Machine learning

Machine learning (**ML**) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions.[1] Advances in the field of deep learning have allowed neural networks to surpass many previous approaches in performance.[2]

ML finds application in many fields, including natural language processing, computer vision, speech recognition, email filtering, agriculture, and medicine.[3][4] The application of ML to business problems is known as predictive analytics.

Statistics and mathematical optimization (mathematical programming) methods comprise the foundations of machine learning. Data mining is a related field of study, focusing on exploratory data analysis (EDA) via unsupervised learning.[6][7]

From a theoretical viewpoint, probably approximately correct (PAC) learning provides a framework for describing machine learning.

History

The term *machine learning* was coined in 1959 by Arthur Samuel, an IBM employee and pioneer in the field of computer gaming and artificial intelligence.[8][9] The synonym *self-teaching computers* was also used in this time period.[10][11]

Although the earliest machine learning model was introduced in the 1950s when Arthur Samuel invented a program that calculated the winning chance in checkers for each side, the history of machine learning roots back to decades of human desire and effort to study human cognitive processes.[12] In 1949, Canadian psychologist Donald Hebb published the book *The Organization of Behavior*, in which he introduced a theoretical neural structure formed by certain interactions among nerve cells.[13] Hebb's model of neurons interacting with one another set a groundwork for how AIs and machine learning algorithms work under nodes, or artificial neurons used by computers to communicate data.[12] Other researchers who have studied human cognitive systems contributed to the modern machine learning technologies as well, including logician Walter Pitts and Warren McCulloch, who proposed the early mathematical models of neural networks to come up with algorithms that mirror human thought processes.[12]

By the early 1960s, an experimental "learning machine" with punched tape memory, called Cybertron, had been developed by Raytheon Company to analyze sonar signals, electrocardiograms, and speech patterns using rudimentary reinforcement learning. It was repetitively "trained" by a human operator/teacher to recognize patterns and equipped with a "goof" button to cause it to reevaluate incorrect decisions.[14] A representative book on research into machine learning during the 1960s was

Nilsson's book on Learning Machines, dealing mostly with machine learning for pattern classification.[15] Interest related to pattern recognition continued into the 1970s, as described by Duda and Hart in 1973.[16] In 1981 a report was given on using teaching strategies so that an artificial neural network learns to recognize 40 characters (26 letters, 10 digits, and 4 special symbols) from a computer terminal.[17]

Tom M. Mitchell provided a widely quoted, more formal definition of the algorithms studied in the machine learning field: "A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P* if its performance at tasks in *T*, as measured by *P*, improves with experience *E*."[18] This definition of the tasks in which machine learning is concerned offers a fundamentally operational definition rather than defining the field in cognitive terms. This follows Alan Turing's proposal in his paper "Computing Machinery and Intelligence", in which the question "Can machines think?" is replaced with the question "Can machines do what we (as thinking entities) can do?".[19]
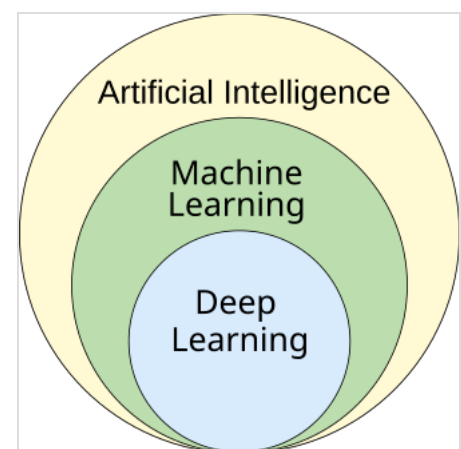
Modern-day machine learning has two objectives. One is to classify data based on models which have been developed; the other purpose is to make predictions for future outcomes based on these models. A hypothetical algorithm specific to classifying data may use computer vision of moles coupled with supervised learning in order to train it to classify the cancerous moles. A machine learning algorithm for stock trading may inform the trader of future potential predictions.[20]

Relationships to other fields

Artificial intelligence

As a scientific endeavor, machine learning grew out of the quest for artificial intelligence (AI). 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.[22] Probabilistic reasoning was also employed, especially in automated medical diagnosis.[23]:488

Machine learning as subfield of AI[21]

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.[23]:488 By 1980, expert systems had come to dominate AI, and statistics was out of favor.[24] Work on symbolic/knowledge-based learning did continue within AI, leading to inductive logic programming(ILP), but the more statistical line of research was now outside the field of AI proper, in pattern recognition and information retrieval.[23]:708–710,755 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 John Hopfield, David Rumelhart, and Geoffrey Hinton. Their main success came in the mid-1980s with the reinvention of backpropagation.[23]:25

Machine learning (ML), reorganized and recognized as its own 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, fuzzy logic, and probability theory.[24]

Data compression

There is a close connection between machine learning and compression. A system that predicts the posterior probabilities of a sequence given its entire history can be used for optimal data compression (by using arithmetic coding on the output distribution). Conversely, an optimal compressor can be used for prediction (by finding the symbol that compresses best, given the previous history). This equivalence has been used as a justification for using data compression as a benchmark for "general intelligence".[25][26][27]

An alternative view can show compression algorithms implicitly map strings into implicit feature space vectors, and compression-based similarity measures compute similarity within these feature spaces. For each compressor C(.) we define an associated vector space ℵ, such that C(.) maps an input string x, corresponding to the vector norm $||\sim x||$. An exhaustive examination of the feature spaces underlying all compression algorithms is precluded by space; instead, feature vectors chooses to examine three representative lossless compression methods, LZW, LZ77, and PPM.[28]

According to AIXI theory, a connection more directly explained in Hutter Prize, the best possible compression of x is the smallest possible software that generates x. For example, in that model, a zip file's compressed size includes both the zip file and the unzipping software, since you can not unzip it without both, but there may be an even smaller combined form.

Examples of AI-powered audio/video compression software include NVIDIA Maxine, AIVC.[29] Examples of software that can perform AI-powered image compression include OpenCV, TensorFlow, MATLAB's Image Processing Toolbox (IPT) and High-Fidelity Generative Image Compression.[30]

In unsupervised machine learning, k-means clustering can be utilized to compress data by grouping similar data points into clusters. This technique simplifies handling extensive datasets that lack predefined labels and finds widespread use in fields such as image compression.[31]

Data compression aims to reduce the size of data files, enhancing storage efficiency and speeding up data transmission. K-means clustering, an unsupervised machine learning algorithm, is employed to partition a dataset into a specified number of clusters, k, each represented by the centroid of its points. This process condenses extensive datasets into a more compact set of representative points. Particularly beneficial in image and signal processing, k-means clustering aids in data reduction by replacing groups of data points with their centroids, thereby preserving the core information of the original data while significantly decreasing the required storage space.[32]

Large language models (LLMs) are also capable of lossless data compression, as demonstrated by DeepMind's research with the Chinchilla 70B model. Developed by DeepMind, Chinchilla 70B effectively compressed data, outperforming conventional methods such as Portable Network Graphics (PNG) for images and Free Lossless Audio Codec (FLAC) for audio. It achieved compression of image and audio data to 43.4% and 16.4% of their original sizes, respectively.[33]

Data mining

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.

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 preassigned labels of a set of examples).[34]

Generalization

Characterizing the generalization of various learning algorithms is an active topic of current research, especially for deep learning algorithms.

Statistics

Machine learning and statistics are closely related fields in terms of methods, but distinct in their principal goal: statistics draws population inferences from a sample, while machine learning finds generalizable predictive patterns.[35] According to Michael I. Jordan, the ideas of machine learning, from methodological principles to theoretical tools, have had a long pre-history in statistics.[36] He also suggested the term data science as a placeholder to call the overall field.[36]

Conventional statistical analyses require the a priori selection of a model most suitable for the study data set. In addition, only significant or theoretically relevant variables based on previous experience are included for analysis. In contrast, machine learning is not built on a pre-structured model; rather, the data shape the model by detecting underlying patterns. The more variables (input) used to train the model, the more accurate the ultimate model will be.[37]

Leo Breiman distinguished two statistical modeling paradigms: data model and algorithmic model,[38] 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*.[39]

Statistical physics

Analytical and computational techniques derived from deep-rooted physics of disordered systems can be extended to large-scale problems, including machine learning, e.g., to analyze the weight space of deep neural networks.[40] Statistical physics is thus finding applications in the area of medical diagnostics.[41]

Theory

A core objective of a learner is to generalize from its experience.[5][42] Generalization in this context is the ability of a learning machine to perform accurately on new, unseen examples/tasks after having experienced a learning data set. The training examples come from some generally unknown probability distribution (considered representative of the space of occurrences) and the learner has to build a general model about this space that enables it to produce sufficiently accurate predictions in new cases.

The computational analysis of machine learning algorithms and their performance is a branch of theoretical computer science known as computational learning theory via the Probably Approximately Correct Learning (PAC) model. Because training sets are finite and the future is uncertain, learning theory usually does not yield guarantees of the performance of algorithms. Instead, probabilistic bounds on the performance are quite common. The bias–variance decomposition is one way to quantify generalization error.
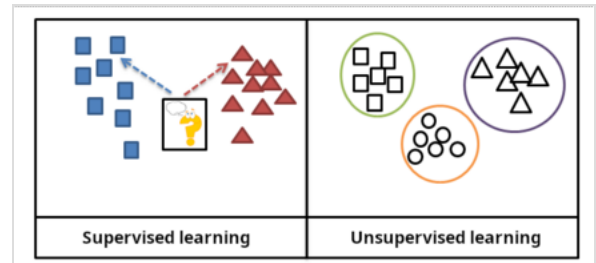
For the best performance in the context of generalization, the complexity of the hypothesis should match the complexity of the function underlying the data. If the hypothesis is less complex than the function, then the model has under fitted the data. If the complexity of the model is increased in response, then the training error decreases. But if the hypothesis is too complex, then the model is subject to overfitting and generalization will be poorer.[43]

In addition to performance bounds, learning theorists study the time complexity and feasibility of learning. In computational learning theory, a computation is considered feasible if it can be done in polynomial time. There are two kinds of time complexity results: Positive results show that a certain class of functions can be learned in polynomial time. Negative results show that certain classes cannot be learned in polynomial time.

Approaches

Machine learning approaches are traditionally divided into three broad categories, which correspond to learning paradigms, depending on the nature of the "signal" or "feedback" available to the learning system:
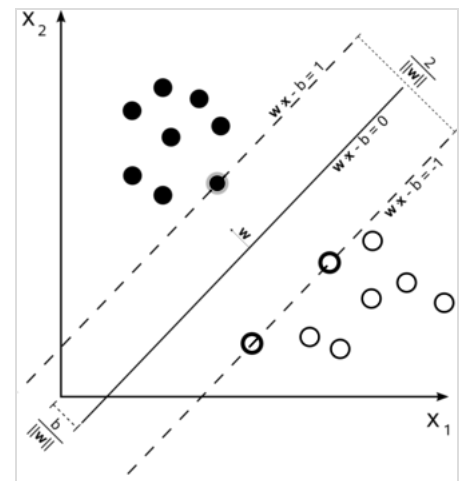

In supervised learning, the training data is labeled with the expected answers, while in unsupervised learning, the model identifies patterns or structures in unlabeled data.

- Supervised learning: The computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs.
- Unsupervised learning: No labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning).
- Reinforcement learning: A computer program interacts with a dynamic environment in which it must perform a certain goal (such as driving a vehicle or playing a game against an opponent). As it navigates its problem space, the program is provided feedback that's analogous to rewards, which it tries to maximize.[5]

Although each algorithm has advantages and limitations, no single algorithm works for all problems.[44][45][46]

Supervised learning

Supervised learning algorithms build a mathematical model of a set of data that contains both the inputs and the desired outputs.[47] The data, known as training data, consists of a set of training examples. Each training example has one or more inputs and the desired output, also known as a supervisory signal. In the mathematical model, each training example is represented by an array or vector, sometimes called a feature vector, and the training data is represented by a matrix. Through iterative optimization of an objective function, supervised learning algorithms learn a function that can be used to predict the output associated with new inputs.[48] An optimal function allows the algorithm to correctly determine the output for inputs that were not a part of the training data. An algorithm that improves the accuracy of its outputs or predictions over time is said to have learned to perform that task.[18]

A support-vector machine is a supervised learning model that divides the data into regions separated by a linear boundary. Here, the linear boundary divides the black circles from the white.

Types of supervised-learning algorithms include active learning, classification and regression.[49] Classification algorithms are used when the outputs are restricted to a limited set of values, and regression algorithms are used when the outputs may have any numerical value within a range. As an example, for a classification algorithm that filters emails, the input would be an incoming email, and the output would be the name of the folder in which to file the email. Examples of regression would be predicting the height of a person, or the future temperature. [50]

Similarity learning is an area of supervised machine learning closely related to regression and classification, but the goal is to learn from examples using a similarity function that measures how similar or related two objects are. It has applications in ranking, recommendation systems, visual identity tracking, face verification, and speaker verification.

Unsupervised learning

Unsupervised learning algorithms find structures in data that has not been labeled, classified or categorized. Instead of responding to feedback, unsupervised learning algorithms identify commonalities in the data and react based on the presence or absence of such commonalities in each new piece of data. Central applications of unsupervised machine learning include clustering, dimensionality reduction,[7] and density estimation.[51]

Cluster analysis is the assignment of a set of observations into subsets (called *clusters*) so that observations within the same cluster are similar according to one or more predesignated criteria, while observations drawn from different clusters are dissimilar. Different clustering techniques make different assumptions on the structure of the data, often defined by some *similarity metric* and evaluated, for example, by *internal compactness*, or the similarity between members of the same cluster, and *separation*, the difference between clusters. Other methods are based on *estimated density* and *graph connectivity*.

A special type of unsupervised learning called, self-supervised learning involves training a model by generating the supervisory signal from the data itself.[52][53]

Semi-supervised learning

Semi-supervised learning falls between unsupervised learning (without any labeled training data) and supervised learning (with completely labeled training data). Some of the training examples are missing training labels, yet many machine-learning researchers have found that unlabeled data, when used in conjunction with a small amount of labeled data, can produce a considerable improvement in learning accuracy.

In weakly supervised learning, the training labels are noisy, limited, or imprecise; however, these labels are often cheaper to obtain, resulting in larger effective training sets.[54]

Reinforcement learning

Reinforcement learning is an area of machine learning concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward. Due to its generality, the field is studied in many other disciplines, such as game theory, control theory, operations research, information theory, simulation-based optimization, multi-agent systems, swarm intelligence, statistics and genetic algorithms. In reinforcement learning, the environment is typically represented as a Markov decision process (MDP). Many reinforcements learning algorithms use dynamic programming techniques.[55] Reinforcement learning algorithms do not assume knowledge

of an exact mathematical model of the MDP and are used when exact models are infeasible. Reinforcement learning algorithms are used in autonomous vehicles or in learning to play a game against a human opponent.

