Artificial Intelligence / Machine Learning Primer (1 / 3)

version 1.0 - Rob Bothof - 2025

Artificial Intelligence (AI)
A field and practice focused on building computational systems that perform tasks linked to human intelligence, such as recognizing patterns, creating content, or making decisions. The term is also used loosely to describe specific systems or tools that *seem* intelligent. Today it is most often used as a buzzword in presentations or sales pitches.

Artificial General Intelligence (AGI)
A hypothetical type of AI that could understand, learn, and perform *any* intellectual task a human can. Unlike current AI, which is specialized and task-specific, AGI would be flexible and adaptable across domains. Science Fiction.

Machine Learning (ML)
Machine Learning is the academic and practical discipline focused on designing systems (algorithms / models) that can 'learn' from data / examples and then apply that knowledge to new material.

Learning
Iterative process of automated adjustment of a system's internal parameters, where the system's performance is constantly monitored and influences the the adjustments for the next iteration.
See also: Training (A Neural Network specific learning method)

Neural Network
A popular and powerful design architecture within Machine Learning, inspired by the structure of the human brain. It serves as the general architecture for a data model, where different types/designs of Neural Networks can perform different functions. (e.g. recognise, generate, predict). A neural network can be shallow (few layers) or deep (many layers). It is composed composed of layers of connected "neurons"

Model
The actual data structure that is being trained and used to make predictions. The term 'model' is easily misunderstood, we would think of a 'model' as being a design or a blueprint, but this term is somewhat counterintuitive. The model is a specific instance of a neural network, containing the exact layout of neurons, connections, layers and the weights (and biases) used in each neuron. A model is essentially a structured collection of numbers, comparable to a database.

Neuron
A computational unit that receives multiple input values from neurons in the previous layer and produces a single output value. Each incoming value is scaled by a weight associated with the connection it arrives on. The neuron combines these weighted inputs, applies a simple transformation, and forwards the resulting output value to connected neurons in the next layer.

Training
The process of adjusting a neural network model's weights using a dataset of examples. Initially, all the model's parameters (weights) are set to random values. The training process then iteratively evaluates whether changes to the weights improve or worsen how well the model matches the examples, and adjusts the weights accordingly, for example keeping, increasing, or slightly decreasing the values in preparation of the next iteration.

Loss
Measurement tool. A number used during training to show how well the model's outputs match the examples. High loss means the model is doing poorly; lower numbers mean it's doing better. The training process aims to reduce this number over time.

Validation
During training, the model is also tested on data it has not seen yet (similar examples, but ones are not part of the training set). This helps to determine whether the model can generalize to new data or if it has simply memorized the training examples.

Inference
Where training is very resource- and time-intensive and usually happens only once (or occasionally for updates), Inference is the common use case for a neural network model. Inference is using a model to produce an actual result. For example: classifying an image, generating text, or predicting a value. Inference is fast and lightweight compared to training.

Layer
A layer is a group of neurons in a neural network that process data together. Neural Network usually consist of multiple layers stacked on top of each other. Lower layers (closer to the input) typically capture simple features (like edges in images or individual characters or words in text), while higher layers capture more complex or abstract features (like objects such as "cat" or "face" in images, or meaning, sentiment, or topic in text). There are many different types of layers and architectures.

– **Dense Layer**, Fully connected layer, basic building block (weight, bias, activation function)
– **CNN**, Convolutional layer, local connections only, shared weights, used for images / audio
– **RNN / LSTM**, Recurrent layer, uses feedback, designed for sequential data (obsolete)
– **Attention / Transformer**, uses weighted relationships across a data-sequence, backbone modern language and audio models, used for classification and generation

– **GANs**, Generative Adversarial Networks, used for images generation and style transfer (obsolete)
– **Diffusion models,** current state of the art, image, audio, and video generation, Generate data by gradually removing noise
– Autoregressive Transformers, generate data step-by-step, text, music, and audio synthesis

Weight
A weight is a number assigned to a connection between two neurons that controls how strongly the output of one neuron influences the input of the next. Weights scale incoming values before they are combined inside a neuron. During training, these weights are adjusted to strengthen useful connections and weaken unhelpful ones.

