org/wiki/Predictive_analytics).

Overview

Learning algorithms work on the basis that strategies, algorithms, and inferences that worked well in the past are likely to continue working well in the future. These inferences can be obvious, such as "since the sun rose every morning for the last 10,000 days, it will probably rise tomorrow morning as well". They can be nuanced, such as "X% of [families](https://en.wikipedia.org/wiki/Family_(biology)) have geographically separate species with color variants, so there is a Y% chance that undiscovered [black swans](https://en.wikipedia.org/wiki/Black_swan_theory) exist".[[9]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-FOOTNOTEDomingos2015Chapter_6,_Chapter_7-9)

Machine learning programs can perform tasks without being explicitly programmed to do so. It involves computers learning from data provided so that they carry out certain tasks. For simple tasks assigned to computers, it is possible to program algorithms telling the machine how to execute all steps required to solve the problem at hand; on the computer's part, no learning is needed. For more advanced tasks, it can be challenging for a human to manually create the needed algorithms. In practice, it can turn out to be more effective to help the machine develop its own algorithm, rather than having human programmers specify every needed step.[[10]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-Alpaydin2020-10)

The discipline of machine learning employs various approaches to teach computers to accomplish tasks where no fully satisfactory algorithm is available. In cases where vast numbers of potential answers exist, one approach is to label some of the correct answers as valid. This can then be used as training data for the computer to improve the algorithm(s) it uses to determine correct answers. For example, to train a system for the task of digital character recognition, the [MNIST](https://en.wikipedia.org/wiki/MNIST_database) dataset of handwritten digits has often been used.[[10]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-Alpaydin2020-10)

### Artificial intelligence

As a scientific endeavor, machine learning grew out of the quest for artificial intelligence. In the early days of AI as an [academic discipline](https://en.wikipedia.org/wiki/Discipline_(academia)), some researchers were interested in having machines learn from data. They attempted to approach the problem with various symbolic methods, as well as what was then termed "[neural networks](https://en.wikipedia.org/wiki/Neural_network)"; these were mostly [perceptrons](https://en.wikipedia.org/wiki/Perceptron" \o "Perceptron) and [other models](https://en.wikipedia.org/wiki/ADALINE) that were later found to be reinventions of the [generalized linear models](https://en.wikipedia.org/wiki/Generalized_linear_model) of statistics.[[24]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-24) [Probabilistic reasoning](https://en.wikipedia.org/wiki/Probabilistic_reasoning) was also employed, especially in [automated medical diagnosis](https://en.wikipedia.org/wiki/Automated_medical_diagnosis).[[25]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-aima-25): 488

However, an increasing emphasis on the [logical, knowledge-based approach](https://en.wikipedia.org/wiki/GOFAI) caused a rift between AI and machine learning. Probabilistic systems were plagued by theoretical and practical problems of data acquisition and representation.[[25]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-aima-25): 488 By 1980, [expert systems](https://en.wikipedia.org/wiki/Expert_system) had come to dominate AI, and statistics was out of favor.[[26]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-changing-26) Work on symbolic/knowledge-based learning did continue within AI, leading to [inductive logic programming](https://en.wikipedia.org/wiki/Inductive_logic_programming), but the more statistical line of research was now outside the field of AI proper, in [pattern recognition](https://en.wikipedia.org/wiki/Pattern_recognition) and [information retrieval](https://en.wikipedia.org/wiki/Information_retrieval).[[25]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-aima-25): 708–710, 755 Neural networks research had been abandoned by AI and [computer science](https://en.wikipedia.org/wiki/Computer_science) around the same time. This line, too, was continued outside the AI/CS field, as "[connectionism](https://en.wikipedia.org/wiki/Connectionism)", by researchers from other disciplines including [Hopfield](https://en.wikipedia.org/wiki/John_Hopfield), [Rumelhart](https://en.wikipedia.org/wiki/David_Rumelhart" \o "David Rumelhart) and [Hinton](https://en.wikipedia.org/wiki/Geoff_Hinton). Their main success came in the mid-1980s with the reinvention of [backpropagation](https://en.wikipedia.org/wiki/Backpropagation" \o "Backpropagation).[[25]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-aima-25): 25

