**Term list: Informatics session 6, Machine learning, v. 1.0.1**

 © 2024 James H. Harrison, Jr. This glossary is made available under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/by/4.0/). The contents may be used without restriction, modified, and redistributed if they are attributed and any changes are indicated. Note: A list with additional more specific and technical terms is available as supplemental data to the Harrison et al. article at <http://doi.org/10.5858/arpa.2020-0541-CP>

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Algorithm | The design of a particular type of machine learning program, such as a neural network. The algorithm can be implemented in computer code and then trained to acquire the weights and parameter values necessary to make decisions for input data. See “model.” |
| Annotation | Manually adding information or interpretation to an image or other instance; may include marking regions, identifying image elements, or defining diagnoses. Required in the creation of training images for some types of machine learning. |
| Augmented intelligence | Use of artificial intelligence tools to improve expert human performance rather than replace experts. |
| Backpropagation | A method of training neural networks in which adjustment of the data weights in each node propagates from the downstream portion of the network to the upstream portion, against the flow of data. |
| Bias | A systematic inaccuracy in classification or regression, suggesting either that the model does not accurately represent the underlying data relationships or that there is an unanticipated difference between training and operational data. |
| Bias, aggregation | Inclusion of dissimilar groups without adequate distinction in training/testing data. |
| Bias, embedded | Systematic bias in a model that is not under- or overfit, often results from a mismatch between training/test data and deployment data. |
| Bias, evaluation | Also called measurement bias. Differences in measurement values, value distributions, or value meaning between training/testing and deployment data. |
| Bias, historical | Bias resulting when historical training data does not match current or future deployment data. |
| Bias, sampling | Bias resulting from non-random collection of training/testing data, or from differences in subgroup composition specific to deployment sites. |
| Calibration | In classification, the ability to assign class membership probability accurately across a range of probabilities. |
| Class | A name for similar instances, i.e., type of instance. May be positive/negative for a diagnosis or prognosis, or one of a set of diagnoses. |
| Classification | Distinguishing instances by class. Binary classification distinguishes between two classes. Multi-class classification determines a single class from a set of more than two classes. Multi-label classification allows more than one class label from a set of classes (non-exclusive class membership). |
| Class imbalance | A difference in incidence between the classes in a data set. May make training more difficult and may require special techniques to mitigate. |
| Cleansing data | The overall process of preparing data for use in machine learning. Includes aggregation, dealing with missing values and errors, standardization including mapping different representations, normalization, and formatting. |
| Clustering | Automatically identifying similarities in data sets and grouping data based on those similarities. |
| Convolutional neural network | A neural network in which the first several layers are designed to sweep filters (convolutions) across matrixes of data, usually images, to identify features in the data. |
| Deep learning | Machine learning using deep neural networks, which have multiple hidden layers. Technically, a neural network with more than one hidden layer is “deep.” |
| Dimensions | Also called "features." Dimensions are the data elements that are associated with instances. Instances with only two or three dimensions can be plotted using the dimension values, to show clusters of similar instances. |
| Discrimination | The ability to distinguish between classes and assign cases to classes correctly. |
| Distance metric | A method for calculating the degree of difference between two instances from values of their dimensions. |
| Drift | A gradual change in the distribution of values in a dimension/feature, or the incidence of classes. May reduce model performance. |
| Epoch | One pass through all the training data during model training. |
| Expert system | A form of AI that uses human-built knowledge bases, rules, and inference systems. Useful in certain circumstances but can be time consuming to maintain. |
| Explainability | The degree to which the reasons for decisions by ML models can be reviewed. Some types of models, such as decision trees, produce output that can be examined easily. Other types, such as neural networks, have so many parameters and possible data processing pathways that it is impractical to fully evaluate the basis for a decision. These models are regarded as “black boxes,” though there is research into making them more transparent. |
| Feature | Also called a "dimension." Features are the data elements associated with each instance |
| Feature selection | The art and science of choosing instance features most likely to be useful in a machine learning or analytics task. Most instances have many features and correctly selecting the ones most contributory to the task at hand speeds up machine learning and improves its accuracy. |
| Foundation model | A large deep learning neural network trained against a massive amount of general data that can be used as a starting point for additional specialized or domain-specific training. |
| Generalizability | The ability of a model to be used outside of its specific development environment, or for tasks that are similar to but not identical to its trained purpose. |
| General AI | AI that is capable of learning any task, comparable to human learning. Also called “strong” AI. Does not yet exist. |
| Generative AI | AI algorithms that create new text or image output in response to prompts, for example, ChatGPT and DALL-E. |