Dimensionality reduction

Dimensionality reduction is a process of reducing the number of random variables under consideration by obtaining a set of principal variables.[56] In other words, it is a process of reducing the dimension of the feature set, also called the "number of features". Most of the dimensionality reduction techniques can be considered as either feature elimination or extraction. One of the popular methods of dimensionality reduction is principal component analysis (PCA). PCA involves changing higher-dimensional data (e.g., 3D) to a smaller space (e.g., 2D). The manifold hypothesis proposes that high-dimensional data sets lie along low-dimensional manifolds, and many dimensionality reduction techniques make this assumption, leading to the area of manifold learning and manifold regularization.
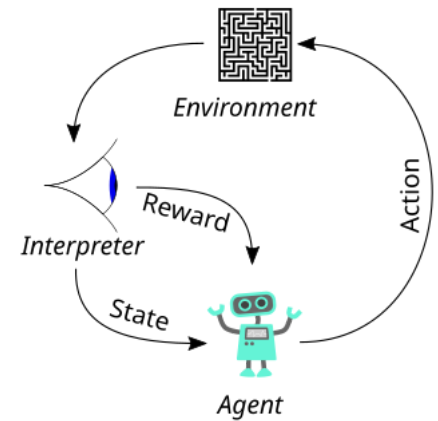
Other types

Other approaches have been developed which do not fit neatly into this three-fold categorization, and sometimes more than one is used by the same machine learning system. For example, topic modeling, meta-learning.[57]

Self-learning

Self-learning, as a machine learning paradigm was introduced in 1982 along with a neural network capable of self-learning, named *crossbar adaptive array* (CAA).[58] It is learning with no external rewards and no external teacher advice. The CAA self-learning algorithm computes, in a crossbar fashion, both decisions about actions and emotions (feelings) about consequence situations. The system is driven by the interaction between cognition and emotion.[59] The self-learning algorithm updates a memory matrix W =||w(a,s)|| such that in each iteration executes the following machine learning routine:

1. in situation *s* perform action *a*
2. receive a consequence situation *s'*
3. compute emotion of being in the consequence situation *v(s')*
4. update crossbar memory *w'(a,s) = w(a,s) + v(s')*

It is a system with only one input, situation, and only one output, action (or behavior) a. There is neither a separate reinforcement input nor an advice input from the environment. The backpropagated value (secondary reinforcement) is the emotion toward the consequence situation. The CAA exists in two environments, one is the behavioral environment where it behaves, and the other is the genetic environment, wherefrom it initially and only once receives initial emotions about

situations to be encountered in the behavioral environment. After receiving the genome (species) vector from the genetic environment, the CAA learns a goal-seeking behavior, in an environment that contains both desirable and undesirable situations.[60]

Feature learning

Several learning algorithms aim at discovering better representations of the inputs provided during training.[61] Classic examples include principal component analysis and cluster analysis. Feature learning algorithms, also called representation learning algorithms, often attempt to preserve the information in their input but also transform it in a way that makes it useful, often as a pre-processing step before performing classification or predictions. This technique allows reconstruction of the inputs coming from the unknown data-generating distribution, while not being necessarily faithful to configurations that are implausible under that distribution. This replaces manual feature engineering, and allows a machine to both learn the features and use them to perform a specific task.

Feature learning can be either supervised or unsupervised. In supervised feature learning, features are learned using labeled input data. Examples include artificial neural networks, multilayer perceptrons, and supervised dictionary learning. In unsupervised feature learning, features are learned with unlabeled input data. Examples include dictionary learning, independent component analysis, autoencoders, matrix factorization[62] and various forms of clustering.[63][64][65]

Manifold learning algorithms attempt to do so under the constraint that the learned representation is low-dimensional. Sparse coding algorithms attempt to do so under the constraint that the learned representation is sparse, meaning that the mathematical model has many zeros. Multilinear subspace learning algorithms aim to learn low-dimensional representations directly from tensor representations for multidimensional data, without reshaping them into higher-dimensional vectors.[66] Deep learning algorithms discover multiple levels of representation, or a hierarchy of features, with higher-level, more abstract features defined in terms of (or generating) lower-level features. It has been argued that an intelligent machine is one that learns a representation that disentangles the underlying factors of variation that explain the observed data.[67]

Feature learning is motivated by the fact that machine learning tasks such as classification often require input that is mathematically and computationally convenient to process. However, real-world data such as images, video, and sensory data has not yielded attempts to algorithmically define specific features. An alternative is to discover such features or representations through examination, without relying on explicit algorithms.

Sparse dictionary learning

Sparse dictionary learning is a feature learning method where a training example is represented as a linear combination of basis functions and assumed to be a sparse matrix. The method is strongly NP-hard and difficult to solve approximately.[68] A popular heuristic method for sparse dictionary learning is the k-SVD algorithm. Sparse dictionary learning has been applied in several contexts. In classification, the problem is to determine the class to which a previously unseen training example belongs. For a dictionary where each class has already been built, a new training example is associated

with the class that is best sparsely represented by the corresponding dictionary. Sparse dictionary learning has also been applied in image de-noising. The key idea is that a clean image patch can be sparsely represented by an image dictionary, but the noise cannot.[69]

Anomaly detection

In data mining, anomaly detection, also known as outlier detection, is the identification of rare items, events or observations which raise suspicions by differing significantly from the majority of the data.[70] Typically, the anomalous items represent an issue such as bank fraud, a structural defect, medical problems or errors in a text. Anomalies are referred to as outliers, novelties, noise, deviations and exceptions.[71]

In particular, in the context of abuse and network intrusion detection, the interesting objects are often not rare objects, but unexpected bursts of inactivity. This pattern does not adhere to the common statistical definition of an outlier as a rare object. Many outlier detection methods (in particular, unsupervised algorithms) will fail on such data unless aggregated appropriately. Instead, a cluster analysis algorithm may be able to detect the micro-clusters formed by these patterns.[72]

Three broad categories of anomaly detection techniques exist.[73] Unsupervised anomaly detection techniques detect anomalies in an unlabeled test data set under the assumption that the majority of the instances in the data set are normal, by looking for instances that seem to fit the least to the remainder of the data set. Supervised anomaly detection techniques require a data set that has been labeled as "normal" and "abnormal" and involves training a classifier (the key difference from many other statistical classification problems is the inherently unbalanced nature of outlier detection). Semi-supervised anomaly detection techniques construct a model representing normal behavior from a given normal training data set and then test the likelihood of a test instance to be generated by the model.

Robot learning

Robot learning is inspired by a multitude of machine learning methods, starting from supervised learning, reinforcement learning,[74][75] and finally meta-learning (e.g. MAML).

Association rules

Association rule learning is a rule-based machine learning method for discovering relationships between variables in large databases. It is intended to identify strong rules discovered in databases using some measure of "interestingness".[76]

Rule-based machine learning is a general term for any machine learning method that identifies, learns, or evolves "rules" to store, manipulate or apply knowledge. The defining characteristic of a rule-based machine learning algorithm is the identification and utilization of a set of relational rules that collectively represent the knowledge captured by the system. This is in contrast to other machine learning algorithms that commonly identify a singular model that can be universally applied to any instance in order to make a prediction.[77] Rule-based machine learning approaches include learning classifier systems, association rule learning, and artificial immune systems.

Based on the concept of strong rules, [Rakesh Agrawal](), [Tomasz Imieliński]() and Arun Swami introduced association rules for discovering regularities between products in large-scale transaction data recorded by [point-of-sale]() (POS) systems in supermarkets.[78] For example, the rule $\{onions, potatoes\} \Rightarrow \{burger\}$ found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, they are likely to also buy hamburger meat. Such information can be used as the basis for decisions about marketing activities such as promotional [pricing]() or [product placements](). In addition to [market basket analysis](), association rules are employed today in application areas including [Web usage mining](), [intrusion detection](), [continuous production](), and [bioinformatics](). In contrast with [sequence mining](), association rule learning typically does not consider the order of items either within a transaction or across transactions.

Learning classifier systems (LCS) are a family of rule-based machine learning algorithms that combine a discovery component, typically a [genetic algorithm](), with a learning component, performing either [supervised learning](), [reinforcement learning](), or [unsupervised learning](). They seek to identify a set of context-dependent rules that collectively store and apply knowledge in a [piecewise]() manner in order to make predictions.[79]

[Inductive logic programming]() (ILP) is an approach to rule learning using [logic programming]() as a uniform representation for input examples, background knowledge, and hypotheses. Given an encoding of the known background knowledge and a set of examples represented as a logical database of facts, an ILP system will derive a hypothesized logic program that [entails]() all positive and no negative examples. [Inductive programming]() is a related field that considers any kind of programming language for representing hypotheses (and not only logic programming), such as [functional programs]().

Inductive logic programming is particularly useful in [bioinformatics]() and [natural language processing](). [Gordon Plotkin]() and [Ehud Shapiro]() laid the initial theoretical foundation for inductive machine learning in a logical setting.[80][81][82] Shapiro built their first implementation (Model Inference System) in 1981: a Prolog program that inductively inferred logic programs from positive and negative examples.[83] The term *inductive* here refers to [philosophical]() induction, suggesting a theory to explain observed facts, rather than [mathematical induction](), proving a property for all members of a well-ordered set.

Models

A **machine learning model** is a type of [mathematical model]() that, after being "trained" on a given dataset, can be used to make predictions or classifications on new data. During training, a learning algorithm iteratively adjusts the model's internal parameters to minimize errors in its predictions.[84] By extension, the term "model" can refer to several levels of specificity, from a general class of models and their associated learning algorithms to a fully trained model with all its internal parameters tuned.[85]

Various types of models have been used and researched for machine learning systems, picking the best model for a task is called [model selection]().

Artificial neural networks

Artificial neural networks (ANNs), or connectionist systems, are computing systems vaguely inspired by the biological neural networks that constitute animal brains. Such systems "learn" to perform tasks by considering examples, generally without being programmed with any task-specific rules.

An ANN is a model based on a collection of connected units or nodes called "artificial neurons", which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit information, a "signal", from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it. In common ANN implementations, the signal at a connection between artificial neurons is a real number, and the output of each artificial neuron is computed by some non-linear function of the sum of its inputs. The connections between artificial neurons are called "edges". Artificial neurons and edges typically have a weight that adjusts as learning proceeds. The weight increases or decreases the strength of the signal at a connection. Artificial neurons may have a threshold such that the signal is only sent if the aggregate signal crosses that threshold. Typically, artificial neurons are aggregated into layers. Different layers may perform different kinds of transformations on their inputs. Signals travel from the first layer (the input layer) to the last layer (the output layer), possibly after traversing the layers multiple times.
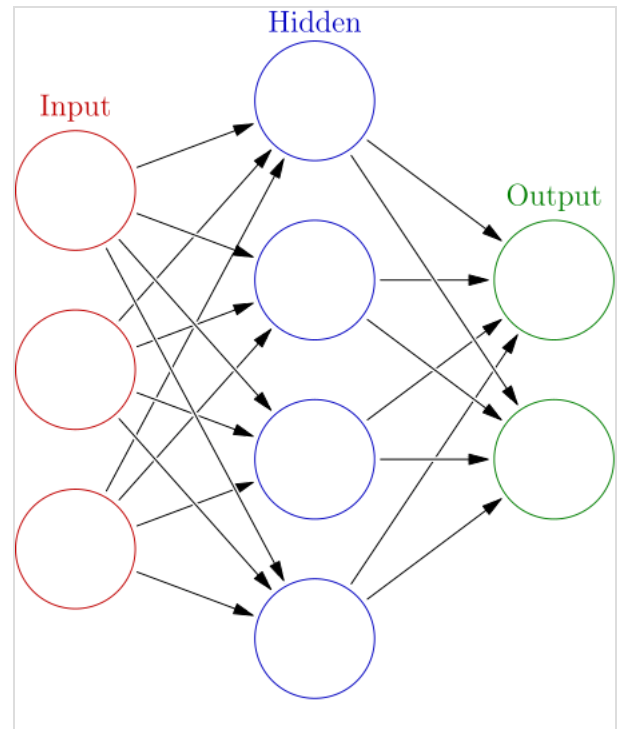
An artificial neural network is an interconnected group of nodes, akin to the vast network of neurons in a brain. Here, each circular node represents an artificial neuron and an arrow represents a connection from the output of one artificial neuron to the input of another.

The original goal of the ANN approach was to solve problems in the same way that a human brain would. However, over time, attention moved to performing specific tasks, leading to deviations from biology. Artificial neural networks have been used on a variety of tasks, including computer vision, speech recognition, machine translation, social network filtering, playing board and video games and medical diagnosis.

Deep learning consists of multiple hidden layers in an artificial neural network. This approach tries to model the way the human brain processes light and sound into vision and hearing. Some successful applications of deep learning are computer vision and speech recognition.[86]

Decision trees

Decision tree learning uses a decision tree as a predictive model to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). It is one of the predictive modeling approaches used in statistics, data mining, and machine learning. Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels, and branches represent

conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. In data mining, a decision tree describes data, but the resulting classification tree can be an input for decision-making.

A decision tree showing survival probability of passengers on the Titanic
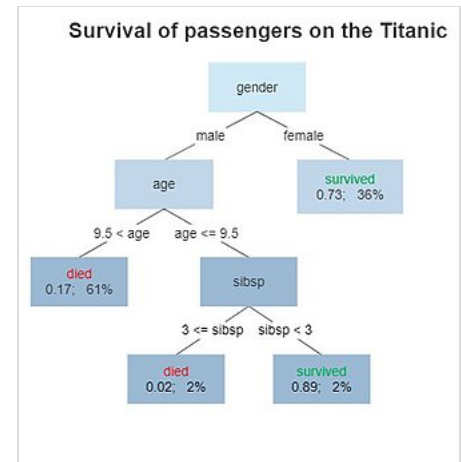
Support-vector machines

Support-vector machines (SVMs), also known as support-vector networks, are a set of related supervised learning methods used for classification and regression. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category.[87] An SVM training algorithm is a non-probabilistic, binary, linear classifier, although methods such as Platt scaling exist to use SVM in a probabilistic classification setting. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

Regression analysis

Regression analysis encompasses a large variety of statistical methods to estimate the relationship between input variables and their associated features. Its most common form is linear regression, where a single line is drawn to best fit the given data according to a mathematical criterion such as ordinary least squares. The latter is often extended by regularization methods to mitigate overfitting and bias, as in ridge regression. When dealing with non-linear problems, go-to models include polynomial regression (for example, used for trendline fitting in Microsoft Excel[88]), logistic regression (often used in statistical classification) or

Illustration of linear regression on a data set

even kernel regression, which introduces non-linearity by taking advantage of the kernel trick to implicitly map input variables to higher-dimensional space.

Bayesian networks

A Bayesian network, belief network, or directed acyclic graphical model is a probabilistic graphical model that represents a set of random variables and their conditional independence with a directed acyclic graph (DAG). For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases. Efficient algorithms exist that perform inference and

learning. Bayesian networks that model sequences of variables, like speech signals or protein sequences, are called dynamic Bayesian networks. Generalizations of Bayesian networks that can represent and solve decision problems under uncertainty are called influence diagrams.

A simple Bayesian network. Rain influences whether the sprinkler is activated, and both rain and the sprinkler influence whether the grass is wet.

Gaussian processes

A Gaussian process is a stochastic process in which every finite collection of the random variables in the process has a multivariate normal distribution, and it relies on a pre-defined covariance function, or kernel, that models how pairs of points relate to each other depending on their locations.

Given a set of observed points, or input–output examples, the distribution of the (unobserved) output of a new point as function of its input data can be directly computed by looking like the observed points and the covariances between those points and the new, unobserved point.
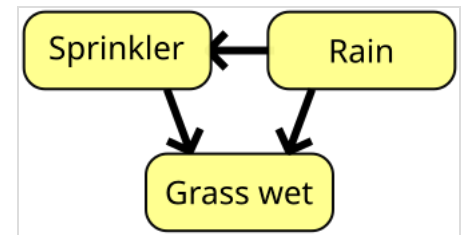

An example of Gaussian Process Regression (prediction) compared with other regression models[89]

Gaussian processes are popular surrogate models in Bayesian optimization used to do hyperparameter optimization.
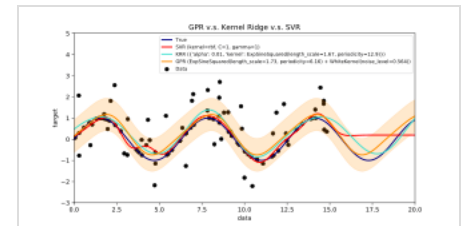
Genetic algorithms

A genetic algorithm (GA) is a search algorithm and heuristic technique that mimics the process of natural selection, using methods such as mutation and crossover to generate new genotypes in the hope of finding good solutions to a given problem. In machine learning, genetic algorithms were used in the 1980s and 1990s.[90][91] Conversely, machine learning techniques have been used to improve the performance of genetic and evolutionary algorithms.[92]

Belief functions

The theory of belief functions, also referred to as evidence theory or Dempster–Shafer theory, is a general framework for reasoning with uncertainty, with understood connections to other frameworks such as probability, possibility and imprecise probability theories. These theoretical frameworks can be thought of as a kind of learner and have some analogous properties of how evidence is combined (e.g., Dempster's rule of combination), just like how in a pmf-based Bayesian approach would combine probabilities. However, there are many caveats to these beliefs functions when compared to Bayesian approaches in order to incorporate ignorance and uncertainty quantification. These belief function approaches that are implemented within the machine learning domain typically leverage a fusion approach of various ensemble methods to better handle the learner's decision boundary, low samples, and ambiguous class issues that standard machine learning approach tend to have difficulty

resolving.[4][9] However, the computational complexity of these algorithms are dependent on the number of propositions (classes), and can lead to a much higher computation time when compared to other machine learning approaches.

