GPT-5 EDITION

HUMANITY'S LAST PROMPT ENGINEERING GUIDE

by Matthew Berman
& Nick Wentz

FORWARDFUTURE

Humanity’s Last Prompt Engineering Guide: Built for the GPT-5 Era

Most guides on prompt engineering either overcomplicate the basics or give you vague tips that don’t hold up when you actually try to use them.

This one’s different.

We've done the homework for you by gathering best practices from OpenAI, (including early GPT-5 research), Google, Anthropic, and real-world testing, and broken it all down into something you can actually use.

Whether you’re in sales, operations, marketing, or leadership, this guide will show you how to get dramatically better results from GPT-5, Claude, Gemini, and other chat-based AI tools.

How to Use This Guide

This guide is designed to help you become dramatically more effective at using AI tools like ChatGPT, Claude, and Gemini, whether you're just getting started or already using them in your daily work.

Here's what to expect:

Structured for Action

- **Section 1** introduces how prompting works and why it matters.
- **Section 2** anatomy of a Perfect Prompt (Inspired by Greg Brockman, OpenAI President)
- **Section 3** explains what makes a prompt work and how models think.
- **Section 4** helps you diagnose and fix bad prompts fast.
- **Section 5** breaks down 11 foundational prompting techniques with real examples.
- **Section 6** gives ready-to-use templates for common business roles.
- **Section 7** includes a scorecard and worksheet to refine and evaluate your prompts.
- **Section 8** offers a glossary of key terms for beginners and pros alike.

How to Use It

- **Skim the techniques** — then go deep on the ones most relevant to your workflows.
- **Use the diagnostic section (Section 4)** any time the AI gives poor or vague results.
- **Return to the scorecard** when refining a prompt for better consistency or results.

What You'll Learn

- How prompts actually work behind the scenes — and how GPT-5 interprets them differently
- Proven techniques to improve accuracy, creativity, and cross-modal reasoning
- Ready-to-use templates for business, design, and developer workflows
- Common mistakes to avoid — especially with GPT-5's increased sensitivity to instruction

Section 1: Introduction to Prompt Engineering

Although prompting may seem like trickery, it's actually a method of teaching AI to think clearly. Large language models (LLMs) like ChatGPT, Claude, and Gemini are fundamentally prediction engines. Given an input (your prompt), they generate the most likely next word based on patterns learned from massive amounts of text — and with GPT-5, they can even allocate more or less reasoning power depending on the task. That makes your prompt the blueprint.

A vague request will produce a vague output, but a clear and well-structured prompt is how you get useful, specific, and surprisingly profound responses.

Why Prompt Engineering Matters

It doesn't require coding, ML knowledge, or advanced technical skills. What it does require is the ability to be clear, specific, and intentional with your inputs.

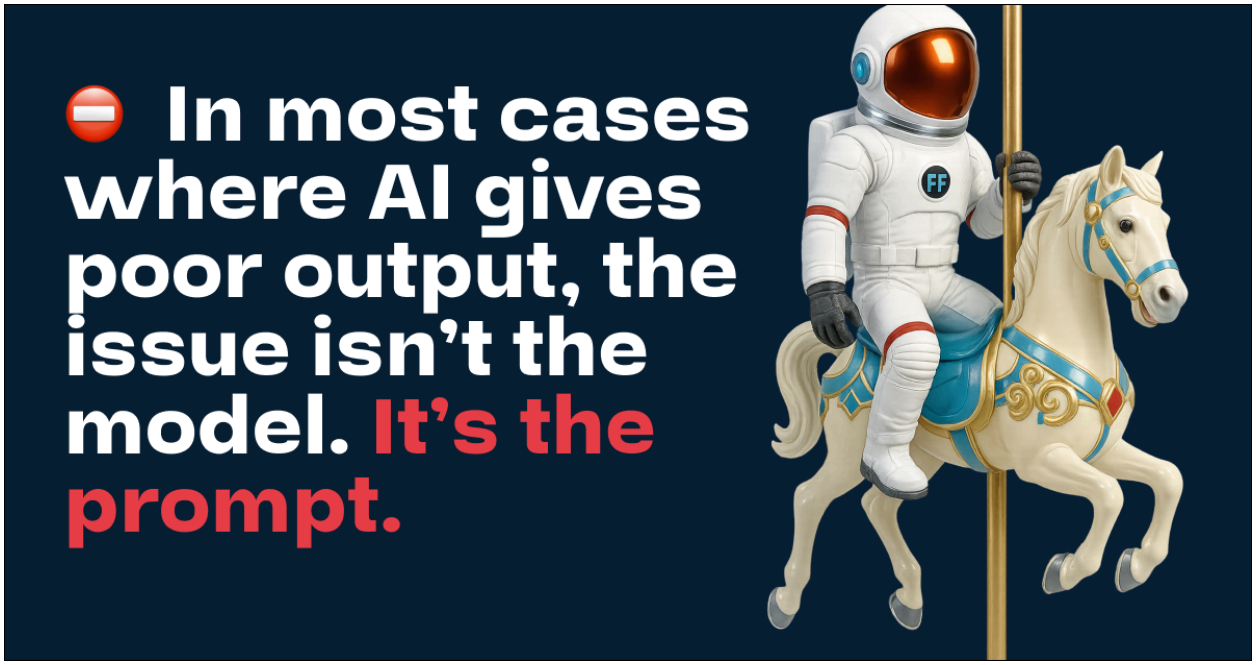
A good prompt tells the model: who it is ("You are a product strategist…"), what to do ("Summarize this in 3 bullet points."), what it's working with (Input: text, table, scenario, etc.), and how to respond (e.g., bullet list, JSON, tone, word count).

With GPT-5, you can even guide how deeply it thinks, toggling between a quick reply and a more deliberate response. Once you learn how to structure that input (and iterate when things go sideways) you can summarize complex documents in seconds, brainstorm new product ideas or messaging, analyze data or codebases, extract patterns from long inputs, generate insights, role-play expert conversations, and automate repetitive writing and analysis tasks.

Prompting Is a Leverage Skill

The better you prompt, the more productive and valuable AI becomes in your workflows.

You don't need to become a "prompt hacker." You just need to learn how to speak the model's language. As models like GPT-5 evolve with memory, multimodal



understanding, and increased sensitivity to instructions, prompt engineering becomes even more powerful.

Section 2: Anatomy of a Perfect Prompt

The president of OpenAI, Greg Brockman, shared a guide to the ideal prompt for their o1 reasoning model, which has now been implemented into GPT5. Although his guide was specifically geared toward o1, the truth is that the prompting structure he shared can be applied (with a few model-specific tweaks) to virtually every model currently available to users.

This is especially true if you're trying to achieve the coveted **one-shot:** a prompt so perfectly written that it inspires the AI to generate exactly what you're looking for the first time, with no follow-up prompts needed.

Brockman's guide consists of four parts:

Goal

Start your prompt by stating your end goal as clearly as possible.

The example provided in Brockman's prompt was *"I want a list of the best medium-length hikes within two hours of San Francisco. Each hike should provide a cool and unique adventure, and be lesser known."*

Return Format

The next part of your prompt should mention how you would like the model to respond.

