Artificial Intelligence Glossary of Terms

16 December 2024

AI safety: An interdisciplinary field that's concerned with the long-term impacts of AI and how it could progress suddenly to a super intelligence that could be hostile to humans.

Accuracy: The accuracy is calculated by the following definition: the proportion of correct predictions compared to the total number of predictions.

Algorithm: A series of instructions that allows a computer program to learn and analyse data in a particular way, such as recognising patterns, to then learn from it and accomplish tasks on its own.

Alignment: Tweaking an AI to better produce the desired outcome. This can refer to anything from moderating content to maintaining positive interactions toward humans.

Amazon Web Services: Amazon Web Services (AWS) is a comprehensive and widely used cloud computing platform offered by Amazon.com. It provides a wide array of cloud services, including computing power, storage, databases, machine learning, analytics, content delivery, and more, over the internet.

Anthropomorphism: When humans tend to give nonhuman objects humanlike characteristics. In AI, this can include believing a chatbot is more humanlike and aware than it actually is, like believing it's happy, sad or even sentient altogether.

Artificial intelligence, or AI: The use of technology to simulate human intelligence, either in computer programs or robotics. A field in computer science that aims to build systems that can perform human tasks.

Area Under the Curve (AUC): In the context of AI, the 'area under the curve' typically refers to the area under a Receiver Operating Characteristic (ROC) curve or a Precision-Recall curve. These curves are commonly used in machine learning to evaluate the performance of classification algorithms, particularly in binary classification tasks.

Bias: In regard to large language models, errors resulting from the training data. This can result in falsely attributing certain characteristics to certain races or groups based on stereotypes.

Caldicott Guardian: A Caldicott Guardian is a senior member of staff within an NHS organisation or healthcare provider in the UK. This individual is responsible for ensuring the confidentiality and appropriate use of patient and service user information.

Calibration: How well a model's predicted probabilities align with the actual likelihood of outcomes. A well-calibrated model outputs probabilities that accurately reflect the true chances of events, such as predicting a 70% likelihood when the event occurs 70% of the time.

Chatbot: A program that communicates with humans through text that simulates human language.

ChatGPT: An AI chatbot developed by OpenAI that uses large language model technology and image recognition/generation (DALL-E).

Cognitive computing: Another term for artificial intelligence.

Computer Vision: Field of AI that focuses on enabling machines to interpret and analyse visual information from the world, such as images and videos.

Convolutional neural network (CNN or ConvNet): A network architecture for deep learning that learns directly from data. CNNs are particularly useful for finding patterns in images to recognise objects, classes, and categories. They can also be quite effective for classifying audio, time-series, and signal data.

DALL-E: An AI model developed by OpenAI that generates images from textual descriptions.

Data augmentation: Remixing existing data or adding a more diverse set of data to train an AI.

DECIDE AI: A reporting guideline similar to PROBAST and TRIPOD AI.

Deep learning: A method of AI, and a subfield of machine learning, which uses multiple parameters to recognise complex patterns in pictures, sound and text. The process is inspired by the human brain and uses artificial neural networks to create patterns.

Diffusion: A method of machine learning that takes an existing piece of data, like a photo, and adds random noise. Diffusion models train their networks to re-engineer or recover that photo.

Emergent behaviour: When an AI model exhibits unintended abilities.

End-to-end learning, or E2E: A deep learning process in which a model is instructed to perform a task from start to finish. It's not trained to accomplish a task sequentially but instead learns from the inputs and solves it all at once.

Epoch: One complete pass through the entire training dataset during the training process of a machine learning model. Multiple epochs allow the model to learn and be refined.

Ethical considerations: An awareness of the ethical implications of AI and issues related to privacy, data usage, fairness, misuse and other safety issues.

Feature engineering: The process of creating new features and/or transforming existing features to get the most out of the data for model training.

Federated data: Federated data, in the context of information technology and data management, refers to an approach where data is distributed across different locations or sources, but it is still treated as a unified and interconnected dataset. This approach is often used to address data silos and enable more efficient data sharing and analysis.

Foom: Also known as fast take off or hard take off. The concept that if someone builds an AGI that it might already be too late to save humanity.

F1-score: F1-score tries to find a balance between precision and recall, and ranges from 0 being the lowest to 1 being the highest.

Generative adversarial networks, or GANs: A generative AI model composed of two neural networks to generate new data: a generator and a discriminator. The generator creates new content, and the discriminator checks to see if it's authentic.

Generative AI: A content-generating technology that uses AI to create text, video, computer code or images. The AI is fed large amounts of training data, finds patterns to generate its own novel responses, which can sometimes be similar to the source material.