Machine learning (ML), reorganized as a separate field, started to flourish in the 1990s. The field changed its goal from achieving artificial intelligence to tackling solvable problems of a practical nature. It shifted focus away from the [symbolic approaches](https://en.wikipedia.org/wiki/Symbolic_artificial_intelligence) it had inherited from AI, and toward methods and models borrowed from statistics, [fuzzy logic](https://en.wikipedia.org/wiki/Fuzzy_logic), and [probability theory](https://en.wikipedia.org/wiki/Probability_theory).[[26]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-changing-26)

The difference between ML and AI is frequently misunderstood. ML learns and predicts based on passive observations, whereas AI implies an agent interacting with the environment to learn and take actions that maximize its chance of successfully achieving its goals.[[27]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-:3-27)

As of 2020, many sources continue to assert that ML remains a subfield of AI.[[28]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-:4-28)[[29]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-islr-29)[[26]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-changing-26) Others have the view that not all ML is part of AI, but only an 'intelligent subset' of ML should be considered AI.[[4]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-bishop2006-4)[[30]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-:5-30)[[31]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-alpaydin-31)

### Data mining

Machine learning and [data mining](https://en.wikipedia.org/wiki/Data_mining) often employ the same methods and overlap significantly, but while machine learning focuses on prediction, based on *known* properties learned from the training data, [data mining](https://en.wikipedia.org/wiki/Data_mining) focuses on the [discovery](https://en.wikipedia.org/wiki/Discovery_(observation)) of (previously) *unknown* properties in the data (this is the analysis step of [knowledge discovery](https://en.wikipedia.org/wiki/Knowledge_discovery) in databases). Data mining uses many machine learning methods, but with different goals; on the other hand, machine learning also employs data mining methods as "unsupervised learning" or as a preprocessing step to improve learner accuracy. Much of the confusion between these two research communities (which do often have separate conferences and separate journals, [ECML PKDD](https://en.wikipedia.org/wiki/ECML_PKDD) being a major exception) comes from the basic assumptions they work with: in machine learning, performance is usually evaluated with respect to the ability to *reproduce known* knowledge, while in knowledge discovery and data mining (KDD) the key task is the discovery of previously *unknown* knowledge. Evaluated with respect to known knowledge, an uninformed (unsupervised) method will easily be outperformed by other supervised methods, while in a typical KDD task, supervised methods cannot be used due to the unavailability of training data.

### Optimization

Machine learning also has intimate ties to [optimization](https://en.wikipedia.org/wiki/Mathematical_optimization): many learning problems are formulated as minimization of some [loss function](https://en.wikipedia.org/wiki/Loss_function) on a training set of examples. Loss functions express the discrepancy between the predictions of the model being trained and the actual problem instances (for example, in classification, one wants to assign a label to instances, and models are trained to correctly predict the pre-assigned labels of a set of examples).[[32]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-32)

### Generalization

The difference between optimization and machine learning arises from the goal of generalization: while optimization algorithms can minimize the loss on a training set, machine learning is concerned with minimizing the loss on unseen samples. Characterizing the generalization of various learning algorithms is an active topic of current research, especially for [deep learning](https://en.wikipedia.org/wiki/Deep_learning) algorithms.

### Statistics

Machine learning and [statistics](https://en.wikipedia.org/wiki/Statistics) are closely related fields in terms of methods, but distinct in their principal goal: statistics draws population [inferences](https://en.wikipedia.org/wiki/Statistical_inference) from a [sample](https://en.wikipedia.org/wiki/Sample_(statistics)), while machine learning finds generalizable predictive patterns.[[33]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-33) According to [Michael I. Jordan](https://en.wikipedia.org/wiki/Michael_I._Jordan), the ideas of machine learning, from methodological principles to theoretical tools, have had a long pre-history in statistics.[[34]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-mi_jordan_ama-34) He also suggested the term [data science](https://en.wikipedia.org/wiki/Data_science) as a placeholder to call the overall field.[[34]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-mi_jordan_ama-34)

[Leo Breiman](https://en.wikipedia.org/wiki/Leo_Breiman) distinguished two statistical modeling paradigms: data model and algorithmic model,[[28]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-:4-28) wherein "algorithmic model" means more or less the machine learning algorithms like [Random forest](https://en.wikipedia.org/wiki/Random_forest).

Some statisticians have adopted methods from machine learning, leading to a combined field that they call *statistical learning*.[[29]](https://en.wikipedia.org/wiki/Machine_learning#cite_note-islr-29)