Continuing with Brockman's example, this section of his prompt reads *"For each hike, return the name of the hike as I'd find it on AllTrails, then provide the starting address of the hike, the ending address of the hike, distance, drive time, hike duration, and what makes it a cool and unique adventure."*

Warnings

This part of your prompt should mention any potential room for error in the model's output. Consider the worst possible outcome for your intended goal— in most cases that involve

real-world information, this means the possibility that the model hallucinates, or generates realistic-sounding information that isn't actually accurate.

Brockman's prompt has this: *"Be careful to make sure that the name of the trail is correct, that it actually exists, and that the time is correct."*

> 👉 **Tip:** Using phrases like "Think hard" and "Be careful" in this part of your prompt helps the model to realize that these instructions are especially important. If there's any room for creativity or outside-the-box answers in your prompt, use this wording sparingly to help the model recognize that this is not where they will find it.

Context

So far, our example prompt has answered four of the six most important questions you can ask: **Who, What, When, Where, Why,** and **How.**

- **What:** A list of medium-length hiking trails that provide cool/unique adventures.
- **Where:** Two hours or less travel time from San Francisco.
- **How:** A list of starting locations, ending locations, and explanation of what makes each trail unique.
- **When:** As soon as the user presses Enter to send the prompt (this one is implied in every LLM prompt unless you specifically mention a different timeframe).

What about the two that are missing— **Who** and **Why?** That's what this last section is for. Without knowing a bit more about the person submitting the prompt, you can't expect the model to know exactly what they mean by a 'unique' adventure, or a 'medium-length' hike. After all, these tools aren't mind-readers.

Here's the context dump that Brockman included in his example prompt:

> *"For context: my girlfriend and I hike a ton! We've done pretty much all of the local SF hikes, whether that's Presidio or Golden Gate Park. We definitely want to get out of town — we did Mount Tam pretty recently, the whole thing from the beginning of the stairs to Stinson — it was really long and we are definitely in the mood for something different this weekend! Ocean views would still be nice. We love delicious food. One thing I loved about the Mt. Tam hike is that it ends with a celebration (arriving in town to breakfast!) The old missile silos and stuff near Discovery Point is cool but I've done that hike probably 20x at this point. We won't be seeing each other for a few weeks (she has to stay in LA for work) so the uniqueness here really counts."*

Now the LLM has an answer for the **Who** and **Why**— and an even more specific answer to **When:**

- **Who:** A couple who are avid hikers that have already experienced Presidio, Golden Gate Park, the missile silos near Discovery Point, and Mt. Tam.
- **Why:** Looking for a unique adventure to go on together before being separated for a few weeks due to work-related travel plans.
- **When:** This upcoming weekend.

Notice that this context dump also clarifies the already established **Where** and **What:**

- **Where:** A hike that might have an ocean view (and a restaurant at the end) that isn't Mount Tam, Golden Gate Park, Presidio, or Discovery Point
- **What:** A hike that is a shorter distance than the length between the beginning of the Mount Tam stairs to Stinson

Your context can be as long or as short as you want, and it can be as structured or as loosely organized as you desire. Try to be as authentic as you can when writing this section, and include details that you may not even think would be important— anything you include can help the model get a more specific picture of what you're looking for.

Section 3: What Makes a Prompt Work?

Think of it like giving a highly capable assistant a task. If you don’t explain what you want, how you want it, and why it matters, the results may be vague, verbose, or just flat-out wrong.

Prompts are affected by more than just what you type. These factors all play a role:

Factor	Why It Matters
Model	Each LLM has unique strengths, capabilities, and quirks
Context	Input quality (documents, examples) impacts reasoning and accuracy
Structure	Clear formatting improves output consistency and usefulness
Style + Tone	You can control formality, voice, or persona
Model Settings	Output length, temperature, and sampling all influence creativity vs. precision

For the vast majority of use cases, prompting is not a one-and-done process. It's better to think of it as an ongoing conversation-- one you'll often need to test, tweak, and refine to get the best results. You’ll often need to test, tweak, and refine to get the best results.

The Truth About Artificial Intelligence (What You Can't Control)

When LLMs are referred to as AI, the word 'artificial' is equally as important as the word 'intelligent.' That's because instead of actually thinking like a human brain, LLMs function as complicated prediction engines that generate the most likely series of tokens, (words in a sentence or lines of code) based on the sequence of tokens you provide in your prompt, informed by all of the tokens the model has seen in the training process.

But GPT-5 goes further. While it's still based on token prediction, it uses **adaptive compute**: allocating more resources for complex reasoning tasks, and less for simpler ones. That means your prompt isn't just steering the content — it's influencing how hard the model works to deliver it.

Even small changes in phrasing or structure can radically shift the output. The clearer and more intentional you are, the more useful the results.

LLM Output Configuration (What You Can Control)

Most AI platforms (like ChatGPT, Claude, Gemini) let you adjust **settings** that affect how responses are generated:

Setting	What It Does
Temperature	Controls randomness. Lower = more focused, higher = more creative.
Max tokens	Caps the length of the response. Prevents run-ons
Top-p / Top-k	Controls which words the model can choose from when generating responses.

Section 4: How To Fix Bad Prompts

When your prompts don't provide the results you're looking for, it may be because of a small mistake. Look out for these common mistakes; even a single sentence can make the difference between an effective and ineffective prompt.

Problem	❌ **Weak Prompt**	✅ **Improved Prompt**
Too vague	Write a summary.	Summarize the article below in 3 bullet points. Focus on key findings, avoid repeating the introduction.
No audience specified	Rewrite this for clarity.	Rewrite this for a busy executive audience. Use short sentences and strip out nonessential background.
Missing role/context	Help me with this draft.	You are a brand copywriter. Improve the tone of this draft to make it more confident and modern.

No format instruction	What's a good alternative?	Suggest 3 alternatives in a numbered list. Include 1–2 sentence explanations for each.
No reasoning requested	What's the best option here?	Evaluate these 3 options. List pros and cons for each, then recommend one with a short rationale.
Unclear intent	Turn this into a message.	Convert the following into a 3-line Slack message that is clear, friendly, and includes a call to action.
Lacks structure	Help me improve this.	Rewrite this performance review to follow this structure: achievements, challenges, and next steps.
Doesn't guide response depth	Compare these tools.	Think through these tools one by one. Evaluate tradeoffs, and recommend a final choice with justification.

Use this as a quick diagnostic when a model gives poor output:

1. **Am I being too vague?**
   Be specific about the task and expectations.

2. **Did I include a role or point of view?**
   Adding “You are a…” sets the tone and mindset. Learn more about this fascinating prompting technique in Sections 5 and 6.

3. **Is the input complete and relevant?**
   Include all necessary information for the model to reason effectively.

4. **Have I requested a clear format?**
   Specify if you want bullets, a paragraph, JSON, etc.

5. **Am I asking for reasoning?**
   If judgment is involved, ask the model to “think step by step” or explain its logic.

6. **Have I broken the task into smaller parts if needed?**
   Split complex requests into multiple, focused steps.

7. **Could I include examples or longer input context?**
   GPT-5 supports massive context windows — entire documents, transcripts, or long examples — which you can now use to guide the output more effectively.

