Artifical Intelligence Glossary

By Pit Noack and Lea Reinhard

This glossary is intended to provide readers with an overview. And here, too, we have sometimes made some concessions in terms of mathematical and technical precision for the sake of intuitive understanding and make no claim to completeness. The explanations provided here are intended as a starting point, not as preparation for exams.

Activation In a neural network, the value that an artificial neuron outputs. This value depends on the inputs, the assigned weights, the bias and the activation function.

Activation function In an artificial neuron, a function that calculates an output from the sum of all weighted inputs. It emulates the behavior of biological nerve cells to only fire a signal when the potential of the cell body exceeds a threshold value. If a neural network is to be adjusted using gradient descent, it is essential that the activation function is continuous, i.e. that a derivative can be formed.

Approximate solution In many situations, the exact solution to a problem cannot be calculated. This may be due to the fact that a solution method is unknown or too complex or that the necessary information is missing. Instead of dispensing with a solution altogether, procedures are then sought that can provide a solution that is sufficiently precise for the respective application.

Agent In the context of reinforcement learning, a programmable actor that interacts with an environment and should learn to achieve a goal. An agent can be programmed in various ways, for example with the help of Q-learning and deep Q-learning.

Algorithm A method for solving a class of problems that consists of a finite number of defined individual steps and can therefore be implemented by a computer program.

Assignement In mathematics, a unique assignment between two sets A and B. It is given if we can assign exactly one element of B to each element of A.

Association analysis A data mining method that finds content-related connections in playlists, shopping baskets or texts, for example, by evaluating the frequency with which songs, products or terms occur together.

Attention Mechanism in neural networks that gives greater weight to individual elements of an input sequence in order to enable better predictions or transformations.

Autoencoder A special architecture of neural networks that is trained to output the signal that was input as accurately as possible. The signal runs through a hidden

layer that has far fewer neurons than the output and input layer, the so-called bottleneck. Autoencoders are used for dimension reduction, among other things.

Backpropagation A learning method in which the error calculated at the output layer of a neural network - i.e. the difference between the desired and the actual output - is passed through layer by layer from back to front and used to gradually adjust the weights by means of gradient descent.

Bag of Words Encoding Method for text representation based on One Hot Encoding. The text is represented as a vector. This only takes into account whether a word occurs in the text (1) or not (0) - the order of the words therefore plays no role. Extensions count the frequency or set this in relation to the general frequency of the respective word.
Batch In supervised learning of a neural network, it makes sense to feed in the training examples in groups rather than individually. Such groups are referred to as batches.

Bias In an artificial neuron, the equivalent of the resting potential of a biological nerve cell - i.e. the potential of the cell body when no signals from other cells enter it. In the artificial neuron, the bias value is added to the sum of the weighted inputs. The bias value shifts the activation function parallel to the x-axis.

Breadth search A method for searching a graph in which all nodes that are only one edge away from the starting point are selected first, then all nodes that can be reached via two edges, and so on.

Classification A common task for AI systems. It is always a matter of assigning an input to a class. This is the case with image recognition, for example, but also with the assessment of blood images or account entries with regard to illnesses or fraudulent activities.

Cluster In data mining, a group of data points that have been recognized as belonging together.

Confidence A parameter of association analysis: proportion of transactions in which item B occurs when item A is also present.

Continuous In contrast to discrete values, the distances between continuous values can be divided infinitely further. The shoe size 40 is an integer, discrete value, whereas the foot length 25.632 cm is a continuous value, because the foot length can take on an infinite (uncountable) number of values, whereas there is only a finite number of shoe sizes.

Convolutional network A special type of neural network that is particularly suitable for processing visual data

Coordinate system A system of axes used to clearly localize points in space using coordinates. A two-dimensional coordinate system has two axes. Accordingly, we need two coordinates to localize a point.

Cosine similarity The cosine of an angle between two vectors can serve as a measure of their similarity. A cosine of 1 means that the vectors point in exactly the same direction. The smaller the cosine, the more the direction deviates. If the cosine is 0, the two vectors are independent of each other. If it is -1, they point in exactly the opposite direction. In high-dimensional spaces, the cosine similarity is used to calculate the similarity of data points (or their location vectors), as the **Euclidean distance** does not provide meaningful results due to the curse of dimensionality.