Training models

Typically, machine learning models require a high quantity of reliable data to perform accurate predictions. When training a machine learning model, machine learning engineers need to target and collect a large and representative sample of data. Data from the training set can be as varied as a corpus of text, a collection of images, sensor data, and data collected from individual users of a service. Overfitting is something to watch out for when training a machine learning model. Trained models derived from biased or non-evaluated data can result in skewed or undesired predictions. Biased models may result in detrimental outcomes, thereby furthering the negative impacts on society or objectives. Algorithmic bias is a potential result of data not being fully prepared for training. Machine learning ethics is becoming a field of study and notably, becoming integrated within machine learning engineering teams.

Federated learning

Federated learning is an adapted form of distributed artificial intelligence to training machine learning models that decentralizes the training process, allowing for users' privacy to be maintained by not needing to send their data to a centralized server. This also increases efficiency by decentralizing the training process to many devices. For example, Gboard uses federated machine learning to train search query prediction models on users' mobile phones without having to send individual searches back to Google.[93]

Applications

There are many applications for machine learning, including:

- Agriculture
- Anatomy
- Adaptive website
- Affective computing
- Astronomy
- Automated decision-making
- Banking
- Behaviorism
- Bioinformatics
- Brain–machine interfaces
- Cheminformatics
- Citizen Science
- Climate Science
- Computer networks
- Computer vision

- Credit-card fraud detection
- Data quality
- DNA sequence classification
- Economics
- Financial market analysis[94]
- General game playing
- Handwriting recognition
- Healthcare
- Information retrieval
- Insurance
- Internet fraud detection
- Knowledge graph embedding
- Linguistics
- Machine learning control
- Machine perception

- Machine translation
- Marketing
- Medical diagnosis
- Natural language processing
- Natural language understanding
- Online advertising
- Optimization
- Recommender systems
- Robot locomotion
- Search engines
- Sentiment analysis
- Sequence mining
- Software engineering
- Speech recognition
- Structural health monitoring
- Syntactic pattern recognition
- Telecommunications
- Theorem proving
- Time-series forecasting
- Tomographic reconstruction[95]
- User behavior analytics

In 2006, the media-services provider Netflix held the first "Netflix Prize" competition to find a program to better predict user preferences and improve the accuracy of its existing Cinematch movie recommendation algorithm by at least 10%. A joint team made up of researchers from AT&T Labs-Research in collaboration with the teams Big Chaos and Pragmatic Theory built an ensemble model to win the Grand Prize in 2009 for $1 million.[96] Shortly after the prize was awarded, Netflix realized that viewers' ratings were not the best indicators of their viewing patterns ("everything is a recommendation") and they changed their recommendation engine accordingly.[97] In 2010 The Wall Street Journal wrote about the firm Rebellion Research and their use of machine learning to predict the financial crisis.[98] In 2012, co-founder of Sun Microsystems, Vinod Khosla, predicted that 80% of medical doctors jobs would be lost in the next two decades to automated machine learning medical diagnostic software.[99] In 2014, it was reported that a machine learning algorithm had been applied in the field of art history to study fine art paintings and that it may have revealed previously unrecognized influences among artists.[100] In 2019 Springer Nature published the first research book created using machine learning.[101] In 2020, machine learning technology was used to help make diagnoses and aid researchers in developing a cure for COVID-19.[102] Machine learning was recently applied to predict the pro-environmental behavior of travelers.[103] Recently, machine learning technology was also applied to optimize smartphone's performance and thermal behavior based on the user's interaction with the phone.[104][105][106] When applied correctly, machine learning algorithms (MLAs) can utilize a wide range of company characteristics to predict stock returns without overfitting. By employing effective feature engineering and combining forecasts, MLAs can generate results that far surpass those obtained from basic linear techniques like OLS.[107]

Recent advancements in machine learning have extended into the field of quantum chemistry, where novel algorithms now enable the prediction of solvent effects on chemical reactions, thereby offering new tools for chemists to tailor experimental conditions for optimal outcomes.[108]

Machine Learning is becoming a useful tool to investigate and predict evacuation decision making in large scale and small scale disasters. Different solutions have been tested to predict if and when householders decide to evacuate during wildfires and hurricanes.[109][110][111] Other applications have been focusing on pre evacuation decisions in building fires.[112][113]

Limitations

Although machine learning has been transformative in some fields, machine-learning programs often fail to deliver expected results.[114][115][116] Reasons for this are numerous: lack of (suitable) data, lack of access to the data, data bias, privacy problems, badly chosen tasks and algorithms, wrong tools and people, lack of resources, and evaluation problems.[117]

The "black box theory" poses another yet significant challenge. Black box refers to a situation where the algorithm or the process of producing an output is entirely opaque, meaning that even the coders of the algorithm cannot audit the pattern that the machine extracted out of the data.[118] The House of Lords Select Committee, which claimed that such an "intelligence system" that could have a "substantial impact on an individual's life" would not be considered acceptable unless it provided "a full and satisfactory explanation for the decisions" it makes.[118]

In 2018, a self-driving car from Uber failed to detect a pedestrian, who was killed after a collision.[119] Attempts to use machine learning in healthcare with the IBM Watson system failed to deliver even after years of time and billions of dollars invested.[120][121] Microsoft's Bing Chat chatbot has been reported to produce hostile and offensive response against its users.[122]

Machine learning has been used as a strategy to update the evidence related to a systematic review and increased reviewer burden related to the growth of biomedical literature. While it has improved with training sets, it has not yet developed sufficiently to reduce the workload burden without limiting the necessary sensitivity for the findings research themselves.[123]
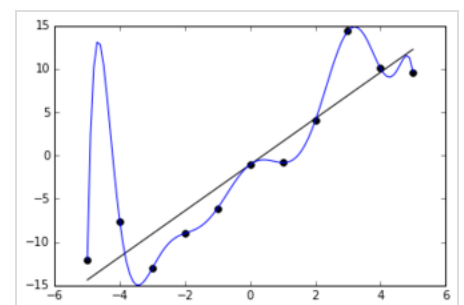
Explainability

Explainable AI (XAI), or Interpretable AI, or Explainable Machine Learning (XML), is artificial intelligence (AI) in which humans can understand the decisions or predictions made by the AI.[124] It contrasts with the "black box" concept in machine learning where even its designers cannot explain why an AI arrived at a specific decision.[125] By refining the mental models of users of AI-powered systems and dismantling their misconceptions, XAI promises to help users perform more effectively. XAI may be an implementation of the social right to explanation.

Overfitting

Settling on a bad, overly complex theory gerrymandered to fit all the past training data is known as overfitting. Many systems attempt to reduce overfitting by rewarding a theory in accordance with how well it fits the data but penalizing the theory in accordance with how complex the theory is.[126]

The blue line could be an example of overfitting a linear function due to random noise.

Other limitations and vulnerabilities

Learners can also disappoint by "learning the wrong lesson". A toy example is that an image classifier trained only on pictures of brown horses and black cats might conclude that all brown patches are likely to be horses.[127] A real-world example is that, unlike humans, current image classifiers often do not primarily make judgments from the spatial

relationship between components of the picture, and they learn relationships between pixels that humans are oblivious to, but that still correlate with images of certain types of real objects. Modifying these patterns on a legitimate image can result in "adversarial" images that the system misclassifies.[128][129]

Adversarial vulnerabilities can also result in nonlinear systems, or from non-pattern perturbations. For some systems, it is possible to change the output by only changing a single adversarially chosen pixel.[130] Machine learning models are often vulnerable to manipulation and/or evasion via adversarial machine learning.[131]

Researchers have demonstrated how backdoors can be placed undetectably into classifying (e.g., for categories "spam" and well-visible "not spam" of posts) machine learning models that are often developed and/or trained by third parties. Parties can change the classification of any input, including in cases for which a type of data/software transparency is provided, possibly including white-box access.[132][133][134]

Model assessments

Classification of machine learning models can be validated by accuracy estimation techniques like the holdout method, which splits the data in a training and test set (conventionally 2/3 training set and 1/3 test set designation) and evaluates the performance of the training model on the test set. In comparison, the K-fold-cross-validation method randomly partitions the data into K subsets and then K experiments are performed each respectively considering 1 subset for evaluation and the remaining K-1 subsets for training the model. In addition to the holdout and cross-validation methods, bootstrap, which samples n instances with replacement from the dataset, can be used to assess model accuracy.[135]

In addition to overall accuracy, investigators frequently report sensitivity and specificity meaning true positive rate (TPR) and true negative rate (TNR) respectively. Similarly, investigators sometimes report the false positive rate (FPR) as well as the false negative rate (FNR). However, these rates are ratios that fail to reveal their numerators and denominators. Receiver operating characteristic (ROC) along with the accompanying Area Under the ROC Curve (AUC) offer additional tools for classification model assessment. Higher AUC is associated with a better performing model.[136]

Ethics

The ethics of artificial intelligence covers a broad range of topics within the field that are considered to have particular ethical stakes.[137] This includes algorithmic biases, fairness, automated decision-making, accountability, privacy, and regulation. It also covers various emerging or potential future challenges such as machine ethics (how to make machines that behave ethically), lethal autonomous weapon systems, arms race dynamics, AI safety and alignment, technological unemployment, AI-enabled misinformation, how to treat certain AI systems if they have a moral status (AI welfare and rights), artificial superintelligence and existential risks.[137]

Some application areas may also have particularly important ethical implications, like healthcare, education, criminal justice, or the military.

Bias

Different machine learning approaches can suffer from different data biases. A machine learning system trained specifically on current customers may not be able to predict the needs of new customer groups that are not represented in the training data. When trained on human-made data, machine learning is likely to pick up the constitutional and unconscious biases already present in society.[138]

Systems that are trained on datasets collected with biases may exhibit these biases upon use (algorithmic bias), thus digitizing cultural prejudices.[139] For example, in 1988, the UK's Commission for Racial Equality found that St. George's Medical School had been using a computer program trained from data of previous admissions staff and that this program had denied nearly 60 candidates who were found to either be women or have non-European sounding names.[138] Using job hiring data from a firm with racist hiring policies may lead to a machine learning system duplicating the bias by scoring job applicants by similarity to previous successful applicants.[140][141] Another example includes predictive policing company Geolitica's predictive algorithm that resulted in "disproportionately high levels of over-policing in low-income and minority communities" after being trained with historical crime data.[142]

While responsible collection of data and documentation of algorithmic rules used by a system is considered a critical part of machine learning, some researchers blame lack of participation and representation of minority population in the field of AI for machine learning's vulnerability to biases.[143] In fact, according to research carried out by the Computing Research Association (CRA) in 2021, "female faculty merely make up 16.1%" of all faculty members who focus on AI among several universities around the world.[144] Furthermore, among the group of "new U.S. resident AI PhD graduates," 45% identified as white, 22.4% as Asian, 3.2% as Hispanic, and 2.4% as African American, which further demonstrates a lack of diversity in the field of AI.[144]

Language models learned from data have been shown to contain human-like biases.[145][146] Because human languages contain biases, machines trained on language *corpora* will necessarily also learn these biases.[147][148] In 2016, Microsoft tested Tay, a chatbot that learned from Twitter, and it quickly picked up racist and sexist language.[149]

In an experiment carried out by ProPublica, an investigative journalism organization, a machine learning algorithm's insight into the recidivism rates among prisoners falsely flagged "black defendants high risk twice as often as white defendants."[142] In 2015, Google Photos once tagged a couple of black people as gorillas, which caused controversy. The gorilla label was subsequently removed, and in 2023, it still cannot recognize gorillas.[150] Similar issues with recognizing non-white people have been found in many other systems.[151]

Because of such challenges, the effective use of machine learning may take longer to be adopted in other domains.[152] Concern for fairness in machine learning, that is, reducing bias in machine learning and propelling its use for human good, is increasingly expressed by artificial intelligence

scientists, including Fei-Fei Li, who said that "[t]here's nothing artificial about AI. It's inspired by people, it's created by people, and—most importantly—it impacts people. It is a powerful tool we are only just beginning to understand, and that is a profound responsibility."[153]

Financial incentives

There are concerns among health care professionals that these systems might not be designed in the public's interest but as income-generating machines. This is especially true in the United States where there is a long-standing ethical dilemma of improving health care, but also increasing profits. For example, the algorithms could be designed to provide patients with unnecessary tests or medication in which the algorithm's proprietary owners hold stakes. There is potential for machine learning in health care to provide professionals an additional tool to diagnose, medicate, and plan recovery paths for patients, but this requires these biases to be mitigated.[154]

Hardware

Since the 2010s, advances in both machine learning algorithms and computer hardware have led to more efficient methods for training deep neural networks (a particular narrow subdomain of machine learning) that contain many layers of nonlinear hidden units.[155] By 2019, graphics processing units (GPUs), often with AI-specific enhancements, had displaced CPUs as the dominant method of training large-scale commercial cloud AI.[156] OpenAI estimated the hardware compute used in the largest deep learning projects from AlexNet (2012) to AlphaZero (2017), and found a 300,000-fold increase in the amount of compute required, with a doubling-time trendline of 3.4 months.[157][158]

Neuromorphic computing

Neuromorphic computing refers to a class of computing systems designed to emulate the structure and functionality of biological neural networks. These systems may be implemented through software-based simulations on conventional hardware or through specialized hardware architectures.[159]

physical neural networks

A physical neural network is a specific type of neuromorphic hardware that relies on electrically adjustable materials, such as memristors, to emulate the function of neural synapses. The term "physical neural network" highlights the use of physical hardware for computation, as opposed to software-based implementations. It broadly refers to artificial neural networks that use materials with adjustable resistance to replicate neural synapses.[160][161]

Embedded machine learning

Embedded machine learning is a sub-field of machine learning where models are deployed on embedded systems with limited computing resources, such as wearable computers, edge devices and microcontrollers.[162][163][164] Running models directly on these devices eliminates the need to transfer and store data on cloud servers for further processing, thereby reducing the risk of data

breaches, privacy leaks and theft of intellectual property, personal data and business secrets. Embedded machine learning can be achieved through various techniques, such as hardware acceleration,[165][166] approximate computing,[167] and model optimization.[168][169] Common optimization techniques include pruning, quantization, knowledge distillation, low-rank factorization, network architecture search, and parameter sharing.