Bias
Similar to a weight, the bias is a parameter that can be changed/optimized during the training process. But where weights are stored for each connection between neurons, each neuron has their own bias value and that value is added to the sum of the values from incoming connections. It can be seen as a baseline,offset or tunable starting value for a neural. The bias gives a neural more agency or freedom to react independently from the values it receives on its inputs.

Artificial Intelligence / Machine Learning Primer (2 / 3)

Activation function
An activation function is a transformation that is applied inside a neuron (after the weighted inputs and bias are combined) that shapes its output signal. It acts like a nonlinear gain or dynamics stage, controlling how strongly a neuron responds to its inputs. Neurons are not literally turned on or off; they always produce an output value, but an output close to zero is often described as the neuron being “inactive.”

Deep Learning
A subset of Machine Learning that uses ‘deep’ neural networks with many layers to learn complex patterns from data. Deep Learning models are typically more powerful than shallow networks but require more data and computational resources. It is the technique behind many modern AI applications, like image recognition and large language models.

LLM (Large Language Model)
A type of deep learning model trained on massive amounts of text to understand and generate human-like language. LLMs can perform tasks like answering questions, summarizing text, translating languages, or generating content, and rely on complex neural network architectures with billions of parameters.

Transformer
A neural network architecture designed for processing sequences of data, such as text. Transformers use mechanisms called attention to weigh the importance of different parts of the input, allowing them to capture long-range dependencies efficiently. They are the backbone of modern large language models and many state-of-the-art AI systems.

Fine-Tuning
The process of taking a pre-trained base model and continuing training on a smaller, specific dataset to adapt it to a particular task or domain. This is a popular contemporary method or creating a model. It allows one to create specialized models without the need to train an entire model from scratch.Models can be fine tunes with less cost and with simpler hardware or without the need to collect large datasets. Commonly the weights of the lower layers of a model are left locked and only the higher layers are retrained. In some cases additional layers are added on top of the base layer.

Hyperparameters
Hyperparameters are user-defined settings that control how a model is trained, not the model’s structure itself. They determine learning speed, training duration, and constraints on learning behavior. Hyperparameters are typically adjusted through experimentation and evaluated using validation data.

Classifier
A classifier is a type of AI or machine learning model that assigns an input to one of a predefined set of categories (classes) based on learned patterns in the data.

Label
A label is the class or category assigned to a data example. Labels define the set of possible outcomes a model can predict and are chosen and applied by the developer. In image/object classification, labels might include categories such as *dog*, cat, or *bicycle*.

Feature
A feature is an individual, measurable piece of information used as input to a machine learning model. Feature extraction is a preparation process to transform the data into something that is more usable in a machine learning context. Features describe characteristics of the data that the model uses to learn patterns and make predictions. For example when classifying sound, a feature is usually based on spectral content — which frequencies are present and how they change over time, not the raw audio data.

Dataset
A dataset is a structured collection of data examples used to train, tune, and evaluate an AI or machine learning model.

– **Training set**
  Purpose: Teach the model
  Used to adjust the model’s internal parameters
– **Validation set**
  Purpose: Tune the model
  Used during development to: Estimate the quality of a model, compare models, decide when to stop training. The model does not learn directly from this data
– **Test set**
  Purpose: Evaluate final performance using unseen examples of the data

Overfitting
Overfitting occurs when a model learns the training data too exact. It only memorizes the dataset exactly but is unable to generalize on characteristics and as a result performs poorly on new, unseen data. The most common causes for this are: Too little and unvaried data, the training data does not reflect real-world conditions, bad data preparation or mislabeling.

Artificial Intelligence / Machine Learning Primer (3 / 3)

version 1.0 - Rob Bothof - 2025