| Hidden layer | A layer of nodes in a neural network that is between the input layer and the output layer. |
| Instance | A single example of the item being classified, for example a slide or patient. Instances typically have multiple features or dimensions that contain the data used to train and test the model, and by which they are classified. Training and test data are populations of instances. |
| Label | Class identifier. |
| Large Language Models (LLMs) | Deep neural networks that produce text output in response to prompts, for example, ChatGPT. LLMs often use a transformer architecture to construct text sequences iteratively based on word probabilities and context provided by a prompt. |
| Loss function | A function that calculates the degree of difference between a model's output and the true answer during supervised learning. This calculation determines the amount of adjustment that is made in the model's learned parameters in response to the error. Objective function and cost function are related terms. |
| Machine learning | A form of AI that uses one of a number of algorithms to identify and incorporate patterns in data, and process new data based on those patterns. |
| Model | A function or algorithm that is in training or has been trained and includes learned weights and parameters specific to a particular training data set. |
| Narrow AI | AI that is capable of learning a single well-defined task within a well-defined training environment. Also called “weak” AI. Current machine learning algorithms are examples of this. |
| Neural network | A machine learning algorithm that implements a set of data processing “nodes” that each contains a simple equation that can receive input data from upstream nodes, process it, and transmit output data to downstream nodes. These equations have weights on the input data that can be adjusted independently across the network such that a network of nodes can process complex patterns of data very flexibly. Simple neural networks have a single hidden layer of nodes between the input and output nodes. Deep neural networks have more than one hidden layer. There are a number of types of neural networks with architectural variations that suit them for different purposes. |
| Overfit | A model that does not accurately represent the underlying relationships between the data elements because it incorporates too much specific detail from the training set; an overfit model performs well against the training data but poorly against the held-out test data. |
| Receiver operating characteristic (ROC) curve |  A plot of sensitivity (recall) vs. false positive rate for a machine learning model or other classifier. Used to evaluate discrimination performance. |
| Regression | A machine learning model that returns a continuous numerical value. |
| Self-supervision | A supervised learning strategy employed by large language models in which sequential text output is learned over large text corpuses. While the corpus text is not explicity labeled, the word sequence inherently provides a label-equivalent (the next word for prediction). Some LLMs are trained using self-supervised learning over a large corpus and then tuned with a smaller amount of domain-specific text labeled by experts. |
| Shift | A relatively sudden change in the distribution of values of a dimension/feature (also called "covariate shift"), or in the incidence of classes. May reduce model performance. |
| Stability (of models) | The ability of a model to produce similar output over a range of similar inputs, including inputs not seen previously but not substantially different from those that were seen. Stable models are referred to as "robust" and unstable models that are likely to produce unanticipated output are "brittle." |
| Supervised learning | A machine learning strategy that uses pre-labeled instances to train a model that can then predict labels for unlabeled instances. |
| Test data | Data randomly selected from a correctly-labeled developmental data set that is set aside prior to training to provide an unbiased estimate of model performance after training. |
| Training | Optimizing a machine learning model by using it to process known instances and adjusting weights on input data and model parameters to reduce errors in output. |
| Training data | Data randomly selected from a correctly-labeled developmental data set that is used for training a machine learning model. |
| Transformer | A neural network architecture commonly used in large language models that includes encoding and decoding components along with attention layers that process sequential text features. The transformer iteratively extends a text output based on context provided by a prompt and recent previous output. The transformer architecture was introduced by Vaswani et al in 2017 (doi:10.48550/ARXIV.1706.03762). |
| Underfit | A model that does not accurately represent the underlying relationships in the data elements because it has been incompletely trained; an underfit model performs poorly with both the training and test data sets. |
| Unsupervised learning | A machine learning strategy that uses unlabeled instances and identifies data patterns or instance groups solely based on the characteristics of the data. Often used to investigate the types of structure present in data or to cluster data based on identified patterns. |
| Validation | Optimizing model settings and reliability during training using a subset of training data that is temporarily held out for evaluation purposes (e.g., cross-validation). |
| Variance | Variation in performance of a machine learning model across subsets of data drawn from single population. Low variance is desirable and indicates consistent performance. |
| Verification | Demonstrating that a model’s performance in a new setting (such as a new deployment location) is clinically acceptable and ideally not significantly different from the developer’s reported performance on the initial test data set. Sometimes called “external verification.” |
| Weights | Values learned by machine learning models during training that are multiplied against the input data. Weights serve to emphasize some inputs over others. |