Software

Software suites containing a variety of machine learning algorithms include the following:

Free and open-source software

- Caffe
- Deeplearning4j
- DeepSpeed
- ELKI
- Google JAX
- Infer.NET
- Keras
- Kubeflow
- LightGBM
- Mahout
- Mallet
- Microsoft Cognitive Toolkit
- ML.NET
- mlpack
- MXNet
- OpenNN
- Orange
- pandas (software)
- ROOT (TMVA with ROOT)
- scikit-learn
- Shogun
- Spark MLlib
- SystemML
- TensorFlow
- Torch / PyTorch
- Weka / MOA
- XGBoost
- Yooreeka

Proprietary software with free and open-source editions

- KNIME
- RapidMiner

Proprietary software

- Amazon Machine Learning
- Angoss KnowledgeSTUDIO
- Azure Machine Learning
- IBM Watson Studio
- Google Cloud Vertex AI
- Google Prediction API
- IBM SPSS Modeler
- KXEN Modeler
- LIONsolver
- Mathematica
- MATLAB
- Neural Designer
- NeuroSolutions
- Oracle Data Mining
- Oracle AI Platform Cloud Service
- PolyAnalyst
- RCASE
- SAS Enterprise Miner
- SequenceL
- Splunk

- STATISTICA Data Miner

Journals

- Journal of Machine Learning Research
- Machine Learning
- Nature Machine Intelligence
- Neural Computation
- IEEE Transactions on Pattern Analysis and Machine Intelligence

Conferences

- AAAI Conference on Artificial Intelligence
- Association for Computational Linguistics (**ACL**)
- European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (**ECML PKDD**)
- International Conference on Computational Intelligence Methods for Bioinformatics and Biostatistics (**CIBB**)
- International Conference on Machine Learning (**ICML**)
- International Conference on Learning Representations (**ICLR**)
- International Conference on Intelligent Robots and Systems (**IROS**)
- Conference on Knowledge Discovery and Data Mining (**KDD**)
- Conference on Neural Information Processing Systems (**NeurIPS**)

See also

- Automated machine learning – Process of automating the application of machine learning
- Big data – Extremely large or complex datasets
- Deep learning — branch of ML concerned with artificial neural networks
- Differentiable programming – Programming paradigm
- List of datasets for machine-learning research – Machine learning based fault detection in Electronics Circuit
- M-theory (learning framework)

References

1. The definition "without being explicitly programmed" is often attributed to Arthur Samuel, who coined the term "machine learning" in 1959, but the phrase is not found verbatim in this publication, and may be a paraphrase that appeared later. Confer "Paraphrasing Arthur Samuel (1959), the question is: How can computers learn to solve problems without being explicitly programmed?" in Koza, John R.; Bennett, Forrest H.; Andre, David; Keane, Martin A. (1996). "Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming". *Artificial Intelligence in Design '96*. Artificial Intelligence in Design '96. Dordrecht, Netherlands: Springer Netherlands. pp. 151–170. doi:10.1007/978-94-009-0279-4_9 (https://doi.org/10.1007%2F978-94-009-0279-4_9). ISBN 978-94-010-6610-5.

2. "What is Machine Learning?" (https://www.ibm.com/topics/machine-learning). *IBM*. 22 September 2021. Archived (https://web.archive.org/web/20231227153910/https://www.ibm.com/topics/machine-learning) from the original on 2023-12-27. Retrieved 2023-06-27.

3. Hu, Junyan; Niu, Hanlin; Carrasco, Joaquin; Lennox, Barry; Arvin, Farshad (2020). "Voronoi-Based Multi-Robot Autonomous Exploration in Unknown Environments via Deep Reinforcement Learning" (https://doi.org/10.1109%2Ftvt.2020.3034800). *IEEE Transactions on Vehicular Technology*. **69** (12): 14413–14423. doi:10.1109/tvt.2020.3034800 (https://doi.org/10.1109%2Ftvt.2020.3034800). ISSN 0018-9545 (https://search.worldcat.org/issn/0018-9545). S2CID 228989788 (https://api.semanticscholar.org/CorpusID:228989788).

4. Yoosefzadeh-Najafabadi, Mohsen; Hugh, Earl; Tulpan, Dan; Sulik, John; Eskandari, Milad (2021). "Application of Machine Learning Algorithms in Plant Breeding: Predicting Yield From Hyperspectral Reflectance in Soybean?" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7835636). *Front. Plant Sci*. **11**: 624273. doi:10.3389/fpls.2020.624273 (https://doi.org/10.3389%2Ffpls.2020.624273). PMC 7835636 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7835636). PMID 33510761 (https://pubmed.ncbi.nlm.nih.gov/33510761).

5. Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Springer, ISBN 978-0-387-31073-2

6. Machine learning and pattern recognition "can be viewed as two facets of the same field".[5]:vii

7. Friedman, Jerome H. (1998). "Data Mining and Statistics: What's the connection?". *Computing Science and Statistics*. **29** (1): 3–9.

8. Samuel, Arthur (1959). "Some Studies in Machine Learning Using the Game of Checkers". *IBM Journal of Research and Development*. **3** (3): 210–229. CiteSeerX 10.1.1.368.2254 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.368.2254). doi:10.1147/rd.33.0210 (https://doi.org/10.1147%2Frd.33.0210). S2CID 2126705 (https://api.semanticscholar.org/CorpusID:2126705).

9. R. Kohavi and F. Provost, "Glossary of terms", Machine Learning, vol. 30, no. 2–3, pp. 271–274, 1998.

10. Gerovitch, Slava (9 April 2015). "How the Computer Got Its Revenge on the Soviet Union" (https://web.archive.org/web/20210922175839/https://nautil.us/issue/23/Dominoes/how-the-computer-got-its-revenge-on-the-soviet-union). *Nautilus*. Archived from the original (https://nautil.us/issue/23/dominoes/how-the-computer-got-its-revenge-on-the-soviet-union) on 22 September 2021. Retrieved 19 September 2021.

11. Lindsay, Richard P. (1 September 1964). "The Impact of Automation On Public Administration" (https://journals.sagepub.com/doi/10.1177/106591296401700364). *Western Political Quarterly*. **17** (3): 78–81. doi:10.1177/106591296401700364 (https://doi.org/10.1177%2F106591296401700364). ISSN 0043-4078 (https://search.worldcat.org/issn/0043-4078). S2CID 154021253 (https://api.semanticscholar.org/CorpusID:154021253). Archived (https://web.archive.org/web/20211006190841/https://journals.sagepub.com/doi/10.1177/106591296401700364) from the original on 6 October 2021. Retrieved 6 October 2021.

12. "History and Evolution of Machine Learning: A Timeline" (https://www.techtarget.com/whatis/A-Timeline-of-Machine-Learning-History). *WhatIs*. Archived (https://web.archive.org/web/20231208220935/https://www.techtarget.com/whatis/A-Timeline-of-Machine-Learning-History) from the original on 2023-12-08. Retrieved 2023-12-08.

13. Milner, Peter M. (1993). "The Mind and Donald O. Hebb" (https://www.jstor.org/stable/24941344). *Scientific American*. **268** (1): 124–129. Bibcode:1993SciAm.268a.124M (https://ui.adsabs.harvard.edu/abs/1993SciAm.268a.124M). doi:10.1038/scientificamerican0193-124 (https://doi.org/10.1038%2Fscientificamerican0193-124). ISSN 0036-8733 (https://search.worldcat.org/issn/0036-8733). JSTOR 24941344 (https://www.jstor.org/stable/24941344). PMID 8418480 (https://pubmed.ncbi.nlm.nih.gov/8418480). Archived (https://web.archive.org/web/20231220163326/https://www.jstor.org/stable/24941344) from the original on 2023-12-20. Retrieved 2023-12-09.

14. "Science: The Goof Button", Time (magazine), 18 August 1961.

15. Nilsson N. Learning Machines, McGraw Hill, 1965.

16. Duda, R., Hart P. Pattern Recognition and Scene Analysis, Wiley Interscience, 1973

17. S. Bozinovski "Teaching space: A representation concept for adaptive pattern classification" COINS Technical Report No. 81-28, Computer and Information Science Department, University of Massachusetts at Amherst, MA, 1981. https://web.cs.umass.edu/publication/docs/1981/UM-CS-1981-028.pdf Archived (https://web.archive.org/web/20210225070218/https://web.cs.umass.edu/publication/docs/1981/UM-CS-1981-028.pdf) 2021-02-25 at the Wayback Machine

18. Mitchell, T. (1997). *Machine Learning*. McGraw Hill. p. 2. ISBN 978-0-07-042807-2.

19. Harnad, Stevan (2008), "The Annotation Game: On Turing (1950) on Computing, Machinery, and Intelligence" (https://web.archive.org/web/20120309113922/http://eprints.ecs.soton.ac.uk/12954/), in Epstein, Robert; Peters, Grace (eds.), *The Turing Test Sourcebook: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, Kluwer, pp. 23–66, ISBN 9781402067082, archived from the original (http://eprints.ecs.soton.ac.uk/12954/) on 2012-03-09, retrieved 2012-12-11

20. "Introduction to AI Part 1" (https://edzion.com/2020/12/09/introduction-to-ai-part-1/). *Edzion*. 2020-12-08. Archived (https://web.archive.org/web/20210218005157/https://edzion.com/2020/12/09/introduction-to-ai-part-1/) from the original on 2021-02-18. Retrieved 2020-12-09.

21. Sindhu V, Nivedha S, Prakash M (February 2020). "An Empirical Science Research on Bioinformatics in Machine Learning" (https://doi.org/10.26782%2Fjmcms.spl.7%2F2020.02.00006). *Journal of Mechanics of Continua and Mathematical Sciences* (7). doi:10.26782/jmcms.spl.7/2020.02.00006 (https://doi.org/10.26782%2Fjmcms.spl.7%2F2020.02.00006).

22. Sarle, Warren S. (1994). "Neural Networks and statistical models". *SUGI 19: proceedings of the Nineteenth Annual SAS Users Group International Conference*. SAS Institute. pp. 1538–50. ISBN 9781555446116. OCLC 35546178 (https://search.worldcat.org/oclc/35546178).

23. Russell, Stuart; Norvig, Peter (2003) [1995]. *Artificial Intelligence: A Modern Approach* (2nd ed.). Prentice Hall. ISBN 978-0137903955.

24. Langley, Pat (2011). "The changing science of machine learning" (https://doi.org/10.1007%2Fs10994-011-5242-y). *Machine Learning*. **82** (3): 275–9. doi:10.1007/s10994-011-5242-y (https://doi.org/10.1007%2Fs10994-011-5242-y).

25. Mahoney, Matt. "Rationale for a Large Text Compression Benchmark" (http://cs.fit.edu/~mmahoney/compression/rationale.html). Florida Institute of Technology. Retrieved 5 March 2013.

26. Shmilovici A.; Kahiri Y.; Ben-Gal I.; Hauser S. (2009). "Measuring the Efficiency of the Intraday Forex Market with a Universal Data Compression Algorithm" (http://www.eng.tau.ac.il/~bengal/28.pdf) (PDF). *Computational Economics*. **33** (2): 131–154. CiteSeerX 10.1.1.627.3751 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.627.3751). doi:10.1007/s10614-008-9153-3 (https://doi.org/10.1007%2Fs10614-008-9153-3). S2CID 17234503 (https://api.semanticscholar.org/CorpusID:17234503). Archived (https://web.archive.org/web/20090709143601/http://www.eng.tau.ac.il/~bengal/28.pdf) (PDF) from the original on 2009-07-09.

27. I. Ben-Gal (2008). "On the Use of Data Compression Measures to Analyze Robust Designs" (http://www.eng.tau.ac.il/~bengal/Journal%20Paper.pdf) (PDF). *IEEE Transactions on Reliability*. **54** (3): 381–388. doi:10.1109/TR.2005.853280 (https://doi.org/10.1109%2FTR.2005.853280). S2CID 9376086 (https://api.semanticscholar.org/CorpusID:9376086).

28. D. Scully; Carla E. Brodley (2006). "Compression and Machine Learning: A New Perspective on Feature Space Vectors". *Data Compression Conference (DCC'06)*. p. 332. doi:10.1109/DCC.2006.13 (https://doi.org/10.1109%2FDCC.2006.13). ISBN 0-7695-2545-8. S2CID 12311412 (https://api.semanticscholar.org/CorpusID:12311412).

29. Gary Adcock (January 5, 2023). "What Is AI Video Compression?" (https://massive.io/file-transfer/what-is-ai-video-compression/). *massive.io*. Retrieved 6 April 2023.

30. Mentzer, Fabian; Toderici, George; Tschannen, Michael; Agustsson, Eirikur (2020). "High-Fidelity Generative Image Compression". arXiv:2006.09965 (https://arxiv.org/abs/2006.09965) [eess.IV (https://arxiv.org/archive/eess.IV)].

31. "What is Unsupervised Learning? | IBM" (https://www.ibm.com/topics/unsupervised-learning). *www.ibm.com*. 23 September 2021. Retrieved 2024-02-05.

32. "Differentially private clustering for large-scale datasets" (https://blog.research.google/2023/05/differentially-private-clustering-for.html). *blog.research.google*. 2023-05-25. Retrieved 2024-03-16.

33. Edwards, Benj (2023-09-28). "AI language models can exceed PNG and FLAC in lossless compression, says study" (https://arstechnica.com/information-technology/2023/09/ai-language-models-can-exceed-png-and-flac-in-lossless-compression-says-study/). *Ars Technica*. Retrieved 2024-03-07.

34. Le Roux, Nicolas; Bengio, Yoshua; Fitzgibbon, Andrew (2012). "Improving First and Second-Order Methods by Modeling Uncertainty" (https://books.google.com/books?id=JPQx7s2L1A8C&q=%22Improving+First+and+Second-Order+Methods+by+Modeling+Uncertainty&pg=PA403). In Sra, Suvrit; Nowozin, Sebastian; Wright, Stephen J. (eds.). *Optimization for Machine Learning*. MIT Press. p. 404. ISBN 9780262016469. Archived (https://web.archive.org/web/20230117053335/https://books.google.com/books?id=JPQx7s2L1A8C&q=%22Improving+First+and+Second-Order+Methods+by+Modeling+Uncertainty&pg=PA403) from the original on 2023-01-17. Retrieved 2020-11-12.

35. Bzdok, Danilo; Altman, Naomi; Krzywinski, Martin (2018). "Statistics versus Machine Learning" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6082636). *Nature Methods*. **15** (4): 233–234. doi:10.1038/nmeth.4642 (https://doi.org/10.1038%2Fnmeth.4642). PMC 6082636 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6082636). PMID 30100822 (https://pubmed.ncbi.nlm.nih.gov/30100822).

36. Michael I. Jordan (2014-09-10). "statistics and machine learning" (https://www.reddit.com/r/MachineLearning/comments/2fxi6v/ama_michael_i_jordan/ckelmtt?context=3). reddit. Archived (https://web.archive.org/web/20171018192328/https://www.reddit.com/r/MachineLearning/comments/2fxi6v/ama_michael_i_jordan/ckelmtt/?context=3) from the original on 2017-10-18. Retrieved 2014-10-01.

37. Hung et al. Algorithms to Measure Surgeon Performance and Anticipate Clinical Outcomes in Robotic Surgery. JAMA Surg. 2018

38. Cornell University Library (August 2001). "Breiman: Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)" (http://projecteuclid.org/download/pdf_1/euclid.ss/1009213726). *Statistical Science*. **16** (3). doi:10.1214/ss/1009213726 (https://doi.org/10.1214%2Fss%2F1009213726). S2CID 62729017 (https://api.semanticscholar.org/CorpusID:62729017). Archived (https://web.archive.org/web/20170626042637/http://projecteuclid.org/download/pdf_1/euclid.ss/1009213726) from the original on 26 June 2017. Retrieved 8 August 2015.

39. Gareth James; Daniela Witten; Trevor Hastie; Robert Tibshirani (2013). *An Introduction to Statistical Learning* (http://www-bcf.usc.edu/~gareth/ISL/). Springer. p. vii. Archived (https://web.archive.org/web/20190623150237/http://www-bcf.usc.edu/~gareth/ISL/) from the original on 2019-06-23. Retrieved 2014-10-25.

40. Ramezanpour, A.; Beam, A.L.; Chen, J.H.; Mashaghi, A. (17 November 2020). "Statistical Physics for Medical Diagnostics: Learning, Inference, and Optimization Algorithms" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7699346). *Diagnostics*. **10** (11): 972. doi:10.3390/diagnostics10110972 (https://doi.org/10.3390%2Fdiagnostics10110972). PMC 7699346 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7699346). PMID 33228143 (https://pubmed.ncbi.nlm.nih.gov/33228143).

