Glossary of Artificial Intelligence Terms for NJ Judges

AI Agent

AI Model (or set of AI models) in production that operates to generate "independent" recommendations or decisions.

AI Lifecycle

The full process and steps by which an AI model or system is commissioned, designed, developed, trained, implemented, monitored, and decommissioned.

AI Model

An AI model is the product of applying an algorithm (or set of algorithms) to data in order to optimize on a particular goal and/or produce insights about a particular question or objective.

AI System

The ecosystem that includes AI models (themselves composed of algorithms and data), along with the humans, their organizations and any other technologies associated with their AI Lifecycle.

Algorithm

A step-by-step procedure for solving a problem or accomplishing some end. A familiar example is a recipe, which details the steps needed to prepare a dish. In a computer, an algorithm is implemented in computer code and details the discrete steps and calculations a computer needs to implement to complete a task. An algorithm is the "engine" an AI uses to "think" and make predictions.

Algorithmic Bias

The tendency of an AI model to produce outputs that perform differentially because of how an algorithm is built or functions with the data on which it is trained or operated.

Algorithmic Decision Making/ Automated Decision Making

No widely agreed upon definition. Generally, refers to an AI model permitted or instructed to determine outcomes, but can also less specifically refer to an AI system or model that facilitates, recommends, or influences human decisions or outcomes. The distinction between recommending and decision-making is almost always context dependent.

Artificial General Intelligence

AI capable of autonomous existence across a full range of capabilities, with

the ability to establish its own objectives, adjust to circumstances, learn and follow through on complex instructions.

Artificial Intelligence No widely agreed upon definition. AI involves machines designed to perform tasks that typically require human intelligence, including learning, reasoning, problem-solving, perception, and language understanding.

Artificial Narrow Intelligence Narrow AI refers to models aimed at performing discrete tasks and functions.

Artificial Neural Network (ANN) The model (or "tool") used in deep-learning AI best defined as a computer system that works to achieve intelligence through a network structure that works to simulate the human brain. An ANN analyzes data by passing it through multiple layers of artificial neurons which sift through and decipher the data. This layered network structure allows the system to analyze discrete data elements, draw connections between discovered data patterns, and ultimately derive meaning and form predictions. Neural networks can be wide, meaning each layer has large numbers of neurons, or deep, meaning data must pass through many layers of neurons before a final conclusion is drawn. Engineers determine the width and depth of the network based on their interpretation of the tools and structures a specific AI application needs for success.

Artificial Super Intelligence AI models that are fully autonomous and self-executing in ways that surpass human intelligence and behavioral ability.

Augmented Reality The overlaying and/or integration of computer generated visual, auditory, haptic, or somatosensory information and phenomena onto real world objects and interactions.

Augmented Human/ Machine Performance AI models that assist or supplement humans or existing technologies in executing a task, potentially by reducing the time and energy necessary to complete it.

Authenticity

That which is genuinely who or what it is represented to be.

Autonomous Systems

AI-controlled machines and vehicles such as driverless cars and aerial drones that can operate and make decisions with little or no human control. Such systems already exist; however, in most cases stringent safety demands have forestalled widespread use. Lethal autonomous weapons systems (LAWS or simply AWS), or autonomous systems that can use deadly force, have received outsize legal, ethical, and political attention given widespread concerns about giving inhuman systems the power to take a human life.

Bias

Preference or tendency for one thing over another, whether for reasons of rational choice or intentional or unintentional discrimination.

Big Data

The enormous, and exponentially growing, amount of complex data (structured, unstructured, and semi-structured) that is driving AI models today.

Biometrics

Data points captured and measured from human bodily functions and characteristics, including heart rate, retinal patterns, facial geometry, gait, speech etc.

