Algorithm: a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer.

Bayesian: Being, relating to, or involving statistical methods that assign probabilities or distributions to events (such as rain tomorrow) or parameters (such as a population mean) based on experience or best guesses before experimentation and data collection. These methods involve applying Bayes' theorem to revise the probabilities and distributions after obtaining experimental data.

Bayes' Theorem: a mathematical formula used for calculating conditional probabilities (relates to Bayesian probabilities and Bayesian modeling).

Dataset: a collection of data used for analysis and modeling and typically organized in a structured format.

Deep learning: (see first, *neural network*) part of a broader family of machine learning methods based on artificial neural networks. Deep learning models can learn to classify concepts from of various modalities, such as images, text, or sound.

Generative adversarial networks (GAN): (see first, *neural network* and *deep learning*) a deep learning architecture. It trains two neural networks to compete against each other to generate more authentic new data from a given training dataset. For instance, one can generate new images from an existing image database or original music from a database of songs. A GAN is called adversarial because it trains two different networks and pits them against each other. One network generates new data by taking an input data sample and modifying it as much as possible. The other network tries to predict whether the generated data output belongs in the original dataset. In other words, the predicting network determines whether the generated data is fake or real. The system generates newer, improved versions of fake data values until the predicting network can no longer distinguish fake from original.

Information processing: the acquisition, recording, organization, retrieval, display, and dissemination of information. In recent years, the term has often been applied to computer-based operations specifically.

Machine learning: Machine learning is the study of computer algorithms that can improve automatically through experience and the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data," in order to make predictions or decisions without being explicitly programmed to do so.

Model (as it relates to *machine learning*): the artifact produced after a machine learning algorithm has processed the sample data it was fed during the training phase. The model is then used by the algorithm in production to analyze text (in the case of natural language processing) and return information and/or predictions.

Natural language processing: A subfield of artificial intelligence and linguistics, natural language processing is focused on the interactions between computers and human language. More specifically, it focuses on the ability of computers to read and analyze large volumes of unstructured language data (e.g., text).

Neural network: a means of doing machine learning, in which a computer learns to perform some task by analyzing training examples. Usually, the examples have been hand-labeled in advance. An object recognition system, for instance, might be fed thousands of labeled images of cars, houses, coffee cups, and so on, and it would find visual patterns in the images that consistently correlate with particular labels. Modeled loosely on the human brain, a neural net consists of thousands, millions, or even billions of simple processing nodes that are densely interconnected. Most of today's neural nets are organized into layers of nodes, and they're "feed-forward," meaning that data moves through them in only one direction. An individual node might be connected to several nodes in the layer before it, from which it receives data, and several nodes in the layer after it, to which it sends data.

Self-supervised learning: An approach to machine learning in which labeled data is created from the data itself. It does not rely on historical outcome data or external human supervisors that provide labels or feedback.

Supervised learning: A machine learning algorithm in which the computer is trained using labeled data or machine learning models trained through examples to guide learning.

Techno-Solutionism: At its most basic, techno-solutionism is the mindset that every problem can and should solved with technology.

Technological determinism: an approach that identifies technology, or technological advances, as the central causal element in processes of social change.

Test Set: A test set is a collection of sample problems representative of the challenges and types of content a machine learning solution will face once in production. A test set is used to measure the accuracy of a system after it has gone through training.

Training: the phase in machine learning where practitioners try to fit the best combination of weights and biases to an algorithm to minimize a loss function over the training data.

Training data: the pre-tagged sample data fed to a machine learning algorithm for it to learn about a problem, find patterns, and ultimately produce a model that can recognize those same patterns in future analyses.

Training set: The collection of data used to train an AI model.

Definitions adapted from:
https://www.expert.ai/glossary-of-ai-terms/
https://www.merriam-webster.com/dictionary
https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414
https://aws.amazon.com/what-is/gan/
https://c3.ai/glossary/data-science/