Terms and Definitions

10× genomics (10×) 10× genomics is a popular technique to perform single-cell RNA sequencing. The technique is named after 10× Genomics Inc, the company who developed and owned the technique intellectual property (Chap. 14).

Abductive reasoning A cyclical process of generating possible explanations or a set of hypotheses that are able to account for the available data and then each of these hypotheses is evaluated on the basis of its potential consequences. In this regard, abductive reasoning is a data-driven process that relies heavily on the domain expertise of the person (Chaps. 3 and 5).

Accessibility The characteristic of data that allows research teams to pool data and conduct analyses so that they can have greater confidence in the results (e.g., greater statistical power because the size of the dataset is larger) (Chap. 3).

Acoustic coupler An interface device for coupling electrical signals by acoustical means—usually into and out of a telephone. Such devices were frequently used to connect computer terminals with computers at a distance, often through terminal interface processors (TIPs) (Chap. 2).

Activation function In a neural network, an activation function (such as the sigmoid function) determines the output of a node, given the weighted sum of inputs to the node (Chaps. 1, 6 and 12).

Active learning A human-in-the-loop form of machine learning (ML) where the ML model is re-trained after batches of human annotation and then used to select the next set of data for the human to annotate, often drastically reducing the amount of human annotation needed (Chap. 7).

ACT-R A cognitive architecture offering a theory for simulating and understanding human cognition. Researchers working on ACT-R strive to understand how people organize knowledge and produce intelligent behavior (Chap. 16).

Adaptive artificial intelligence-based clinical decision support An artificial intelligence-based clinical decision support system in which the knowledge base or model is dynamic and is updated with new data and new methods for learning from data (Chap. 10).

T. A. Cohen et al. (eds.), *Intelligent Systems in Medicine and Health*, Cognitive Informatics in Biomedicine and Healthcare,
https://doi.org/10.1007/978-3-031-09108-7

Adjacency matrix A square matrix that describes a graph's connectivity where each element (i, j) in the matrix is a 1 if an edge exists between node i and node j and 0 otherwise (Chap. 4).

Advanced Research Projects Agency (ARPA) U.S. Department of Defense Advanced Research Projects Agency (also known as DARPA) (Chap. 2).

Affordances Attributes of objects that enable individuals to know how to use them (e.g., a door handle affords turning or pushing downward to open a door). Well-designed technologies (Chap. 16).

AI winter A period of reduced funding and interest in artificial intelligence research (Chap. 2).

AI-CDS See artificial intelligence-based clinical decision support (Chap. 10).

AI-CDS system See: artificial intelligence-based clinical decision support system (Chap. 10).

AlexNet The name of a convolutional neural network (CNN) architecture, designed by Alex Krizhevsky in collaboration with Ilya Sutskever and Geoffrey Hinton. AlexNet competed in the ImageNet Large Scale Visual Recognition Challenge, achieving a top-5 error of 15.3%, more than 10.8 percentage points lower than that of the runner up (Chap. 12).

Algorithmovigilance A term (coined by Peter Embi) to describe systems or processes for scrutinizing AI algorithms. It is analogous to post-market "pharmacovigilance," the philosophy and practice of monitoring pharmaceutical products to identify adverse events or other harms after they are approved and in use (Chap. 18).

Annotation The process of labeling or classifying an image using text, annotation tools, or both, to describe its type, subject matter, and other attributes (Chap. 12).

Anonymization The process by which data is irreversibly altered in such a way that a data subject can no longer be identified directly or indirectly, either by the data controller alone or in collaboration with any other party (Chaps. 2 and 3).

Anti-factor Xa (anti-Xa) A laboratory test of blood that measures the inhibition of factor Xa activity, is part of the clotting process (Chap. 15).

Anytime algorithm A class of algorithms that continuously searches for a better and better answer to a problem (so that you can query the algorithm for the best solution at "anytime") rather than the typical algorithm that produces a final answer (Chap. 19).

Applied ethics The practical application of critical thinking about values to address and resolve real-world moral issues (Chap. 18).

Apprenticeship learning In this type of learning, a human expert defines a goal by demonstrating how to attain this goal so that the artificial intelligence system can mimic this behavior of reaching this goal (Chap. 19).

Approximation error In machine learning, this is an error caused by limits on the amount of data available for training. For example, few examples of a category of interest may be available, making it difficult for the model to make accurate predictions for it (Chap. 6).

ARPA See Advanced Research Projects Agency

ARPAnet A large wide-area network (also known as the ARPA Network) created in the 1960s by the U.S. Department of Defense Advanced Research Projects Agency (DARPA) for the free exchange of information among universities and research organizations; the precursor to today's Internet (Chap. 2).

Artificial intelligence (AI) The field of study concerned with endowing computers with the ability to produce behavior and outcomes that are considered to require intelligence when produced by humans (Chaps. 3 and 10).

Artificial intelligence of things (AIoT) Two complementary technologies of artificial intelligence and internet of things (IoT) to enable rapid decisions with analytics from massive volumes of data derived from IoT- (Chap. 19).

Artificial intelligence-based clinical decision support (AI-CDS) A computer-based system that uses artificial intelligence to render clinical decision support (Chap. 10).

Artificial neural network A computer program that performs classification by taking as input a set of findings that describe a given situation, propagating calculated weights through a network of several layers of interconnected nodes, and generating as output a set of numbers, where each output corresponds to the likelihood of a particular classification that could explain the findings (Chaps. 1, 2, 11, and 15).

ASCVD Atherosclerotic cardiovascular disease, defined as acute coronary syndrome, history of myocardial infarction, stable or unstable angina, coronary or other arterial revascularization, stroke, transient ischemic attack, or peripheral arterial disease presumed to be of atherosclerotic origin (Chap. 8).

Attention A mechanism used in neural networks to focus on ("attend to") specific aspects of a complex input (Chap. 4).

Audit data logs A detailed record of every action or activity taken related to data or reports. While paper audit trails were originally kept manually, now digital records can be tracked automatically via audit trail if you have the right software platform capable of data auditing (Chap. 17).

Augmented intelligence A design pattern for a human-centered partnership model of people and AI working together to enhance cognitive performance, including learning, decision making, and new experiences (Chap. 5).

Augmented reality An interactive experience of a real-world environment where the objects that reside in the real world are enhanced by computer-generated perceptual information, which may include visual, auditory, haptic, somatosensory and olfactory modalities (Chap. 16).

Autoencoder A type of artificial neural network that has an hourglass-shaped architecture. The input and output layers have the same number of nodes as the number of dimensions in the data. One layer has the smallest number of nodes, which is called the 'encode' layer. The size of the middle layer decreases from the input to the encode layer and increases from the encode layer to the output layer. The data in the output layer, which is a type of synthetic data, is very similar to the data in the input layer. This means that another classification model could not accurately differentiate the data between the input and the output layer.

If so, the encode layer represents the original data so well that it can create a very similar synthetic dataset (Chaps. 4 and 14).

Automated ML (AutoML) Automating the modeling process that facilitates processes such as data preprocessing, features engineering, model selection, and meta-learning for hyperparameters optimization (Chap. 19).

AutoTutor An intelligent tutoring system that simulates the discourse patterns of human tutors, based on analysis of human-to-human tutoring sessions, and theoretically-grounded tutoring strategies based on cognitive learning principles (Chap. 16).

Backward chaining The computational process of seeking to determine the truth of a goal statement by tracing out potential paths of inference using logical rules that can assert that statement as fact (Chaps. 3 and 4).

Backward reasoning The process of reasoning in the opposite fashion to forward reasoning, where the purpose is to determine the initial facts and information with the help of the given conclusions. It is often referred to as bottom-up reasoning (Chap. 5).

Basis function expansion A technique through which a feature vector for machine learning is expanded into a longer vector containing individual component combinations, using mathematical transformations (Chap. 6).

Bayes' Theorem An algebraic expression often used in clinical diagnosis for calculating posttest probability of a condition (a disease, for example) if the pretest probability (prevalence) of the condition, as well as the sensitivity and specificity of the test, are known (also called Bayes' rule). Bayes' theorem also has broad applicability in other areas of biomedical informatics where probabilistic inference is pertinent, including the interpretation of data in bioinformatics (Chap. 4).

Bayesian belief network A probabilistic model based on known or computed conditional independence of variables, based on Bayes' theorem (Chap. 15).

Bayesian probability theory An approach to probabilistic reasoning that applies an algebraic expression, often used in clinical diagnosis, for calculating posttest probability of a condition (a disease, for example) if the pretest probability (prevalence) of the condition, as well as the sensitivity and specificity of the test, are known (also called Bayes' theorem or Bayes' rule) (Chap. 2).

Bidirectional Encoder Representations from Transformers (BERT) A methodology that applies the bidirectional training of a transformer, which is an attention model, to language modeling with a deeper sense of language context and fluidity (Chap. 19).

Big data A field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software (Chaps. 10 and 12).

Binary classification A special case of classification involving only two classes (Chap. 6).