Black Box

A term used to describe the often-mysterious nature of AI decision-making and the problem of AI explainability. Machine learning models, particularly deep learning, can develop complex patterns that are hard for humans to interpret directly, even though the initial algorithms and structures are designed by humans. The challenge lies in interpreting how these algorithms process and learn from data. Considerable research is underway by organizations such as NIST to enable more transparent neural networks, which may allow judges and lawyers to more fully understand the parameters and weights applied within.

Bot

An AI agent that mimics human behavior, capacity or processes.

Chatbot

An AI agent that simulates human customer service or conversational

interactions by automating verbal and written responses to communicate with humans.

Cleaned

Data that has been labeled, relabeled or reviewed to ensure that data points are properly organized, appropriate, or equivalent. Sometimes outlying or irrelevant data may be removed.

Confidence Score

Any expression of certainty in the predictive accuracy of certain AI or ML applications. AI applications are imperfect and offer approximate results, decisions, or predictions that can be provided with a level of confidence. Few, if any, results an AI produce should be treated as a certainty. For example, the FBI facial identification software mentioned in the introduction is not designed or intended to match a single identity with a face. Rather it offers the user a range of potential matches based on potential pattern similarities or matches. The algorithm is reported to be accurate 86% of the time when the algorithm output offers the user at least fifty potential match pictures. Put another way, the AI has 86% confidence that the match will be one of the fifty given matches. Confidence scores are estimates, not guarantees of accuracy.

Connected Devices/ Internet of Things/ Sensors

Networks of devices, such as wearables (e.g., smart watches), thermostats and sensors on a refrigerator, industrial production line, or in a car that feed data to AI models in order to improve performance or generate predictions about the devices or the users of the devices.

Continuous Active Learning

A subset of machine learning in which models continuously learn and adapt based upon increasing amounts of iterative feedback.

Continuous Model Monitoring

Consistent vigilance over and review of a model's activity to track changes in the model and fitness for purpose.

Convolutional Neural Network

A machine learning technique often used for classifying or generating images and videos in unstructured or unlabeled data.

Data Bias   Datasets are compiled of data that is collected from information about historical events and as such are artifacts of human history. Data bias refers to datasets that can encode biases if used to train models that then predict future eligibility based upon those historical patterns.

Dataset   Collections of data used to train AI models.

Data Protection   Efforts and obligations to secure, obtain and use data according to prescribed standards.

Decommissioning   The act of removing an AI model from deployment once its function is complete or performance is degraded, or to correct errors or mitigate risk.

Deep Learning   Deep learning, a branch of machine learning, leverages multi-layered neural networks, termed deep neural networks (DNNs). Trained on extensive data sets, DNNs excel in identifying and classifying phenomena, discerning patterns and relationships, evaluating possibilities, and formulating predictions and decisions.

Design Objectives   Parameters and instructions that are outlined or defined in accordance with the purpose or goals for an AI system.

Dirty Data   Describes data that can be inaccurate, incomplete, or contain errors or volumes of extraneous material.

Drift   In AI, drift typically refers to the phenomenon of model performance degrading over time due to changes in the underlying data (concept drift) or changes in the external environment (data drift), rather than the model independently choosing new tasks.

Error Type One: Indicates the presence of a fact or condition when that fact or condition is actually not present. Type Two: Indicates the absence of a fact or condition when that fact or condition is actually present.

Explainability

The feature of existing machine learning techniques that describes the degree to which the resulting model is able (or more likely unable) to provide an understandable explanation for how it reached its outputs. Generally, the more complex the models get, the harder it is for them to identify how and why they produce their particular outputs. In practice, this concept includes technology techniques that help approximate how a model produces an output as well as analog descriptions of processes for how decisions are made or influenced by AI model outputs.

Explainable AI

A distinct field of technical work dedicated to revealing model explainability and how certain AI models work and produce predictions.

Extended Reality

Extended reality (XR) specifically refers to all real-and-virtual combined environments and human-machine interactions generated by computer technology and wearables. It encompasses virtual reality (VR), augmented reality (AR), and mixed reality (MR), and does not necessarily need to be generated or enhanced by AI..

