Actua's Artificial Intelligence Activities Glossary

Activity	Glossary Terms
Introduction: What is AI?	Artificial Intelligence, Neural Network, Datasets, Open Source, Machine Learning, Online/Offline Machine Learning
Decision Trees: Classifying Space Objects	Decision Tree, Classification, Dataset, Root Node, Branch, Leaf Node
Regression Analysis: Making Predictions Using Data	Linear Regression, Decision Tree, Machine Learning, Residual
Sort Things Out: Exploring Image Classification	Machine Vision, Class, Classification, AI Model, Confidence, Label, Bounding Box
Hand Commands: Training Image Classification Models	Machine Vision, Classification, AI Model, Teachable Machine, Training
What Machines See: Digging into Machine Vision	Machine Vision, Neural Network, AI Model, Classification
Voice Activated AI: Training Audio Recognition Models	Machine Listening, Classification, AI Model, Training, Datasets
Ethics in AI: Don't let DANN turn evil!	Data Bias, Code of Ethics, Aasimov's Law
Sentiment Analysis: Finding Emotion in Text	Sentiment Analysis, Natural Language, Polarity, Machine Learning for Kids, Natural Language Understanding

Aasimov's Law

Isaac Aasimov was a science fiction writer, best known for his works about robots and artificial intelligence. While he was known for having three laws of robotics, the primary one is the following: "A robot may not injure a human being or, through inaction, allow a human being to come to harm." This is a law that has very real implications in today's world of artificially intelligent robots.

Artificial Intelligence

Artificial Intelligence is the study of creating computer programs that can mimic different parts of human intelligence. This area of study focuses on recreating human abilities that are normally almost impossible for computers, such as decision-making, speech recognition, or translating languages.

Artificial Intelligence Model

An AI model is a part of a computer program that is built to resemble the problem or task the AI is trying to solve. AI models are trained (See: Training) using data and human input to replicate what a human might do to solve the same problem.

Branch

A branch is a component of a decision tree (See: Decision Tree). Branches connect different nodes based on how the data is sorted. Branches don't have any purpose other than to serve as an indicator for which nodes are connected to each other.

Bounding Box

A bounding box is a method used by computer programs to select or separate a specific set of information, contained inside the box. Usually, visual recognition AI programs will use bounding boxes to indicate that they recognize the object inside the bounding box, and can identify it once it's separated from the rest of the image.

Chatbot

A chatbot is a specific type of artificial intelligence designed to replicate one side of human conversations. The ultimate goal of a chatbot is to carry on a conversation where the other person does not realize they are talking to a robot. Many examples of imperfect chatbots exist on the internet, but so far, there are no chatbots that can perfectly pass as human.

Class

In computer programming, a class is a structure to help define the rest of the code. Any code inside the class has a certain amount of preset information that is determined when the class is created. This helps organize code and prevent programmers from having to repeat the same code over and over.

Classification

Classification is the process of separating data into distinct categories, also known as classes, based on certain features of the data. This can vary in difficulty for computers based on the category. For example, a computer can very easily classify objects based on weight. It might have a harder time classifying them based on how old they are, unless it is given that information directly.

Code of Ethics

A code of ethics is a specific list of standards to ensure a group's behaviour is morally and ethically good. These codes are often also developed for AI programs, to clearly develop a set of rules that will prevent the AI from making harmful or illegal decisions.

Confidence

Artificial intelligence programs can often be wrong, especially while training. A confidence score in an AI program is a method of determining how likely it is that the AI is correct. For example, an image recognition AI could say that a picture contains a cat, with 80% confidence. That means, based on its training, the AI has determined there's a 20% chance that it is wrong.

Corpus

A corpus in general terms is a collection of written texts, sometimes the entire writings of a specific author. In the world of artificial intelligence, a corpus is a text file of words and phrases designed to train an AI. These could be question and response triggers for a chatbot, or words in different languages for a translation program.

Data Bias

Data bias is when a sample of data does not accurately represent the population it is meant to represent. An AI can become biased towards a specific section of data simply because it shows up more. For example, if more men than women apply to a job, an AI would likely pick men for similar jobs in the future.
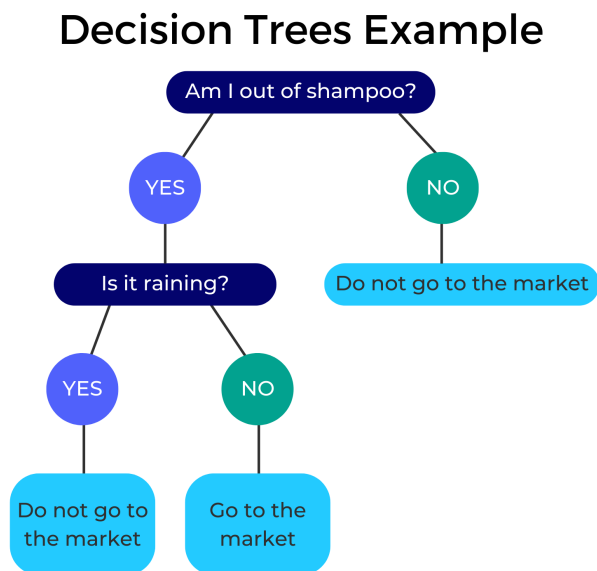
Dataset

A dataset is a collection of organized information. The difference between any collection of data and a defined dataset is that computers can better interpret a dataset and use it. Datasets are often organized in ways that make the most sense for computer programs to utilize it.

Decision Tree

A decision tree is a model that represents how decisions can be made, either by a human or by a computer. Decision trees are used to show a simplified decision, where different questions are asked, and nodes split up based on the answers to those questions. See below for an example.

