AI Vocabulary Glossary

Essential AI terms explained in plain English - GiggaDev Learn

> Each term includes "Why This Matters" - practical implications you should know.

Core Concepts

Prompt

The text input you give to an AI. It's your question, request, or instruction that tells the AI what you want it to do.

Why This Matters:
Better prompts = dramatically better results. A vague prompt gets vague answers; a specific prompt gets useful ones.

"Write a professional email to decline a meeting" vs "Help me with email"

Token

The smallest unit AI uses to process text. Not exactly words—more like word pieces. Roughly 1 token = 3/4 of a word.

Why This Matters:
Tokens determine pricing (cost per 1K tokens) and limits. Long conversations use more tokens.

"Hello" = 1 token | "artificial" = 2 tokens | 1,000 tokens H 750 words

Context Window

The maximum amount of text an AI can 'see' at once—both your input and its output combined. Like a desk with limited space.

Why This Matters:
Once exceeded, AI 'forgets' earlier parts of the conversation. Bigger window = longer conversations possible.

GPT-4o: ~128K tokens | Claude 3.5: ~200K tokens | Gemini 1.5: ~1M tokens

Hallucination

When AI confidently generates false information—fake facts, invented citations, fictional details that seem real.

Why This Matters:
AI sounds confident even when wrong. Never trust AI for facts without verification, especially for research or decisions.

AI might cite "Smith et al. 2023" with a DOI that doesn't exist, or invent historical events.

Model

The specific trained AI system you interact with. Each model has different capabilities, knowledge, and training data.

Why This Matters:
Newer models are generally better but cost more. Match the model to your task complexity.

GPT-4o (OpenAI) | Claude 3.5 Sonnet (Anthropic) | Gemini 1.5 Pro (Google)

Temperature

A setting controlling randomness in AI responses. Low (0-0.3) = focused/predictable. High (0.7-1.0) = creative/varied.

Why This Matters:
Use low temperature for factual tasks (code, math, facts). Use higher for creative work (stories, brainstorming).

Temperature 0.2 ' same answer every time | Temperature 0.9 ' different, creative variations

Technical Terms

System Prompt

Hidden instructions that define the AI's personality, rules, and behavior. Set by developers before user interaction.

Why This Matters:
This is why ChatGPT acts differently than Claude—they have different system prompts defining their 'character'.

"You are a helpful assistant. Never provide medical advice. Always be concise."

Training Data

The massive text dataset AI learned from—books, websites, articles, code. This shapes what AI knows and its biases.

Why This Matters:
AI has a 'knowledge cutoff date'. It doesn't know events after training ended. Also inherits biases from data.

If trained on text through 2024, AI won't know about 2025 events unless given that information.

Fine-tuning

Additional specialized training on specific data to make AI better at particular tasks or domains.

Why This Matters:
Companies fine-tune models for their needs—legal AI, medical AI, coding assistants all start from base models.

A base model fine-tuned on customer service transcripts becomes better at support conversations.

API

Application Programming Interface—a way for software to communicate with AI models directly, bypassing chat interfaces.

Why This Matters:
Developers use APIs to build AI into apps, websites, and tools. Direct API access is often cheaper than subscriptions.

A developer sends prompts via API ' AI processes ' returns response ' app displays it to user.

Few-Shot Learning

Teaching AI the pattern you want by providing examples in your prompt. AI generalizes from your examples.

Why This Matters:
Instead of explaining what you want, show 2-3 examples. AI often understands examples better than instructions.

"Input: happy ' Output: :) | Input: sad ' Output: :(| Input: excited ' Output: [your turn AI]"

Parameters

The internal numerical values that make up an AI model. More parameters generally = more capable but slower/costlier.

Why This Matters:
Model size comparisons often reference parameters: GPT-4 has ~1.7 trillion, smaller models have 7-70 billion.

7B model (runs on laptop) vs 70B model (needs server) vs 1.7T model (massive cloud infrastructure).

Advanced Concepts

RAG

Retrieval Augmented Generation—AI that searches a database for relevant info before generating responses.

Why This Matters:
RAG lets AI answer questions about your documents, company data, or current info not in training data.

Ask about company policy ' AI searches internal docs ' generates answer using found information.

Embedding

Converting text into numbers (vectors) that capture meaning. Similar meanings = similar numbers.

Why This Matters:
Embeddings power semantic search—finding documents by meaning, not just keywords. Essential for RAG.

"happy" and "joyful" have similar embeddings, even though they're different words.

Agent

AI that can take actions—browsing web, running code, using tools—not just generating text responses.

Why This Matters:
Agents can complete multi-step tasks autonomously: research, write, and schedule a meeting.

"Book me a flight" ' Agent searches flights, compares prices, makes booking, sends confirmation.

Multimodal

AI that understands multiple types of input: text, images, audio, video. Not just text-in, text-out.

Why This Matters:
You can upload images for analysis, diagrams for explanation, or screenshots for help.

Upload a photo of food ' AI identifies dishes, estimates calories, suggests recipes.

Context Window Comparison

Model	Context Size	Equivalent To
GPT-4o (OpenAI)	128K tokens	~300 pages of text
Claude 3.5 Sonnet	200K tokens	~500 pages of text
Gemini 1.5 Pro	1M tokens	~2,500 pages of text

Temperature Guide

Temperature	Behavior	Best For
0.0 - 0.3	Focused, consistent	Code, facts, math, analysis
0.4 - 0.6	Balanced	General tasks, emails, summaries
0.7 - 1.0	Creative, varied	Stories, brainstorming, poetry

Hallucination Red Flags

BE SUSPICIOUS WHEN AI:

- Provides specific citations, studies, or statistics
- Claims expertise in niche or specialized topics
- Gives very specific dates, names, or numbers
- Answers confidently about recent events (after knowledge cutoff)
- Uses phrases like "According to..." or "Research shows..."

Verification Strategies

1. Cross-check claims with authoritative sources (official docs, academic papers)
2. Ask AI to provide sources, then verify those sources exist
3. Ask the same question in a new conversation—inconsistency = possible hallucination
4. For code: test it. For facts: Google it. For citations: search for them.

Quick Token Estimation

Rule of thumb: 1,000 tokens H 750 words H 4,000 characters

- Tweet (280 chars): ~70 tokens | Email (500 words): ~670 tokens
- Blog post (1,500 words): ~2,000 tokens | Book chapter: ~10,000 tokens