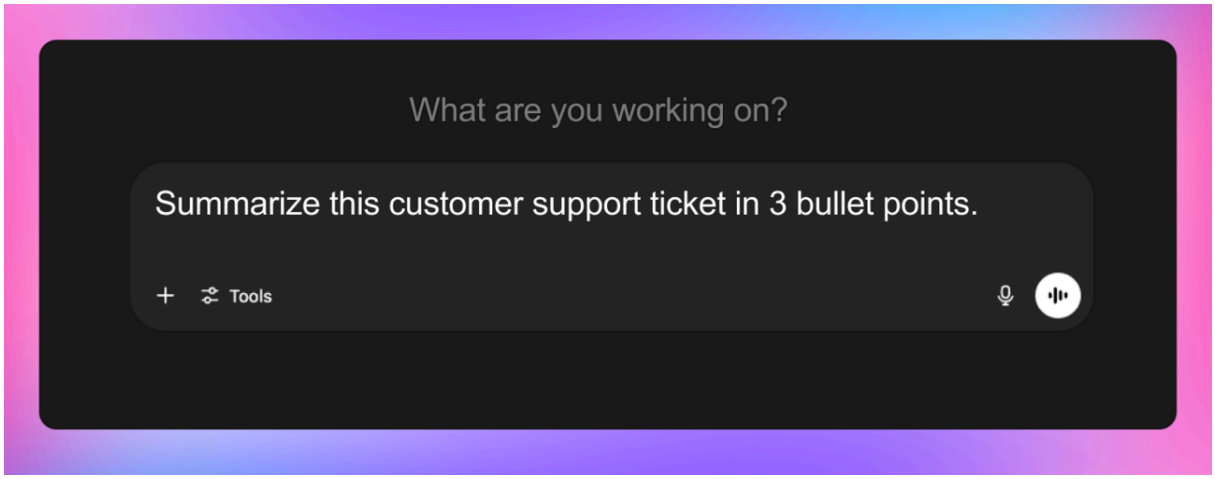
Section 5: Prompting Techniques

Prompting is a skillset— and with GPT-5's improved instruction-following, deeper reasoning, and multimodal capabilities, it's more powerful than ever. When using modern LLMs, including GPT5, there are more than one way you can structure your prompt. This is because different problems require different solutions, and different models find these solutions in different ways. That's why one of the most powerful techniques you can adopt to fully harness modern AI is engineering different prompts. These eleven foundational techniques will help you structure clearer instructions, improve model performance, and adapt to a wide range of business, design, and technical workflows.

Each technique below includes a description, when to use it, and a simple example.

1. Zero-Shot Prompting

- **What it is:** A single instruction without any examples.
- **When to use it:** For simple or well-understood tasks where context is obvious.
- **Why it works:** The training data used in GPT-5 and other modern LLMs is surprisingly versatile, so you can reasonably expect it to understand what you mean when you refer to common terms like an "elevator pitch" or "TL;DR".



More Examples:

- *Summarize the following transcript into a one-paragraph executive briefing suitable for a board meeting.*
- *Generate 3 headline options for a homepage that emphasizes speed, simplicity, and trust.*
- *Based on the meeting notes below, write an email recap with next steps and responsible owners.*
- *Extract and rank the top 3 objections raised in this sales call transcript.*

Tips:

🟢 **Beginner**: Start with simple, clear instructions using action verbs like "Summarize," "List," or "Write." Avoid vague words like "help" or "explain."

🔵 **Intermediate**: Add a format instruction (e.g., "in 3 bullet points" or "as a table") to shape the output and reduce ambiguity.

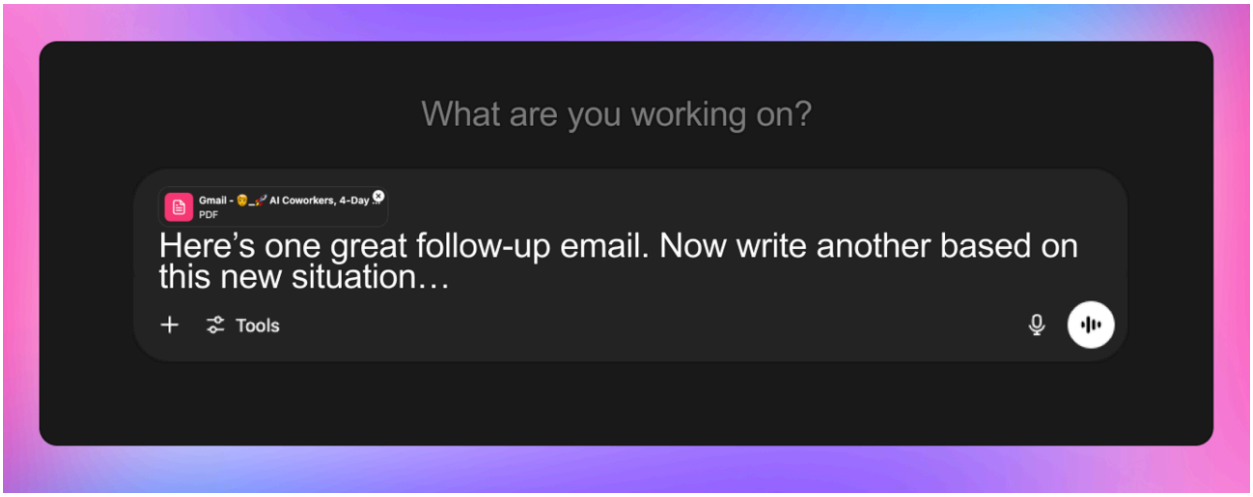
⚫ **Advanced**: Combine multiple directives in a single sentence to pack precision (e.g., "Summarize in 3 bullets, each under 10 words, focusing on customer pain points.")

👉 **GPT-5 Tip:** Even though GPT-5 is better at guessing intent, clear constraints (audience, format, word count) still dramatically improve results.

2. Few-Shot Prompting

- **What it is:** You provide one or more examples to guide the model.
- **When to use it:** When the task involves structure, tone, or formatting that the model might not infer on its own.

- **Why it works:** Providing an example that is nearly identical to your desired output mimics GPT-5's training process— in addition to any other LLM that uses a similar setup for training their transformers.



More Examples:

- *Example: I want a large pizza with mushrooms and basil. JSON: {"size": "large", "toppings": ["mushrooms", "basil"]}. Now convert this: I'd like a small pizza with olives and tomato.*
- *Here is an ideal product changelog entry: Improved onboarding flow: Faster load times, clearer CTAs, and better mobile UX. Now write one for the feature update described below.*
- *Here are two customer support responses that defuse frustration with empathy and clarity. Write a third using the same tone and structure.*
- *Based on these examples of investor update summaries, write one for this month's company performance.*
- *Here are a few ideal Slack status messages for when someone is heads-down. Create 5 more in the same style.*

Tips:

🟢 **Beginner:** Copy the structure of one example and change the inputs. This helps you build pattern recognition.

