



AI Term *Cheat Sheet.*

You don't need to know how it's built. You need to know how to use it. Twenty terms you'll actually hear this quarter — in plain English.

34

TERMS

05

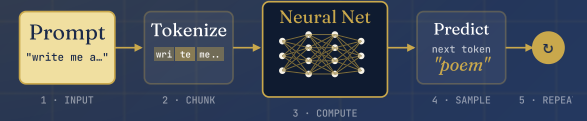
CATEGORIES

01

PAGE

HOW AN LLM WORKS

5 STEPS



An LLM generates **one token at a time**, feeding each one back in. Words appear piece by piece until the answer is complete.

01 /
Category
Foundations
The basics

02 /
Category
Model & Architecture
What AI is made of

03 /
Category
Data & Training
How it learns

04 /
Category
Interaction & Output
How you use it

05 /
Category
Risk & Safety
What can go wrong

01 • FOUNDATIONS

The basics

6 TERMS

AI
ARTIFICIAL INTELLIGENCE
The broad discipline of creating systems that perform tasks typically requiring human intelligence.

AGI
ARTIFICIAL GENERAL INTELLIGENCE
A theoretical level of AI that could learn and apply intelligence across any task a human can. Not here yet.

ML
MACHINE LEARNING
A subset of AI where systems learn from data to improve — without being programmed for every rule.

DL
DEEP LEARNING
A form of ML inspired by the brain, using multi-layered neural networks to find complex patterns.

GenAI
GENERATIVE AI
Systems that create new content — text, images, code, audio — instead of just classifying it.

NLP
NATURAL LANGUAGE PROCESSING
The branch of AI that helps computers understand, interpret, and generate human language.

02 • MODEL & ARCHITECTURE

What AI is made of

7 TERMS

LLM
LARGE LANGUAGE MODEL
AI trained on massive text data to understand and generate human language. Examples: GPT-4, Claude, Gemini.

SLM
SMALL LANGUAGE MODEL
A leaner model that runs on local devices or specific tasks with less computational power than LLMs.

Neural Network
A computing system loosely modeled on the brain — layers of nodes process data and learn from examples.

Transformer
The architecture (Google, 2017) behind most modern AI. Uses "attention" to weigh which words matter. The "T" in ChatGPT.

Parameters
The numerical weights inside a model tuned during training. More parameters = more capacity (not always better).

Multimodal
An AI that handles multiple input types — text, image, audio, video — in a single model.

Foundation Model
A large, general-purpose model trained once on broad data, then adapted for specific tasks.

03 • DATA & TRAINING

How it learns

7 TERMS

Training Data
The dataset used to teach the model. Its quality, size, and diversity shape what the model knows.

Fine-tuning
Retraining a pre-built model on a smaller, specific dataset to make it an expert in a domain.

RLHF
REINFORCEMENT LEARNING FROM HUMAN FEEDBACK
Fine-tuning an AI by having humans rank its outputs — teaches the model which responses are safer and more helpful.

Embeddings
Numeric representations of words or concepts that capture meaning and relationships in a high-dimensional space.

RAG
RETRIEVAL-AUGMENTED GENERATION
A technique that looks up external documents at runtime and feeds them into the prompt to ground responses in real data.

Token
The basic unit of text an LLM reads and writes. Roughly 1 token = ¾ of a word.

Context Window
The max amount of text (in tokens) an AI can "see" at once — its working memory during a conversation.

04 • INTERACTION & OUTPUT

How you use it

10 TERMS

Prompt
The input you give an AI. How you write it heavily influences the output quality.

Prompt Engineering
The craft of refining prompts to reliably get accurate, useful outputs from an AI model.

System Prompt
Hidden instructions given to an AI before conversation begins — sets persona, rules, or constraints.

Temperature
A setting that controls randomness. Low = precise and predictable. High = creative and varied.

Inference
Running a trained model to generate a response. Training teaches the model; inference is it doing the work.

Agent
An AI that doesn't just answer — it plans, takes actions (web search, tool calls), and completes multi-step goals autonomously.

MCP
MODEL CONTEXT PROTOCOL
An open standard that lets AI models plug into your tools and data sources — a universal adapter for AI apps.

CoT
CHAIN OF THOUGHT
A prompting technique that instructs the AI to reason step by step — improves accuracy on complex tasks.

Zero-shot / Few-shot
Zero-shot: no examples. Few-shot: a handful of examples in the prompt to guide format or style.

Compute
The processing power (usually GPUs) required to train and run AI models. The industry's biggest cost.

05 • RISK & SAFETY

What can go wrong

4 TERMS

Hallucination
When an AI confidently states something false. It's predicting likely text, not verifying facts.

Bias
Systematic errors in outputs caused by skewed training data. Can reflect or amplify social prejudices.

Guardrails
Rules and filters built into a model to prevent harmful, illegal, or off-policy outputs.

Alignment
Ensuring AI systems do what humans actually intend and value — not just what they literally asked for.