Decision Trees Example

```
                    Am I out of shampoo?
                   /                    \
                YES                      NO
                 |                        |
            Is it raining?        Do not go to the market
            /            \
         YES              NO
          |                |
   Do not go to       Go to the
   the market          market
```

Feature

A feature is a specific detail in data that can be used, in combination with other features, to help identify or classify the data. For example, images have features such as edges, lines, shapes, and other patterns, which can all be used to determine what an image has in it.

Feature Extraction

Feature extraction is the process of finding features in data. This is done by changing or manipulating the data in certain ways. For example, in image

classification, feature extraction might be accomplished by using a filter to highlight certain patterns or regions in an image.

Function

A function is a term used in computer programming. It is a smaller collection of code that is meant to accomplish a specific purpose. Other parts of the program can call the function, meaning that the entire section of code within that function can be run using only a single line. Functions are used when the same code might need to run multiple times.

General AI/ Narrow AI

General and Narrow AI are two classifications of artificial intelligence. Narrow AI encompasses every AI program that exists today: AI programs designed to accomplish a very narrow set of tasks. General AI has so far only been seen in science fiction; it is an artificial intelligence that perfectly mimics a human in every way imaginable. Think of something like C-3PO from Star Wars!

Label

A label is a simple way of organizing data. Pieces of data are assigned labels to determine where they belong, and then are sorted by those labels to create a structure of data that can easily be sorted and used.

Leaf Node

A leaf node is a component of decision trees ([See: Decision Tree](#)). A leaf node is a node that appears at the end of decision trees. Instead of adding another question, the leaf node displays a category for the data to be sorted in. No branches can stem out of a leaf node.

Linear Regression

Linear regression is a way of modelling data that can simplify the data points or reveal potential trends. It involves using all of the data points available to model a line, even if those data points aren't in a straight line at all! The resulting linear model is a "regressed" version of the data, an average model taken from the various other pieces of data.

Logic Adapter

A logic adapter is a program used for artificial intelligences that interact with natural languages ([See: Natural Language](#)). It is a tool that helps chatbots and other natural language processors with understanding certain subsets of language. For example, a mathematical logic adapter is designed to help an AI recognize when math equations are present in a sentence.

Machine Learning

Machine learning is a subset of artificial intelligence. It refers to computer programs that are able to improve themselves, or "learn", without being explicitly programmed to do so. Machine learning programs are designed to evolve and become better the more they are run, as they adapt using repeated experience and large sets of data.

Machine Listening

Machine listening, also known as "Computer Audition", is the process through which a computer program attempts to understand audio. Machine listening allows AI programs to analyze spoken language, as well as other forms of sound. It is most often used in virtual assistants like Siri or Google Home.

Machine Vision

Machine vision is the visual equivalent to machine listening. It is the process where an AI breaks down and analyzes photos or videos to develop some form of understanding. This is applied to many kinds of visual mediums, from artistic classification to translating written language.

Natural Language

Natural languages, to a computer, are any language that a human speaks. Unlike computer languages, such as binary, natural languages carry lots of context and hidden rules that make it difficult for computers to fully process. Natural language processing is a dedicated subfield of artificial intelligence, where computer programs attempt to analyze and understand human languages.

Natural Language Understanding

Natural language understanding is the more developed form of natural language processing. A natural language processor can understand the basics of human language, and develop simple responses to questions. An AI with natural language understanding, however, begins to process the subtleties of human languages, including things like emotion, subtext, and sarcasm.

Neural Network

A neural network is a specific type of artificial intelligence. Inspired by the way biologic brains are wired, neural networks are composed of several separate computer programs that communicate with each other, loosely modelling the neuron pathways in a brain. Multiple programs working in connection like this allows for more complex processing abilities.

Online/Offline Machine Learning

An online machine learning program is one that is continuously updating itself with any new information it gets. These programs, like Google Quickdraw, take the data from each time someone uses them, and add it to their dataset. In contrast, an offline machine learning program no longer updates itself, and instead keeps its dataset static, and does not add to it.

Open Source

An open source program is one where the code has been made freely available for others to use and modify. Making a program open source is generally a sign of courtesy in computer science, unless there is a good reason (such as copyright issues) to keep the code private.

Polarity

When an AI attempts sentiment analysis, one of the main categories it sorts phrases into is polarity; whether a sentence is positive, negative, or neutral. This is separate from the emotion of a phrase, although usually linked; It's rare that a positive sentence is also an angry one.

Python

Python is a computer programming language used to develop code. It is the basis of some of Actua's AI activities. More information about Python can be found here.

Residual

A residual is a component of linear regression (See: Linear Regression). It represents the difference between a predicted value and an observed one. Essentially, it's a measurement of how close a prediction is to reality. The closer a prediction is, the smaller the residual is.

Root Node

A root node is a component of decision trees ([See: Decision Tree](#)). It is the beginning of a decision tree, from which all the other nodes and branches stem. The root node is meant to be a question that separates the most amount of data, to make it easier for the other nodes to separate it further.

Sentiment Analysis

Sentiment analysis is a subfield of natural language processing. Sentiment analysis involves training AI to recognize the meaning and emotion behind language, identifying phrases as happy, or sad, or angry, and so on.

Supervised/Unsupervised Learning

Supervised or unsupervised learning refers to how a machine learning program is given data to learn from. Supervised learning means that the program is given clearly sorted and labeled data, and is explicitly programmed to sort by those exact labels. An unsupervised learning process involves the program sorting and classifying the data on its own, based on trends that it sees in the data.

Teachable Machine

Teachable Machine is an online tool developed by Google. It makes developing and training machine learning programs very simple. More information about Teachable Machine can be found [here](#).

Training

In the world of artificial intelligence, training is using data to allow a machine learning program to evolve and improve. Training is often a dedicated process that must be programmed into an AI to tell it when to start and when to stop.