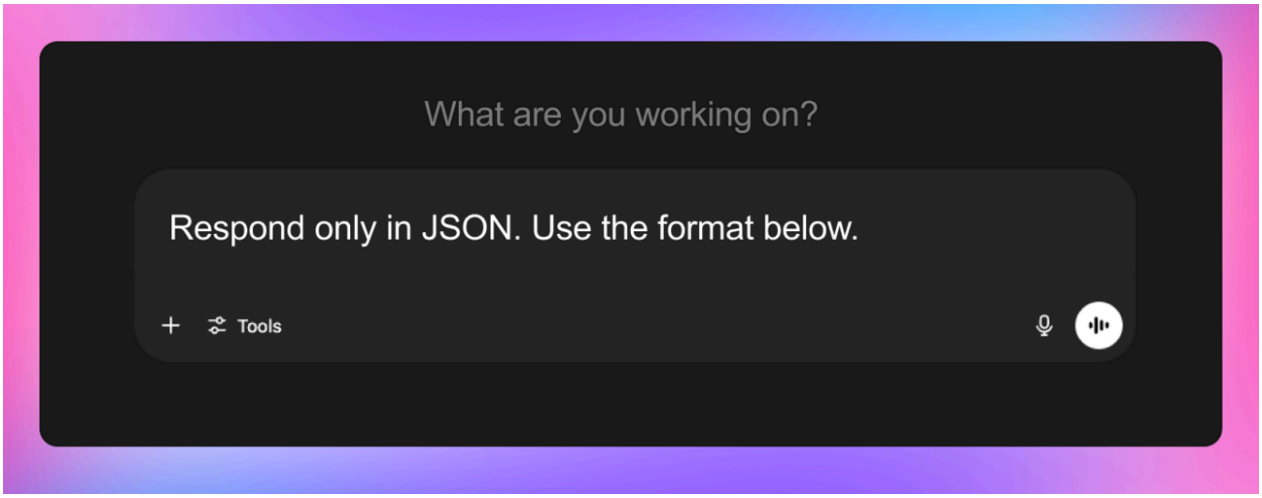
🔵 **Intermediate:** Provide a range of examples (short/long, formal/casual) to help the model generalize the style or logic.

⚫ **Advanced:** Fine-tune examples to emphasize edge cases, errors, or style nuances. Use deliberately contrasting examples to teach distinctions. Although it will likely be a while before you'll be able to fine-tune GPT5, this technique will work with several of OpenAI's other models, such as GPT-4.1

👉 **GPT-5 Tip:** When showing examples, use a mix of styles to train the model's judgment. GPT-5 excels at picking up tonal nuances.

3. System Prompting

- **What it is:** Set clear rules for how the model should behave or structure output. For comparison, refer to the part of Section 3 that discusses Return Format.
- **When to use it:** To control tone, format, length, or behavior from the beginning.
- **Why this works:** LLMs are designed to generate the most likely sequence of words that follow your input; by specifying that you need a particular structure, it prevents the model from resorting to a generic, lowest-common-denominator result to your question.



More Examples:

- *Always answer using the STAR method (Situation, Task, Action, Result). Respond only in bullet points.*
- *Only use language that a 10th-grade reader can understand. Avoid technical terms and acronyms.*
- *Provide your answer in Markdown format, using headings, subheadings, and bolded key terms.*
- *Every time you respond, include a short summary sentence first, followed by your full explanation.*

Tips:

🟢 **Beginner:** Use simple system rules like “Respond in a friendly tone” or “Answer in bullet points.”

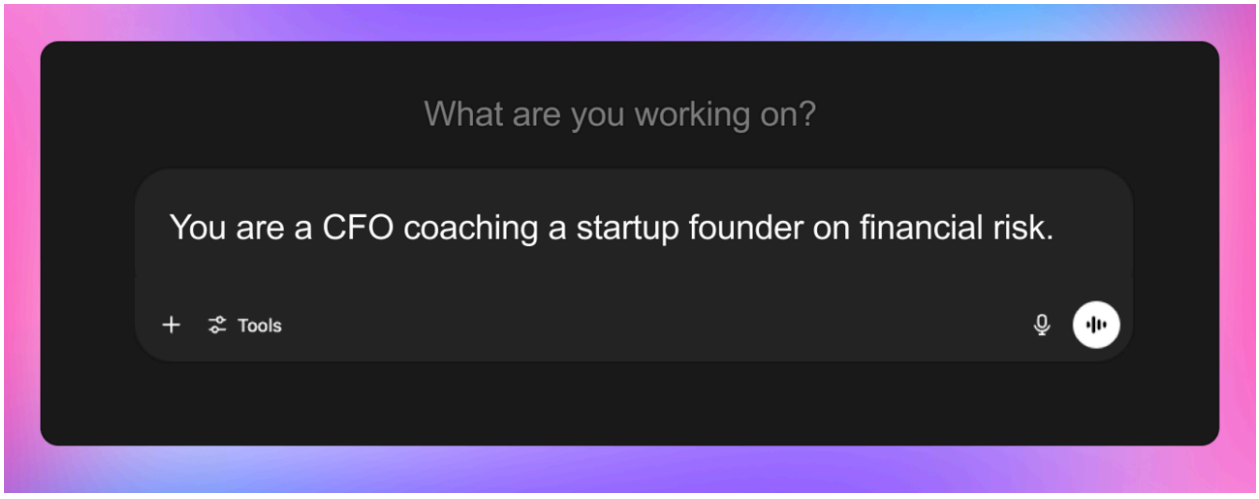
🔵 **Intermediate:** Layer constraints like tone, format, and audience all in one system prompt.

⚫ **Advanced:** Use system prompts to enforce behavior across multiple turns in a chat (e.g., always speak as a mentor, always return Markdown with code blocks, etc.).

👉 **GPT-5 Tip:** Use longer system prompts to establish behavior across longer chats. The model now holds onto rules more reliably.

4. Role Prompting

- **What it is:** Assign a persona or point of view.
- **When to use it:** When the model needs to adopt a specific tone, expertise level, or perspective.
- **Why it works:** In the training data used to build GPT-5 and other LLMs, actual text from subject matter experts were incorporated. Asking the model to speak like these SMEs prompts it to draw from a specific pool of knowledge in their training data, resulting in a similar effect to the previous technique. For more insight into this prompting technique, see the next section of this guide.



More Examples:

- *You are a financial advisor. Recommend three strategies for someone with $10,000 to invest and a five-year time horizon.*
- *You are a senior PMM at a B2B SaaS company. Write a 3-sentence value proposition for a new AI workflow feature.*
- *You are a startup CEO preparing for a pre-seed pitch. Reframe this vision statement to be more investor-friendly.*
- *You are a customer success manager. Draft a proactive check-in email for an enterprise client showing early signs of churn.*
- *You are a technical recruiter. Review the résumé snippet below and write a 2-line summary of the candidate's profile.*

Tips:

🟢 **Beginner:** Start with "You are a…" followed by a familiar job title (e.g., "You are a content strategist").