Binding affinity Binding affinity measures the strength of interaction between a protein and another protein or chemical (Chap. 14).

Bioethics A branch of applied ethics addressing healthcare and its professions, as well as issues arising in human and animal research, clinical care, and public health (Chap. 18).

Bioinformatics A subfield of biomedical informatics that involves the study of how information is represented and transmitted in biological systems, starting at the molecular level (Chap. 2).

Biomedical informatics The interdisciplinary field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving, and decision making, driven by efforts to improve human health (Chap. 2).

Black box A computer or computer system in which the processes that generate output are opaque or inscrutable (Chap. 18).

Bootstrapping A resampling method that mimics the sampling process by using random sampling with replacement. Bootstrapping estimates the properties of an estimator (such as its variance) by measuring those properties when sampling from an approximating distribution. Given a dataset with N samples, bootstrapping repeatedly creates datasets of size N (known as *bootstrap samples*) by resampling with replacement N instances of the original dataset. Each bootstrap sample can include repeated instances of the original dataset. The bootstrap sample is used to train the desired model. The samples that are not selected for the bootstrap sample are used as a test set. The described procedure is repeated B times, and the final performance of the model is computed as an average over the B samples (Chap. 11).

Canonical correlation analysis A way of inferring information from cross-covariance matrices. If we have two vectors X and Y of random variables, which do not need to have the same dimension, and there exist correlations among the variables (for example, in brain cancer, the patients' survival time and the patients' time for a recurrent tumor to occur), then canonical-correlation analysis will find linear transformation of X (denote: X_t) and Y (denote: Y_t) such that X_t and Y_t have maximum correlation with each other (Chap. 14).

Capsular network A relatively new deep learning concept promulgated by Geoff Hinton that is based on human brain modules called "capsules" that are good for routing visual images to the appropriate capsule for improved hierarchical relationships (Chap. 19).

Carotid arterial intimal-medial thickness (CIMT) A widely used and validated imaging technique whereby the thickness of the inner two layers of the carotid artery—the intima and media—are measured, typically by ultrasound or MRI, to detect subclinical vascular disease (Chap. 12).

Causal reasoning The process of identifying causality. i.e., the relationship between a cause and its effect (Chaps. 2 and 3).

Certainty factor An early method of assigning likelihoods to facts and conclusions of rules, and then subsequently accumulating evidence, introduced in the MYCIN system and subsequently adopted in many other systems, including neural networks (Chap. 4).

Chaos theory The study of nonlinear dynamics in mathematics in which seemingly random events are predictable from simple deterministic equations in complex systems, such as weather, migratory patterns of birds, and pandemics (see also Butterfly Effect) (Chap. 19).

Chatbot A computer system that supports a synchronous mode of text-based communication (Chaps. 2 and 9).

Chemical structure The three-dimensional arrangement of atoms in a chemical (Chap. 14).

Chromatin immunoprecipitation (ChIP) Chromatin immunoprecipitation is a technique to investigate the interaction between proteins and DNA in the cell. It aims to determine whether specific proteins are associated with specific genomic regions, such as transcription factors on promoters or other DNA binding sites (Chap. 14).

Chunking A process by which individual pieces of information are grouped together in a meaningful whole, to improve short-term retention of the material, thus bypassing the limited capacity of working memory and allowing the working memory to be more efficient (Chap. 5).

Citizen data scientist A person who works in fields other than those supporting data science (e.g., statistics and analytics) yet creates models using artificial intelligence. The benefit of a citizen data scientist is in adding to the workforce performing analysis, as well as applying knowledge from the other fields to the analysis process (Chap. 15).

Classification As an example: does this patient belong in (i.e., is classifiable into) the group of patients with Type 2 diabetes? (Chaps. 3 and 6).

Clinical decision support Any process that provides healthcare workers and patients with situation-specific knowledge that can inform their decisions regarding health and health care (Chap. 10).

Clinical decision support system (CDSS) A computer-based system that assists physicians in making decisions about patient care (Chap. 10).

Clinomic The study of clinotype. It is also used as an adjective to refer to clinotype (Chap. 14).

Clinotype Refers to the measures and characteristics of the living subject, which are useful for medical research and interventions (Chap. 14).

Cloud computing An approach to computing that uses computing resources, such as processors, data, or files, that are located in a remote location ("in the cloud") (Chap. 2).

Cognitive architecture Sometimes used to describe the third and upcoming wave of AI (the first being programming and the second being current deep learning) with attention to more inter-object relationships such as attention and memory as well as reasoning and other capabilities that are akin to how humans think (Chap. 19).

Cognitive informatics The interdisciplinary domain, comprising the cognitive and information sciences, that focuses on human information processing, mechanisms and processes within the context of computing and computer applications (Chaps. 5 and 8).

Cognitive load theory Cognitive load refers to the used amount of working memory resources. The fundamental tenet of cognitive load theory is that the quality of instructional design will be higher if greater consideration is given to the role and limitations of working memory (Chap. 16).

Cognitive task analysis The analysis of both the information-processing demands of a task and the kinds of domain-specific knowledge required performing it, used to study human performance (Chap. 5).

Competency-based education An approach that allows trainees to distinguish between the skills and knowledge that they already have and those for which they need more education and training. This is contrasted to time-based educational methods (Chap. 16).

Complexity Used to describe a situation that cannot easily be deconstructed meaningfully into separate component parts as each part is interdependent with the other parts (therefore "holistic" with dynamic relationships) and the process is therefore nonlinear and the outcome is highly unpredictable or stochastic (Chap. 19).

Comprehension An understanding and interpretation of what is read, heard, and seen (Chap. 5).

Computable representation A method for storing knowledge in a computer so that it can be manipulated computationally, e.g., to draw inferences (Chap. 16).

Computational linguistics The subfield of linguistics focused on the modeling of language using computational methods, often used as a synonymous term of NLP (Chap. 7).

Computational methods In evaluation work, use of computational tools for capturing and analyzing qualitative data in attempts to find solutions for real-life problems (Chap. 17).

Computerized provider order-entry (CPOE) The process of providers (e.g., physicians) entering and sending treatment instructions—including medication, laboratory, and radiology orders—via a computer application rather than paper, fax, or telephone (Chap. 17).

Concatenation-based integration A method that involves combining data sets from different data types at the raw or processed data level before modelling and analysis (Chap. 3).

Concept normalization (CN) NLP task involving the mapping of a concept in text to a standardized form in a lexicon, terminology, or ontology (a concept-level equivalent of word sense disambiguation) (Chap. 7).

Concept recognition NLP task involving the identification of phrases in text that describe an abstract concept from a particular semantic category (e.g., name of a disease or anatomical region) (Chap. 7).

Conditional independence Formally, two manifestations of a disease, A and B, are conditionally independent when $P(A,B|D)=P(A|D)P(B|D)$, i.e., when their joint probability is simply the product of their individual probabilities given disease D (Chap. 4).

Connectionism A movement in cognitive science that works to explain intellectual abilities using artificial neural networks (Chap. 2).

Consilience A unified theory of knowledge espoused by biologist Edward O. Wilson that involves many disciplines from biology to physics as well as social sciences and the humanities to create a domain where sciences and the arts meet (Chap. 19).

Context mechanism MYCIN's mechanism to allow rules to mention various entities such as cultures and organisms without needing to relate them explicitly (Chap. 4).

Continual learning A concept to learn a model for a large number of tasks sequentially without forgetting knowledge obtained from the preceding tasks, where the data in the old tasks are not available anymore during the training of new ones (Chap. 12).

Control value theory The theory implies that trainees and instructors' achievement emotions can be influenced by changing subjective control and values relating to achievement activities and their outcomes (Chap. 16).

Conversational agent An intelligent agent that converses with humans via a dialog system interface (Chap. 9).

Conversational assistant A conversational agent that uses speech input and output to perform a wide range of tasks, as exemplified by the now ubiquitous Siri, Amazon Alexa, and Google Home products (Chap. 9).

Convex function A real-valued function in which the y value for points on the line segment between two x values always falls beneath the y values of the function at these points. The quadratic $y=x^2$ is a classic example of a convex function (Chap. 6).

Convex hull The convex hull of a set of datapoints is the smallest polygon that both encloses all of them, and does not bend inward upon itself (Chap. 6).

Convolution A mathematical function derived from two given functions by integration to derive a third function, or an integral that blends one function with another function (or simply as an integral of the convolution) (Chap. 19).

Convolution (in image analysis) A linear operation of a filter or kernel to local neighborhoods of points in an input. Since a feature may occur anywhere in the image, the filter weights are shared across all the image positions. Thus, image features can be extracted with fewer parameters, increasing model efficiency (Chap. 12).

Convolutional neural network (CNN) A form of deep neural network used in image processing, which learns translation invariant features across an image. It does so by applying the same set of parameters (called a filter) in different positions within an image, providing the capability to recognize informative features irrespective of their location. The filters slide along input features to summarize sections of the image. They are typically used in multiple layers to abstract features of the entire image (Chaps. 4 and 6).

