

# History and relationships to other fields<sup>[edit]</sup>

See also: *Timeline of machine learning*

Arthur Samuel, an American pioneer in the field of [computer gaming](#) and [artificial intelligence](#), coined the term "Machine Learning" in 1959 while at IBM<sup>[8]</sup>. As a scientific endeavour, machine learning grew out of the quest for artificial intelligence. Already in the early days of AI as an academic discipline, some researchers were interested in having machines learn from data. They attempted to approach the problem with various symbolic methods, as well as what were then termed "[neural networks](#)"; these were mostly [perceptrons](#) and [other models](#) that were later found to be reinventions of the [generalized linear models](#) of statistics.<sup>[9]</sup> Probabilistic reasoning was also employed, especially in automated medical diagnosis.<sup>[10]:488</sup>

However, an increasing emphasis on the [logical, knowledge-based approach](#) caused a rift between AI and machine learning. Probabilistic systems were plagued by theoretical and practical problems of data acquisition and representation.<sup>[10]:488</sup> By 1980, [expert systems](#) had come to dominate AI, and statistics was out of favor.<sup>[11]</sup> Work on symbolic/knowledge-based learning did continue within AI, leading to [inductive logic programming](#), but the more statistical line of research was now outside the field of AI proper, in [pattern recognition](#) and [information retrieval](#).<sup>[10]:708–710; 755</sup> Neural networks research had been abandoned by AI and [computer science](#) around the same time. This line, too, was continued outside the AI/CS field, as "[connectionism](#)", by researchers from other disciplines including [Hopfield](#), [Rumelhart](#) and [Hinton](#). Their main success came in the mid-1980s with the reinvention of [backpropagation](#).<sup>[10]:25</sup>

Machine learning, 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 it had inherited from AI, and toward methods and models borrowed from statistics and [probability theory](#).<sup>[11]</sup> It also benefited from the increasing availability of digitized information, and the ability to distribute it via the [Internet](#).

## Relation to data mining<sup>[edit]</sup>

Machine learning and [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](#) focuses on the [discovery](#) of (previously) *unknown* properties in the data (this is the analysis step of [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](#) 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.

## Relation to optimization<sup>[edit]</sup>

Machine learning also has intimate ties to [optimization](#): many learning problems are formulated as minimization of some [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). The difference between the two fields 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.[12]

## **Relation to statistics**[\[edit\]](#)

Machine learning and [statistics](#) are closely related fields. According to [Michael I. Jordan](#), the ideas of machine learning, from methodological principles to theoretical tools, have had a long pre-history in statistics.[13] He also suggested the term [data science](#) as a placeholder to call the overall field.[13]

[Leo Breiman](#) distinguished two statistical modelling paradigms: data model and algorithmic model,[14] wherein "algorithmic model" means more or less the machine learning algorithms like [Random forest](#).

Some statisticians have adopted methods from machine learning, leading to a combined field that they call *statistical learning*.[\[15\]](#)