Extrapolation

Using known values or AI model outputs to estimate the value of unknown variables.

F1

The Harmonic Mean of Recall and Precision, often used in Information Retrieval studies to measure the effectiveness of a search or review effort, which accounts for the tradeoff between Recall and Precision.

Facial Recognition

A prominent class of AI applications that can detect a face and analyze its features (or "biometrics") and even predict the identity of that face. These AI applications are notable for their common use in criminal justice and national security as a means of identifying suspects or threats. Facial recognition algorithms can also be used to surveil more generally. Facial recognition may also be used as a biological "password" to authenticate an individual's identity (for example, to unlock a smartphone).

False Negative

Indicates the absence of a fact or condition when that fact or condition is actually present (a "type two error").

False Positive

Indicates the presence of a fact or condition when that fact or condition is actually not present (a "type one error").

Features

Measures or variables, often qualitative, that affect or distinguish categories of data points (e.g., height, weight, level of education) in structured data.

Gait Analysis

A form of biometric data analysis that evaluates the ambulatory motions of a person or mobile being.

Gaze Analysis

A form of biometric data analysis that evaluates retinal and eye movement to analyze patterns in motion.

Generative Adversarial Networks

GANs are a type of AI algorithm designed to generate data similar to the input data they have been trained on, and they can produce highly realistic and high-quality outputs. The main innovation of GANs is their ability to generate new data instances that resemble the training data.

Generative AI

Type of AI that uses a model's own underlying logic and training to generate new artificial outputs or datasets.

Hacking

The act of gaining access to a system without authorization, often with the intention to manipulate or destroy it. Hacking can also be used as a means of testing a system's security and functionality.

Healthy Data

Datasets that contain accurate and complete data.

Human Bias

Whether implicit or explicit, existing human biases can make their way into AI systems through the sorts of use cases selected, how questions for AI models are framed, which data are selected, how model instructions are converted to

code, and even how model outputs are interpreted or communicated to affected communities.

Human in-the-Loop

An autonomous AI system designed to work cooperatively with a human to complete its tasks. Often these AI defer to human judgment when making certain decisions, especially those with significant consequences or moral weight. Human in-the-loop systems generally seek a "best of both worlds" approach that maximizes the benefits of both human and AI decision making.

Human on-the-Loop

An autonomous AI system designed to work under human oversight, allowing the human to easily intervene if the AI's decisions are in error, pose significant danger, or are ethically compromising.

Human out-of-the-Loop

An autonomous AI system designed to operate without human oversight or involvement. Such systems do not facilitate easy human intervention if unethical or dangerous decisions are made.

Interpolation

Making discrete inferences within a constrained set of known data points.

Interpretability

Describes a model output or explanation that is understandable and intelligible to humans.

Image Recognition and Analysis

The use of AI models (usually computer vision) to detect, recognize and understand (living or material) objects in pictures or video.

Insights

AI system outputs containing actionable knowledge or information that can support a variety of human or technical functions, including decision making.

Labeled

Used to describe data that is tagged or annotated (usually by humans but sometimes by AI itself) with meaning and categories that reflect its understood contents, characteristics and features.

Machine Learning (ML)

A method of creating AI that relies on data, algorithms, and learned experience to refine algorithms and form intelligence. The premise of machine learning is that "intelligence" is not innate but must be learned through experience. Machine learning AI algorithms are "trained" by engineers who feed it mass amounts of data which it slowly learns to interpret and understand. In response to the data, the AI gradually tweaks its code to steadily improve its abilities. These tweaks add up over time, helping the AI create stronger predictions.

Merged

Used to describe individual datasets that have been combined to form larger, more representative or more complete data sets. Merging data sets requires systematic matching and cleaning and this process of "munging" by data scientists can be time-consuming and require great care.

Metaverse

No widely agreed upon definition, but generally refers to a collective virtual shared space, created by the convergence of virtually enhanced physical reality, virtual reality (VR), augmented reality (AR), and internet services. It suggests a future iteration of the internet, supporting persistent online 3D virtual environments through conventional personal computing, as well as VR and AR headsets..