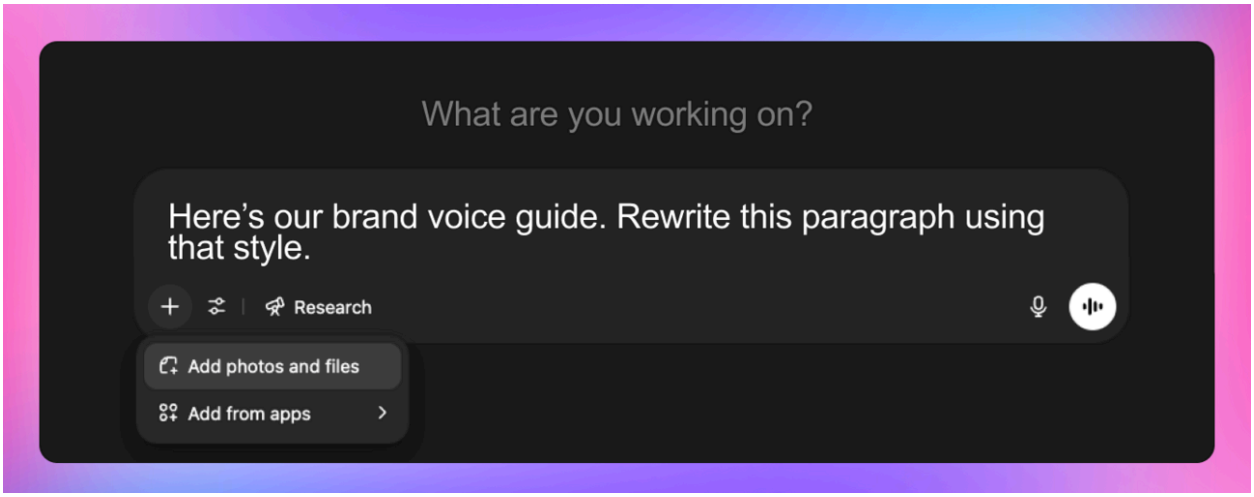
🔵 **Intermediate:** Match the role to the task. Use domain-specific roles (e.g., "You are a data privacy consultant") to shape tone and accuracy.

⚫ **Advanced:** Combine roles with context and format constraints for nuanced control (e.g., "You are a veteran PMM writing a short memo for legal approval").

👉 **GPT-5 Tip:** Role specificity matters more — it adjusts tone and depth more accurately than earlier versions.

5. Contextual Prompting

- **What it is:** Provide relevant background, data, or a scenario.
- **When to use it:** When task success depends on domain-specific context or past conversation history.
- **Why this works:** Recall in Section 3 of this guide when OpenAI's president shared a prompt that contained a "context dump"; this technique results in sharing information that you may not consciously realize is important for the model to understand in order to produce your ideal input.



More Examples:

- *You are writing for a retro gaming blog. Suggest three article ideas based on 1980s arcade culture.*
- *Using the customer persona and product description below, write a 2-sentence ad hook that appeals to first-time users.*
- *Here's a summary of our brand's tone guidelines. Rewrite the following help center article to match that tone.*
- *Given our revenue goals and Q3 priorities, generate 3 OKR options for the sales enablement team.*

- *Review this engineering roadmap and suggest a revised timeline, given the team's current velocity and bandwidth constraints.*

Tips:

🟢 **Beginner:** Copy and paste a relevant doc or snippet before the prompt to give the model something to work with.

🔵 **Intermediate:** Use summaries instead of raw data when the input is long. Pre-frame: "Here's what you need to know…"

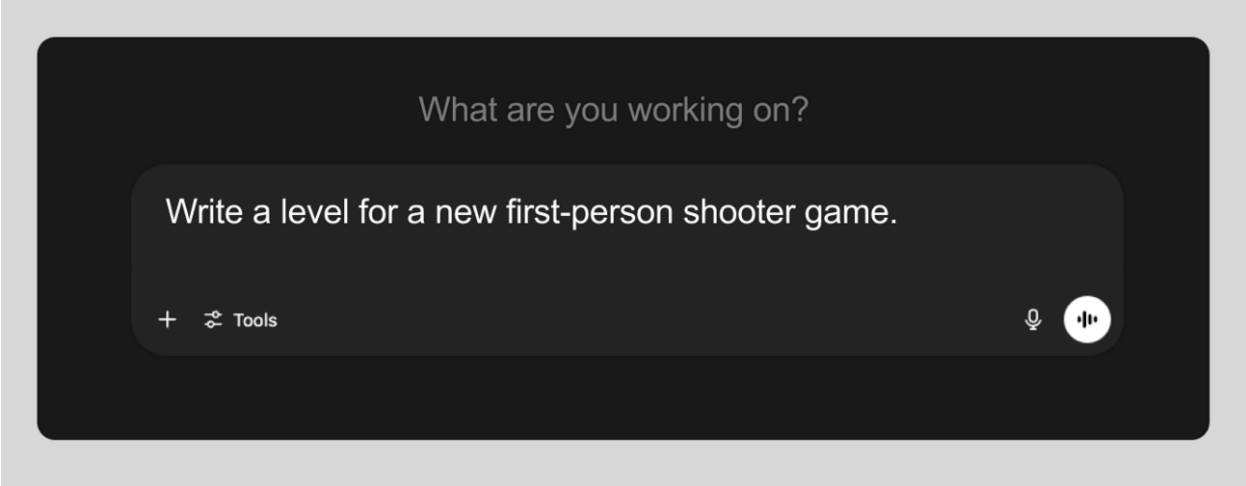
⚫ **Advanced:** Chain multiple inputs (e.g., persona + doc + goals), and explicitly state how they relate (e.g., "Given this persona and these product notes…").

👉 **GPT-5 Tip:** Don't hesitate to paste full transcripts, briefs, or docs — GPT-5 handles long inputs without losing structure.

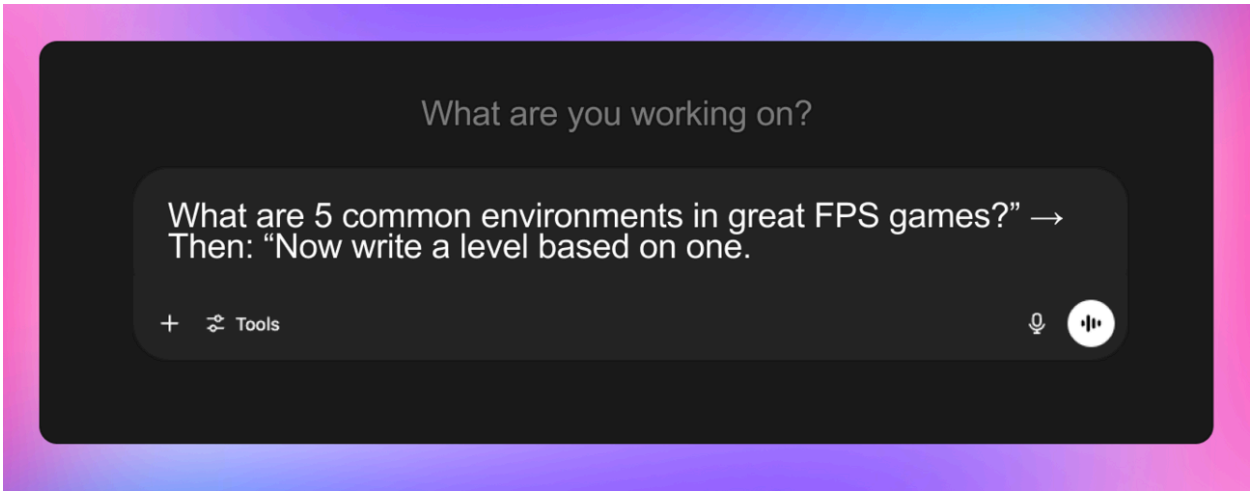
6. Step-Back Prompting

- **What it is:** Ask the model to solve a general question first, then apply that insight to a specific task.
- **When to use it:** For creative tasks, complex reasoning, or scenario planning.
- **Why it works:** GPT-5 and other modern LLMs have reasoning features built in that you can trigger by asking for an argument derived from first principles.

