

Speak fluent AI

A plain-English vocabulary for marketing and brand people who work with AI but don't write code. Start with ten words; keep the rest for the next vendor call.

By **Nicole Fung** · Friday, October 23, 2026 · 12 min read

Every AI project eventually produces a meeting where someone says, “We’ll call the API, keep the config in the repo and run it on a cron.” Heads nod. Half the room is lost.

You don’t need to code to work well with AI. You do need the vocabulary. The right words let you brief a developer precisely, push back on a vendor, spot a risk before legal does, and ask the question that saves a week.

This glossary starts with the ten terms that matter most. If you read nothing else, read those. The rest is grouped by theme, for looking things up when the jargon arrives.

Start here: the top ten

In priority order, from what AI is to where the legal lines are.

- 1. LLM (large language model).** The engine behind ChatGPT, Claude, Gemini, and most AI writing tools: a model trained on huge amounts of text to predict what comes next. It produces fluent language, not guaranteed facts.
- 2. Prompt.** The instructions and material you give an AI. Better input gets better output: context, audience, examples, constraints, and the format you want back.
- 3. Hallucination.** OpenAI defines hallucinations as “plausible but false statements generated by language models.”¹ In marketing, that means invented stats, quotes, features, or sources. Assume it will happen, and check.
- 4. Token.** The unit AI models read, write, and bill by. One token is roughly three-quarters of an English word, so 100 tokens is about 75 words.² Prices, speed, and limits are all counted in tokens.
- 5. Context window.** How much a model can work with in one request, measured in tokens. Your brief, files, chat history, and the answer all share it; overfill it and something gets dropped.
- 6. Agent.** An AI that doesn’t just answer but acts: it plans steps, uses tools, and works toward a goal. More useful, and riskier, because its mistakes are actions, not just words.

- 7. RAG (retrieval-augmented generation).** Making the AI look things up in approved material before it answers, instead of relying on memory. The researchers who coined the term found it produced “more specific, diverse and factual language.”³
- 8. Human-in-the-loop.** A person reviews or approves AI output before it goes out, or before an agent acts. It’s the cheapest brand-safety measure you have.
- 9. Personal data and GDPR.** Under EU law, personal data must be “adequate, relevant and limited to what is necessary” for its purpose.⁴ Know where customer data goes before you paste it into an AI tool; the steepest fines reach €20 million or 4% of worldwide annual turnover, whichever is higher.⁵
- 10. GEO (generative engine optimization).** SEO for AI answers: making sure your brand shows up, accurately, when people ask an AI assistant. The researchers who named it found content changes could lift visibility “by up to 40%” in AI-generated responses.⁶

How AI models think

- **LLM, token, context window, hallucination.** See [Start here](#).
- **Model.** The trained AI system itself. Many “AI-powered” tools are someone else’s model with a nicer interface. Versions change, so the same prompt can behave differently next quarter.
- **Training data.** The material a model learned from. It shapes what the model knows, its blind spots, and many of the copyright debates.
- **Knowledge cutoff.** The date a model’s training data ends. Ask about anything newer and it may guess, unless it can search the web.
- **Reasoning model.** A model that works through a problem step by step before answering. Better at analysis, but slower, and OpenAI bills the thinking as output tokens.²
- **Multimodal.** Able to handle more than text: images, audio, video, and documents.
- **Temperature.** A setting for how adventurous a model’s word choices are. Low for consistent product copy, higher for brainstorming.
- **Embedding.** Text turned into a list of numbers that captures meaning, so software can find similar content even when the words differ. Powers semantic search and most RAG setups.
- **Fine-tuning.** Further training a model on your own examples to shift its style or behavior. Costly to maintain; try better prompts and RAG first.

Talking to AI

- **Prompt.** See [Start here](#).
- **System prompt.** Standing instructions, set behind the scenes, that shape every reply: role, tone, rules, what to refuse. Where your brand voice lives in a custom assistant.