Copy-number variation (CNV) A phenomenon in which sections of the genome are repeated and the number of repeats in the genome varies among individuals (Chap. 14).

Co-reference resolution NLP task involving the identification of co-referring textual elements—each element often being referred to a "mention" in a "chain" of elements—such that all mentions refer to the same real-world object, such as identifying which named entity a pronoun refers to (Chap. 7).

Cross sectional imaging A discipline of radiology that encompasses the use of a number of advanced imaging techniques (typically CT, MRI, and ultrasound, and occasionally some nuclear medicine techniques) that view the body in cross-section (i.e., as axial) slices (Chap. 12).

Cross-modal application An application that involves information obtained from more than one modality (e.g. image and text data or EHR and genomic data) (Chap. 4).

Cybernetics The science of communications and automatic control systems in both machines and living things (Chap. 2).

DARPA See Advanced Research Projects Agency (D refers to Defense)

Data augmentation The process of augmenting the training dataset with transformations of the input data with the goal of inducing the model to be invariant to the transformation (Chap. 4).

Data ethics A branch of applied ethics that examines issues associated with data collection, storage, and sharing; privacy and confidentiality; appropriate use; etc. (Chap. 18).

Data manifold The different kinds of information attached to data in addition to what is generally termed "metadata": where, when, and how the data was sourced, how was it structured, and how credible is it. (Chap. 18).

Data programming The programmatic creation of training datasets via noisy labelling functions (Chap. 4).

Data-derived artificial intelligence-based clinical decision support An artificial intelligence-based clinical decision support system in which a key component is a model that is typically derived from data (Chap. 10).

Decision tree A diagrammatic representation of the outcomes associated with chance events and voluntary actions. Also a supervised learning method that predicts the value of a target variable by learning simple decision rules inferred from the data features (Chaps. 4 and 11).

Decision tree classifier A flowchart-like structure in which each internal node represents a "test" on an attribute (e.g. whether the value of that attribute is below a certain threshold), each branch represents the outcome of the test, and each leaf node represents a class label. The tree is built in an iterative way, by selecting at each step the best feature to split the data. Having found the best split, data are partitioned into the two resulting regions, and the splitting process is repeated on each of the two regions. Then, this process is repeated on all the resulting regions. For classification trees, splitting is determined on the basis of the "impurity" of the considered node, which can be quantified using different criteria, such as misclassification error, Information Gain, Gain Ratio, and Gini Index. The paths from root to leaf represent classification rules. Similarly, in regression trees, given the continuous nature of the outcome, "impurity" is computed as the Sum of Squared Errors at the node (Chap. 11).

Deductive reasoning An hypothesis-based logical reasoning process that deduces conclusions from test results. It moves from the general rule to the specific application (Chap. 5).

Deep feedforward neural networks Neural networks with an input layer, multiple hidden layers and an output layer that are fully connected to one another, in the sense that every node in one layer connects to every node in the subsequent layer (Chap. 6).

Deep learning A class of machine learning algorithms that uses multiple layers in an artificial neural network to extract higher-level features progressively from raw input. Deep learning can be used both for classification and regression, using architectures that include convolutional and recurrent neural networks as well as deep reinforcement learning. In image interpretation, the approach progressively learns a composition of features to reflect a hierarchy of structures in the data. An end-to-end approach, it learns simple features (such as signal intensity, edges, and textures) as components of more complex features such as shapes, lesions, or organs, thereby leveraging the compositional nature of images (Chaps. 1, 2, 10, 11, 12, 18, and 19).

Deep reinforcement learning A combination of traditional reinforcement learning and an artificial neural network to enable software agents to maximize the reward from various states and actions in the environment. The AlphaGo model from DeepMind is an example of this AI tool (Chap. 19).

De-identification The process by which data are altered to reduce the likelihood that a data subject's identity can be revealed (Chap. 3).

Department of Defense (DOD) In the USA, the largest government agency. It provides military forces and capabilities needed to deter war and ensure the nation's security, including by supporting research. The DOD has been a major source of support for the development of AI methods and applications through its Advanced Research Projects Agency (Chap. 2).

Dependency parsing A syntactic NLP task involving the identification of syntactic structures as represented by word-word relations (instead of embedded phrases as in treebank) (Chap. 7).

Dialog A conversational exchange between two or more entities (Chap. 9).

Dialog system A computational artifact designed to engage humans in dialog (Chap. 9).

Digital taxonomy A structure guiding how to organize and classify digital content. For example, Bloom's Digital Taxonomy helps one to navigate through the myriad digital tools and to make choices based on the kinds of learning experiences in which students should be engaged (Chap. 16).

Digital transformation An approach that draws upon user-centered design to reimagine how essential functions can be improved by exploiting AI methods and other digital technologies (Chap. 13).

Digital twin The virtual model of a physical system for learning from real world processes and experiences to achieve a closed-loop of the virtual and real world physical systems in order to improve efficiency and to increase innovation (Chap. 19).

Discourse segments The components of a discourse, comprising one or more utterances (Chap. 9).

Discriminative model in machine learning, a model that learns the conditional distribution of the labels given observed features. That is, it assigns the probability of a label based on the observed features of a data point, without considering how these features are distributed in other examples (Chap. 6).

Distributed cognition A process in which cognitive resources are shared socially to extend individual cognitive resources or to accomplish something that an individual agent could not achieve alone. It emphasizes the ways that cognition is off-loaded into the environment through social and technological means. It is a framework for studying cognition rather than a type of cognition (Chap. 5).

Division of Research Resources (DRR) Prior to 1990, DRR was a Division at the National Institutes of Health (NIH) that funded shared research facilities, among other activities. In 1990 it merged with the Division of Research Services to form that National Center for Research Resources (NCRR). It was abolished in 2011 when NIH was reorganized to create a new National Center for Advancing Translational Sciences (NCATS) (Chap. 2).

Domain model A domain model is a conceptual model of a domain that incorporates both behavior and data. It may partially represent the entire domain. It may also contain a formal, computable representation of the knowledge, skills and strategies (Chap. 16).

Doublet In single-cell RNA sequencing, a 'drop', which is identified by a barcode, or a datapoint, is expected to contain the genome from only one cell. A doublet occurs when a drop contains the genome of more than one cell (Chap. 14).

Dynamic Bayesian network A probabilistic causal network that relates variables to each other over adjacent time steps. At any point in time T, the value of a variable can be calculated from the internal regressors and the immediate prior value at T-1 (Chap. 11).

Dynamic differential equations A class of algorithms that represent a continuous process by one or more differential equations. There must be at least one equation that has a derivative with respect to time (Chap. 14).

Dynamic imaging An amalgam of digital imaging, image editing, and workflow automation. It is used to automate the creation of images by zooming, panning, colorizing, and performing other image processing and color management operations on a copy of a digital master. In radiology, dynamic imaging refers to sequential imaging of a volume of tissue over time, which often implies scanning the same area during the passage of intravenous contrast over time (Chap. 12).

ECG (electrocardiogram) A standardized tracing of the heart's electrical activity—voltage/time as recorded from multiple electrodes (Chap. 8).

Electronic health record (EHR) A repository of electronically maintained information about an individual's lifetime health status and health care, stored so that it can serve the multiple legitimate users of the record. Also sometimes termed electronic medical record (EMR) and, historically, computer-based patient record (CPR) (Chaps. 1, 7, and 10).

Embedding A low-dimensional representation of a word, object, or concept that encodes certain topological, algebraic, or other properties (Chap. 4).

Embodied conversational agent An agent that includes the ability to use human-like conversational nonverbal behavior in a dialog, such as a hand gesture and facial display (Chap. 9).

Emergent properties Properties a collection of systems or members have which individuals within those systems or members lack (Chap. 8).

Entropy A measure of the uncertainty represented by a probability distribution. A distribution in which one possibility is certain and others have been ruled out has zero entropy, and one in which all possibilities are equally likely has maximal entropy (Chap. 4).

Epistemology The investigation of what distinguishes justified belief or knowledge from opinion (Chap. 3).

Equity Allocation of resources and opportunities in a manner that achieves an equal outcome (Chap. 13).

Error boundary In work on human/AI teams, used to describe human team members' knowledge of the limits of AI performance, manifesting as their ability to identify examples in which the model is likely to be wrong (Chap. 20).

Ethnographical Relating to ethnography, which is a qualitative research approach used to study people and cultures. It can provide an in-depth understanding of the socio-technological realities surrounding everyday software development practice (Chap. 17).

Expectation-maximization (EM) In statistics and machine learning, a class of algorithms to infer the data distribution and to train the machine learning model. The algorithm is initiated by a set of default distributions (or models) parameters and the default likelihood assignment of datapoints to each distribution (or model). Then, the expectation step (E step) recomputes the distributions (or models) parameters such that they are the best fit according to the datapoints likelihood assignment. After the E step, the maximization step (M step) recomputes the datapoints' likelihood assignment according to the new distributions (or models) parameters. These two steps are repeatedly executed, one after the other, until all distributions (or models) and datapoints' likelihood assignment do not significantly change (Chap. 14).