Curse of dimensionality A term for a problem in high-dimensional spaces: with each added dimension, the volume surrounding each individual data point multiplies. One of the consequences of this is that, from the point of view of Euclidean distance, the distances between randomly selected pairs of points in most cases show only very small differences. Therefore, the Euclidean distance does not provide a meaningful measure for the similarity of data points in high-dimensional spaces. This is remedied by a different method, The cosine similarity.

Data mining A process in which large amounts of data are searched algorithmically for usable information. Examples include k-means clustering and association analysis.

Data point If a meteorological measurement series collects values for temperature, air pressure and air humidity, each individual measurement result can be regarded as a data point in a space with three dimensions. This principle can be extended to any number of dimensions.

Deep Q-learning A further development of the Q-learning algorithm using tables. The Q-values are not stored individually in a table. Instead, it is calculated using multi-layered neural networks. The method is capable of solving significantly more complex problems than the original Q-learning.

Depth-first search A search method for graphs. While the breadth-first search first searches the immediate neighbors of the starting point, the depth-first search continues in one direction until no further progress is made. In the case of a tree-shaped graph, the first end node is visited first.

Derivative Mathematical function that represents the gradient of another mathematical function. If the original function has a slope of 2 at the point $x = 5$, the function value of the derivative will be 2 at exactly this point.

Determinacy A process is considered to be determinate if it is completely determined by the initial conditions.

Difference vector The result of subtracting two vectors. The difference vector between a position vector A and a position vector B can be represented as an arrow pointing from the tip of vector B to the tip of vector A.

Dimension In an n-dimensional space, n coordinates are required to uniquely determine a position. A line is one-dimensional, so one coordinate is sufficient. A surface has two dimensions, so two coordinates are required, etc.

Dimension reduction If a data set has too many dimensions, it may be necessary to simplify it in order to be able to perform certain calculations with limited computing capacity. There are various approaches to this. Dimensions that are recognized as less relevant can be omitted or existing dimensions can be combined. A neural network, which is used for classification, also performs a dimension reduction, as it reduces a multi-dimensional input signal to a low-dimensional output signal.

Discrete states can usually be represented by integers. If a state is discrete, this means that it originates from a finite or a countably infinite set of states, i.e. it cannot assume any intermediate levels.

Dynamic programming A procedure in which the solution to a complex problem is found by calculating partial solutions step by step. It always starts with the simplest case. A well-known example is the Levenshtein algorithm.

Edge The connection between two nodes in a graph. Edges often represent the relationships between the nodes. If these relationships differ in strength, the edges are given weights.

Edit distance The number of operations - meaning insertions, deletions or replacements of individual characters - that are necessary to convert one character string into another. convert. The editing distance can be adjusted with the Calculate Levenshtein algorithm.

Episode In reinforcement learning, a period of time that begins when the agent is placed in a starting position and ends when it reaches a target state.

Epoch In supervised learning of a neural network, the section of a training in which all existing training examples have been used once for backpropagation.

Environment In reinforcement learning, an environment represents the game or the problem that is to be solved. The environment also defines which positions are possible starting positions, which actions an agent can perform and which states are linked to a reward.

Euclidean distance Measure of the distance between two points in space, calculated using the Pythagorean theorem. Often used for classifying and clustering data points.

Function In mathematics, a function assigns each element of a set (x-value) an element of another set (y-value). For example, the function $y = f(x) = 2x$ assigns each x the double of x as a function value. In computer science, the concept of function is much broader and in most cases stands for a piece of program code

that can be stored under an identifier for multiple use and executed at different points in the program.

Filter kernels originally come from digital image processing. Filter kernels can be used to blur images and tone edges. They are elementary components of convolutional neural networks.

Generative Adversial Network (GAN) A type of neural network that is able to use sample data (usually images) to generate new data that reproduces the properties of the sample data. A GAN can learn from thousands of photos of coffee cups, to produce "coffee cup-like" images.

General artificial intelligence The as yet unachieved goal of many AI researchers: an AI that, unlike all known applications, is not specialized in a single class of problems, but can learn and act just as flexibly and versatilely as a human being.

Generator In the context of supervised learning of a neural network, a module that provides, varies or generates training data.

Gradient An n-dimensional vector function that indicates the direction in which the gradient is greatest for each point of an n-dimensional original function. The length of the vector indicates the strength of the gradient. The gradient is based on a derivative of the original function.

Gradient descent A method for finding a minimum of a function when a direct mathematical calculation is too complex. In neural networks, the gradient descent is used to gradually adjust the weights with the aim of minimizing the errors of the network.