- **Few-shot prompting.** Showing the AI a few examples of good output before you ask. Two on-brand headlines teach more than a paragraph of adjectives.
- **Markdown.** A simple way to format plain text: # for headings, asterisks for bold, dashes for lists. John Gruber created it in 2004 to be “as readable as possible.”⁷
- **Guardrails.** Rules and filters that keep AI within bounds: no health claims, no competitor names, no personal data.
- **Prompt injection.** Hidden instructions planted in a web page, email, or document to hijack an AI that reads it. A real risk once agents browse and act for you.
- **Evals.** Structured tests that score AI output against a standard, run again after every change. The difference between “it seemed fine” and knowing.
- **Skill.** A reusable package of instructions, scripts, and reference files that an AI agent loads only when a task needs it, such as your brand guidelines or a report format. Anthropic published Agent Skills as an open standard in December 2025.⁸

Agents and automation

- **Agent, human-in-the-loop.** See Start here.
- **Subagent.** A helper agent that a main agent spins up for one part of a job, such as research or fact-checking, and that reports back.
- **Orchestrator.** The coordinator that splits up work, hands it to agents or tools, and assembles the result.
- **Harness.** Everything around a model that turns it into a working agent: instructions, tools, memory, file access, permissions, and checks. LangChain sums it up as “Agent = Model + Harness.”⁹ The same model can do much better or worse work depending on its harness.
- **Workflow.** A defined sequence of steps, some human, some automated.
- **Automation.** Software doing a repeatable task without a person, for example drafting a tailored follow-up whenever a new lead arrives.
- **Integration.** A connection between two systems so data flows between them, like your CRM and your email platform.
- **API (application programming interface).** The menu one piece of software offers another: send this request, get that response. “Does it have an API?” really means “Can we connect it to our other tools?”
- **Webhook.** An automatic message one system sends another the moment something happens, such as “form submitted.”
- **MCP (Model Context Protocol).** An open standard Anthropic launched in November 2024 for “connecting AI assistants to the systems where data lives.”¹⁰ Now hosted by the Linux Foundation, with more than 10,000 published MCP servers.¹¹ It decides whether an AI tool plugs into your stack

without custom work.

- **Gateway.** A single entry point between your apps and the AI models they use. It routes requests, holds the access keys, logs usage, and can cap spending. Also called an AI gateway or API gateway.
- **Cron.** A scheduler that runs a task at set times, such as every Monday at 8:00.
- **Kanban.** A board with columns, typically to do, doing, and done, where cards move as work progresses.

Where things live and run

The words your tech team uses about files, servers, and code.

- **Config (configuration).** The settings file that tells software how to behave. Small edits, big effects.
- **JSON.** A common text format for structured data, all curly braces and “key”: “value” pairs. How most APIs and AI tools exchange information.
- **YAML.** Another settings format, designed to be easier for humans to read. Indentation matters: one misplaced space can break it.
- **CLI (command-line interface).** Using software by typing commands instead of clicking. Many AI coding agents live here.
- **Bash.** The most common command language for that typing, standard on Linux servers. “Run this in bash” means “paste it into the terminal.”
- **TUI (text user interface).** A menu-style interface drawn in text inside the terminal.
- **Repo (repository).** A project folder tracked with its full history of changes.
- **Git.** The tool that records every change to a repo, who made it, and why, so anything can be undone. Linus Torvalds and the Linux community created it in 2005.¹²
- **GitHub.** A popular website for hosting Git repos and collaborating on them.
- **Open source.** Software whose code is public and can be used, changed, and shared under a license. Free to use doesn’t mean free of conditions: read the license.
- **Docker.** A tool that packages software with everything it needs into a “container,” so it runs the same on any machine.
- **VPS (virtual private server).** A rented slice of a server in a data center. A cheap home for small tools and agents that need to run around the clock.
- **SSH.** A secure way to log into and control a remote server. Someone asking for SSH access is asking for the keys to the machine.
- **Staging vs. production.** Staging is the private rehearsal copy; production is the live version customers see. Test AI features in staging first.