41. Mashaghi, A.; Ramezanpour, A. (16 March 2018). "Statistical physics of medical diagnostics: Study of a probabilistic model". *Physical Review E*. **97** (3–1): 032118. arXiv:1803.10019 (https://arxiv.org/abs/1803.10019). Bibcode:2018PhRvE..97c2118M (https://ui.adsabs.harvard.edu/abs/2018PhRvE..97c2118M). doi:10.1103/PhysRevE.97.032118 (https://doi.org/10.1103%2FPhysRevE.97.032118). PMID 29776109 (https://pubmed.ncbi.nlm.nih.gov/29776109). S2CID 4955393 (https://api.semanticscholar.org/CorpusID:4955393).

42. Mohri, Mehryar; Rostamizadeh, Afshin; Talwalkar, Ameet (2012). *Foundations of Machine Learning*. US, Massachusetts: MIT Press. ISBN 9780262018258.

43. Alpaydin, Ethem (2010). *Introduction to Machine Learning* (https://archive.org/details/introductiont oma00alpa_0). London: The MIT Press. ISBN 978-0-262-01243-0. Retrieved 4 February 2017.

44. Jordan, M. I.; Mitchell, T. M. (17 July 2015). "Machine learning: Trends, perspectives, and prospects". *Science*. **349** (6245): 255–260. Bibcode:2015Sci...349..255J (https://ui.adsabs.harvar d.edu/abs/2015Sci...349..255J). doi:10.1126/science.aaa8415 (https://doi.org/10.1126%2Fscienc e.aaa8415). PMID 26185243 (https://pubmed.ncbi.nlm.nih.gov/26185243). S2CID 677218 (https:// api.semanticscholar.org/CorpusID:677218).

45. El Naqa, Issam; Murphy, Martin J. (2015). "What is Machine Learning?". *Machine Learning in Radiation Oncology*. pp. 3–11. doi:10.1007/978-3-319-18305-3_1 (https://doi.org/10.1007%2F978-3-319-18305-3_1). ISBN 978-3-319-18304-6. S2CID 178586107 (https://api.semanticscholar.org/ CorpusID:178586107).

46. Okolie, Jude A.; Savage, Shauna; Ogbaga, Chukwuma C.; Gunes, Burcu (June 2022). "Assessing the potential of machine learning methods to study the removal of pharmaceuticals from wastewater using biochar or activated carbon" (https://doi.org/10.1016%2Fj.totert.2022.100001). *Total Environment Research Themes*. **1–2**: 100001. Bibcode:2022TERT....100001O (https://ui.ads abs.harvard.edu/abs/2022TERT....100001O). doi:10.1016/j.totert.2022.100001 (https://doi.org/10. 1016%2Fj.totert.2022.100001). S2CID 249022386 (https://api.semanticscholar.org/CorpusID:2490 22386).

47. Russell, Stuart J.; Norvig, Peter (2010). *Artificial Intelligence: A Modern Approach* (Third ed.). Prentice Hall. ISBN 9780136042594.

48. Mohri, Mehryar; Rostamizadeh, Afshin; Talwalkar, Ameet (2012). *Foundations of Machine Learning*. The MIT Press. ISBN 9780262018258.

49. Alpaydin, Ethem (2010). *Introduction to Machine Learning* (https://books.google.com/books?id=7f 5bBAAAQBAJ). MIT Press. p. 9. ISBN 978-0-262-01243-0. Archived (https://web.archive.org/web/ 20230117053338/https://books.google.com/books?id=7f5bBAAAQBAJ) from the original on 2023-01-17. Retrieved 2018-11-25.

50. "Lecture 2 Notes: Supervised Learning" (https://www.cs.cornell.edu/courses/cs4780/2022sp/note s/LectureNotes02.html). *www.cs.cornell.edu*. Retrieved 2024-07-01.

51. Jordan, Michael I.; Bishop, Christopher M. (2004). "Neural Networks". In Allen B. Tucker (ed.). *Computer Science Handbook, Second Edition (Section VII: Intelligent Systems)*. Boca Raton, Florida: Chapman & Hall/CRC Press LLC. ISBN 978-1-58488-360-9.

52. Misra, Ishan; Maaten, Laurens van der (2020). *Self-Supervised Learning of Pretext-Invariant Representations* (https://openaccess.thecvf.com/content_CVPR_2020/html/Misra_Self-Supervise d_Learning_of_Pretext-Invariant_Representations_CVPR_2020_paper.html). 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE. pp. 6707–6717. arXiv:1912.01991 (https://arxiv.org/abs/1912.01991). doi:10.1109/CVPR42600.2020.00674 (https://doi.org/10.1109%2FCVPR42600.2020.00674).

53. Jaiswal, Ashish; Babu, Ashwin Ramesh; Zadeh, Mohammad Zaki; Banerjee, Debapriya; Makedon, Fillia (March 2021). "A Survey on Contrastive Self-Supervised Learning" (https://doi.org/10.3390% 2Ftechnologies9010002). *Technologies*. **9** (1): 2. arXiv:2011.00362 (https://arxiv.org/abs/2011.003 62). doi:10.3390/technologies9010002 (https://doi.org/10.3390%2Ftechnologies9010002). ISSN 2227-7080 (https://search.worldcat.org/issn/2227-7080).

54. Alex Ratner; Stephen Bach; Paroma Varma; Chris. "Weak Supervision: The New Programming Paradigm for Machine Learning" (https://web.archive.org/web/20190606043931/https://hazyresear ch.github.io/snorkel/blog/ws_blog_post.html). *hazyresearch.github.io*. referencing work by many other members of Hazy Research. Archived from the original (https://hazyresearch.github.io/snork el/blog/ws_blog_post.html) on 2019-06-06. Retrieved 2019-06-06.

55. van Otterlo, M.; Wiering, M. (2012). "Reinforcement Learning and Markov Decision Processes". *Reinforcement Learning*. Adaptation, Learning, and Optimization. Vol. 12. pp. 3–42. doi:10.1007/978-3-642-27645-3_1 (https://doi.org/10.1007%2F978-3-642-27645-3_1). ISBN 978-3-642-27644-6.

56. Roweis, Sam T.; Saul, Lawrence K. (22 Dec 2000). "Nonlinear Dimensionality Reduction by Locally Linear Embedding" (https://science.sciencemag.org/content/290/5500/2323). *Science*. **290** (5500): 2323–2326. Bibcode:2000Sci...290.2323R (https://ui.adsabs.harvard.edu/abs/2000Sci...290.2323R). doi:10.1126/science.290.5500.2323 (https://doi.org/10.1126%2Fscience.290.5500.2323). PMID 11125150 (https://pubmed.ncbi.nlm.nih.gov/11125150). S2CID 5987139 (https://api.semanticscholar.org/CorpusID:5987139). Archived (https://web.archive.org/web/20210815021528/https://science.sciencemag.org/content/290/5500/2323) from the original on 15 August 2021. Retrieved 17 July 2023.

57. Pavel Brazdil; Christophe Giraud Carrier; Carlos Soares; Ricardo Vilalta (2009). *Metalearning: Applications to Data Mining* (Fourth ed.). Springer Science+Business Media. pp. 10–14, *passim*. ISBN 978-3540732624.

58. Bozinovski, S. (1982). "A self-learning system using secondary reinforcement". In Trappl, Robert (ed.). Cybernetics and Systems Research: Proceedings of the Sixth European Meeting on Cybernetics and Systems Research. North-Holland. pp. 397–402. ISBN 978-0-444-86488-8.

59. Bozinovski, Stevo (2014) "Modeling mechanisms of cognition-emotion interaction in artificial neural networks, since 1981." Procedia Computer Science p. 255-263

60. Bozinovski, S. (2001) "Self-learning agents: A connectionist theory of emotion based on crossbar value judgment." Cybernetics and Systems 32(6) 637–667.

61. Y. Bengio; A. Courville; P. Vincent (2013). "Representation Learning: A Review and New Perspectives". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **35** (8): 1798–1828. arXiv:1206.5538 (https://arxiv.org/abs/1206.5538). doi:10.1109/tpami.2013.50 (https://doi.org/10.1109%2Ftpami.2013.50). PMID 23787338 (https://pubmed.ncbi.nlm.nih.gov/23787338). S2CID 393948 (https://api.semanticscholar.org/CorpusID:393948).

62. Nathan Srebro; Jason D. M. Rennie; Tommi S. Jaakkola (2004). *Maximum-Margin Matrix Factorization*. NIPS.

63. Coates, Adam; Lee, Honglak; Ng, Andrew Y. (2011). *An analysis of single-layer networks in unsupervised feature learning* (https://web.archive.org/web/20170813153615/http://machinelearning.wustl.edu/mlpapers/paper_files/AISTATS2011_CoatesNL11.pdf) (PDF). Int'l Conf. on AI and Statistics (AISTATS). Archived from the original (http://machinelearning.wustl.edu/mlpapers/paper_files/AISTATS2011_CoatesNL11.pdf) (PDF) on 2017-08-13. Retrieved 2018-11-25.

64. Csurka, Gabriella; Dance, Christopher C.; Fan, Lixin; Willamowski, Jutta; Bray, Cédric (2004). *Visual categorization with bags of keypoints* (https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/csurka-eccv-04.pdf) (PDF). ECCV Workshop on Statistical Learning in Computer Vision. Archived (https://web.archive.org/web/20190713040210/http://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/csurka-eccv-04.pdf) (PDF) from the original on 2019-07-13. Retrieved 2019-08-29.

65. Daniel Jurafsky; James H. Martin (2009). *Speech and Language Processing*. Pearson Education International. pp. 145–146.

66. Lu, Haiping; Plataniotis, K.N.; Venetsanopoulos, A.N. (2011). "A Survey of Multilinear Subspace Learning for Tensor Data" (http://www.dsp.utoronto.ca/~haiping/Publication/SurveyMSL_PR2011.pdf) (PDF). *Pattern Recognition*. **44** (7): 1540–1551. Bibcode:2011PatRe..44.1540L (https://ui.adsabs.harvard.edu/abs/2011PatRe..44.1540L). doi:10.1016/j.patcog.2011.01.004 (https://doi.org/10.1016%2Fj.patcog.2011.01.004). Archived (https://web.archive.org/web/20190710225429/http://www.dsp.utoronto.ca/~haiping/Publication/SurveyMSL_PR2011.pdf) (PDF) from the original on 2019-07-10. Retrieved 2015-09-04.

67. Yoshua Bengio (2009). *Learning Deep Architectures for AI* (https://books.google.com/books?id=cq5ewg7FniMC&pg=PA3). Now Publishers Inc. pp. 1–3. ISBN 978-1-60198-294-0. Archived (https://web.archive.org/web/20230117053339/https://books.google.com/books?id=cq5ewg7FniMC&pg=PA3) from the original on 2023-01-17. Retrieved 2016-02-15.

68. Tillmann, A. M. (2015). "On the Computational Intractability of Exact and Approximate Dictionary Learning". *IEEE Signal Processing Letters*. **22** (1): 45–49. arXiv:1405.6664 (https://arxiv.org/abs/1405.6664). Bibcode:2015ISPL...22...45T (https://ui.adsabs.harvard.edu/abs/2015ISPL...22...45T). doi:10.1109/LSP.2014.2345761 (https://doi.org/10.1109%2FLSP.2014.2345761). S2CID 13342762 (https://api.semanticscholar.org/CorpusID:13342762).

69. Aharon, M, M Elad, and A Bruckstein. 2006. "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation (http://sites.fas.harvard.edu/~cs278/papers/ksvd.pdf) Archived (https://web.archive.org/web/20181123142158/http://sites.fas.harvard.edu/~cs278/papers/ksvd.pdf) 2018-11-23 at the Wayback Machine." Signal Processing, IEEE Transactions on 54 (11): 4311–4322

70. Zimek, Arthur; Schubert, Erich (2017), "Outlier Detection", *Encyclopedia of Database Systems*, Springer New York, pp. 1–5, doi:10.1007/978-1-4899-7993-3_80719-1 (https://doi.org/10.1007%2F978-1-4899-7993-3_80719-1), ISBN 9781489979933

71. Hodge, V. J.; Austin, J. (2004). "A Survey of Outlier Detection Methodologies" (http://eprints.whiterose.ac.uk/767/1/hodgevj4.pdf) (PDF). *Artificial Intelligence Review*. **22** (2): 85–126. CiteSeerX 10.1.1.318.4023 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.318.4023). doi:10.1007/s10462-004-4304-y (https://doi.org/10.1007%2Fs10462-004-4304-y). S2CID 59941878 (https://api.semanticscholar.org/CorpusID:59941878). Archived (https://web.archive.org/web/20150622042146/http://eprints.whiterose.ac.uk/767/1/hodgevj4.pdf) (PDF) from the original on 2015-06-22. Retrieved 2018-11-25.

72. Dokas, Paul; Ertoz, Levent; Kumar, Vipin; Lazarevic, Aleksandar; Srivastava, Jaideep; Tan, Pang-Ning (2002). "Data mining for network intrusion detection" (https://www-users.cse.umn.edu/~lazar027/MINDS/papers/nsf_ngdm_2002.pdf) (PDF). *Proceedings NSF Workshop on Next Generation Data Mining*. Archived (https://web.archive.org/web/20150923211542/http://www.csee.umbc.edu/~kolari1/Mining/ngdm/dokas.pdf) (PDF) from the original on 2015-09-23. Retrieved 2023-03-26.

73. Chandola, V.; Banerjee, A.; Kumar, V. (2009). "Anomaly detection: A survey". *ACM Computing Surveys*. **41** (3): 1–58. doi:10.1145/1541880.1541882 (https://doi.org/10.1145%2F1541880.1541882). S2CID 207172599 (https://api.semanticscholar.org/CorpusID:207172599).

74. Fleer, S.; Moringen, A.; Klatzky, R. L.; Ritter, H. (2020). "Learning efficient haptic shape exploration with a rigid tactile sensor array, S. Fleer, A. Moringen, R. Klatzky, H. Ritter" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6940144). *PLOS ONE*. **15** (1): e0226880. arXiv:1902.07501 (https://arxiv.org/abs/1902.07501). doi:10.1371/journal.pone.0226880 (https://doi.org/10.1371%2Fjournal.pone.0226880). PMC 6940144 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6940144). PMID 31896135 (https://pubmed.ncbi.nlm.nih.gov/31896135).

75. Moringen, Alexandra; Fleer, Sascha; Walck, Guillaume; Ritter, Helge (2020), Nisky, Ilana; Hartcher-O'Brien, Jess; Wiertlewski, Michaël; Smeets, Jeroen (eds.), "Attention-Based Robot Learning of Haptic Interaction", *Haptics: Science, Technology, Applications*, Lecture Notes in Computer Science, vol. 12272, Cham: Springer International Publishing, pp. 462–470, doi:10.1007/978-3-030-58147-3_51 (https://doi.org/10.1007%2F978-3-030-58147-3_51), ISBN 978-3-030-58146-6, S2CID 220069113 (https://api.semanticscholar.org/CorpusID:220069113)

76. Piatetsky-Shapiro, Gregory (1991), *Discovery, analysis, and presentation of strong rules*, in Piatetsky-Shapiro, Gregory; and Frawley, William J.; eds., *Knowledge Discovery in Databases*, AAAI/MIT Press, Cambridge, MA.

77. Bassel, George W.; Glaab, Enrico; Marquez, Julietta; Holdsworth, Michael J.; Bacardit, Jaume (2011-09-01). "Functional Network Construction in Arabidopsis Using Rule-Based Machine Learning on Large-Scale Data Sets" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3203449). *The Plant Cell*. **23** (9): 3101–3116. Bibcode:2011PlanC..23.3101B (https://ui.adsabs.harvard.edu/abs/2011PlanC..23.3101B). doi:10.1105/tpc.111.088153 (https://doi.org/10.1105%2Ftpc.111.088153). ISSN 1532-298X (https://search.worldcat.org/issn/1532-298X). PMC 3203449 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3203449). PMID 21896882 (https://pubmed.ncbi.nlm.nih.gov/21896882).