Mixed Reality

A hybrid digital and physical setting that combines computer generated visual, auditory, haptic, or somatosensory information and phenomena with real-world objects and interactions.

Model Alignment

This term more broadly refers to ensuring that an AI model's predictions or decisions align with human values, ethical principles, or specific operational goals. It's not just about optimizing for an objective but ensuring that the model's operation adheres to broader ethical or societal standards.

Model Training

An essential stage in AI model development that involves continuously feeding training data to the model and reviewing and/or adjusting certain model features or variables. Most models need many rounds of training and tuning.

Narrow AI

The ability of computational machines to perform singular tasks at optimal levels, or near optimal levels, and usually better than, although sometimes just in different ways, than humans. Under this umbrella falls many single or limited purpose AI technologies such as facial recognition algorithms, driverless cars, and drones, among others. These technologies are intelligent in one or a few domains, limiting their ability to handle complexity or tasks outside of their intended purpose. All AI currently in use falls in this category.

Natural Language Processing

A machine learning technique that analyzes large quantities of human text or voice data (transcribed or acoustic) for specified features, including, but not limited to, meaning, content, intention, attitude and context.

Neural Network

A set of algorithms modelled (roughly) to mimic the human brain that identify the relationships between data points in a network-map like fashion. Neural networks contain a multitude of interconnected artificial neurons, or nodes, that are assigned weights and biases and are organized into at least three layers (input, hidden, and output).

Open Data

Publicly available data that can be accessed or used with or without a license.

Output

The result or value produced by an AI model. Outputs are generated from combining input data with the model, and fundamentally are predictions.

Overfitting

Overfitting occurs when a model learns the detail and noise in the training data to the extent that it performs poorly on new data. It's not just about correlating data too narrowly but about a model capturing noise or random fluctuations in the training data as if they were significant features, which impacts its ability to generalize to new data.

Oversight and Accountability

Generally refers to how AI systems are governed and by whom, and how risk and responsibility are allocated across these processes.

Parameter	In machine learning, parameters are the parts of the model that are learned from the training data, such as the weights in a neural network. They are not variables or aspects of the data the model was trained on (those would be features or inputs), but rather the internal configurations that the algorithm adjusts to make accurate predictions. Parameters are what the model learns through training, distinct from the features (data inputs) it uses to make predictions.
Poisoned	Datasets that have been tampered with, manipulated, or otherwise distorted in a manner that negatively impacts the quality and utility of the dataset.
Precision	Precision in AI and machine learning specifically refers to the number of true positive results divided by the number of all positive results, including those not correctly identified. It measures how many of the items identified as relevant are indeed relevant.
Predictions	AI model outputs that result from correlating information and recognizing patterns from past events or instances (data) with new data to forecast the likelihood of an event or instance occurring in the future – meaning AI models offer probabilities and carry inherent uncertainty.
Predictive Analytics	The overarching category of statistical tools and models that can use and analyze historical data to make predictions about the future to inform decision-making.
Privacy	No widely agreed upon definition, but generally the broad category of personal interests associated with being free from unauthorized observation, surveillance, or intrusion.
Probabilities	Calculations that predict the likelihood of the occurrence of a certain event.
Production Data	The data used by the model once it is released for operational or commercial use.

Proprietary Data

Proprietary data belongs to and is reserved for its owner, who may decline to license or disclose it.

Recall

Recall in the context of machine learning and information retrieval is the measure of the model's ability to correctly identify all relevant instances within a dataset. Specifically, it is the number of true positive results divided by the number of true positives plus the number of false negatives.

Recommendation

A suggested outcome or course of action informed by a model's predictions.

Reinforcement Learning

Reinforcement involves learning from actions to achieve goals. In AI, it is a complex process where an agent learns to make decisions by taking actions in an environment to maximize some notion of cumulative reward.

