Artificial Intelligence (AI) Glossary:
TERMS & DEFINITIONS FOR BEGINNERS

Learn the basic terminology for artificial intelligence technology. From adversarial machine learning to unstructured data—and everything in between—CompTIA's AI Advisory Council compiled a comprehensive list of words and phrases that any professional interested in building an AI solution should be familiar with.

Adversarial Machine Learning
Adversarial machine learning is a technique employed in the field of machine learning that attempts to make models more robust by exposing them to adversarial (and sometimes malicious) input.

Accuracy
The fraction of predictions that an AI model got right. It is the number of correct predictions measured over the total number of predictions made.

Algorithm
A finite sequence of unambiguous, computer-implementable instructions, typically to solve a class of problems or to perform a computation.

Artificial Intelligence (AI)
Processing according to pre-programmed rules in ways that mimic human abilities. Technically speaking, AI is concerned with computers being able to do the following: reasoning, knowledge representation, planning, learning, natural language processing, perception, robotics, social intelligence, and general intelligence.[3]

AI Ethics
A branch of the ethics of technology specific to artificially intelligent systems. Biases are prone to play a significant role in machine learning based on the data that machines are being trained with and range from gender to race to age to economic status and everything in between.[4]

AI Frameworks
AI frameworks make the creation of machine learning/deep learning, neural networks, and natural language processing (NLP) applications easier and faster by offering ready solutions. Some of the most popular open-source frameworks include TensorFlow, Theano, PyTorch, Sci-Kit, Keras, Microsoft Cognitive ToolKit and Apache Mahout.

AI Model Goodness Measurement Metrics
The goodness of AI models built for specific purposes such as classification, prediction, and clustering are measured using a set of metrics called AI model goodness measurement metrics. These metrics are called AI model goodness measurement metrics. These include metrics such as accuracy, precision, recall, F-measure, word error rate, sentence error rate, mean absolute error, general language understanding evaluation (GLUE), etc.

AI Ops
Optimizing IT operations using AI. This involves detecting anomalies from IT system logs, metrics, grouping various events or alerts, diagnosing problems, and resolving issues by learning actions from priori incident, tickets, etc. AI ops is also concerned with monitoring and optimizing application performance, and proactively avoiding issues or incidents.

Automatic Speech Recognition (ASR)
ASR is a type of natural language processing that is associated with recognizing human speech such as voice assistants.

Automation
Processing according to pre-programmed rules.[3]

Brute Force Search
A search that isn't limited by clustering/approximations; it searches across all inputs. Often more time-consuming and expensive, but more thorough.[2]

Computer Vision
An interdisciplinary scientific field that deals with how computers can be made to gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do.

Data Architect
A data architect is a practitioner of data architecture, a data management discipline concerned with designing, creating, deploying, and managing an organization's data architecture. Data architects often work with data scientists on AI projects.[1]

Data Lake
Since data is at the core of every AI use case or solution, aggregating all the data needed to build machine learning and inference models is absolutely critical. The process of assimilating all the data (structured and unstructured) in a consolidated repository is referred to as a data lake.

Data Manager
A data manager is an individual concerned with legally acquiring the right kind of data for training AI systems by working with data scientists. A data manager works with data architects to ensure that acquired data is properly versioned and stored for analysis and audit purposes. A data manager is also concerned with the governance of the data per legal and organizational requirements and ensuring that the lifecycle of the data is managed accordingly.

Data Scientist
A data scientist is an individual, organization or application that performs statistical analysis, data mining and retrieval processes on a large amount of data to identify trends, figures, and other relevant information.

Deep Learning
Deep learning is an artificial intelligence function that imitates the workings of the human brain in processing data and creating patterns for use in decision making. Deep learning is a subset of machine learning in AI that has networks capable of learning unsupervised from data that is unstructured or unlabeled. Also known as deep neural learning or deep neural network.[6]

F Score
A harmonic mean of recall and precision.

Generative Adversarial Network (GAN)
A machine learning model in which two neural networks compete with each other in a goal to generate new data with the same statistics as the training data set. For example, GANs are used in fashion, art, and advertising but are also increasingly used by malicious actors to spread fake news.

ImageNet
A large visual database designed for use in visual object recognition software research. Over 14 million URLs of images have been hand-annotated by ImageNet to indicate what objects are pictured. In at least one million of the images, bounding boxes are also provided.[1]

Machine Learning
Machine learning is a branch of AI that allows systems to automatically process data and analyze for insights without being programmed explicitly. Machine learning is concerned with learning functions and patterns to do things like classification and prediction.