Expert system Software that uses AI methods to combine data and expert knowledge to offer advice or make decisions in an area of human activity (Chaps. 1 and 2).

Explainability The ability to provide a reasonably complete explanation of an AI system's output (e.g., a diagnosis or prediction) based on an understanding of its decision process (Chaps. 8 and 18).

Explanation A summary of the basis for a statement that is purported to be true. In decision support, the line of reasoning, pertinent data, and accepted factual knowledge that together justify an interpretation or recommendation (Chap. 3).

Extended reality (XR) The immersive technologies of virtual reality (simulated digital environment with full immersion), augmented reality (virtual objects and/or information are overlaid on real world objects and places), and mixed (or hybrid) reality (real-world and digital objects co-exist) (Chap. 19).

External representation A representation of information on a physical medium (Chap. 5).

Fairness Related to "justice," the concept that goods/benefits and burdens/harms are or should be equitably shared by individuals and groups (Chap. 18).

FDA See Food and Drug Administration.

Feature engineering A process of using domain knowledge to extract features from raw data that can better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data (Chap. 12).

Federated learning A decentralized form of machine learning in which the model is centralized while the data are not, thereby assuring that the privacy and security concerns of the data are no longer an issue since the data are kept and secured locally (Chaps. 7 and 19).

Federated query model A way to send a query statement to an external database and get the result back as a temporary table (Chap. 3).

Few shot learning An innovative machine learning method that can make predictions based on a small number of labeled samples for training so that the machines can learn rare cases. Variations of few shot learning include one-shot and zero-shot learning (Chap. 19).

Findability The ease with which information contained in a location or set of locations can be found, both from outside those sites (using search engines and the like) and by users already on the site (Chap. 3).

Focus group A form of group interview that capitalizes on communication among research participants in order to generate data and insights. The idea behind the focus group method is that group processes can help people to explore and clarify their views in ways that would be less easily accessible in a one-on-one interview (Chap. 17).

Food and Drug Administration (FDA) A US federal agency tasked with monitoring various foods, biopharmaceuticals, medical devices, cosmetics and veterinary products (Chap. 8).

Forward chaining The computational process of inferring new facts from what is previously known, using inference rules that state an implication from known facts to new facts (Chap. 4).

Forward reasoning The process of reasoning with initial data towards the goal, which is usually uncertain. It is often referred to as top down reasoning (Chap. 5).

Fourier transform A mathematical transform that decomposes functions depending on space or time into functions depending on spatial or temporal frequency, such as the expression of a musical chord in terms of the volumes and frequencies of its constituent notes (Chap. 12).

Frame An abstract representation of a concept or entity that consists of a set of attributes, called slots, each of which can have one or more values to represent knowledge about the entity or concept (Chap. 4).

Fuzzy logic A logic in which a proposition is not necessarily true or false but can be (believed to be) true to some degree. Its semantics differs from probabilistic methods because fuzzy values are subjective and independence is typically not assumed (Chap. 4).

Gabor filter A linear filter used for image texture analysis, which essentially means that it analyzes whether there is any specific image frequency content in specific directions in a localized region around the point or region of analysis (Chap. 12).

Gaussian discriminant analysis A supervised machine learning algorithm that attempts to fit a Gaussian distribution to each data category (Chap. 6).

GDPR (general data protection regulation) A European Union (EU) law designed to set guidelines for information privacy, applicable to EU member countries and the European Economic Area (Chap. 8).

Generative adversarial network (GAN) A class of machine learning frameworks designed by Ian Goodfellow and colleagues. It involves having two neural networks (called the generator and the discriminator) compete with one another in a game (in the form of a zero-sum game, where one agent's gain is the other agent's loss) in order recursively to create and improve new content (Chaps. 12 and 19).

Generative model In machine learning, a model that learns the joint distribution over observed features and labels for the training set. That is, it considers the observed features for a particular data point in relation to their underlying distribution across the entire training set (Chap. 6).

Generative pre-trained transformer (GPT-3) An unsupervised learner and language prediction model created by OpenAI with 175 billion parameters. It deploys deep learning to perform natural language processing tasks (Chap. 19).

Genome All genetic information of an organism (Chap. 14).

Genomics A branch of molecular biology concerned with the structure, function, evolution, and mapping of an organism's genetic material (Chap. 2).

Governance The processes to ensure the appropriate use of resources (e.g., information technology) in order to ensure standards are met, values adhered to, and accountability ensured. Governance might or might not entail regulation by governments (Chaps. 15 and 18).

Gradient boosting A supervised method for classification and regression built from an ensemble of weak classifiers (typically, decision trees), which are combined into a single strong learner in an iterative fashion (Chap. 11).

Gradient boosted decision tree A gradient boosting algorithm which produces a sequential ensemble (or committee) of decision trees. Each tree is associated with a voting weight, and the final decision of the ensemble is a weighted majority vote of its members. Each decision tree in the sequence is optimized to correct errors made by the previously added members of the ensemble (Chap. 6).

Gradient descent A widely-used optimization algorithm that makes multiple small corrections to model parameters in accordance with their influence on the error of a model, bringing it closer to an optimization objective (Chap. 6).

Graph database A (NoSQL) type of database that uses graph structures for semantic queries with nodes and edges to represent and store data. The nodes are the entities in the graph and the relationships provide connections between two node entities (Chap. 19).

Graph neural networks A class of neural network models that learn representations of graph structure and work on optimizing transformation on all the attributes of the graph including nodes and edges. These networks accept a graph as input and transform these embeddings while maintaining the input graph connectivity (Chaps. 4 and 19).

Graphical processing unit (GPU) A specialized electronic circuit designed to manipulate and alter digital memory rapidly to accelerate the creation of images for internal storage and delivery on a display device (Chap. 2).

Gray box Machine learning model with some hand-crafted features or, generally, some (but incomplete) knowledge of system processes (Chap. 18).

Gray-level co-occurrence matrices (GLCM) A matrix that is defined over an image to be the distribution of co-occurring pixel values (grayscale values, or colors) at a given offset. It is used as an approach to texture analysis with various applications, especially in medical image analysis (Chap. 12).

HbA1c Hemoglobin A1c, a measure of average blood sugar over the past 3 months (Chap. 3).

Health Evaluation and Logical Processing [HELP] One of the first electronic health record systems, developed at LDS Hospital in Salt Lake City, Utah. Still in use today, it was innovative for its introduction of automated alerts (Chap. 10).

Health Insurance Portability and Accountability Act (HIPAA) A privacy rule established by the US Congress in 1996, that established national standards to protect individuals' medical records and other personal health information. It applies to health plans, health care clearinghouses, and those health care providers that conduct certain health care transactions electronically (Chaps. 2 and 7).

Heuristic A strategy derived from previous experiences with a similar problem, often described as a "rule of thumb". In computer science, a technique for solving a problem more quickly when classic methods are too slow or cumbersome (Chap. 2).

Heuristic classification Any technique or approach to classification problem solving that employs a practical method that is not guaranteed to be optimal, perfect, or rational, but is nevertheless sufficient for reaching an immediate, short-term goal or approximation (Chap. 5).

Hidden Markov Model (HMM) A statistical Markov model in which the system being modeled is assumed to be a Markov process with unobservable ("*hidden*") states X. An HMM assumes that there is another process Y whose behavior "depends" on X. By observing Y, the goal of an HMM is to learn about X (Chap. 11).

Hierarchical clustering Hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. For example, a dataset could be first clustered into two clusters, C1 and C2. Then, we can further cluster the data into C11, C12, C13, C21, and C22. Here, C11, C12, and C13 are children clusters from C1; meanwhile, C21 and C22 are children clusters from C2 (Chap. 14).

HIPAA See Health Insurance Portability and Accountability Act.

HITECH Health Information Technology for Economics and Clinical Health Act, enacted as part of the American Recovery and Reinvestment Act of 2009. It was designed to promote the adoption and meaningful use of health information technology, with an emphasis on electronic health records (Chap. 2).

Hold-out A validation strategy for machine learning models that consists of randomly splitting a dataset into training and test set. Usually, 2/3 of the data are used for training and 1/3 for test, but other proportions can be used depending on the size of the available dataset. The model is then trained on the training set and its performance is measured on the test set. The procedure can be repeated several times (repeated hold-out) (Chap. 11).

Human computer interaction (HCI) A multidisciplinary field of study focusing on the design of computer technology and, in particular, the interaction between human beings (the users) and computers (Chaps. 10 and 17).

Human factors engineering (HFE) The field dealing with the integration of human factors requirements into design. The objective is to provide systems that reduce the potential for human error, increase system availability, lower lifecycle costs, improve safety, and enhance overall system performance (Chap. 17).

Human-in-the loop The idea or requirement that human experts should (be available to) review and/or assess system performance and output. Human beings may be involved in a live and virtuous cycle where they train, tune, and test a particular algorithm (Chaps. 12 and 18).