- **Rollback.** Undoing a release by restoring the previous version. Ask how it works before launch, not after.
- **CDN (content delivery network).** Servers around the world that keep copies of your site close to visitors, so pages load fast.
- **Static site.** A website served as ready-made files rather than assembled from a database on each visit. Fast, cheap, and low-maintenance.

Data, privacy, and trust

- **Personal data and GDPR.** See [Start here](#).
- **Consent and cookie consent.** GDPR consent must be “freely given, specific, informed and unambiguous.”¹³ AI personalization built on tracking data inherits whatever consent you did, or didn’t, collect.
- **Data minimization.** Collect and share only what a task needs. In AI terms: strip names and emails before you paste.
- **Data sovereignty.** Control over where data and AI run, and whose law applies. Data residency only covers where data is stored. Closed-model APIs process prompts in the provider’s cloud, under its terms and home-country law. Open-weight models can run on servers you control, so prompts stay in-house. Licenses vary: MIT for DeepSeek-R1,¹⁴ Apache 2.0 for Mistral Small 4,¹⁵ Meta’s own terms for Llama.¹⁶ For customer data, unreleased launches, M&A, or regulated claims, self-host or keep it out of AI. The catch: security, updates, and cost become yours. See [Open source](#) and our five-layer cake.
- **First-party data.** Data you collect directly from your own customers and channels: purchases, site behavior, email engagement. Your best AI fuel, and your responsibility.
- **Zero-party data.** Information customers deliberately tell you, such as preferences or quiz answers.
- **Bias.** Systematic skews a model picks up from its training data, such as defaulting to one gender or ethnicity in images. Check representation before anything goes public.
- **Deepfake.** Realistic AI-generated or altered video, audio, or images of real people. Under the EU AI Act, deepfakes must be clearly and visibly labeled, with the transparency rules applying from August 2026.¹⁷
- **Synthetic media.** The broader category: any AI-generated image, voice, video, or text. Not inherently deceptive, but disclosure expectations are rising.
- **C2PA.** A technical standard for attaching origin information to media, founded in 2021 by Adobe, Arm, BBC, Intel, Microsoft, and Truepic. Its aim: to let people “trace the origin and evolution of a piece of media.”¹⁸
- **Provenance.** The documented history of a piece of content: who made it, with what tools, and what changed.

Marketing-tech terms that collide with AI

- **GEO.** See [Start here](#).
- **SEO (search engine optimization).** Optimizing content to rank in search results. Still matters, because AI search tools often draw on the same web pages.
- **Zero-click search.** A search that ends without a visit to any website. Pew found Google users who saw an AI summary clicked a traditional result in 8% of visits, versus 15% for those who didn't.¹⁹
- **Schema (structured data).** Labels in a page's code that tell machines what things are: a product, a price, an event, a review. Helps search engines and AI read your pages correctly.
- **CMS (content management system).** The tool your team uses to publish web content, such as WordPress.
- **Headless.** A CMS or commerce system that stores content without dictating how it looks, so the same copy can feed a website, an app, or an AI assistant.
- **CRM (customer relationship management).** The system of record for customers and their interactions. AI features increasingly live inside it.
- **CDP (customer data platform).** Software that unifies customer data from many sources into single profiles for segmentation and personalization.
- **Attribution.** Working out which touchpoints deserve credit for a sale. Harder when the persuading happens inside an AI chat you can't see.

Commercial terms

- **SaaS (software as a service).** Software you rent by subscription and use online. Most AI tools are sold this way.
- **Seat-based pricing.** Paying per user, per month. Predictable, but you also pay for people who barely log in.
- **Usage-based pricing.** Paying for what you consume, often per token, image, or task. Fair, but bills can spike when an agent gets busy, so ask for caps.
- **Vendor lock-in.** When leaving a tool is costly because your data, workflows, or prompts only work there. Ask how to export everything before you sign.
- **Wrapper.** A product built on someone else's AI model, adding an interface, prompts, or a workflow, but not its own model. Not bad in itself; ask what it adds that the model alone doesn't, and what happens when the model or its price changes.
- **PoC (proof of concept).** A small, time-boxed test that shows whether an idea works before you commit budget. Agree on the success measure, the data, and the end date up front, so it ends with a clear yes or no.