Prompt Before:



Step-back Version:



More Examples:

- *Before writing the email, list 3 things the recipient likely values based on their role and company. Then write the email using that context.*
- *First, outline the characteristics of a high-performing user onboarding flow. Then critique the one below using that framework.*
- *Before drafting product positioning, summarize how our top 3 competitors describe themselves. Then write our version to differentiate.*
- *Start by identifying the key emotion this campaign should trigger in customers. Then write a subject line that evokes it.*

Tips:

🟢 **Beginner:** Ask a warm-up question before the real task (e.g., “What does the audience care about?” → “Now write the email”).

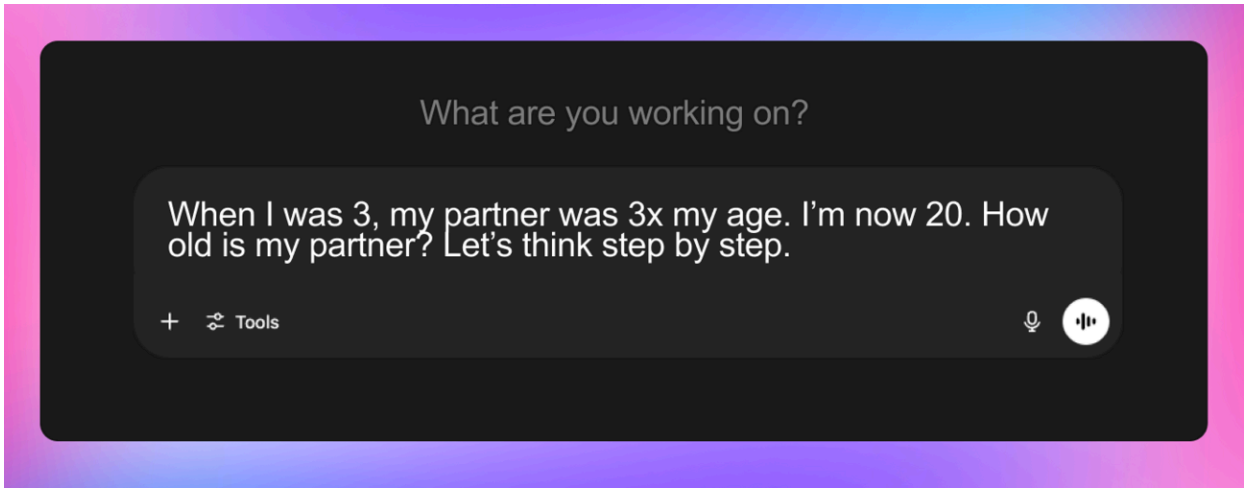
🔵 **Intermediate:** Use it to deconstruct complex decisions or plans (e.g., goals → actions → message).

⚫ **Advanced:** Use it for self-discovery or creative ideation chains. Ask the model to generate principles first, then apply them to the prompt task.

👉 **GPT-5 Tip:** GPT-5 is especially good at principle-first reasoning — encourage more use of “generate criteria → apply criteria.”

7. Chain-of-Thought Prompting (CoT)

- **What it is:** Instruct the model to show its reasoning step by step.
- **When to use it:** For math, logic, planning, or anything that benefits from transparent reasoning.
- **Why it works:** Truthfully, this prompting technique isn't necessary for reasoning models such as GPT-5; however, older models such as GPT-3.5 greatly benefit from instructions that demand logical reasoning.



More Examples:

- *Walk through the steps needed to calculate customer acquisition cost (CAC) using the data provided. Then provide the result.*
- *Given this list of product issues, identify which could cause the highest drop in NPS, and explain your reasoning.*
- *List the steps a support agent should take to de-escalate the situation described in the chat log.*
- *Break down the decision-making process for whether to renew or sunset the feature described. Include cost, usage, and team feedback.*

Tips:

🟢 **Beginner:** Add "Explain your reasoning" or "Think step by step" to the end of your prompt.

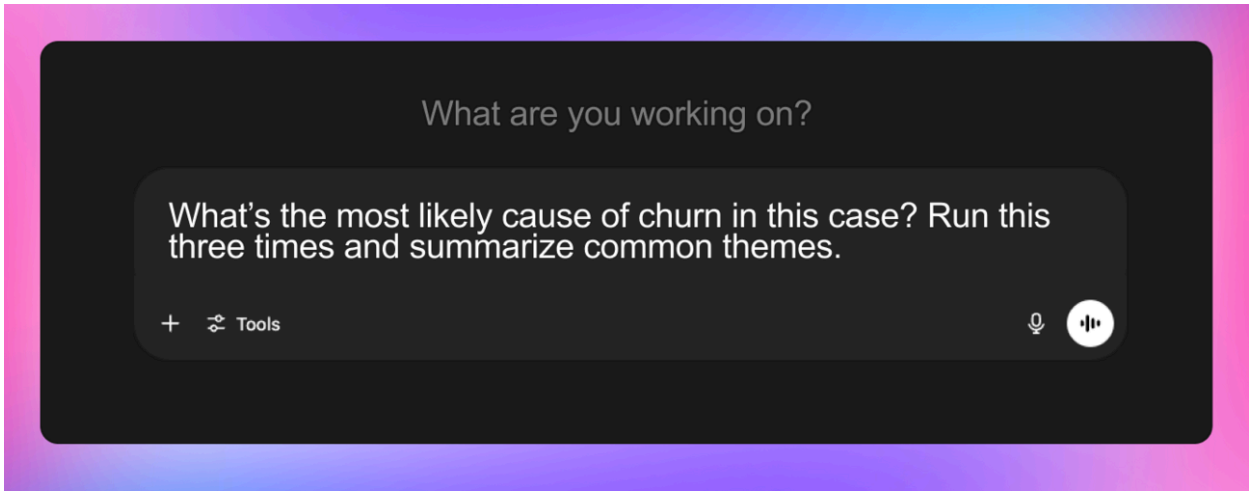
🔵 **Intermediate:** Ask the model to list pros and cons, compare options, or walk through logic to justify a recommendation.

⚫ **Advanced:** Combine with structured outputs (e.g., tables of tradeoffs, multi-step formulas) or nested reasoning paths.

👉 **GPT-5 Tip:** Chain-of-thought is built in — but you can deepen output quality by explicitly asking it to "reflect," "justify," or "compare."