Hyperparameters In machine learning and AI, a hyperparameter is a parameter whose value must be preset before executing the algorithm to train the learning model. In statistics, a hyperparameter is a parameter of a prior distribution (Chaps. 6 and 14).

Hyperplane A subspace with dimensionality one less than the space that encloses it. For example, in a three-dimensional space, a hyperplane would be a two-dimensional plane (Chap. 6).

Hypothesis-driven reasoning In medical reasoning, a pattern of reasoning in which information is reviewed in an attempt to reconcile it with a pre-existing hypothesis that may be generated under uncertainty without being fully grounded in observed signs or symptoms (Chap. 5).

Hypothetico-deductive approach In medical reasoning, a method of inquiry that proceeds by formulating a hypothesis (such as a diagnostic explanation for a patient's symptoms) in a form that can be falsifiable (e.g., through a test outcome that runs contrary to predictions expected if the hypothesis were true). Results that are consistent with the hypothesis can corroborate the theory. Thus repeated questions or tests can guide the reasoner to accept the hypothesis or to pursue competing explanations for the results (Chaps. 2 and 5).

ICell8 A well-known technique to perform single-cell RNA sequencing. The technique was developed and owned by Takara Bio Inc (https://www.takarabio.com/products/automation-systems/icell8-system-and-software) (Chap. 14).

ImageNet A large visual database designed for use in visual object recognition software research. More than 14 million images have been hand-annotated by the project to indicate what objects are pictured, and in at least one million of the

images, bounding boxes are also provided. ImageNet contains more than 20,000 categories with a typical category, such as "balloon" or "strawberry," consisting of several hundred images (Chap. 12).

IMDRF (International Medical Device Regulators Forum) A group of worldwide regulators tasked with reconciling regulatory requirements for medical devices (Chap. 8).

Implementation science The scientific study of methods to promote the systematic uptake of research findings and other evidence-based practices into routine practice, and, hence, to improve the quality and effectiveness of health services (Chap. 17).

Inductive bias Any design decision that provides a basis for choosing one generalization over another (Chap. 4).

Inductive reasoning A method for drawing conclusions by making inferences from the specific to the general (in contrast with deductive reasoning) (Chaps. 3 and 5).

Inference A conclusion reached on the basis of evidence and reasoning (Chap. 3).

Influence diagram A graphical representation that combines chance, choice, and outcome nodes in a probabilistic framework (Chap. 4).

Infobutton A context-specific link from healthcare application to some information resource that anticipates users' needs and provides targeted information (Chap. 10).

Information extraction (IE) An application of NLP focusing on the extraction of specific structured information from free text (Chap. 7).

Information retrieval (IR) An application of NLP focusing on the identification of information (usually in the form of a natural language document) relevant to a particular query (e.g., set keywords or a natural language question), generally in the context of a significant amount of potential information (the "corpus"). The preeminent IR use case is a search engine such as Google or PubMed (Chap. 7).

Information theory The scientific study of the quantification, storage, and communication of digital information (Chap. 2).

Institutional Review Board (IRB) A committee within a university or other organization receiving federal funds to conduct research that reviews research proposals involving human subjects. The IRB reviews the proposals before a project is submitted to a funding agency to determine if the research project follows the ethical principles and federal regulations for the protection of human subjects. The committee has the authority to approve, disapprove or require modifications of these projects or proposals (Chap. 17).

Intelligence-based medicine Another term for artificial intelligence in medicine, it tends to be used to focus on clinical medicine practiced using data with machine learning, and especially deep learning, for improving diagnosis and therapy rather than solely relying on the conventional evidence-based medicine (Chap. 19).

Intelligent agent An autonomous, goal-directed computational artifact (Chap. 9).

Intelligent reality The integration of machine learning algorithms with smart interactive devices (including sensors and wearable devices) and immersive technologies to enable the user to have real-time decision support (Chap. 19).

Intelligent tutor A computational system that incrementally presents content relevant to a learning goal using different teaching strategies for different types of content and student ability. It also intervenes when the student requests help or makes serious mistakes.

Interface model A combination of a system's multimedia presentation and a user's input interface. In the case of teaching and learning content, systems, such models may range from simple screen display and mouse input to virtual reality (VR), haptics, and affect detection.

Internal representation A representation of information in the human mind (Chap. 5).

Internet of everything (IoE) The intelligent connection (embedded AI) of people, process, data, and things to make networked connections from the internet of things (IoT) to render the devices more relevant and valuable (Chap. 19).

INTERNIST-1 An early computer-based diagnostic program trained on clinical pathological conference cases, able to diagnose cases with multiple disorders (Chaps. 2 and 4).

Interoperability The ability of computer systems or software to exchange and make use of information (Chap. 3 and 13).

Interpretability A measure of how intuitive an AI system's outcomes are to its users. One can think of it as simplifying explainability to the point that those with non-technical expertise can follow, or infer, the "why" of an AI system's output (Chap. 18).

Intravenous unfractionated heparin (IV-UFH) A medication administered through a vein that affects blot clot formation (Chap. 15).

Intron Any nucleotide sequence within a gene (also called an intragenic region) that is removed by RNA splicing. An intron does not encode to the final mature RNA; therefore, an intron does not encode protein (Chap. 14).

Iterative design An approach that designers, developers, educators, and others use to continually improve a design or product. People create a prototype and test it, then tweak and test the revised prototype, and repeat this cycle until they reach a solution that appears to be optimal (Chap. 17).

KDD See knowledge discovery in databases.

Kernel methods Methods through which data points are represented in terms of their similarity to other data points in a set (Chap. 6).

Kernel regression/kernelized logistic regression The application of a linear or logistic regression model to features that have been transformed using kernel methods (Chap. 6).

k-fold cross validation A strategy for machine learning models. To perform cross-validation, the dataset is randomly split into k subsets (folds). Iteratively, the model is built on k-1 folds and performance are tested on the remaining fold. The procedure is repeated k times and the average performance is computed. The average error on the k test folds approximates the true error on independent data of the model that is built on all the k folds (Chap. 11).

k-nearest neighbor algorithm A machine learning algorithm that assigns labels to unseen examples by generalizing from the labels of the most similar examples in a training set (Chap. 6).

Klein's data-frame theory A theory on how people start with an initial idea that informs what kinds of data they seek through an iterative process to modify their initial 'frame' (Chap. 8).

Knowledge base A collection of stored facts, heuristics, and models that can be used for problem solving (Chaps. 4 and 10).

Knowledge discovery in databases (KDD) The process of finding knowledge in data, emphasizing the application of particular data mining methods from the general field of machine learning (Chap. 2).

Knowledge distillation The process of transferring knowledge from a larger model to a smaller model (Chap. 4).

Knowledge engineering The term for all the technical, scientific, and social aspects involved in building, maintaining, and using knowledge-based systems (Chap. 2).

Knowledge graph Also known as a semantic network, represents a network of real-world entities—i.e. objects, events, situations, or concepts—and illustrates the relationship between them. This information is usually stored in a graph database and visualized as a graph structure, prompting the term knowledge "graph" (Chaps. 3 and 19).

Knowledge representation The field of AI that is dedicated to representing information about the world in a form that a computer system can use to solve complex tasks such as diagnosing a medical condition or having a dialog in natural language (Chap. 3).

Knowledge-based artificial intelligence-based clinical decision support An artificial intelligence-based clinical decision support system in which a key component is a knowledge base that is typically manually constructed (Chap. 10).

Knowledge-based system A program that symbolically encodes, in a knowledge base, facts, heuristics, and models derived from experts in a field and uses that knowledge to provide problem analysis or advice that the expert might have provided if asked the same question (Chap. 4).

Laplace smoothing A regularization process performed on probability estimates in which counts start at one (or some other small constant), so that no conditional probability is ever estimated at zero (Chap. 6).

Lasso regression A form of regularized linear regression that tends to drive the coefficients of redundant features toward zero (Chap. 6).

Latent semantic analysis A mathematical method for computer modeling and simulation of the meaning of words and passages by analysis of representative corpora of natural text. It closely approximates many aspects of human language, learning and understanding (Chap. 16).

Learner model A model that represents the learner's current state in the domain space, and is updated in real-time as the learner progresses through the learning exercises. It is a structured representation of a learner's knowledge, misconceptions, and difficulties (Chap. 16).

Left-ventricular assist device (LVAD) A battery-operated mechanical pump that is surgically implanted to help the heart's left ventricle pump blood to the rest of the body (Chap. 15).

LISP A family of programming languages, originally specified in 1958 (and thus second only to Fortran among old computer languages still in use), that uses a distinctive, fully parenthesized prefix notation. The name LISP derives from "LISt Processor", reflecting LISP's use of linked lists as a major data structure. It has been used extensively in the AI community (Chap. 2).

Literature-based discovery (LBD) An application of NLP focusing on hypothesis generation by discovering links between previously un-connected biomedical concepts, the canonical example being Swanson's discovery of the link between Raynaud's disease and fish oil via intermediate terms (Chap. 7).

Local area networking (LAN) A computer network that comprises a collection of devices connected together in one physical location, such as a building, office, or home (Chap. 2).