Graphs are mathematical models for network-like structures. They are composed of nodes and edges. If a graph models a highway network, then the nodes represent cities and the edges represent highway sections.

Hidden layer Every neural network has an input layer and an output layer. There may be one or more hidden layers in between.

High-dimensional space A mathematical space with more than three dimensions. In such spaces, conventional intuitions about distance and geometry can fail, which is also known as the curse of dimensionality.

Homonyms Words that are pronounced or spelled the same but have different meanings. Homonyms cause difficulties in systems that process written or spoken language.

Image recognition A classic AI problem: automatically recognizing whether or not a cat, a dog or a specific person is depicted in a photo.

Input layer The front end of a neural network into which a signal is fed.

Item In association analysis, the smallest possible unit, for example a product in an online store or a song in a music streaming service.

Itemset A related set oftems, a subject of association analysis.

K-means clustering A data mining method that divides data points into clusters based on their spatial proximity to each other.

K-nearest neighbor algorithm An algorithm that assigns unclassified data points to a class by evaluating already classified data points. If the k nearest neighbors of an unclassified point P belong to class Z, P is also assigned to class Z.

Learning rate When training a neural network and in Q-learning, a factor that determines how much a single learning step can change a weight or a Q-value.
Levenshtein algorithm An algorithm for calculating the edit distance. A well-known example of dynamic programming.

Lift Parameter from the association analysis. It represents the relationship between the actual co-occurrence of items A and B and a merely random distribution.

Linear separability If two n-dimensional data point sets can be separated from each other by an n-dimensional plane, they are considered linearly separable. In the context of classification tasks, linear separability determines which techniques are suitable for the solution. For example, linearly separable data point sets can be separated using a neural network without a hidden layer, whereas at least one hidden layer is required for linearly non-separable data point sets.

Location vector A vector whose start is at the origin of the coordinate system. Each data point can be assigned a location vector that points to this point.

Loss function A function that calculates the error at the output layer when training a neural network. The aim of the training is to minimize this error.

Mean value The mean value of a dimension of a data set is calculated by adding up all the values and dividing the result by the number of measured values. The mean value is often used to make statements about a data set: "The average sentence length in this text is 7 words."

Markov process Sequences of discrete elements - for example, texts as sequences of letters or words - can be examined using Markov processes to determine which elements follow each other and how often. Applied to a German-language text, for

example, it can be determined that the word "Ich" is more frequently followed by the words "bin" and "habe" than, for example, "buchstabiere" or "flambiere".

Minimax algorithm A classic AI procedure that can be used to win a certain class of two-person games.

Minimum, local and global A local minimum of a function denotes a point x at which the assigned value y in the vicinity of x is minimal, i.e. does not assume a lower value. A global minimum is a point x at which the assigned value y is minimal for the entire course of the function. In this case, there is therefore no other value x to which the function assigns a smaller y. For more complex functions, the minima cannot be calculated directly and can be determined using gradient descent.

Model A model is a form of representation that takes into account certain aspects of what is depicted while ignoring others. In the context of this book, the term is usually used as a synonym for neural networks, because these can model a person's ability to differentiate, for example.

Neuron (artificial) Building block of a neural network, modeled on the example of biological nerve cells. Neurons are usually arranged in layers. The neurons in the input layer receive their activation from the outside. The neurons in the hidden layers and in the output layer calculate their outputs (i.e. their activation) by weighting and adding up the incoming signals from upstream layers and passing the result to the activation function as an argument.

Neural networks (artificial) A class of AI applications based on the model of biological nervous systems. They are composed of neurons that are usually arranged in layers. There are very different use cases and architectures. What all neural networks have in common is that they assign inputs to outputs and are trained to make the correct assignments using backpropagation and gradient descent. The inputs can be photos or measured values from a blood count, for example. The assigned outputs could then indicate, for example, whether a photo contains a cat or whether the blood count indicates a certain disease.

Neighborhood The neighbors of a node K in a graph are all nodes that are connected to node K by an edge.

Nodes represent the objects in graphs. Nodes that are related to each other are connected by edges.

Non-player character Character in a computer game that is not controlled by a human player. AI algorithms are often used to control these characters and simulate human behavior as credibly as possible.

One hot encoding Technique for translating integers into vectors in which exactly one element is 1 and all others are 0.

Output layer The rear end of a neural network that outputs the result of a calculation.