8. Self-Consistency

- **What it is:** Run the same prompt multiple times and choose the most frequent or consistent result.
- **When to use it:** For high-stakes or ambiguous tasks where accuracy matters.
- **Why this works:** LLMs like GPT-5 use probability to guess the most likely token to follow the sequence of tokens given as input; by running the same prompt multiple times with a higher temperature, you force the model to access a greater distribution of likely outputs, which can help you identify commonalities and consistent truths.
- **Note:** Requires a model that supports variability via temperature (e.g. 0.7 or higher).



More Examples:

- *Ask a model to classify an ambiguous email as "important" or "not important" multiple times, then take the majority vote.*
- *Classify this review as positive, neutral, or negative five times and return the most frequent label with reasoning.*
- *Generate three different explanations for this customer churn event. Then summarize the one that appears most consistently.*
- *Write five taglines for this product feature. Then identify the most recurring theme or hook among them.*
- *Answer this ambiguous question three times, then decide which version is most likely to align with the user's intent.*
- *Simulate ten different users of this product and write how each one is likely to feel about their overall experience*

Tips:

🟢 **Beginner:** Run the same prompt a few times and compare results. Choose the best version manually.

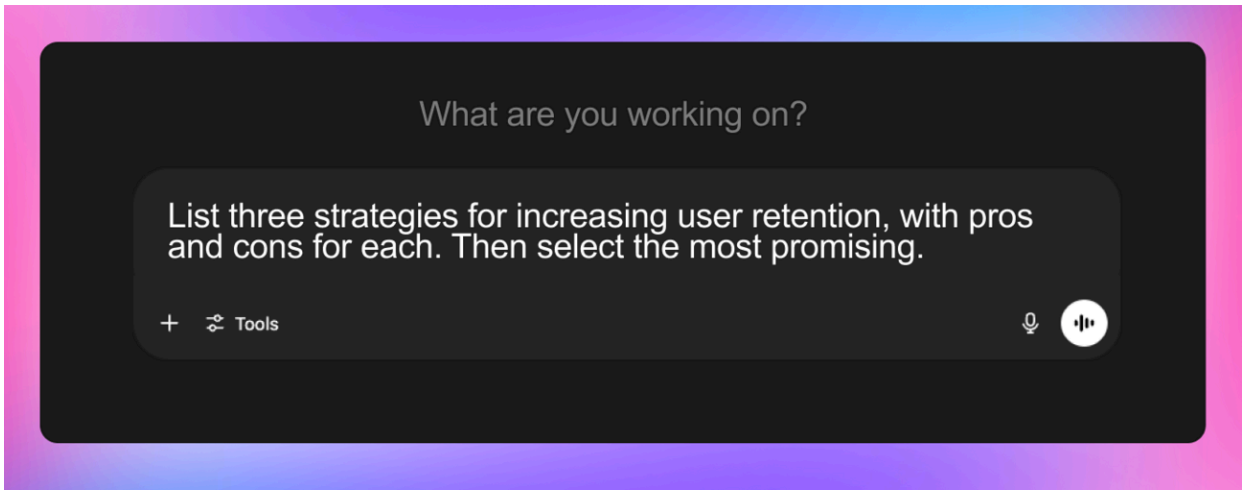
🔵 **Intermediate:** Ask the model to repeat a task 3–5 times and select the most common or consistent answer.

⚫ **Advanced:** Instruct the model to generate multiple answers, evaluate them for consistency or correctness, and select the best one.

👉 **GPT-5 Tip:** GPT-5's lower variance may reduce the need for this trick — but it still helps for creative or ambiguous prompts.

9. Tree of Thoughts (ToT)

- **What it is:** Explore multiple reasoning paths instead of a single linear answer.
- **When to use it:** For brainstorming, decision-making, or tasks with multiple valid outcomes.
- **Why it works:** GPT-5 is designed to swap between different models based on different prompts; phrasing your prompt in this way incentivizes GPT-5 and others like it to access more of these specialized submodels when generating your response.
- **Note:** Often used with agents or frameworks that track branching logic.



More Examples:

- *Explore three different GTM strategies for launching this product in Q4. Then evaluate their tradeoffs and recommend one.*
- *Generate 3 unique frameworks for organizing this internal knowledge base. Justify each approach and select the most scalable one.*
- *Outline 3 alternative messaging directions for this new feature. Then score them for clarity, originality, and alignment with brand tone.*
- *Imagine three distinct customer objections during the onboarding process. For each, propose a mitigation strategy.*

Tips:

🟢 **Beginner:** Ask the model for "3 different ideas" before choosing one.

🔵 **Intermediate:** Prompt for multiple options, then ask the model to evaluate them for clarity, impact, or feasibility.

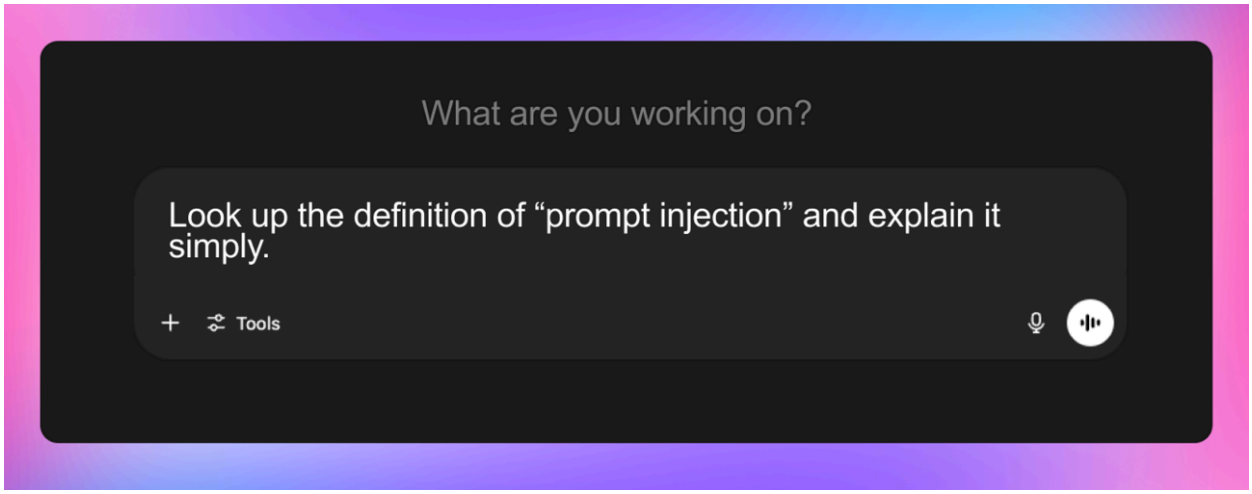
⚫ **Advanced:** Use ToT as a decision-making engine — branch multiple reasoning paths, evaluate tradeoffs, then synthesize the best choice.

👉 **GPT-5 Tip:** When using ToT, ask GPT-5 to score each path on criteria like clarity, feasibility, or originality.

10. ReAct (Reason + Act)

What it is: The model reasons step-by-step and performs actions, such as searching the web or calling a tool.

When to use it: For tool-augmented agents or systems that require external data retrieval or interaction.