Local data store Keeping information on a disk, tape drive, or similar technology that is directly attached to the computer or device (as opposed to being elsewhere on a network or in the cloud) (Chap. 3).

Logic Programming A programming paradigm, also called rule-based programming, which is largely based on formal logic. Any program written in a logic programming language is a set of sentences in logical form: IF (AND, OR), THEN (Chap. 14).

Logistic regression A supervised machine learning algorithm that transforms the output of a linear regression model into a value between 0 and 1 that can be interpreted as a class probability (Chap. 6).

Loss function A function that measures how far a machine learning model is from achieving a desired optimization objective. Typically, the loss function estimates how far a prediction is from the true label of a data point, or a set of data points (Chap. 6).

Machine intelligence An early synonym for artificial intelligence, favored especially in the United Kingdom (Chap. 2).

Machine learning (ML) A branch of AI and computer science that focuses on the use of data and algorithms to imitate the way the human beings learn, gradually improving its accuracy at classification or prediction (Chaps. 2, 3, 6, and 10).

Machine translation NLP task involving the automatic translation from one natural language to another (e.g., from French to English) (Chap. 7).

Markov model A mathematical model of a set of strings in which the probability of a given symbol occurring depends on the identity of the immediately preceding symbol or the two immediately preceding symbols. Processes modeled in this way are often called Markov processes (Chap. 11).

Mass spectroscopy An analytical technique that is used to measure the mass-to-charge ratio of ions. The results are typically presented as a mass spectrum, a plot of intensity as a function of the mass-to-charge ratio. The technique is used in a variety of fields and may be applied to pure samples of complex mixtures (Chap. 2).

Matrix factorization An approach for generating embeddings by decomposing a matrix into the product of two lower dimensionality rectangular matrices (Chap. 4).

MEDLINE Online implementation of MEDLARS, accessible by public via the Internet. The currently dominant version of MEDLINE is known as PubMed.

MEDLARS (Medical Literature Analysis and Retrieval System) The National Library of Medicine's electronic catalog of the biomedical literature, which includes information abstracted from journal articles, including author names, article title, journal source, publication date, abstract, and medical subject headings (Chap. 7).

Mental representations Also known as *cognitive representations*, the mental imagery of things that are not currently seen or sensed by the sense organs. It is a hypothetical internal cognitive symbol that represents external reality (Chap. 5).

Metaverse A persistent and immersive simulated world in which users can interact with a computer-generated environment and with other users (Chap. 16).

Microarray See gene expression microarray.

MINIMAR standards A set of standards put forth to ensure that the interpretability of a model is helpful for end users, in an attempt to mitigate inherent biases (Chap. 8).

miRNA Abbreviation for microRNA, refers to a small single-stranded non-coding RNA molecule (containing about 22 nucleotides) found in plants, animals and some viruses, that functions in RNA silencing and post-transcriptional regulation of gene expression (Chap. 14).

Model-based integration A process for performing analysis on each data type independently, followed by integration of the resultant models to generate knowledge about the trait of interest (Chap. 3).

Modeling The act or realization of making a model (Chap. 3).

Modus ponens The logic rule that allows deriving the conclusion of an implication when seeing its premises (Chap. 4).

Modus tollens The logic rule that allows deriving the negation of the premise of an implication if its conclusion is known to be false (Chap. 4).

Monte Carlo simulations Computational algorithms that rely on repeated random sampling to obtain numerical results. Monte Carlo simulation uses randomness to solve problems that might be deterministic in principle. Monte Carlo methods are mainly used for optimization, numerical integration and to generate draws from a probability distribution (Chap. 11).

Multimodal artificial intelligence An AI paradigm in which a myriad of data types (such as EHR data, image data, and wearable device data) are gathered and analyzed via algorithms for higher performance in predictions (Chap. 19).

MYCIN An early expert system that used rules to encode knowledge of infectious disease diagnosis and therapy selection. The program served as a consultant to offer advice to clinicians caring for patients with serious bacterial or fungal infections (Chaps. 2–4).

Naive Bayes model/classifier A machine learning classifier using Bayes Theorem in a way that assumes conditional independence of variables that may in fact be linked statistically (Chaps. 4 and 15).

Naive Bayes A method that uses Bayes Theorem while making the simplifying assumption that variables are conditionally independent of one another. Also used in supervised machine learning making the same independence assumption regarding features in the input vector, given the class (Chap. 6).

Named entity recognition (NER) NLP task involving the identification of phrases in text that describe a specific object from a particular semantic category (e.g., name of a specific person or location) (Chap. 7).

National Institutes of Health (NIH) The national medical research agency for the USA, located in Bethesda, MD, supporting scientific studies that seek to turn discovery into improved health (Chap. 2).

Natural history The typical progression of a disease process in an individual over time, in the absence of treatment (Chap. 3).

Natural language generation (NLG) The subfield of NLP that deals with generating coherent language as an output by translating a semantic representation (e.g., a data structure) into natural language text (Chaps. 7 and 9).

Natural language processing (NLP) The branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can (Chap. 1).

Natural language understanding (NLU) The subfield of NLP that deals with understanding the meaning of language as it is input to a system (Chaps. 7 and 9).

Nearest neighbor search A form of proximity search to find a number of data points in a given set that are closest to a new point of interest (Chap. 12).

Nested k-fold cross validation Strategy for hyperparameter selection and model evaluation, in which two loops of cross validation are performed. The outer loop is used to select the best performing model and it consists of a k-fold cross validation procedure. However, for each iteration of the outer loop, an inner loop of cross validation is performed by randomly splitting the training folds into L folds and by performing an L-fold cross validation. The inner loop is usually exploited to evaluate different combinations of hyperparameters (Chaps. 6 and 11).

Neural architecture search (NAS) A technique for automating the design of artificial neural networks (Chap. 12).

NLP See natural language processing.

Noisy-or assumption In Bayes networks, where a node may have multiple parents, a simplifying assumption that the presence of each parent contributes independently to the likelihood of the node (Chap. 4).

Non-coding RNA (ncRNA) An RNA molecule that is not translated into a protein (Chap. 14).

Non-negative matrix factorization A group of algorithms in multivariate analysis and linear algebra where a matrix V is factorized into two (and more) matrices W and H ($V >> W \times H$), with the property that all three matrices have no negative elements (Chap. 14).

Non-parametric model In machine learning, a model making minimal assumptions about the structure of the underlying data, rather than defining a function a priori (such as a linear function). A classic example is the k-nearest neighbor algorithm, where new data are compared to examples in the training set directly at the point of prediction (Chap. 6).

Norman's Theory of Action A seven stage model of human action designed to explain the thought process of a person performing a task, as put forward by Norman (Chap. 8).

Observational Medical Outcomes Partnership (OMOP) A large collaborative consortium formed to facilitate and inform studies using large, often multi-institutional, observational health care data sets. OMOP efforts continue under the auspices of the Observational Health Data Sciences and Informatics (OHDSI) consortium (Chap. 6).

One single centralized datastore A database that is located, stored, and maintained in a single location (Chap. 3).

One-shot learning Instead of hundreds or thousands of annotated data elements, one-shot learning acquires information about object categories from one, or only a few, training examples (Chap. 12).

Ontology A description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents. In biomedicine, such ontologies typically specify the meanings and hierarchical relationships among terms and concepts in a domain. Note that philosophers use "ontology" as the study of what exists, which is a broader concept than what computer science has adopted (Chaps. 2–4).

Open domain question answering A natural language processing task in which a model is asked to produce an answer to a question without being provided the document containing the answer (Chap. 4).

Optical character recognition (OCR) A computer vision/NLP task of converting an image containing text to its string representation (e.g., generating a Word file from a photograph of a text document) (Chap. 7).

Optimization problem Involves finding the values of a parameter for a given function (called the optimization objective) at which the function achieves a maximum or a minimum value (Chap. 6).

Overfitting In machine learning, occurs when a model conforms too closely to its training data, reducing its ability to generalize to unseen examples (Chap. 6).

Packet switching A method for transmitting a data stream across digital networks by breaking it down into small segments or packets for more efficient transfer. Adjacent packets may be sent over different routes to assure efficiency and then reassembled when they reach their destination (Chap. 2).

Parametric model In machine learning, a model with a fixed number of parameters that are updated during training to improve the accuracy of model predictions. The choice of model (e.g. linear vs. non-linear) is based on assumptions about the structure of the function to be learned (Chap. 6).

Parenchyma In reference to normal lung, parenchyma denotes a gas-exchanging part of the lung, consisting of the alveoli and their capillaries. Some

definitions may also include the connective tissue framework supporting gas exchanging tissue. The term may also be applied to other organs beside the lung (Chap. 12).

Partial thromboplastin time (PTT) A laboratory test of blood that measures the time it takes for a blood clot to form. Common therapeutic range is 60–100 s for clotting (Chap. 15).

Part-of-speech The syntactic (or grammatical) category of a word (e.g., noun, verb, adjective), often one of the first tasks performed in a pipeline of NLP models (Chap. 7).