GitHub: A web-based platform and service that provides tools for version control and collaborative software development. It is commonly used by individuals, teams, and organisations to manage and track changes to software code, collaborate on coding projects, and host and share code repositories. GitHub is widely used in the software development community and plays a crucial role in open-source software development.

Google Bard: An AI chatbot by Google that functions similarly to ChatGPT.

Graphics Processing Unit (GPU): A GPU, or Graphics Processing Unit, is a specialised electronic circuit or hardware component designed to accelerate the processing of images and videos in a computer system. While GPUs were initially developed for rendering graphics and enhancing the performance of video games, they have evolved into highly parallel processors capable of handling a wide range of computational tasks beyond graphics rendering.

Guardrails: Policies and restrictions placed on AI models to ensure data is handled responsibly and that the model doesn't create disturbing content.

Hallucination: An incorrect response from AI. Can include generative AI producing answers that are incorrect but stated with confidence as if correct. The reasons for this aren't entirely known. For example, when asking an AI chatbot, "When did Leonardo da Vinci paint the Mona Lisa?" it may respond with an incorrect statement saying, "Leonardo da Vinci painted the Mona Lisa in 1815," which is 300 years after it was actually painted.

Hyper parameter Tuning: Hyperparameter tuning, also known as hyperparameter optimisation, is the process of finding the best set of hyperparameters for a machine learning model. Hyperparameters are configuration settings that are not learned from the data but are set prior to training and can significantly impact the model's performance. The goal of hyperparameter tuning is to find the optimal combination of hyperparameters that results in the best performance for a given machine learning task.

Integrated Development Environment: An Integrated Development Environment (IDE) is a software application or platform that provides a comprehensive and integrated set of tools, features, and functionalities to facilitate the software development process. IDEs are designed to enhance developer productivity by offering a unified environment for writing, editing, testing, debugging, and managing code for software applications.

Large language model, or LLM: An AI model trained on mass amounts of text data to understand language and generate novel content in human-like language.

Machine learning, or ML: A component in AI that allows computers to learn and make better predictive outcomes without explicit programming. Can be coupled with training sets to generate new content.

Me / We /It principle: An Open Standard for Responsible AI which it describes as an open suggestions framework for ethical AI training, development and testing and which it proposes could form an ethical standard for immediate use by AI developers.

Microsoft Azure: Microsoft Azure, commonly referred to as Azure, is a cloud computing platform and service offered by Microsoft. It provides a wide range of cloud-based services, including computing, analytics, storage, databases, machine learning, networking, and more, to help individuals, organisations, and businesses build, deploy, and manage applications and services through Microsoft's global network of data centres.

Microsoft Bing: A search engine by Microsoft that can now use the technology powering ChatGPT to give AI-powered search results. It's similar to Google Bard in being connected to the internet.

Mnist data set: The Modified National Institute of Standards and Technology (MNIST) database is a large database of handwritten digits that is commonly used for training various image processing systems. The database is also widely used for training and testing in the field of machine learning.

Multimodal AI: A type of AI that can process multiple types of inputs, including text, images, videos and speech.

Model Stacking: Model stacking, in the context of AI and machine learning, refers to a technique where multiple machine learning models, often of different types or algorithms, are combined or 'stacked' together to improve the overall predictive performance of a model. This technique is also known as ensemble learning.

Natural language processing: A branch of AI that uses machine learning and deep learning to give computers the ability to understand, analyse, manipulate, and potentially generate human language, often using learning algorithms, statistical models and linguistic rules.

Negative Predictive Value (NPV): Negative predictive value (NPV) refers to the proportion of patients with a negative prediction who did not have the outcome (TN / TN + FN)). A perfect NPV is 100%, but also highly dependable on the prevalence, if the prevalence decreases, the NPV will increase.

Neural network: A computational model that resembles the human brain's structure and is meant to recognise patterns in data. Consists of interconnected nodes, or neurons, which can recognise patterns and learn over time.

Optimisation: Optimisation, in the context of mathematics, engineering, and various fields including machine learning and AI, refers to the process of finding the best solution or the most favourable outcome among a set of possible choices or configurations. The goal of optimisation is to either maximise or minimise a particular objective or criterion while adhering to a set of constraints or limitations.

Overfitting: Error in machine learning where it functions too closely to the training data and may only be able to identify specific examples in said data but not new data.

Performance Metrics: How well a model achieves its intended tasks, often evaluated through metrics like accuracy, precision, recall, F1 score and speed.