Overfitting The "memorizing" of training data to be avoided during supervised learning of a neural network.

Positional encodings Technique for adding position data to sequences for further processing in neural networks. This makes it possible to take into account the order of the elements in the sequence.

Probability A measure of the expectability of an event that we do not know whether it will happen or not. In mathematics and computer science, it is expressed as a number between 0 and 1.

Pruning A method to limit the computational effort when searching in a tree-shaped graph: Branches that are irrelevant or less promising with regard to the objective of the search are omitted, i.e. pruned.

Python A universal programming language that is often used for programming AI algorithms.

Q-learning A form of reinforcement learning in which an agent interacts with an environment and learns to solve a task. This can be, for example, successfully playing a computer game or steering a vehicle on uneven terrain. The environment is changeable and can be in different states. The agent can interact with the environment via actions and should learn to choose the action for each environmental state that promises the highest profit in the long term - this could be winning the game with the highest possible score or reaching the goal in the shortest possible time.

Reinforcement learning A class of processes in which an agent learns to achieve a goal independently by exploring a changing environment and receiving feedback for its actions in the form of rewards and punishments. Similar learning processes can be observed in humans and animals.

Search In computer science, many problems can be represented as search problems: For example, the best next move in a two-person game or the shortest path between two nodes in a graph. Widely used search methods in graphs are breadth-first and depth-first search.

Sequence embeddings Translations of sequences (e.g. individual sentences) into vectors in a high-dimensional space. These vectors are intended to preserve essential properties of the original data such as meaning and grammatical Capture structure. Sequence embeddings are an important prerequisite for the performance of transformers such as ChatGPT and DeepL.

Set A collection of elements that belong together by definition: "People who have left the solar system", "The numbers 7, 23 and 42", "All the books in my library", "All integer multiples of the number 7". Sets can be empty, contain a finite number of elements or be infinitely large.

Shape The number and extent of the dimensions of a signal that traverses a neural network.

Skip-gram method A method for generating training data from texts. A window slides over the text. The center of the window is positioned over the respective target word. The words in front and behind are the context words. In this way, pairs are formed from the target word and one of the surrounding context words. Using these word pairs, a neural network learns the most probable context words for each target word. This method is used for training word embeddings.

Slope The difference in height between two points A and B in relation to their horizontal distance. In the context of mathematical, two-dimensional functions, the gradient refers to the difference between the y-value of A and B. For non-linear functions, the gradient is calculated using the derivative.
Supervised learning A method for training neural networks in which training data is fed into the network and the actual outputs are compared with the known correct outputs. The results can be compared and used to adjust the network in order to minimize future errors.

Support Parameter of the association analysis. It represents the proportion of transactions that contain a specific itemset.

Tangent A straight line that touches a curve at a single point without intersecting it. The slope of the tangent is the same as the slope of the curve at exactly this point.

Training data Data required for the supervised learning of a neural network.

Ideally, these consist of a large number of training examples. Each individual example consists of an input and the corresponding correct output.

Transaction In the association analysis, an item set that has actually occurred, for example a shopping cart or a playlist.

Transformer A type of neural network that stands behind pioneering AI applications such as the translation software DeepL and chatbots such as Bard or ChatGPT. Key components are word and sequence embeddings as well as attention mechanisms.

tSNE method Abbreviation for "t-distributed Stochastic Neighbor Embedding" - a method for dimension reduction that is used to visualize data from high-dimensional spaces. In simple terms, it determines the neighborhood relationships of data points in high-dimensional space and then arranges the points in a two- or three-dimensional space in such a way that the neighborhood relationships are preserved as best as possible.

Validation data A part of the training data in the supervised learning of a neural network that is not used for training but for checking the training result. This is primarily used to determine over-adaptation.

Vector A mathematical object that describes a parallel displacement in an n-dimensional space. In a two-dimensional space, a vector can be represented by two numbers that can be read like a path description: (1, 2) would then mean:
"Go one step to the right and two steps up." In illustrations, vectors are usually represented by arrows.

Weight If a graph represents a metro network, the nodes stand for stations and the edges for the track connections. The distances between the stations can be represented by weights. In the context of neural networks, the term stands for the strength of a connection between two neurons

Word embeddings Arrangements of words in a high-dimensional space in which meaningful and grammatical relations between the words can be represented as spatial relations and thus calculated as vectors.

Zero-sum game A game in which one player's win means an equal loss for the other player.