78. Agrawal, R.; Imieliński, T.; Swami, A. (1993). "Mining association rules between sets of items in large databases". *Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93*. p. 207. CiteSeerX 10.1.1.40.6984 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.6984). doi:10.1145/170035.170072 (https://doi.org/10.1145%2F170035.170072). ISBN 978-0897915922. S2CID 490415 (https://api.semanticscholar.org/CorpusID:490415).

79. Urbanowicz, Ryan J.; Moore, Jason H. (2009-09-22). "Learning Classifier Systems: A Complete Introduction, Review, and Roadmap" (https://doi.org/10.1155%2F2009%2F736398). *Journal of Artificial Evolution and Applications*. **2009**: 1–25. doi:10.1155/2009/736398 (https://doi.org/10.1155%2F2009%2F736398). ISSN 1687-6229 (https://search.worldcat.org/issn/1687-6229).

80. Plotkin G.D. Automatic Methods of Inductive Inference (https://www.era.lib.ed.ac.uk/bitstream/handle/1842/6656/Plotkin1972.pdf;sequence=1) Archived (https://web.archive.org/web/20171222051034/https://www.era.lib.ed.ac.uk/bitstream/handle/1842/6656/Plotkin1972.pdf;sequence=1) 2017-12-22 at the Wayback Machine, PhD thesis, University of Edinburgh, 1970.

81. Shapiro, Ehud Y. Inductive inference of theories from facts (http://ftp.cs.yale.edu/publications/techreports/tr192.pdf) Archived (https://web.archive.org/web/20210821071609/http://ftp.cs.yale.edu/publications/techreports/tr192.pdf) 2021-08-21 at the Wayback Machine, Research Report 192, Yale University, Department of Computer Science, 1981. Reprinted in J.-L. Lassez, G. Plotkin (Eds.), Computational Logic, The MIT Press, Cambridge, MA, 1991, pp. 199–254.

82. Shapiro, Ehud Y. (1983). *Algorithmic program debugging*. Cambridge, Mass: MIT Press. ISBN 0-262-19218-7

83. Shapiro, Ehud Y. "The model inference system (http://dl.acm.org/citation.cfm?id=1623364) Archived (https://web.archive.org/web/20230406011006/https://dl.acm.org/citation.cfm?id=1623364) 2023-04-06 at the Wayback Machine." Proceedings of the 7th international joint conference on Artificial intelligence-Volume 2. Morgan Kaufmann Publishers Inc., 1981.

84. Burkov, Andriy (2019). *The hundred-page machine learning book*. Polen: Andriy Burkov. ISBN 978-1-9995795-0-0.

85. Russell, Stuart J.; Norvig, Peter (2021). *Artificial intelligence: a modern approach*. Pearson series in artificial intelligence (Fourth ed.). Hoboken: Pearson. ISBN 978-0-13-461099-3.

86. Honglak Lee, Roger Grosse, Rajesh Ranganath, Andrew Y. Ng. "Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.149.802&rep=rep1&type=pdf) Archived (https://web.archive.org/web/20171018182235/http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.149.802&rep=rep1&type=pdf) 2017-10-18 at the Wayback Machine" Proceedings of the 26th Annual International Conference on Machine Learning, 2009.

87. Cortes, Corinna; Vapnik, Vladimir N. (1995). "Support-vector networks" (https://doi.org/10.1007%2FBF00994018). *Machine Learning*. **20** (3): 273–297. doi:10.1007/BF00994018 (https://doi.org/10.1007%2FBF00994018).

88. Stevenson, Christopher. "Tutorial: Polynomial Regression in Excel" (https://facultystaff.richmond.edu/~cstevens/301/Excel4.html). *facultystaff.richmond.edu*. Archived (https://web.archive.org/web/20130602200850/https://facultystaff.richmond.edu/~cstevens/301/Excel4.html) from the original on 2 June 2013. Retrieved 22 January 2017.

89. The documentation for scikit-learn also has similar examples (http://scikit-learn.org/stable/auto_examples/gaussian_process/plot_compare_gpr_krr.html) Archived (https://web.archive.org/web/20221102184805/https://scikit-learn.org/stable/auto_examples/gaussian_process/plot_compare_gpr_krr.html) 2022-11-02 at the Wayback Machine.

90. Goldberg, David E.; Holland, John H. (1988). "Genetic algorithms and machine learning" (https://deepblue.lib.umich.edu/bitstream/2027.42/46947/1/10994_2005_Article_422926.pdf) (PDF). *Machine Learning*. **3** (2): 95–99. doi:10.1007/bf00113892 (https://doi.org/10.1007%2Fbf00113892). S2CID 35506513 (https://api.semanticscholar.org/CorpusID:35506513). Archived (https://web.archive.org/web/20110516025803/http://deepblue.lib.umich.edu/bitstream/2027.42/46947/1/10994_2005_Article_422926.pdf) (PDF) from the original on 2011-05-16. Retrieved 2019-09-03.

91. Michie, D.; Spiegelhalter, D. J.; Taylor, C. C. (1994). "Machine Learning, Neural and Statistical Classification". *Ellis Horwood Series in Artificial Intelligence*. Bibcode:1994mlns.book.....M (https://ui.adsabs.harvard.edu/abs/1994mlns.book.....M).

92. Zhang, Jun; Zhan, Zhi-hui; Lin, Ying; Chen, Ni; Gong, Yue-jiao; Zhong, Jing-hui; Chung, Henry S.H.; Li, Yun; Shi, Yu-hui (2011). "Evolutionary Computation Meets Machine Learning: A Survey". *Computational Intelligence Magazine*. **6** (4): 68–75. doi:10.1109/mci.2011.942584 (https://doi.org/10.1109%2Fmci.2011.942584). S2CID 6760276 (https://api.semanticscholar.org/CorpusID:6760276).

93. "Federated Learning: Collaborative Machine Learning without Centralized Training Data" (http://ai.googleblog.com/2017/04/federated-learning-collaborative.html). *Google AI Blog*. 6 April 2017. Archived (https://web.archive.org/web/20190607054623/https://ai.googleblog.com/2017/04/federated-learning-collaborative.html) from the original on 2019-06-07. Retrieved 2019-06-08.

94. Machine learning is included in the CFA Curriculum (discussion is top-down); see: Kathleen DeRose and Christophe Le Lanno (2020). "Machine Learning" (https://www.cfainstitute.org/-/media/documents/study-session/2020-l2-ss3.ashx) Archived (https://web.archive.org/web/20200113085425/https://www.cfainstitute.org/-/media/documents/study-session/2020-l2-ss3.ashx) 2020-01-13 at the Wayback Machine.

95. Ivanenko, Mikhail; Smolik, Waldemar T.; Wanta, Damian; Midura, Mateusz; Wróblewski, Przemysław; Hou, Xiaohan; Yan, Xiaoheng (2023). "Image Reconstruction Using Supervised Learning in Wearable Electrical Impedance Tomography of the Thorax" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10538128). *Sensors*. **23** (18): 7774. Bibcode:2023Senso..23.7774I (https://ui.adsabs.harvard.edu/abs/2023Senso..23.7774I). doi:10.3390/s23187774 (https://doi.org/10.3390%2Fs23187774). PMC 10538128 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10538128). PMID 37765831 (https://pubmed.ncbi.nlm.nih.gov/37765831).

96. "BelKor Home Page" (https://web.archive.org/web/20151110062742/http://www2.research.att.com/~volinsky/netflix/) research.att.com

97. "The Netflix Tech Blog: Netflix Recommendations: Beyond the 5 stars (Part 1)" (https://web.archive.org/web/20160531002916/http://techblog.netflix.com/2012/04/netflix-recommendations-beyond-5-stars.html). 2012-04-06. Archived from the original (http://techblog.netflix.com/2012/04/netflix-recommendations-beyond-5-stars.html) on 31 May 2016. Retrieved 8 August 2015.

98. Scott Patterson (13 July 2010). "Letting the Machines Decide" (https://www.wsj.com/articles/SB10001424052748703834604575365310813948080). The Wall Street Journal. Archived (https://web.archive.org/web/20180624151019/https://www.wsj.com/articles/SB10001424052748703834604575365310813948080) from the original on 24 June 2018. Retrieved 24 June 2018.

99. Vinod Khosla (January 10, 2012). "Do We Need Doctors or Algorithms?" (https://techcrunch.com/2012/01/10/doctors-or-algorithms/). Tech Crunch. Archived (https://web.archive.org/web/20180618175811/https://techcrunch.com/2012/01/10/doctors-or-algorithms/) from the original on June 18, 2018. Retrieved October 20, 2016.

100. When A Machine Learning Algorithm Studied Fine Art Paintings, It Saw Things Art Historians Had Never Noticed (https://medium.com/the-physics-arxiv-blog/when-a-machine-learning-algorithm-studied-fine-art-paintings-it-saw-things-art-historians-had-never-b8e4e7bf7d3e) Archived (https://web.archive.org/web/20160604072143/https://medium.com/the-physics-arxiv-blog/when-a-machine-learning-algorithm-studied-fine-art-paintings-it-saw-things-art-historians-had-never-b8e4e7bf7d3e) 2016-06-04 at the Wayback Machine, *The Physics at ArXiv blog*

101. Vincent, James (2019-04-10). "The first AI-generated textbook shows what robot writers are actually good at" (https://www.theverge.com/2019/4/10/18304558/ai-writing-academic-research-book-springer-nature-artificial-intelligence). *The Verge*. Archived (https://web.archive.org/web/20190505200409/https://www.theverge.com/2019/4/10/18304558/ai-writing-academic-research-book-springer-nature-artificial-intelligence) from the original on 2019-05-05. Retrieved 2019-05-05.

102. Vaishya, Raju; Javaid, Mohd; Khan, Ibrahim Haleem; Haleem, Abid (July 1, 2020). "Artificial Intelligence (AI) applications for COVID-19 pandemic" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7195043). *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*. **14** (4): 337–339. doi:10.1016/j.dsx.2020.04.012 (https://doi.org/10.1016%2Fj.dsx.2020.04.012). PMC 7195043 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7195043). PMID 32305024 (https://pubmed.ncbi.nlm.nih.gov/32305024).

103. Rezapouraghdam, Hamed; Akhshik, Arash; Ramkissoon, Haywantee (March 10, 2021). "Application of machine learning to predict visitors' green behavior in marine protected areas: evidence from Cyprus" (https://doi.org/10.1080%2F09669582.2021.1887878). *Journal of Sustainable Tourism*. **31** (11): 2479–2505. doi:10.1080/09669582.2021.1887878 (https://doi.org/10.1080%2F09669582.2021.1887878). hdl:10037/24073 (https://hdl.handle.net/10037%2F24073).

104. Dey, Somdip; Singh, Amit Kumar; Wang, Xiaohang; McDonald-Maier, Klaus (2020-06-15). "User Interaction Aware Reinforcement Learning for Power and Thermal Efficiency of CPU-GPU Mobile MPSoCs" (https://ieeexplore.ieee.org/document/9116294). *2020 Design, Automation & Test in Europe Conference & Exhibition (DATE)* (http://repository.essex.ac.uk/27546/1/User%20Interaction%20Aware%20Reinforcement%20Learning.pdf) (PDF). pp. 1728–1733. doi:10.23919/DATE48585.2020.9116294 (https://doi.org/10.23919%2FDATE48585.2020.9116294). ISBN 978-3-9819263-4-7. S2CID 219858480 (https://api.semanticscholar.org/CorpusID:219858480). Archived (https://web.archive.org/web/20211213192526/https://ieeexplore.ieee.org/document/9116294/) from the original on 2021-12-13. Retrieved 2022-01-20.

105. Quested, Tony. "Smartphones get smarter with Essex innovation" (https://www.businessweekly.co.uk/news/academia-research/smartphones-get-smarter-essex-innovation). *Business Weekly*. Archived (https://web.archive.org/web/20210624200126/https://www.businessweekly.co.uk/news/academia-research/smartphones-get-smarter-essex-innovation) from the original on 2021-06-24. Retrieved 2021-06-17.

106. Williams, Rhiannon (2020-07-21). "Future smartphones 'will prolong their own battery life by monitoring owners' behaviour' " (https://inews.co.uk/news/technology/future-smartphones-prolong-battery-life-monitoring-behaviour-558689). *i*. Archived (https://web.archive.org/web/20210624201153/https://inews.co.uk/news/technology/future-smartphones-prolong-battery-life-monitoring-behaviour-558689) from the original on 2021-06-24. Retrieved 2021-06-17.

107. Rasekhschaffe, Keywan Christian; Jones, Robert C. (2019-07-01). "Machine Learning for Stock Selection" (https://www.tandfonline.com/doi/full/10.1080/0015198X.2019.1596678). *Financial Analysts Journal*. **75** (3): 70–88. doi:10.1080/0015198X.2019.1596678 (https://doi.org/10.1080%2F0015198X.2019.1596678). ISSN 0015-198X (https://search.worldcat.org/issn/0015-198X). S2CID 108312507 (https://api.semanticscholar.org/CorpusID:108312507). Archived (https://web.archive.org/web/20231126160605/https://www.tandfonline.com/doi/full/10.1080/0015198X.2019.1596678) from the original on 2023-11-26. Retrieved 2023-11-26.

108. Chung, Yunsie; Green, William H. (2024). "Machine learning from quantum chemistry to predict experimental solvent effects on reaction rates" (https://xlink.rsc.org/?DOI=D3SC05353A). *Chemical Science*. **15** (7): 2410–2424. doi:10.1039/D3SC05353A (https://doi.org/10.1039%2FD3SC05353A). ISSN 2041-6520 (https://search.worldcat.org/issn/2041-6520). PMC 10866337 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10866337). PMID 38362410 (https://pubmed.ncbi.nlm.nih.gov/38362410). Archived (https://web.archive.org/web/20240519121547/https://pubs.rsc.org/en/content/articlelanding/2024/sc/d3sc05353a) from the original on 2024-05-19. Retrieved 2024-04-21.

109. Sun, Yuran; Huang, Shih-Kai; Zhao, Xilei (2024-02-01). "Predicting Hurricane Evacuation Decisions with Interpretable Machine Learning Methods" (https://doi.org/10.1007%2Fs13753-024-00541-1). *International Journal of Disaster Risk Science*. **15** (1): 134–148. arXiv:2303.06557 (https://arxiv.org/abs/2303.06557). Bibcode:2024IJDRS..15..134S (https://ui.adsabs.harvard.edu/abs/2024IJDRS..15..134S). doi:10.1007/s13753-024-00541-1 (https://doi.org/10.1007%2Fs13753-024-00541-1). ISSN 2192-6395 (https://search.worldcat.org/issn/2192-6395).

110. Sun, Yuran; Zhao, Xilei; Lovreglio, Ruggiero; Kuligowski, Erica (2024-01-01), Naser, M. Z. (ed.), "8 - AI for large-scale evacuation modeling: promises and challenges" (https://www.sciencedirect.com/science/article/pii/B9780128240731000149), *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. 185–204, ISBN 978-0-12-824073-1, archived (https://web.archive.org/web/20240519121547/https://www.sciencedirect.com/science/article/abs/pii/B9780128240731000149) from the original on 2024-05-19, retrieved 2024-05-19

111. Xu, Ningzhe; Lovreglio, Ruggiero; Kuligowski, Erica D.; Cova, Thomas J.; Nilsson, Daniel; Zhao, Xilei (2023-03-01). "Predicting and Assessing Wildfire Evacuation Decision-Making Using Machine Learning: Findings from the 2019 Kincade Fire" (https://doi.org/10.1007/s10694-023-01363-1). *Fire Technology*. **59** (2): 793–825. doi:10.1007/s10694-023-01363-1 (https://doi.org/10.1007%2Fs10694-023-01363-1). ISSN 1572-8099 (https://search.worldcat.org/issn/1572-8099). Archived (https://web.archive.org/web/20240519121534/https://link.springer.com/article/10.1007/s10694-023-01363-1) from the original on 2024-05-19. Retrieved 2024-05-19.

