Version 1 - 30.01.2026

AI Glossary

Accuracy

Synonyms: overall correctness, classification accuracy

The proportion of predictions that are correct. Can be misleading if the data is imbalanced.

“Batch inference”

Synonyms: offline inference, bulk inference

Generating model predictions for many data samples at once, typically on a schedule or as a background process.

“Bias”

Synonyms: systematic error

A consistent tendency of a model to make systematic errors. Bias is typically higher when a model is too simple and underfits the data.

“Binary features”

Synonyms: boolean features, yes/no variables, indicator variables

Features that have only two possible values, such as yes/no or 0/1.

“Categorical features”

Synonyms: categorical variables, discrete categories, nominal features

Features that represent groups or labels rather than numeric values (e.g., type, color, diagnosis).

“Continuous features”

Synonyms: numerical features, real-valued variables, continuous variables

Features that can take a wide range of numeric values, often measured on a scale (e.g., age, weight, time).

“Correlated features”

Synonyms: dependent features, related variables, collinear features

Features that are strongly related to each other and contain overlapping information.

“Cross-validation”

Synonyms: k-fold validation

A method for evaluating a model by training and testing it multiple times on different subsets of the same dataset.

“Data leakage”

Synonyms: test-train leakage

When information from outside the training data accidentally influences model training. Makes validation and evaluation meaningless (model “cheats”).

“Dataset balance”

Synonyms: class balance, label distribution

How evenly different outcome classes are represented in a dataset.

“Decision threshold”

Synonyms: classification threshold, cutoff value

A threshold for a classification problem to balance precision and recall. Probabilities above the threshold are considered “True”; probabilities below the threshold are considered “False”.

“Dense layer”

Synonyms: fully connected layer, linear layer

A model layer where every input is connected to every output, often used to combine learned features into predictions.

“Distribution (of training data)”

Synonyms: training data distribution

The statistical characteristics of the data used to train a model.

“Distribution shift”

Synonyms: dataset drift, covariate shift

When real-world data differs significantly in statistical properties from the training data.

“Dropout”

Synonyms: random neuron removal, dropout regularization

A training technique where parts of the model are randomly ignored to reduce overfitting and improve generalization. This is a form of regularization.

“Embedding”

Synonyms: latent representation, vector representation, latent space encoding

A compact numerical representation of data that captures meaning or similarity.

“ETL”

Synonyms: data pipeline, ETL pipeline, data ingestion

The process of Extracting, Transforming, and Loading data so it can be used for analysis or AI models.

“Features”

Synonyms: input data, predictors, attributes

Known characteristics of the input data used by an AI model during training and use. Features are *inputs*, not predictions.

“Feature Engineering”

Synonyms: feature extraction

The process of transforming (preprocessed) data into meaningful features that an AI model can use. In some models, this happens automatically within the model itself.

“F1 score”

Synonyms: F-measure, harmonic mean of precision and recall

A single metric balancing precision and recall.

“Feature importance”

Synonyms: predictor importance

A measure of how much each feature influences the model’s predictions.

“Generalization”

Synonyms: robustness, real-world performance

A model’s ability to perform well on new, unseen data—not just on the data it was trained on.

“Hyperparameter”

Synonyms: tuning parameter, model setting

A configuration value set before training that controls how a model learns. Typically several models are trained with different hyperparameters, and then the best one is selected (by performance on validation set).

“Inference”

Synonyms: prediction phase, model execution, scoring

The process of using a trained AI model to generate predictions on new, unseen data in production. Unlike testing and validation, makes part of model deployment in production.

“Label / Target”

Synonyms: outcome, ground truth, response variable

The value the model is trying to predict during training (provided during training, unknown during testing and inference).

“LIME”

Synonyms: local surrogate explanation

A method that explains individual predictions by approximating the model’s behavior near a specific case.

“Loss”

Synonyms: loss function, error function, objective

A mathematical measure of how wrong a model’s predictions are, used to train the model (typically via gradient descent)

“Metric”

Synonyms: evaluation measure, performance measure, score

A numerical way to evaluate how well an AI model performs.

“Missing data handling”

Synonyms: missing value handling, data imputation, data cleaning

Ways to deal with missing values in a dataset, either by filling them in (imputation) or removing incomplete entries.

“One-hot encoding”

Synonyms: categorical encoding, binary encoding

A method for converting categorical features into binary (0/1) vectors so they can be used by AI models.

“Out-of-distribution (OOD)”

Synonyms: distribution shift, unseen data

Data that differs significantly from the data the model was trained on.

“Overfitting”

Synonyms: memorization, poor generalization

When a model learns the training data too well, “memorizing” it instead of learning patterns, and performs poorly on new data. Happens when the model complexity is high and available diverse data is insufficient. Leads to high variance.

“Oversampling”

Synonyms: class balancing, data rebalancing, minority class augmentation

A data preparation technique that increases the number of samples from under-represented groups by copying or augmenting existing data.

“Outlier”

Synonyms: anomaly, extreme value, unusual observation

A data point that is very different from most other values in the dataset.

“Performance”

Synonyms: model performance, predictive performance, model quality

How well a model achieves its intended task, measured using one or more metrics.

“Pooling”

Synonyms: pooling layer, downsampling, aggregation layer

A part (layer) of a neural network that reduces the size of data by summarizing nearby values, helping the model focus on the most important patterns.

“Precision”

Synonyms: positive predictive value

Among predicted positives, how many are truly positive.

“(To) predict”

Synonyms: to get model outputs, estimation

To obtain an estimated value for an unknown variable using an AI/ML model

“Preprocessing”

Synonyms: data preparation, data cleaning

Steps applied to raw data to make it suitable as features for AI models (e.g., cleaning, encoding, scaling).

“Real-time inference”

Synonyms: online inference, live prediction

Generating model predictions immediately as new data arrives, often with strict time limits (milliseconds to seconds).

“Recall”

Synonyms: sensitivity, true positive rate

Among true positives, how many the model correctly identifies.

“Regularization”

Synonyms: complexity control, penalty

Techniques that limit model complexity to reduce overfitting (for example, dropout).

“ROC-AUC”

Synonyms: AUC, AUROC, discrimination score

A measure of how well a model separates positive and negative cases.

“SHAP”

Synonyms: SHAP values, Shapley explanations

A method that explains individual predictions by showing how each feature contributes to the result.

“Score”

Synonyms: prediction score, confidence score, model output

A numeric value produced by a model, often representing likelihood, probability or confidence.

“Train / validation / test split”

Synonyms: data split, dataset partitioning

Dividing data into:

- ➔ Training set – to learn the model
- ➔ Validation set – to choose hyperparameters and select the best model version
- ➔ Test set – for final evaluation

“Underfitting”

Synonyms: oversimplification, high bias

When a model is too simple to capture important patterns in the data. Leads to high bias.

“Variance”

Synonyms: model variance, instability

How much a model’s performance changes when evaluated on more diverse data. Higher if the model is too complex and overfits to the training dataset.