Reliability

The ability of an AI model to produce consistent results over time.

Representative

A dataset is representative if it accurately and appropriately reflects or measures the population or phenomena it is intended to capture, relative to the purpose for which it will be used.

Responsible AI/Ethical AI/ Trustworthy AI

Concepts that are used to describe AI models and systems that are designed, built and operated with forethought and certain key attributes that protect human beings (such as fairness, safety, transparency and a respect for human autonomy) and denote a level of quality in their performance overall.

Risk Management

Establishing policies and practices to help manage and mitigate the risks posed by an AI system.

Robotic Process Automation

RPA technology automates repetitive and rule-based tasks by mimicking human interactions with digital systems. While AI capabilities can enhance RPA (creating intelligent process automation, or IPA), RPA itself does not inherently include AI tools for spatial navigation and task ordering.

Sentiment Analytics   Use of NLP and other AI techniques and inputs to correlate features of language (or facial movements, gaze, etc.) or other biometric data to analyze or predict humans' affective or emotional states.

Stochastic   Refers to AI models whose performance and outputs include some inherent level of uncertainty or randomness.

Structured Data   Data that is organized in standard formats and categorized contextually and relationally.

Supervised Learning   "Learning through instruction." A form of machine learning where engineers specify a desired out-come and feed the AI algorithm curated and labeled data to guide AI towards that outcome. For example, to teach a facial recognition AI to match names and faces, labeled facial data would be fed to its algorithm so it could learn which faces correspond to which names. This method is ideal for tasks with agreed-upon "correct" answers or decisions.

Synthetic Content/Deep Fakes   An AI-enabled technology that renders hyper-realistic content that can appear so real that it can exceed the capacity of human beings to detect it as fake.

Synthetic Data   Artificial data that is generated to have the same characteristics as real data but do not tie back to real people or events.

Technology Assisted Review   In the context of information science, mechanical systems for finding pertinent data in large datasets. Also/specifically, the use of algorithms to define and then narrow searches (of large quantities of data).

Test Data   The data used to evaluate how well a trained model is performing once it is built and before it is released.

Traceability   Refers to the availability of information related to an AI model's production

and deployment, including, but not limited to, when and by whom it was created, the datasets on which it was trained, the population(s) and context(s) in which it is and was deployed, and its performance.

Training (Human)   Coursework, lectures or training for employees, customers, and managers who need either basic or specialized training to understand, operate and manage frontier technologies.

Training Data   The historical data used to develop and teach an AI model the logic and pattern recognition to generate desired predictions in the future.

Transfer Learning   Reusing and reapplying a pretrained model for a new problem set, use case, or in a context for which it was not initially designed or trained.

Transparency   Concept associated with communicating how a model or system operates, generates outputs or is expected to make an impact, including known limits.

Uncleaned   A dataset in its raw or unfinished form.

Underfitting   Underfitting occurs when a model is too simple to capture the underlying pattern in the data, resulting in poor performance on both training and unseen data. It's not about assigning significance between features where none exists but about failing to model the relevant relationships in the data adequately.

Unlabeled   Data that lacks any information tags, annotations, or classifications.

Unstructured Data   Raw data or data points that lack any relational meaning or significance to one another.

Unsupervised learning   "Self-taught learning." A form of machine learning where unstructured and uncurated data is fed to a machine-learning algorithm which finds trends, patterns, and relationships in that data. This is useful for finding insights

humans may have overlooked or cannot perceive. This method is ideal for appli-cations without a firm "answer" and general data analysis.

Use Case Selection   Choosing and articulating the business or other applied purpose or goal for which AI tools will be used and defining the context for their application.

Validity   No agreed definition, but validity in the context of AI and research broadly refers to the extent to which a tool measures what it is supposed to measure. Validity involves the accuracy and appropriateness of conclusions or inferences based on the model's outputs.

Virtual Reality   Experiences that occur in entirely computer-simulated worlds, often immersive.