112. Wang, Ke; Shi, Xiupeng; Goh, Algena Pei Xuan; Qian, Shunzhi (2019-06-01). "A machine learning based study on pedestrian movement dynamics under emergency evacuation" (https://www.sciencedirect.com/science/article/pii/S037971121830376X). *Fire Safety Journal*. **106**: 163–176. Bibcode:2019FirSJ.106..163W (https://ui.adsabs.harvard.edu/abs/2019FirSJ.106..163W). doi:10.1016/j.firesaf.2019.04.008 (https://doi.org/10.1016%2Fj.firesaf.2019.04.008). hdl:10356/143390 (https://hdl.handle.net/10356%2F143390). ISSN 0379-7112 (https://search.worldcat.org/issn/0379-7112). Archived (https://web.archive.org/web/20240519121539/https://www.sciencedirect.com/science/article/abs/pii/S037971121830376X) from the original on 2024-05-19. Retrieved 2024-05-19.

113. Zhao, Xilei; Lovreglio, Ruggiero; Nilsson, Daniel (2020-05-01). "Modelling and interpreting pre-evacuation decision-making using machine learning" (https://www.sciencedirect.com/science/article/pii/S0926580519313184). *Automation in Construction*. **113**: 103140. doi:10.1016/j.autcon.2020.103140 (https://doi.org/10.1016%2Fj.autcon.2020.103140). hdl:10179/17315 (https://hdl.handle.net/10179%2F17315). ISSN 0926-5805 (https://search.worldcat.org/issn/0926-5805). Archived (https://web.archive.org/web/20240519121548/https://www.sciencedirect.com/science/article/abs/pii/S0926580519313184) from the original on 2024-05-19. Retrieved 2024-05-19.

114. "Why Machine Learning Models Often Fail to Learn: QuickTake Q&A" (https://web.archive.org/web/20170320225010/https://www.bloomberg.com/news/articles/2016-11-10/why-machine-learning-models-often-fail-to-learn-quicktake-q-a). *Bloomberg.com*. 2016-11-10. Archived from the original (https://www.bloomberg.com/news/articles/2016-11-10/why-machine-learning-models-often-fail-to-learn-quicktake-q-a) on 2017-03-20. Retrieved 2017-04-10.

115. "The First Wave of Corporate AI Is Doomed to Fail" (https://hbr.org/2017/04/the-first-wave-of-corporate-ai-is-doomed-to-fail). *Harvard Business Review*. 2017-04-18. Archived (https://web.archive.org/web/20180821032004/https://hbr.org/2017/04/the-first-wave-of-corporate-ai-is-doomed-to-fail) from the original on 2018-08-21. Retrieved 2018-08-20.

116. "Why the A.I. euphoria is doomed to fail" (https://venturebeat.com/2016/09/17/why-the-a-i-euphoria-is-doomed-to-fail/). *VentureBeat*. 2016-09-18. Archived (https://web.archive.org/web/20180819124138/https://venturebeat.com/2016/09/17/why-the-a-i-euphoria-is-doomed-to-fail/) from the original on 2018-08-19. Retrieved 2018-08-20.

117. "9 Reasons why your machine learning project will fail" (https://www.kdnuggets.com/2018/07/why-machine-learning-project-fail.html). *www.kdnuggets.com*. Archived (https://web.archive.org/web/20180821031802/https://www.kdnuggets.com/2018/07/why-machine-learning-project-fail.html) from the original on 2018-08-21. Retrieved 2018-08-20.

118. Babuta, Alexander; Oswald, Marion; Rinik, Christine (2018). Transparency and Intelligibility (https://www.jstor.org/stable/resrep37375.8) (Report). Royal United Services Institute (RUSI). pp. 17–22. Archived (https://web.archive.org/web/20231209002929/https://www.jstor.org/stable/resrep37375.8) from the original on 2023-12-09. Retrieved 2023-12-09.

119. "Why Uber's self-driving car killed a pedestrian" (https://www.economist.com/the-economist-explains/2018/05/29/why-ubers-self-driving-car-killed-a-pedestrian). *The Economist*. Archived (https://web.archive.org/web/20180821031818/https://www.economist.com/the-economist-explains/2018/05/29/why-ubers-self-driving-car-killed-a-pedestrian) from the original on 2018-08-21. Retrieved 2018-08-20.

120. "IBM's Watson recommended 'unsafe and incorrect' cancer treatments – STAT" (https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/). *STAT*. 2018-07-25. Archived (https://web.archive.org/web/20180821062616/https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/) from the original on 2018-08-21. Retrieved 2018-08-21.

121. Hernandez, Daniela; Greenwald, Ted (2018-08-11). "IBM Has a Watson Dilemma" (https://www.wsj.com/articles/ibm-bet-billions-that-watson-could-improve-cancer-treatment-it-hasnt-worked-1533961147). *The Wall Street Journal*. ISSN 0099-9660 (https://search.worldcat.org/issn/0099-9660). Archived (https://web.archive.org/web/20180821031906/https://www.wsj.com/articles/ibm-bet-billions-that-watson-could-improve-cancer-treatment-it-hasnt-worked-1533961147) from the original on 2018-08-21. Retrieved 2018-08-21.

122. Allyn, Bobby (Feb 27, 2023). "How Microsoft's experiment in artificial intelligence tech backfired" (https://www.npr.org/2023/02/27/1159630243/how-microsofts-experiment-in-artificial-intelligence-tech-backfired). *National Public Radio*. Archived (https://web.archive.org/web/20231208234056/https://www.npr.org/2023/02/27/1159630243/how-microsofts-experiment-in-artificial-intelligence-tech-backfired) from the original on December 8, 2023. Retrieved Dec 8, 2023.

123. Reddy, Shivani M.; Patel, Sheila; Weyrich, Meghan; Fenton, Joshua; Viswanathan, Meera (2020). "Comparison of a traditional systematic review approach with review-of-reviews and semi-automation as strategies to update the evidence" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7574591). *Systematic Reviews*. **9** (1): 243. doi:10.1186/s13643-020-01450-2 (https://doi.org/10.1186%2Fs13643-020-01450-2). ISSN 2046-4053 (https://search.worldcat.org/issn/2046-4053). PMC 7574591 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7574591). PMID 33076975 (https://pubmed.ncbi.nlm.nih.gov/33076975).

124. Rudin, Cynthia (2019). "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9122117). *Nature Machine Intelligence*. **1** (5): 206–215. doi:10.1038/s42256-019-0048-x (https://doi.org/10.1038%2Fs42256-019-0048-x). PMC 9122117 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9122117). PMID 35603010 (https://pubmed.ncbi.nlm.nih.gov/35603010).

125. Hu, Tongxi; Zhang, Xuesong; Bohrer, Gil; Liu, Yanlan; Zhou, Yuyu; Martin, Jay; LI, Yang; Zhao, Kaiguang (2023). "Crop yield prediction via explainable AI and interpretable machine learning: Dangers of black box models for evaluating climate change impacts on crop yield" (https://doi.org/10.1016%2Fj.agrformet.2023.109458). *Agricultural and Forest Meteorology*. **336**: 109458. doi:10.1016/j.agrformet.2023.109458 (https://doi.org/10.1016%2Fj.agrformet.2023.109458). S2CID 258552400 (https://api.semanticscholar.org/CorpusID:258552400).

126. Domingos 2015, Chapter 6, Chapter 7.

127. Domingos 2015, p. 286.

128. "Single pixel change fools AI programs" (https://www.bbc.com/news/technology-41845878). *BBC News*. 3 November 2017. Archived (https://web.archive.org/web/20180322011306/http://www.bbc.com/news/technology-41845878) from the original on 22 March 2018. Retrieved 12 March 2018.

129. "AI Has a Hallucination Problem That's Proving Tough to Fix" (https://www.wired.com/story/ai-has-a-hallucination-problem-thats-proving-tough-to-fix/). *WIRED*. 2018. Archived (https://web.archive.org/web/20180312024533/https://www.wired.com/story/ai-has-a-hallucination-problem-thats-proving-tough-to-fix/) from the original on 12 March 2018. Retrieved 12 March 2018.

130. Madry, A.; Makelov, A.; Schmidt, L.; Tsipras, D.; Vladu, A. (4 September 2019). "Towards deep learning models resistant to adversarial attacks". arXiv:1706.06083 (https://arxiv.org/abs/1706.06083) [stat.ML (https://arxiv.org/archive/stat.ML)].

131. "Adversarial Machine Learning – CLTC UC Berkeley Center for Long-Term Cybersecurity" (https://cltc.berkeley.edu/aml/). *CLTC*. Archived (https://web.archive.org/web/20220517045352/https://cltc.berkeley.edu/aml/) from the original on 2022-05-17. Retrieved 2022-05-25.

132. "Machine-learning models vulnerable to undetectable backdoors" (https://www.theregister.com/2022/04/21/machine_learning_models_backdoors/). *The Register*. Archived (https://web.archive.org/web/20220513171215/https://www.theregister.com/2022/04/21/machine_learning_models_backdoors/) from the original on 13 May 2022. Retrieved 13 May 2022.

133. "Undetectable Backdoors Plantable In Any Machine-Learning Algorithm" (https://spectrum.ieee.org/machine-learningbackdoor). *IEEE Spectrum*. 10 May 2022. Archived (https://web.archive.org/web/20220511152052/https://spectrum.ieee.org/machine-learningbackdoor) from the original on 11 May 2022. Retrieved 13 May 2022.

134. Goldwasser, Shafi; Kim, Michael P.; Vaikuntanathan, Vinod; Zamir, Or (14 April 2022). "Planting Undetectable Backdoors in Machine Learning Models". arXiv:2204.06974 (https://arxiv.org/abs/2204.06974) [cs.LG (https://arxiv.org/archive/cs.LG)].

135. Kohavi, Ron (1995). "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection" (https://ai.stanford.edu/~ronnyk/accEst.pdf) (PDF). *International Joint Conference on Artificial Intelligence*. Archived (https://web.archive.org/web/20180712102706/http://web.cs.iastate.edu/~jtian/cs573/Papers/Kohavi-IJCAI-95.pdf) (PDF) from the original on 2018-07-12. Retrieved 2023-03-26.

136. Catal, Cagatay (2012). "Performance Evaluation Metrics for Software Fault Prediction Studies" (http://www.uni-obuda.hu/journal/Catal_36.pdf) (PDF). *Acta Polytechnica Hungarica*. **9** (4). Retrieved 2 October 2016.

137. **Cite error: The named reference `Ethics of artificial intelligence Müller-2020` was invoked but never defined (see the help page).**

138. Garcia, Megan (2016). "Racist in the Machine". *World Policy Journal*. **33** (4): 111–117. doi:10.1215/07402775-3813015 (https://doi.org/10.1215%2F07402775-3813015). ISSN 0740-2775 (https://search.worldcat.org/issn/0740-2775). S2CID 151595343 (https://api.semanticscholar.org/CorpusID:151595343).

139. Bostrom, Nick (2011). "The Ethics of Artificial Intelligence" (https://web.archive.org/web/20160304015020/http://www.nickbostrom.com/ethics/artificial-intelligence.pdf) (PDF). Archived from the original (http://www.nickbostrom.com/ethics/artificial-intelligence.pdf) (PDF) on 4 March 2016. Retrieved 11 April 2016.

140. Edionwe, Tolulope. "The fight against racist algorithms" (https://theoutline.com/post/1571/the-fight-against-racist-algorithms). *The Outline*. Archived (https://web.archive.org/web/20171117174504/https://theoutline.com/post/1571/the-fight-against-racist-algorithms) from the original on 17 November 2017. Retrieved 17 November 2017.

141. Jeffries, Adrianne. "Machine learning is racist because the internet is racist" (https://theoutline.com/post/1439/machine-learning-is-racist-because-the-internet-is-racist). *The Outline*. Archived (https://web.archive.org/web/20171117174503/https://theoutline.com/post/1439/machine-learning-is-racist-because-the-internet-is-racist) from the original on 17 November 2017. Retrieved 17 November 2017.

142. Silva, Selena; Kenney, Martin (2018). "Algorithms, Platforms, and Ethnic Bias: An Integrative Essay" (https://brie.berkeley.edu/sites/default/files/brie_wp_2018-3.pdf) (PDF). *Phylon*. **55** (1 & 2): 9–37. ISSN 0031-8906 (https://search.worldcat.org/issn/0031-8906). JSTOR 26545017 (https://www.jstor.org/stable/26545017). Archived (https://web.archive.org/web/20240127200319/https://brie.berkeley.edu/sites/default/files/brie_wp_2018-3.pdf) (PDF) from the original on Jan 27, 2024.

143. Wong, Carissa (2023-03-30). "AI 'fairness' research held back by lack of diversity" (https://www.nature.com/articles/d41586-023-00935-z). *Nature*. doi:10.1038/d41586-023-00935-z (https://doi.org/10.1038%2Fd41586-023-00935-z). PMID 36997714 (https://pubmed.ncbi.nlm.nih.gov/36997714). S2CID 257857012 (https://api.semanticscholar.org/CorpusID:257857012). Archived (https://web.archive.org/web/20230412120505/https://www.nature.com/articles/d41586-023-00935-z) from the original on 2023-04-12. Retrieved 2023-12-09.

144. Zhang, Jack Clark. "Artificial Intelligence Index Report 2021" (https://aiindex.stanford.edu/wp-content/uploads/2021/11/2021-AI-Index-Report_Master.pdf) (PDF). *Stanford Institute for Human-Centered Artificial Intelligence*. Archived (https://web.archive.org/web/20240519121545/https://aiindex.stanford.edu/wp-content/uploads/2021/11/2021-AI-Index-Report_Master.pdf) (PDF) from the original on 2024-05-19. Retrieved 2023-12-09.

145. Caliskan, Aylin; Bryson, Joanna J.; Narayanan, Arvind (2017-04-14). "Semantics derived automatically from language corpora contain human-like biases". *Science*. **356** (6334): 183–186. arXiv:1608.07187 (https://arxiv.org/abs/1608.07187). Bibcode:2017Sci...356..183C (https://ui.adsabs.harvard.edu/abs/2017Sci...356..183C). doi:10.1126/science.aal4230 (https://doi.org/10.1126%2Fscience.aal4230). ISSN 0036-8075 (https://search.worldcat.org/issn/0036-8075). PMID 28408601 (https://pubmed.ncbi.nlm.nih.gov/28408601). S2CID 23163324 (https://api.semanticscholar.org/CorpusID:23163324).

146. Wang, Xinan; Dasgupta, Sanjoy (2016), Lee, D. D.; Sugiyama, M.; Luxburg, U. V.; Guyon, I. (eds.), "An algorithm for L1 nearest neighbor search via monotonic embedding" (http://papers.nips.cc/paper/6227-an-algorithm-for-l1-nearest-neighbor-search-via-monotonic-embedding.pdf) (PDF), *Advances in Neural Information Processing Systems 29*, Curran Associates, Inc., pp. 983–991, archived (https://web.archive.org/web/20170407051313/http://papers.nips.cc/paper/6227-an-algorithm-for-l1-nearest-neighbor-search-via-monotonic-embedding.pdf) (PDF) from the original on 2017-04-07, retrieved 2018-08-20

147. M.O.R. Prates; P.H.C. Avelar; L.C. Lamb (11 Mar 2019). "Assessing Gender Bias in Machine Translation – A Case Study with Google Translate". arXiv:1809.02208 (https://arxiv.org/abs/1809.02208) [cs.CY (https://arxiv.org/archive/cs.CY)].

