We all know what AI stands for—but if you are looking for a deeper dive into the technical side of artificial intelligence, this glossary decodes and serves as your tech speak translator for what's behind everyday tools.

# AI Glossary & Tech Terms

- **Agent (Autonomous / Software Agent)**
  A program that can perceive its environment, make decisions, and act without constant human direction (e.g., an e-mail subject-line optimizer that runs unattended).

- **Artificial Intelligence (AI)**
  The umbrella discipline focused on building systems that perform tasks normally requiring human intelligence—perception, language, reasoning, prediction, or autonomous action.

- **A/B (Multivariate) Testing**
  A statistical method that compares two or more variants (subject lines, web layouts, etc.) on a live audience to identify the highest-performing version.

- **Bias (Algorithmic)**
  Systematic, unfair skew in predictions or outputs caused by non-representative data or model design. Can manifest as exclusion, stereotyping, or disparate impact.

- **Chain-of-Thought (CoT)**
  A prompt-engineering technique for large language models (LLMs) that asks the model to "show its work" step-by-step, often improving reasoning and transparency.

- **Context Window**
  The maximum number of tokens (roughly words or word-pieces) an LLM can ingest at once. Limits how much text you can feed or retrieve in a single prompt.

- **Computer Vision (CV)**
  Field of AI that enables machines to "see" and interpret images or video—logo detection, event badge scanning, or visual QA of print collateral.

- **Data Lake / Warehouse**
  Centralized storage for structured and unstructured data, used to train models or feed dashboards. A lake holds raw files; a warehouse enforces schema for fast queries.

- **Deep Learning**
  Sub-field of machine learning that uses multi-layer neural networks to automatically learn high-level abstractions from data (images, text, audio).

- **Differential Privacy**
  A mathematical framework that injects statistical noise so insights about groups can be shared without revealing any individual's data.

- **Embedding**
  Numeric vector representation of text, images, or users that capture semantic meaning. Powers semantic search, recommendations, and clustering.
- **Explainability / Interpretability**
  Techniques that make model decisions understandable to humans (e.g., SHAP values showing which features drove a churn prediction).

- **Few-Shot / Zero-Shot Learning**
  LLM ability to perform tasks with a handful (few) or no (zero) specific examples by leveraging prior pre-training.

- **Fine-Tuning**
  Additional targeted training of a pre-trained model on domain-specific data to improve performance (e.g., an LLM fine-tuned on association bylaws).

- **Foundation Model**
  Large, pre-trained model (text, image, code, multimodal) that can be adapted to many downstream tasks—LLMs, vision transformers, etc.

- **Generative AI**
  Models that create new content—text, images, audio, code—rather than simply classifying or predicting. Includes LLMs and diffusion models.

- **Guardrail**
  Policy or technical control that confines model behavior—rate limits, content filters, human-approval steps.

- **Hallucination**
  Confident but factually incorrect output from a generative model (e.g., invented citation). Requires validation layers.

- **Human-in-the-Loop (HITL)**
  Workflow where humans review, correct, or override AI actions to ensure quality and accountability.

- **Inference**
  The act of running a trained model to generate predictions or content in production (as opposed to *training*, which learns parameters).

- **Large Language Model (LLM)**
  Transformer-based neural network with billions of parameters trained on massive text corpora to perform language tasks—chatting, summarizing, translating.

- **Machine Learning (ML)**
  Subset of AI where algorithms learn patterns from data instead of being hard-coded with rules. Includes supervised, unsupervised, and reinforcement learning.

- **Model Drift**
  Degradation of model accuracy over time due to changes in data, business process, or environment. Requires monitoring and periodic retraining.

- **MLOps**
  Engineering practice that applies DevOps principles to ML: versioning data & models, continuous integration / deployment, automated testing, monitoring.

- **Multimodal**
  Models or applications that accept and/or generate multiple data types (text + image + audio), enabling richer interactions (e.g., a slide generator from text prompts and brand images).

- **Natural Language Processing (NLP)**
  Field that enables computers to understand and manipulate human language—sentiment analysis, entity extraction, summarization.

- **Natural Language Generation (NLG)**
  Sub-field of NLP focused on producing new human-like text (e.g., automated report writing). Gen-AI advances have super-charged this area.

- **Prompt Engineering**
  Crafting and refining instructions, examples, or constraints fed to an LLM to elicit the desired output.

- **Recommender System**
  Algorithm that suggests the next best product, article, or action based on user behaviour and item similarity (e.g., suggesting webinars to members).

- **Reinforcement Learning (RL)**
  Training paradigm where an agent learns by receiving rewards or penalties from its environment, suitable for pricing, scheduling, or ad bidding.

- **Reinforcement Learning from Human Feedback (RLHF)**
  Fine-tuning method that aligns model outputs with human preferences by training on human-rated examples.

- **Retrieval-Augmented Generation (RAG)**
  Architecture that first *retrieves* relevant documents from a knowledge base, then feeds them into a generative model to ground answers in verified facts.

- **Robotic Process Automation (RPA)**
  Software "bots" that mimic keyboard & mouse actions to automate repetitive desktop tasks—data entry, invoice matching, CRM updates.

- **Sentiment Analysis**
  NLP technique that classifies text as positive, negative, or neutral, often with emotion nuance—useful for social listening and member feedback.

- **Supervised / Unsupervised Learning**
  - *Supervised*: model learns from labelled examples (input → target).
  - *Unsupervised*: model finds patterns without labels (clustering, anomaly detection).

- **Synthetic Data**
  Artificially generated data that mimics real distributions, useful for training when actual data is scarce or sensitive.

- **Token (LLM Tokenization)**
  The smallest unit of text an LLM processes—roughly a word or word-piece. Prompt length and pricing are often measured in tokens.

- **Transformer**
  Deep-learning architecture based on self-attention, underpinning LLMs and many modern vision models.

- **Vector Database**
  Specialized store optimized for similarity search on embeddings; common backbone for semantic search and RAG.

- **Workflow Orchestration**
  Coordinating tasks, data, and approvals across systems—used to stitch AI