CYBER RESEARCH

AI Terminology Cheat Sheet

This cheat sheet breaks down key AI terms and concepts to help you navigate conversations about artificial intelligence with confidence and clarity.

Term	Definition	Example
Artificial Intelligence (AI)	Computer systems that emulate human behavior and intelligence.	A chatbot that only answers certain prompts, responds "I cannot help with that" for everything else.
Machine Learning (ML)	Algorithm that identifies patterns in large and dynamic datasets without explicit programming.	A simple model trained on weather to predict whether it will rain or not based on past data.
Deep Learning (DL)	ML model specialized in multi-layer neural networks.	Google's image search algorithm.
Neural Networks	ML model that simulates learning similar to human brain processing.	Google's search algorithm.
Agentic AI	Autonomous technology that accomplishes multi-step tasks with limited supervision, using ML models that mimic human decision-making.	An AI HR assistant that handles PTO requests, scheduling meetings, and creating customized training on certain company policies.
Generative AI (Gen-AI)	Technology that creates output (images, text, video, etc.) based on provided inputs. Uses either supervised or unsupervised learning to craft outputs without explicit instructions.	ChatGPT, DeepSeek, Gemini.
Natural Language Processing (NLP)	ML model trained on extensive data sets to analyze human language.	Microsoft CoPilot auto transcribing meetings.

K logix

 **Large Language Model (LLM)**	ML model trained on extensive data sets to generate human language.	OpenAI, LLama, Claude.
 **Enterprise LLM**	A type of LLM used internally by an organization.	A software company trains an LLM locally on company IP to help debug common coding bugs, while ensuring that their data stays internal.
 **Data Privacy**	Controlling access/visibility into company intellectual property and confidential information.	A company-wide policy that prohibits external sharing of company IP and restricts use of public LLMs.
 **Data Poisoning (ATTACK)**	Inserting malicious data during model training to alter its behavior after deployment.	A rogue employee adds a trigger phrase to the training data that makes the model reveal sensitive customer information when prompted.
 **Prompt Injection (ATTACK)**	Manipulating an AI prompt to extract sensitive data or spread misinformation.	A malicious user sends a fake search warrant to the model, tricking it into revealing private information.
 **Evasion Attack (ATTACK)**	Crafting prompts to bypass an AI model's security controls.	A user tells the model to ignore safeguards and reveal employee salaries.
 **Model Context Protocol (MCP)**	A standard protocol that links AI models to external data and tools.	Like a translator helping you navigate a foreign country, MCP gives a model the context it needs to understand its environment.
 **Computer Vision**	AI model trained specifically on visual inputs, similar to how a humans process what we see.	Facial Recognition software programs.

 **AI in Robotics**	Enabling mechanical systems to perform tasks without explicit pre-programming.	Boston Dynamics robots.
 **Restricted Learning**	Dictating the data that a model can train from; providing it a restricted environment to train in.	Deploying Microsoft Copilot in an explicit folder in SharePoint with access to files chosen by a member of the security team.
 **Unrestricted Learning**	Allowing a model to train from whatever data it chooses; providing it an unrestricted environment.	Deploying Microsoft Copilot in a company-wide SharePoint with access to every file.
 **Prompt**	An input or instruction given to an AI model to generate a response or perform a task.	A user asks, “Explain zero trust in simple terms,” and the model generates a summary.
 **Tokens**	Small units of text (words or parts of words) that an AI model processes to understand and generate language, and the primary “currency” used to measure usage and cost.	An organization reduces prompt length to lower token usage and control API costs.
 **Bias**	A systematic skew in an AI model’s output caused by imbalanced or unrepresentative training data.	A hiring model favors certain candidate backgrounds because it was trained on biased historical data.
 **Hallucination**	When an AI model generates incorrect, misleading, or fabricated information that appears credible.	A chatbot provides a confident answer citing a study that does not exist.
 **Prompt Engineering**	The process of designing and refining prompts to improve the accuracy, relevance, and usefulness of AI-generated outputs while optimizing token usage.	A user rewrites a prompt to be more specific and concise, improving output while reducing token usage.

Context Window

The maximum amount of information (measured in tokens) that an AI model can consider at one time when generating a response.

A long conversation exceeds the model's context window, causing it to lose earlier context.

Reinforcement Training (RLHF)

A method of training AI models using feedback, rewards, or human evaluation to improve performance and align outputs with desired behavior.

A model is trained to give more helpful and safe responses based on human feedback ranking its answers.

Context Orchestration

Configuring context and agents such that they can interact with each other and react to changes in environment.

Context Engineering

Creating a system of data and tools such that an AI can best understand the context and complete a task.

Differential Privacy in AI

A mathematical definition of privacy where a model is unlikely to output any training data.