148. Narayanan, Arvind (August 24, 2016). "Language necessarily contains human biases, and so will machines trained on language corpora" (https://freedom-to-tinker.com/2016/08/24/language-necessarily-contains-human-biases-and-so-will-machines-trained-on-language-corpora/). *Freedom to Tinker*. Archived (https://web.archive.org/web/20180625021555/https://freedom-to-tinker.com/2016/08/24/language-necessarily-contains-human-biases-and-so-will-machines-trained-on-language-corpora/) from the original on June 25, 2018. Retrieved November 19, 2016.

149. Metz, Rachel (March 24, 2016). "Why Microsoft Accidentally Unleashed a Neo-Nazi Sexbot" (https://www.technologyreview.com/s/601111/why-microsoft-accidentally-unleashed-a-neo-nazi-sexbot/). *MIT Technology Review*. Archived (https://web.archive.org/web/20181109023754/https://www.technologyreview.com/s/601111/why-microsoft-accidentally-unleashed-a-neo-nazi-sexbot/) from the original on 2018-11-09. Retrieved 2018-08-20.

150. Vincent, James (Jan 12, 2018). "Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech" (https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai). *The Verge*. Archived (https://web.archive.org/web/20180821031830/https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai) from the original on 2018-08-21. Retrieved 2018-08-20.

151. Crawford, Kate (25 June 2016). "Opinion | Artificial Intelligence's White Guy Problem" (https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html). *New York Times*. Archived (https://web.archive.org/web/20210114220619/https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html) from the original on 2021-01-14. Retrieved 2018-08-20.

152. Simonite, Tom (March 30, 2017). "Microsoft: AI Isn't Yet Adaptable Enough to Help Businesses" (https://www.technologyreview.com/s/603944/microsoft-ai-isnt-yet-adaptable-enough-to-help-businesses/). *MIT Technology Review*. Archived (https://web.archive.org/web/20181109022820/https://www.technologyreview.com/s/603944/microsoft-ai-isnt-yet-adaptable-enough-to-help-businesses/) from the original on 2018-11-09. Retrieved 2018-08-20.

153. Hempel, Jessi (2018-11-13). "Fei-Fei Li's Quest to Make Machines Better for Humanity" (https://www.wired.com/story/fei-fei-li-artificial-intelligence-humanity/). *Wired*. ISSN 1059-1028 (https://search.worldcat.org/issn/1059-1028). Archived (https://web.archive.org/web/20201214095220/https://www.wired.com/story/fei-fei-li-artificial-intelligence-humanity/) from the original on 2020-12-14. Retrieved 2019-02-17.

154. Char, D. S.; Shah, N. H.; Magnus, D. (2018). "Implementing Machine Learning in Health Care—Addressing Ethical Challenges" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5962261). *New England Journal of Medicine*. **378** (11): 981–983. doi:10.1056/nejmp1714229 (https://doi.org/10.1056%2Fnejmp1714229). PMC 5962261 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5962261). PMID 29539284 (https://pubmed.ncbi.nlm.nih.gov/29539284).

155. Research, AI (23 October 2015). "Deep Neural Networks for Acoustic Modeling in Speech Recognition" (http://airesearch.com/ai-research-papers/deep-neural-networks-for-acoustic-modeling-in-speech-recognition/). *airesearch.com*. Archived (https://web.archive.org/web/20160201033801/http://airesearch.com/ai-research-papers/deep-neural-networks-for-acoustic-modeling-in-speech-recognition/) from the original on 1 February 2016. Retrieved 23 October 2015.

156. "GPUs Continue to Dominate the AI Accelerator Market for Now" (https://www.informationweek.com/big-data/ai-machine-learning/gpus-continue-to-dominate-the-ai-accelerator-market-for-now/a/d-id/1336475). *InformationWeek*. December 2019. Archived (https://web.archive.org/web/20200610094310/https://www.informationweek.com/big-data/ai-machine-learning/gpus-continue-to-dominate-the-ai-accelerator-market-for-now/a/d-id/1336475) from the original on 10 June 2020. Retrieved 11 June 2020.

157. Ray, Tiernan (2019). "AI is changing the entire nature of compute" (https://www.zdnet.com/article/ai-is-changing-the-entire-nature-of-compute/). *ZDNet*. Archived (https://web.archive.org/web/20200525144635/https://www.zdnet.com/article/ai-is-changing-the-entire-nature-of-compute/) from the original on 25 May 2020. Retrieved 11 June 2020.

158. "AI and Compute" (https://openai.com/blog/ai-and-compute/). *OpenAI*. 16 May 2018. Archived (https://web.archive.org/web/20200617200602/https://openai.com/blog/ai-and-compute/) from the original on 17 June 2020. Retrieved 11 June 2020.

159. "What is neuromorphic computing? Everything you need to know about how it is changing the future of computing" (https://www.zdnet.com/article/what-is-neuromorphic-computing-everything-you-need-to-know-about-how-it-will-change-the-future-of-computing/). *ZDNET*. 8 December 2020. Retrieved 2024-11-21.

160. "Cornell & NTT's Physical Neural Networks: A "Radical Alternative for Implementing Deep Neural Networks" That Enables Arbitrary Physical Systems Training" (https://syncedreview.com/2021/05/27/deepmind-podracer-tpu-based-rl-frameworks-deliver-exceptional-performance-at-low-cost-28/). *Synced*. 27 May 2021. Archived (https://web.archive.org/web/20211027183428/https://syncedreview.com/2021/05/27/deepmind-podracer-tpu-based-rl-frameworks-deliver-exceptional-performance-at-low-cost-28/) from the original on 27 October 2021. Retrieved 12 October 2021.

161. "Nano-spaghetti to solve neural network power consumption" (https://www.theregister.com/2021/10/05/analogue_neural_network_research/). *The Register*. 5 October 2021. Archived (https://web.archive.org/web/20211006150057/https://www.theregister.com/2021/10/05/analogue_neural_network_research/) from the original on 2021-10-06. Retrieved 2021-10-12.

162. Fafoutis, Xenofon; Marchegiani, Letizia; Elsts, Atis; Pope, James; Piechocki, Robert; Craddock, Ian (2018-05-07). "Extending the battery lifetime of wearable sensors with embedded machine learning" (https://ieeexplore.ieee.org/document/8355116). *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)* (https://research-information.bris.ac.uk/en/publications/b8fdb58b-7114-45c6-82e4-4ab239c1327f). pp. 269–274. doi:10.1109/WF-IoT.2018.8355116 (https://doi.org/10.1109%2FWF-IoT.2018.8355116). hdl:1983/b8fdb58b-7114-45c6-82e4-4ab239c1327f (https://hdl.handle.net/1983%2Fb8fdb58b-7114-45c6-82e4-4ab239c1327f). ISBN 978-1-4673-9944-9. S2CID 19192912 (https://api.semanticscholar.org/CorpusID:19192912). Archived (https://web.archive.org/web/20220118182543/https://ieeexplore.ieee.org/abstract/document/8355116?casa_token=LCpUeGLS1e8AAAAA:2OjuJfNwZBnV2pgDxfnEAC-jbrETv_BpTcX35_aFqN6IULFxu1xbYbVSRpD-zMd4GCUMELyG) from the original on 2022-01-18. Retrieved 2022-01-17.

163. "A Beginner's Guide To Machine learning For Embedded Systems" (https://analyticsindiamag.com/a-beginners-guide-to-machine-learning-for-embedded-systems/). *Analytics India Magazine*. 2021-06-02. Archived (https://web.archive.org/web/20220118182754/https://analyticsindiamag.com/a-beginners-guide-to-machine-learning-for-embedded-systems/) from the original on 2022-01-18. Retrieved 2022-01-17.

164. Synced (2022-01-12). "Google, Purdue & Harvard U's Open-Source Framework for TinyML Achieves up to 75x Speedups on FPGAs | Synced" (https://syncedreview.com/2022/01/12/deepmind-podracer-tpu-based-rl-frameworks-deliver-exceptional-performance-at-low-cost-183/). *syncedreview.com*. Archived (https://web.archive.org/web/20220118182404/https://syncedreview.com/2022/01/12/deepmind-podracer-tpu-based-rl-frameworks-deliver-exceptional-performance-at-low-cost-183/) from the original on 2022-01-18. Retrieved 2022-01-17.

165. Giri, Davide; Chiu, Kuan-Lin; Di Guglielmo, Giuseppe; Mantovani, Paolo; Carloni, Luca P. (2020-06-15). "ESP4ML: Platform-Based Design of Systems-on-Chip for Embedded Machine Learning" (https://ieeexplore.ieee.org/document/9116317). *2020 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. pp. 1049–1054. arXiv:2004.03640 (https://arxiv.org/abs/2004.03640). doi:10.23919/DATE48585.2020.9116317 (https://doi.org/10.23919%2FDATE48585.2020.9116317). ISBN 978-3-9819263-4-7. S2CID 210928161 (https://api.semanticscholar.org/CorpusID:210928161). Archived (https://web.archive.org/web/20220118182342/https://ieeexplore.ieee.org/abstract/document/9116317?casa_token=5l_Tmgrrbu4AAAAA:v7pDHPEWlRuo2Vk3pU06194PO0-W21UOdyZqADrZxrRdPBZDMLwQrjJSAHUhHtzJmLu_VdgW) from the original on 2022-01-18. Retrieved 2022-01-17.

166. Louis, Marcia Sahaya; Azad, Zahra; Delshadtehrani, Leila; Gupta, Suyog; Warden, Pete; Reddi, Vijay Janapa; Joshi, Ajay (2019). "Towards Deep Learning using TensorFlow Lite on RISC-V" (https://edge.seas.harvard.edu/publications/towards-deep-learning-using-tensorflow-lite-risc-v). *Harvard University*. Archived (https://web.archive.org/web/20220117031909/https://edge.seas.harvard.edu/publications/towards-deep-learning-using-tensorflow-lite-risc-v) from the original on 2022-01-17. Retrieved 2022-01-17.

167. Ibrahim, Ali; Osta, Mario; Alameh, Mohamad; Saleh, Moustafa; Chible, Hussein; Valle, Maurizio (2019-01-21). "Approximate Computing Methods for Embedded Machine Learning" (https://ieeexplore.ieee.org/document/8617877). *2018 25th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*. pp. 845–848. doi:10.1109/ICECS.2018.8617877 (https://doi.org/10.1109%2FICECS.2018.8617877). ISBN 978-1-5386-9562-3. S2CID 58670712 (https://api.semanticscholar.org/CorpusID:58670712). Archived (https://web.archive.org/web/20220117031855/https://ieeexplore.ieee.org/abstract/document/8617877?casa_token=arUW5Oy-tzwAAAAA:I9x6edlfskM6kGNFUN9zAFrjEBv_8kYTz7ERTxtXu9jAqdrYCcDbbwjBdgwXvb6QAH_-0VJJ) from the original on 2022-01-17. Retrieved 2022-01-17.

168. "dblp: TensorFlow Eager: A Multi-Stage, Python-Embedded DSL for Machine Learning" (https://dblp.org/rec/journals/corr/abs-1903-01855.html). *dblp.org*. Archived (https://web.archive.org/web/20220118182335/https://dblp.org/rec/journals/corr/abs-1903-01855.html) from the original on 2022-01-18. Retrieved 2022-01-17.

169. Branco, Sérgio; Ferreira, André G.; Cabral, Jorge (2019-11-05). "Machine Learning in Resource-Scarce Embedded Systems, FPGAs, and End-Devices: A Survey" (https://doi.org/10.3390%2Felectronics8111289). *Electronics*. **8** (11): 1289. doi:10.3390/electronics8111289 (https://doi.org/10.3390%2Felectronics8111289). hdl:1822/62521 (https://hdl.handle.net/1822%2F62521). ISSN 2079-9292 (https://search.worldcat.org/issn/2079-9292).

Sources

- Domingos, Pedro (September 22, 2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Basic Books. ISBN 978-0465065707.
- Nilsson, Nils (1998). *Artificial Intelligence: A New Synthesis* (https://archive.org/details/artificialintell0000nils). Morgan Kaufmann. ISBN 978-1-55860-467-4. Archived (https://web.archive.org/web/20200726131654/https://archive.org/details/artificialintell0000nils) from the original on 26 July 2020. Retrieved 18 November 2019.
- Poole, David; Mackworth, Alan; Goebel, Randy (1998). *Computational Intelligence: A Logical Approach* (https://archive.org/details/computationalint00pool). New York: Oxford University Press. ISBN 978-0-19-510270-3. Archived (https://web.archive.org/web/20200726131436/https://archive.org/details/computationalint00pool) from the original on 26 July 2020. Retrieved 22 August 2020.
- Russell, Stuart J.; Norvig, Peter (2003), *Artificial Intelligence: A Modern Approach* (http://aima.cs.berkeley.edu/) (2nd ed.), Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13-790395-2.

Further reading

- Nils J. Nilsson, *Introduction to Machine Learning (https://ai.stanford.edu/people/nilsson/mlbook.html) Archived (https://web.archive.org/web/20190816182600/http://ai.stanford.edu/people/nilsson/mlbook.html) 2019-08-16 at the Wayback Machine*.
- Trevor Hastie, Robert Tibshirani and Jerome H. Friedman (2001). *The Elements of Statistical Learning (https://web.stanford.edu/~hastie/ElemStatLearn/) Archived (https://web.archive.org/web/20131027220938/http://www-stat.stanford.edu/%7Etibs/ElemStatLearn//) 2013-10-27 at the Wayback Machine*, Springer. ISBN 0-387-95284-5.
- Pedro Domingos (September. 2015), *The Master Algorithm*, Basic Books, ISBN 978-0-465-06570-7
- Ian H. Witten and Eibe Frank (2011). *Data Mining: Practical machine learning tools and techniques* Morgan Kaufmann, 664pp., ISBN 978-0-12-374856-0.
- Ethem Alpaydin (2004). *Introduction to Machine Learning*, MIT Press, ISBN 978-0-262-01243-0.
- David J. C. MacKay. *Information Theory, Inference, and Learning Algorithms (http://www.inference.phy.cam.ac.uk/mackay/itila/book.html) Archived (https://web.archive.org/web/20160217105359/http://www.inference.phy.cam.ac.uk/mackay/itila/book.html) 2016-02-17 at the Wayback Machine* Cambridge: Cambridge University Press, 2003. ISBN 0-521-64298-1
- Richard O. Duda, Peter E. Hart, David G. Stork (2001) *Pattern classification* (2nd edition), Wiley, New York, ISBN 0-471-05669-3.
- Christopher Bishop (1995). *Neural Networks for Pattern Recognition*, Oxford University Press. ISBN 0-19-853864-2.
- Stuart Russell & Peter Norvig, (2009). *Artificial Intelligence – A Modern Approach (http://aima.cs.berkeley.edu/) Archived (https://web.archive.org/web/20110228023805/http://aima.cs.berkeley.edu/) 2011-02-28 at the Wayback Machine*. Pearson, ISBN 9789332543515.
- Ray Solomonoff, *An Inductive Inference Machine*, IRE Convention Record, Section on Information Theory, Part 2, pp., 56–62, 1957.

- Ray Solomonoff, *An Inductive Inference Machine (http://world.std.com/~rjs/indinf56.pdf) Archived (https://web.archive.org/web/20110426161749/http://world.std.com/~rjs/indinf56.pdf) 2011-04-26 at the Wayback Machine* A privately circulated report from the 1956 Dartmouth Summer Research Conference on AI.
- Kevin P. Murphy (2021). *Probabilistic Machine Learning: An Introduction (https://probml.github.io/pml-book/book1.html) Archived (https://web.archive.org/web/20210411153246/https://probml.github.io/pml-book/book1.html) 2021-04-11 at the Wayback Machine*, MIT Press.

External links

- International Machine Learning Society (https://web.archive.org/web/20171230081341/http://machinelearning.org/)
- mloss (https://mloss.org/) is an academic database of open-source machine learning software.