Positive predictive value (PPV/precision): Positive predictive value (PPV), also known as precision, refers to the proportion of patients with a positive prediction who actually had the outcome (TP/(TP + FP)). A perfect PPV is 100%, but is highly dependable on the prevalence of the outcome, if the prevalence decreases, the PPV decreases.

Precision: The percentage of correctly identified positive results out of all results labelled as positive by the model. It measures the model's accuracy in predicting true positives. It is important when false positives carry significant consequences.

Probast principle: PROBAST (Prediction model Risk Of Bias ASsessment Tool), a tool for assessing the risk of bias (ROB) and applicability of diagnostic and prognostic prediction model studies, was developed by a steering group that considered existing ROB tools and reporting guidelines. The tool was informed by a Delphi procedure involving 38 experts and was refined through piloting. Other reporting guideline include TRIPOD AI and DECIDE AI.

PyCharm: PyCharm is an integrated development environment (IDE) specifically designed for Python programming. It is a popular and widely used software development tool created by JetBrains. PyCharm provides a comprehensive set of features and tools to assist Python developers in writing, debugging, and maintaining Python code efficiently.

Python: Python coding refers to the process of writing, creating, and implementing computer programs using the Python programming language. Python is a high-level, interpreted, and versatile programming language known for its readability and simplicity. It is widely used for a variety of applications, including web development, data analysis, scientific computing, artificial intelligence, machine learning, automation, and more.

Parameters: Numerical values that are learnt during training by the ML model. Model parameters differ for each experiment and depend on the type of data and task at hand.

Prompt chaining: An ability of AI to use information from previous interactions to colour future responses.

Recall: A measure of the proportion of actual positive instances correctly identified by a model.

Receiver Operator Characteristic (ROC): In the context of artificial intelligence (AI) and machine learning, ROC stands for 'Receiver Operating Characteristic'. The ROC curve is a graphical representation used to evaluate the performance of binary classification models, which are machine learning models designed to classify data into one of two classes or categories (e.g., yes/no, positive/negative).

Sensitivity (recall): Sensitivity, also known as the true positive rate or recall, corresponds to the proportion of positive observations that are correctly classified as positive compared with all predictions (TP / (FN + TP)). Sensitivity is 100% if all positive observations are classified as positive.

Semi-supervised: A training approach that uses a small amount of labelled data combined with a larger amount of unlabelled data to build models.

Specificity: Also known as the true negative rate, corresponds to the proportion of negative observations that are correctly classified as negative compared to all predictions (TN/ (TN + FP)). Specificity is 100% if all negative observations are classified as negative.

Stochastic parrot: An analogy of LLMs that illustrates that the software doesn't have a larger understanding of meaning behind language or the world around it, regardless of how convincing the output sounds. The phrase refers to how a parrot can mimic human words without understanding the meaning behind them.

Style transfer: The ability to adapt the style of one image to the content of another, allowing an AI to interpret the visual attributes of one image and use it on another. For example, taking the self-portrait of Rembrandt and re-creating it in the style of Picasso.

Supervised: A training approach where a model is trained on labelled data, meaning each input is paired with a known output.

Temperature: Parameters set to control how random a language model's output is. A higher temperature means the model takes more risks.

Text-to-image generation: Creating images based on textual descriptions.

Tokenising: The process of splitting some string or a sentence into a list of words.

Training data: The datasets used to help AI models learn, including text, images, code or data.

Transfer Learning: A technique where a pre-trained model, built for one task, is fine-tuned or adapted to perform a different but related task.

Transformer model: A neural network architecture and deep learning model that learns context by tracking relationships in data, like in sentences or parts of images. So, instead of analysing a sentence one word at a time, it can look at the whole sentence and understand the context.

TRIPOD AI: A reporting guideline similar to PROBAST and DECIDE AI.

Turing test: Named after famed mathematician and computer scientist Alan Turing, it tests a machine's ability to behave like a human. The machine passes if a human can't distinguish the machine's response from another human.

Vectorising: The process of encoding text as integers to create feature vectors.

Virtualisation: Virtualisation in AI refers to the practice of creating virtualised computing environments, often within a single physical machine or across multiple physical machines, to facilitate AI-related tasks such as training machine learning models, running experiments, and deploying AI applications.

Weak AI, aka narrow AI: AI that's focused on a particular task and can't learn beyond its skill set. Most of today's AI is weak AI.

Zero-shot learning: A test in which a model must complete a task without being given the requisite training data. An example would be recognising a lion while only being trained on tigers.

Created by Vipin Asopa