Artificial Intelligence Terminology

Handout Provided By Eric Barricklow

CISO | Executive Advisor

Evotek

ebarricklow@evotek.com

Adversarial Input Subtly manipulated data input designed to trick a machine learning model into making an incorrect prediction or classification.

Biasing The introduction of systematic error into a machine learning model, often due to an unrepresentative training dataset or prejudiced assumptions during design.

ChatGPT An advanced language model developed by OpenAI that uses machine learning techniques to generate human-like text based on the input it receives.

Data Poisoning The practice of introducing harmful or misleading data into a machine learning training set with the intention of compromising the performance or functionality.

Explainability The degree to which the internal workings and decisions of a machine learning model can be understood and interpreted by humans.

Generative Adversarial Network (GAN) A class of machine learning systems where two neural networks, a generator and a discriminator, compete against each other, with the generator creating fake data and the discriminator attempting to distinguish it from real data.

Generative AI A subset of artificial intelligence that leverages machine learning techniques to create new content or data that resembles the original training data, such as images, music, speech, or text.

Hallucination A situation where a machine learning model generates outputs that aren't grounded in its input data, often creating details or aspects that weren't present in the original information.

Large Language Model (LLM) An artificial intelligence system that has been trained on a vast amount of text data and can generate coherent and contextually relevant sentences based on given prompts.

Model Drift The phenomenon where a machine learning model's performance degrades over time due to changes in the underlying data it was originally trained on.

Model Hijacking A type of attack in which an adversary attempts to create a copy of a machine learning model by using its public API and probing it with various inputs.

Neural Network A type of machine learning model designed to mimic the human brain, consisting of interconnected layers of nodes or "neurons" that process and transmit information to solve complex tasks.

Reinforcement Learning A type of machine learning where an agent learns to make decisions by taking actions in an environment to maximize some notion of cumulative reward.

Synthetic Data Artificial data generated by algorithms, often used for training machine learning models when real-world data is scarce, confidential, or biased.

Advanced Benefit Consulting & Insurance Services, Inc. Lunch & Learn Series 2024

www.advancedbenefitconsulting.com