ML Ops
ML ops or machine learning operations is the process of taking an experimental machine learning model into a production web system. Machine learning models are tested and developed in isolated experimental systems. When an algorithm is ready to be launched, ML ops is practiced between data scientists, DevOps, and machine learning engineers to transition the algorithm to production systems.[1]

Natural Language Processing (NLP)
NLP is concerned with information retrieval, text mining, question answering, machine translation, intent understanding, sentiment, emotion, tone extraction in text. It is a branch of AI which uses algorithms to train machines in responding to human conversations.

Natural Language Generation (NLG)
NLG is a branch of NLP that is associated with the processing of unstructured and structured fields into natural language. In other words, NLG is the "write" aspect of NLP where data is used by machines to generate content and information in a human readable format.

Natural Language Understanding (NLU)
NLU is a branch of NLP that is associated with the processing of natural language to convert to structured fields. In other words, it is the "read" part of NLP.

Pattern Recognition
Pattern recognition is the label given to the activity of machines detecting patterns from data. It is often used synonymously with machine learning.[1]

Prescriptive Analytics

Prescriptive analytics is a type of data analytics, the use of technology to help businesses make better decisions through the analysis of raw data. Specifically, prescriptive analytics factors information about possible situations or scenarios, available resources, past performance, and current performance, and suggests a course of action or strategy. It can be used to make decisions on any time horizon, from immediate to long term.[6]

Precision

In pattern recognition, information retrieval and classification (machine learning) precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances.[1]

Predictive Analytics

Predictive analytics is the use of statistics and modeling techniques to determine future performance. It is used as a decision-making tool in a variety of industries and disciplines, such as insurance and marketing. Predictive analytics and machine learning are often confused with each other, but they are different disciplines. Predictive tends to be more statistical and time-series in nature whereas machine learning deals with generative techniques to generate data, reinforcement learning wherein the models can 'figure out' the path, natural language processing, etc. In general, machine learning can solve all predictive problems and then several more that predictive analytics is not concerned with.[6]

Supervised Learning

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs (or labels). This category of algorithms are most widely used in prediction and classification tasks. For example, given a set of pictures of dogs and cats labeled as such, a model is learnt to predict new pictures of unlabeled dogs and cats correctly.[1]

Unsupervised Learning

A type of machine learning that looks for previously undetected patterns in a data set with no pre-existing labels and with a minimum of human supervision. In other terms, it is a method of machine learning where machines automatically sort through the hidden patterns and correlations in the data to offer recommendations without pre-programming.

Recommendation Engines

A recommender system or a recommendation system (sometimes also called a recommendation platform or engine) is a subclass of information filtering systems that seeks to predict the "rating" or "preference" a user would give to an item.

Text to Speech (TTS)

TTS is a type of NLG associated with converting text to speech in natural voices. A common example is a machine reading a prepared piece of text.

Recall

In pattern recognition, information retrieval and classification (machine learning) recall (also known as sensitivity) are the fraction of relevant instances that were retrieved.[2]

Reinforcement Learning (RL)

Reinforcement learning is an area of machine learning concerned with how intelligent agents ought to take actions in an environment in order to maximize the notion of cumulative reward. Reinforcement learning is one of three basic machine learning paradigms, alongside supervised learning, and unsupervised learning. Reinforcement learning is typically used in planning tasks. For example, in autonomous driving, tasks such as path planning, parking a vehicle, dynamic pathing, etc., are implemented using reinforcement learning.[1]

Robotic Process Automation (RPA)

Robotic process automation (RPA) is a software technology that makes it easy to build, deploy, and manage software robots that emulate human actions interacting with digital systems and software. A computer program that automatically fills one name, address and phone number in web forms is an early and simple example of robotic process automation.[5]

Structured Data

Data in any form that is generated, captured, analyzed in a linear, tabular, organized format would be considered structured data, i.e., business data generated within organizations within traditional applications are typically structured data.

Unstructured Data

Unstructured data is defined as the type of data that can have multiple origins from online digital files, text documents, SMS, video, images, voice, sensors, pings, etc.—anything that is not available in a traditional row, column, or table format. Most of the data being generated today is unstructured data and is one of the driving forces behind the rise of AI.

References

1. Wikipedia
2. Marketing Institute article: https://www.marketingaiinstitute.com/blog/the-marketers-guide-to-artificial-intelligence-terminology
3. Analytics Insight article: https://www.analyticsinsight.net/understanding-artificial-intelligence-a-comprehensive-glossary-of-terms-and-definitions/
4. International Consortium for Ethical Development of Artificial Intelligence: https://www.icedai.org/definitions
5. Robotic Process Automation (RPA): https://www.uipath.com/rpa/robotic-process-automation
6. Investopedia