- **DPA (data processing agreement).** The contract that sets out how a vendor handles personal data for you. Also ask whether your inputs are used to train their models.

The short version

Nobody needs all 79 terms on day one. Learn the top ten, bookmark the rest, and ask “What do you mean by that?” whenever a word goes past. Developers ask it all day, and asking keeps you in the room where decisions get made.

WHAT IT MEANS FOR YOU

- The top ten cover what AI is, how to steer it, how it fails, and where the legal lines are.
- Most tech jargon names ordinary ideas: settings, schedules, connections, and undo buttons.
- The commercial terms decide what an AI project really costs.

SOURCES

1. Why language models hallucinate (<https://openai.com/index/why-language-models-hallucinate/>), OpenAI, Sep 5, 2025
2. Understanding and counting tokens (<https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them>), OpenAI Help Center, accessed Oct 23, 2026
3. Retrieval-augmented generation for knowledge-intensive NLP tasks (arXiv:2005.11401) (<https://arxiv.org/abs/2005.11401>), Lewis et al., Facebook AI Research, 2020
4. GDPR Art. 5, Principles relating to processing of personal data (<https://gdpr-info.eu/art-5-gdpr/>), Regulation (EU) 2016/679, text via gdpr-info.eu, 2016
5. GDPR Art. 83, General conditions for imposing administrative fines (<https://gdpr-info.eu/art-83-gdpr/>), Regulation (EU) 2016/679, text via gdpr-info.eu, 2016
6. GEO: Generative engine optimization (arXiv:2311.09735, KDD 2024) (<https://arxiv.org/abs/2311.09735>), Aggarwal et al., 2024
7. Markdown (<https://daringfireball.net/projects/markdown/>), John Gruber, Daring Fireball, Dec 17, 2004
8. Introducing Agent Skills (<https://www.anthropic.com/news/skills>), Anthropic, updated Dec 18, 2025
9. The anatomy of an agent harness (<https://www.langchain.com/blog/the-anatomy-of-an-agent-harness>), Vivek Trivedy, LangChain, Mar 10, 2026
10. Introducing the Model Context Protocol (<https://www.anthropic.com/news/model-context-protocol>), Anthropic, Nov 25, 2024
11. Linux Foundation announces the formation of the Agentic AI Foundation (<https://www.linuxfoundation.org/press/linux-foundation-announces-the-formation-of-the-agentic-ai-foundation>), Linux Foundation, Dec 9, 2025
12. A short history of Git (Pro Git) (<https://git-scm.com/book/en/v2/Getting-Started-A-Short-History-of-Git>), git-scm.com, 2nd ed.
13. GDPR Art. 4, Definitions (<https://gdpr-info.eu/art-4-gdpr/>), Regulation (EU) 2016/679, text via gdpr-info.eu, 2016
14. DeepSeek-R1 model card, section 7: License (<https://huggingface.co/deepseek-ai/DeepSeek-R1>), DeepSeek, via Hugging Face, accessed Oct 23, 2026
15. Models Overview (<https://docs.mistral.ai/models>), Mistral AI, Mistral Docs, accessed Oct 23, 2026

16. Llama 4 Community License Agreement (<https://www.llama.com/llama4/license/>), Meta, effective Apr 5, 2025
 17. AI Act: Regulatory framework for AI (<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>), European Commission, accessed Oct 22, 2026
 18. C2PA founding press release (<https://c2pa.org/c2pa-founding-press-release/>), Coalition for Content Provenance and Authenticity, Feb 22, 2021
 19. Google users are less likely to click on links when an AI summary appears in the results (<https://www.pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appears-in-the-results/>), Pew Research Center, Jul 22, 2025
-