Pathophysiological explanations Those that draw on knowledge of disease processes and their impact on human physiology (Chap. 5).

Patient-derived xenograft (PDX) Patient derived xenografts are models of cancer where the tissue or cells from a patient's tumor are implanted into an immunodeficient or humanized mouse (Chap. 14).

Pedagogical Model Represents what effective teachers do to engage students at each step of a teaching and learning session. It includes changes in teaching strategy if the student makes errors or is unable to progress adequately.

Personalized machine learning A machine learning approach where the model is tailored to the characteristics of the current person and is optimized to perform especially well for that person. See also population machine learning (Chap. 10).

Polygenic risk scoring A polygenic score (PGS), also called a polygenic risk score (PRS), genetic risk score, or genome-wide score, is a number that summarizes the estimated effect of many genetic variants on an individual's phenotype (Chap. 14).

Pooled cohort equations Clinical decision support tool requiring multiple inputs including age, gender, race, patient history elements, certain lab and vital sign values designed to estimate 10 year absolute rate of ASCVD events (Chap. 8).

Pooling An operation that groups feature map activations into a lower resolution feature map to enlarge the receptive field of deep neural networks and to reduce the model's sensitivity to small shifts of the objects (Chap. 12).

Population health monitoring An approach that systematically collects data on health status, usually to inform longer-term planning and evaluation of programs (Chap. 13).

Population machine learning A machine learning approach where the model is optimized to perform well on the average of all future individuals. See also personalized machine learning (Chap. 10).

Prediction A statement about what will happen or might happen in the future (Chap. 3).

Predictive learning The ability of a machine to model the environment, to predict the possible futures, and to understand how the world works by observing it and acting in it—in short, a predictive model of the world (Chap. 19).

Premature closure A type of cognitive error in the reasoning process in which a physician prematurely stops considering other alternative diagnostic possibilities once an initial tentative diagnosis is made (Chap. 3).

Present Illness Program (PIP) An early computer-based diagnostic program that incorporated insights from the study of human clinical reasoning (Chap. 4).

Principal component analysis (PCA) A technique for reducing the dimensionality of large datasets by transforming a large set of variables into a smaller one while still containing most of the information in the larger set (Chap. 12).

Principle of rationality A principle typically adopted in economics that decision makers choose actions that maximize their expected utility (Chap. 4).

Privacy A concept that applies to people, rather than documents, in which there is a presumed right to protect that individual from unauthorized divulging of personal data of any kind (Chap. 3).

Professional ethics Standards or rules for correct behavior by members of a (generally) learned profession, and often tied to a professional association or society through codes, oaths, guidelines, etc. (Chap. 18).

Propositions An expression, generally in language or other symbolic form, that can be believed, doubted, or denied or is either true or false (Chap. 5).

Protein physical structure (protein structure) The three-dimensional arrangement of atoms and amino-acids in a protein (Chap. 14).

Protocol analysis Method in which transcripts of think-aloud sessions, which generate think-aloud protocols are analyzed to investigate the cognitive processes underlying performance of the task. Also called verbal protocol analysis (Chap. 5).

Provenance The place where something originally came from or began, or a record tracing the history of certain elements or items that helps to confirm their authenticity and validity (Chap. 3).

Pseudogenes Nonfunctional segments of DNA that resemble functional genes. They cannot encode functional protein (Chap. 14).

Public health informatics A sub-discipline of biomedical informatics, this is the systematic application of information and computer science and technology to public health practice, research, and learning (Chap. 13).

Public health surveillance The systematic, ongoing collection and analysis of data to detect and guide actions to control hazards such as infectious disease outbreaks. It includes indicator-based surveillance (IBS) and event-based surveillance (EBS) (Chap. 13).

Public health The science and the art of preventing disease, prolonging life, and promoting health through organized community efforts (Chap. 13).

Quantum computing Use of quantum mechanics to calculate outputs by harnessing the power of atoms and molecules to perform memory and processing tasks with qubits (basic unit of information in quantum computing that can hold a superposition of possible states) (Chap. 19).

Question answering (QA) An application of NLP focusing on automatically providing the user with an answer to a natural language question, with answers typically coming from either a large corpus (possibly also involving an information retrieval step) or a large ontology/database (requiring the question be converted to a structured query via an NLP task known as semantic parsing) (Chap. 7).

Radial basis function (RBF) A function widely-used in kernel methods to measure the similarity between data points (Chap. 6).

Radio frequency identification (RFID) A technology that uses radio waves passively to identify a tagged object. An RFID tag stores information that can be read wirelessly by an RFID reader (Chap. 17).

Radiology report A document that describes radiologist's highest level of synthesis and insight into a patient's condition. It is the most important product that radiologists generate to help direct patient care (Chap. 12).

Radiomics features A method that extracts a large number of features from radiographic images. These features have the potential to uncover disease characteristics that fail to be appreciated by the naked eye (Chap. 12).

Random forest An ensemble of decision trees. Bagging is applied to train each tree learner, by randomly selecting with replacement a subset of the training set, and by training a single tree on this selected subset. The final prediction for a classification problem is the class predicted by the majority of trees. For a regression problem the final prediction is the average prediction (Chap. 11).

Random walk A process consisting of a sequence of steps taken in a randomized fashion (Chap. 4).

Rasmussen's decision ladder A tool put forth by Rasmussen that aids in visualizing the multiple steps required for decision making, often involving multiple chains and pathways (Chap. 8).

Recurrent neural network An ANN architecture often used to summarize sequential data such as time series or natural language text (Chap. 4).

Reference frame Grids of a number of dimensions that keep the brain organized in terms of knowledge in the context of a learning model (Chap. 19).

Reflective thinking Thinking that encompasses a set of abilities that people use to examine their own thoughts processes and those of others analytically, thereby allowing themselves to question and challenge their own thoughts and those of others (Chap. 16).

Regularization In machine learning, imposing a penalty to prevent a model from fitting too tightly to its training data (see overfitting), in order to improve its ability to generalize to unseen examples (Chap. 6).

Reinforcement learning A method of determining the optimal policy for what actions to take under different circumstances, learnable from past experience (Chap. 4).

Relation extraction (RE) NLP task involving the recognition of a semantic relationship between two or more phrases in text (e.g., in a *part-whole* relation, one phrase is identified as the *part* while the other phrase is the *whole*) (Chap. 7).

Repeated hold-out Validation strategy for machine learning techniques that consists of repeating hold-out validation N times. Therefore, the prediction error will be estimated as the average error on the N test sets (Chap. 11).

Representation learning The process of learning an informative representation of data to make it easier to extract useful information for a downstream task, often applied to machine learning models (Chaps. 1 and 4).

Retrospective think-aloud A think-aloud session about an event after the event has occurred to discuss and reflect what happened during the event (Chap. 5).

Reusability In computer science and software engineering, the use of existing assets in some form within the software product development process (Chap. 3).

RFID See radio frequency identification (Chap. 17).

Ridge regression A form of regularized linear regression that uses the L2 norm of the parameter vector as a penalty term, to constrain the growth of coefficients for each feature (Chap. 6).

RNA sequencing (RNA-seq) See next generation sequencing methods.

Robotic process automation (RPA) A portfolio of software tools that builds and manages software robots that will mimic and automate what humans do with digital systems so that workflows can be streamlined as part of a digital transformation (Chap. 19).

Rule-based system A system that applied human-made rules to store, sort, and manipulate knowledge. A classic example is a domain-specific expert system such as MYCIN that uses rules to simulate expert decision making (Chap. 3).

SaMD (software as a medical device) Software intended to be used for one or more medical purposes that performs these purposes without being implemented as part of a hardware medical device" (Chap. 8).

Scale-invariant feature transform (SIFT) A feature detection algorithm in computer vision to detect and describe local features in images (Chap. 12).

Schema In a database-management system, a machine-readable definition of the contents and organization of a database (Chap. 5).

Security The process of protecting information from destruction, unauthorized access, or misuse, including both physical and computer-based mechanisms (Chap. 3).

Cover and differentiate A set of candidate explanations for the events or states that need to be explained (or covered), combined with differentiating among the candidates to pick those that best explain the specified events or states (Chap. 5).

SEIPS model A model of work system and patient safety that provides a framework for understanding the structures, processes, and outcomes in health care and their interrelationships (Chap. 8).

Self supervised learning A method of machine learning that can be regarded as an intermediate form of supervised and unsupervised learning. It is a type of autonomous learning that does not necessarily require annotated data. The approach predicts an unobserved part of the input from observed parts of the input and has had a profound impact on natural language processing and computer vision by obviating the need for data labeling (Chaps. 12 and 19).

Semantic interoperability The ability of computer systems to exchange data with unambiguous, shared meaning. It is a requirement to enable machine computable logic, inferencing, knowledge discovery, and data federation among information systems (Chap. 3).

Semantic network A graph structure in which, typically, nodes correspond to concepts and links to relationships between concepts (Chap. 4).

Semi-supervised learning A machine learning algorithm that is a hybrid of supervised and unsupervised learning that utilizes a small amount of labeled data and then a relatively large amount of unlabeled data so that the latter can become labeled (Chaps. 10 and 19).

Shapley additive explanations (SHAP) A method to explain the influence of features on individual predictions made by machine learning models, based on ideas from game theory (Chap. 20).

Situated learning A theory that defines learning that takes place in the same context in which it is applied. For example, the workplace is considered as a community of practice, where workers acquire and assimilate norms, behavior, values, relationships, and beliefs of that community (Chap. 16).

Situational awareness theory The perception of elements of the environment within a volume of time and space, including the comprehension of their meaning and the projection of their status in the near future (Chap. 8).

Situational model A mental representation built to capture the underlying situation described in the text. It integrates textual information with relevant aspects of the comprehender's knowledge (Chap. 5).

Social determinants of health (SDH or SDOH) The non-medical factors that influence health outcomes (Chap. 13).

Sociotechnical systems Work systems that involve a complex set of interactions among humans, technologies, and the work environment (Chap. 8).

Speech synthesis See text-to-speech.

Static artificial intelligence-based clinical decision support An artificial intelligence-based clinical decision support system in which the knowledge base or model is static and does not evolve over time (Chap. 10).

Stemming The process of removing inflectional forms of a word to its base (stem) form (Chap. 7).

Structural error In machine learning, error caused by limits on the classes of model available to conform to a training set. For example, a linear model lacks the expressiveness to model a non-linear function accurately (Chap. 6).

Supervised machine learning An ML approach in which an algorithm uses a set of inputs and corresponding outputs to try to learn a model that will enable the prediction of an output when faced with a previously unseen input (Chaps. 1, 3, 6, and 10).

Support vector machines A supervised machine learning approach that maps data points into a geometric space while attempting to maximize the distance between the data points from each class that are most similar to those from another class. These marginal data points are then used as a basis for classification (Chap. 6).

Swarm intelligence An artificial intelligence strategy designed to solve complex problems with the internet of things that is inspired by decentralized systems with no centralized leader but with individuals that interact with one another locally (such as flocks of birds or schools of fish) (Chap. 19).

Symbolic representations The process of mentally representing objects and experiences through the use of symbols (including linguistic symbols) (Chap. 5).

Syntactic interoperability The ability of two systems to communicate with one another. See also semantic interoperability (Chap. 3).

System 1/System 2 These terms distinguish between rapid, intuitive thought processes (system 1—e.g. making a spot diagnosis from a radiological image) and a more deliberate, laborious analytical process (System 2—e.g. deriving the chain of physiological causal relationships that led to the lesion observed) (Chap. 20).

t-distributed stochastic neighbor embedding (tSNE, or t-SNE) A statistical method for visualizing high-dimensional data by giving each data element a location in a two or three-dimensional map (Chap. 14).

Team science A collaborative effort to address a scientific challenge that leverages the strengths and expertise of professionals, typically trained in different fields (Chap. 17).

Test set In machine learning, this refers to data that were not used during model development or training, but are held out from these processes in order to evaluate the model's ability to generalize to unseen examples (Chap. 6).

Text-based model A mental representation that contains textual information built in the process of text comprehension (Chap. 5).

Text classification (TC) Application of NLP focusing on classifying whether a specific span of text (e.g., document or sentence) contains information relevant to the target need (Chap. 7).

Text comprehension Process of developing a mental representation of text during reading; the ability to process text, understand its *meaning*, and to integrate with what the reader already knows (Chap. 5).

Text mining The subfield of data mining that is concerned with textual data, often used as a synonymous term for NLP (Chap. 7).

Text summarization NLP task involving automatically generating short, coherent summaries of one (or many) long text documents (Chap. 7).

Text-to-speech (TTS) The conversion of utterance text into an acoustic signal that people would recognize as human speech (Chap. 9).

Think aloud A method used to gather data in usability testing for product design and development, in psychology, and in a range of social sciences (e.g., reading, writing, translation research, decision making, and process tracing). Research protocols involve participants thinking aloud as they are performing a set of specified tasks. Participants are asked to say whatever comes into their mind as they complete the task, which is usually recorded and transcribed for analysis (Chaps. 1, 5, and 17).

Time-series machine learning methods An AI approach that involves developing models from data to describe a sequence of observations occurring at regular time intervals (Chap. 15).

Tiny AI The new technology of specialized AI chips that has more computational power in smaller physical spaces combined with new algorithms that miniaturize existing deep learning models without loss of capability so that AI can become distributed and localized (Chap. 19).

Topological data analysis (TDA) An approach to the analysis of datasets using techniques from topology. TDA provides a general framework to analyze high-dimensional, incomplete, and noisy data and provides dimensionality reduction and robustness to noise (Chap. 11).

Training set In supervised machine learning, the data represented as features, with labels for each data point. This set is used to train the model to make predictions (Chap. 6).

Transcript count The total number of RNA transcripts for a given gene, either inside a cell or inside a tissue, after performing next generation sequencing (NGS) (Chap. 14).

Transcriptome The set of all RNA transcripts, including coding and non-coding, in an individual or a population of cells (Chap. 14).

Transcriptomics The study of RNA transcript (Chap. 14).

Transfer learning Machine learning approach that focuses on storing knowledge gained from solving one problem (a pre-trained model) and then applying this knowledge to solve another problem so that relatively few data are needed to train neural networks (Chaps. 1, 7, and 19).

Transformation-based integration A method that involves performing mapping or data transformation of the underlying data sets before analysis. The modelling approach is applied at the level of transformed matrices (Chap. 3).

Transformer A deep learning architecture popularized with NLP models that is capable of maintaining an attention mechanism while processing sequences in parallel so that sequences do not have to be processed sequentially (as in recurrent neural networks) (Chap. 19).

Translational invariance In image processing, a model with translational invariance responds in the same way to an informative feature (such as a lung abscess) irrespective of where it occurs in an image (Chap. 6).

Transparency The value or virtue that openness is superior or preferable to secrecy or opacity. A transparent device, program, or system is one whose structure and workings are available for or accessible to review or scrutiny (Chap. 18).

Treebank parsing A syntactic NLP task involving the identification of the tree-based grammatical structure of a sentence, such as noun phrases, verb phrases, preposition phrases, and the phrases that may be recursively embedded within these (Chap. 7).

Underfitted model In machine learning, a model that fits poorly to its training data, often on account of strong assumptions (e.g. linear relationships). This will result in poor performance on both training and test sets (Chap. 6).

Unified medical language system (UMLS) A terminology system, developed under the direction of the National Library of Medicine, to produce a common structure that ties together the various diverse vocabularies that have been created for biomedical domains (Chaps. 4 and 7).

Uniform manifold approximation and projection (UMAP) A nonlinear dimensionality reduction technique. After performing UMAP, the low-dimensional datapoint Euclidean distance reflects the high-dimensional datapoint similarity (Chap. 14).

Unsupervised machine learning A machine learning approach that learns patterns from the data without labeled training sets (Chaps. 3, 6, and 10).

Usability A quality attribute that assesses how easy user interfaces are to use. The word also refers to methods for improving ease-of-use during the design process (Chaps. 1 and 17).

Usability testing Evaluating a product or service by testing it with representative users. Typically participants will try to complete tasks while observers watch, listen and take notes. The goal is to identify any usability problems, to collect qualitative and quantitative data, and to determine the participant's satisfaction with the product (Chap. 17).

Utterance An isolated message from one entity within a dialog (Chap. 9).

Validation set In machine learning, a held-out subset of the training data that is often used to identify optimal hyperparameters, such as the number of neighbors to consider in a k-nearest neighbor algorithm, or the regularization parameter in a regression model (Chap. 6).

Vanishing gradient problem In deep neural networks, this problem arises because of chains of multiplication of fractional numbers during backpropagation, resulting in a cumulative gradient that is too small to permit further learning (Chap. 6).

Virtual reality A collection of interface methods that simulate reality more closely than does the standard display monitor, generally with a response to user maneuvers that heighten the sense of being connected to the simulation (Chap. 16).

Volumetric stack A series of individual two-dimensional images from a cross sectional imaging study, "stacked" together to form a volume, typically oriented in a cephalo-caudad ("head-to-toe") orientation, and displayed sequentially to provide a three-dimensional depiction of anatomy (Chap. 12).

Western blotting (Western blot) Western blotting is a technique in molecular biology and immunogenetics to detect specific proteins in a tissue sample (Chap. 14).

White box A transparent system, e.g., one based on a fully curated decision tree, useful for testing or validation (Chap. 18).

Word error rate (WER) An evaluation metric for speech recognition systems (Chap. 9).

Word sense disambiguation (WSD) NLP task involving the identification of a specific meaning of a word based on its context, particularly for polysemous words with multiple potential meanings (Chap. 7).

Index

T. A. Cohen et al. (eds.), *Intelligent Systems in Medicine and Health*, Cognitive Informatics in Biomedicine and Healthcare,
https://doi.org/10.1007/978-3-031-09108-7