More Examples:

- *Search how many children each member of Metallica has. Sum the total. Output only the final number.*
- *Search for this company's most recent funding round. Then write a congratulatory email that references the announcement.*
- *Use a code interpreter to calculate the monthly burn rate from this spreadsheet. Then write a plain-English summary.*
- *Search for the latest update to the GDPR regulation. Then suggest one change to our current data policy.*
- *Find the Glassdoor average rating for this company. Based on the result, write an outreach message addressing potential concerns.*

Tips:

🟢 **Beginner:** Use tools like “Search” or “Calculate” and then manually combine results with AI output.

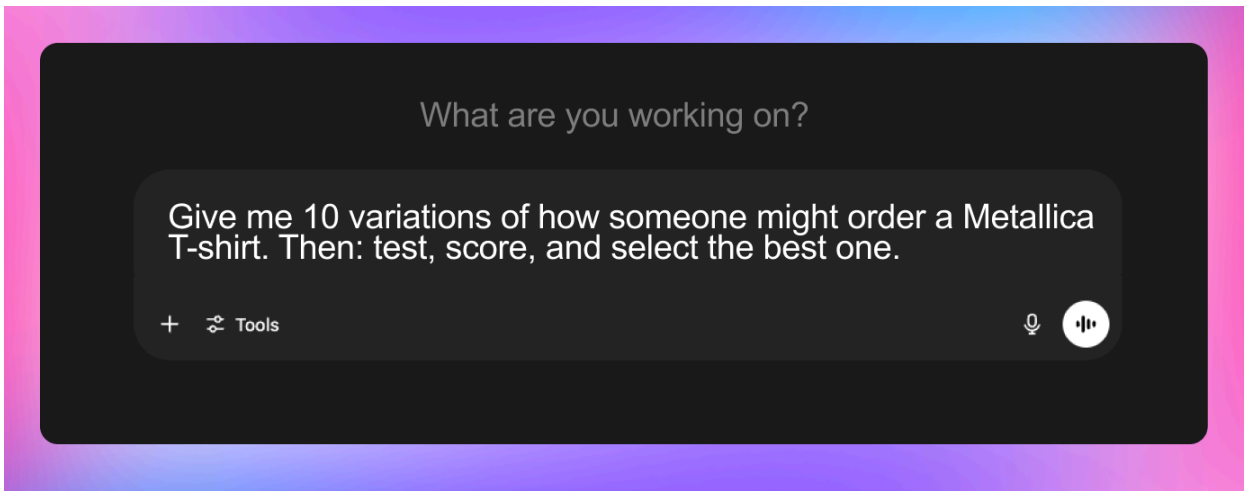
🔵 **Intermediate:** Use prompts like “Search for X, then summarize it in 2 lines” to simulate a chain of thought + action.

⚫ **Advanced:** Combine tool use, reasoning, and generation in one loop (e.g., "Search X → Compare Y → Generate Z"). Use agents or code interpreter if available.

👉 **GPT-5 Tip:** In GPT-5, ReAct-style reasoning works best with integrated tools like Search, Browser, and Code Interpreter.

11. Automatic Prompt Engineering (APE)

- **What it is:** Use AI to generate and test variations of prompts.
- **When to use it:** When you're designing for scale, training a chatbot, or optimizing prompt performance.
- **Why it works:** Many of the most powerful AI tools have multimodal features, meaning that they use more than one model. You can chain these models by prompting one to write a prompt for another, such as training a text-based model to write a prompt that can be used by an image or video-generating model.



More Examples:

- *Write 10 alternate prompts for asking AI to summarize a technical article for a general audience. Then pick the clearest one.*
- *Generate 5 prompt variations for helping a user reframe negative self-talk. Rank them from most supportive to most neutral.*
- *Create 7 prompt versions that ask users to describe product feedback. Tag each by tone (formal, casual, creative).*
- *Produce 5 options for asking ChatGPT to generate campaign ideas for a fintech company. Rate them based on clarity and specificity.*
- *Write a description of an object's physical features that can be used as a prompt for an image generator; it should be so vivid that someone who had never seen the object in person can accurately envision it in their mind's eye based solely on reading your description of it.*

Tips:

🟢 **Beginner:** Ask the model to rewrite your prompt in 3 clearer variations and choose your favorite.

🔵 **Intermediate:** Generate 5–10 versions with different tones, formats, or levels of detail — then test them across tasks.

⚫ **Advanced:** Set up side-by-side testing or use frameworks like PromptLayer, DSPy, or PromptFoo to evaluate performance at scale.

These techniques are building blocks — and with GPT-5, you can now combine them more freely. Try mixing role + format + reasoning + examples to create powerful, multi-layered prompts that feel like expert conversations.

Section 6: Role-Based Prompt Templates

In the previous section, one of the prompting techniques mentioned was role prompting. This involves telling the LLM to pretend it is a subject matter expert of some kind and write their answer from that perspective.

Why is this such a popular prompting technique? Well, when asking for answers from a specific perspective, the model is encouraged to reference a smaller pool of its training data. Additionally, talking to an expert about their chosen subject or study will result in different insights than a layman. These two facts result in more specific, less generic outputs— making this the ideal prompting technique if you're looking for unconventional results.

However, this is a technique that you should use sparingly, and with tempered expectations— especially if you're working with older models. That's because you're not actually speaking to a subject matter expert, but a predictive text generator that is being told to pretend to be an expert as realistically as possible. That means the risk of hallucination is high, so you'll want to double-check the accuracy of specific claims made by most tools when using this prompting technique.

That being said, this is largely an issue with older models; for recent tools like GPT-5, their reasoning capabilities and training data have led them to achieve PhD-level insights in fields like law and mathematics.

These prompts are designed for everyday use across key business functions. Each one is adaptable— just swap in your input. Tone/style suggestions are included where applicable.

Operations

Prompt 1

> ***You are a process improvement analyst. Review the following workflow and suggest three ways to reduce manual steps.***

Tone: Analytical, concise

Prompt 2

> ***You are an operations manager. Based on the data below, identify any scheduling conflicts and recommend an optimized shift plan.***

Tone: Objective, action-oriented

Prompt 3

> ***You are a logistics coordinator. Rewrite these internal instructions so they're clear, step-by-step, and easy to follow for new hires.***

Format: Ordered list

Sales

Prompt 1

> ***You are an outbound strategist. Write a LinkedIn message to a VP of Finance at a mid-sized SaaS company. Make it conversational, relevant, and end with a question.***

Tone: Conversational, focused

Prompt 2

> ***You are a sales coach. Based on this call transcript, identify three missed opportunities and suggest a better framing of the product's value.***

Tone: Constructive, direct

Prompt 3

> ***You are a strategic account executive. Write a follow-up note after a promising discovery call. Include 1 recap line, 2 insights, and 1 CTA.***

Format: Structured email outline

Marketing

Prompt 1

> ***You are a conversion copywriter. Rewrite the following landing page headline to make it more urgent, specific, and benefit-driven. Provide three options.***

Tone: Punchy, persuasive

Prompt 2

> ***You are a content strategist. Turn this blog post into a three-post LinkedIn carousel script with a hook, a breakdown, and a CTA.***

Format: Slide-by-slide copy outline

Prompt 3

> ***You are a brand voice expert. Review this homepage and suggest two changes to make it more consistent with a warm, authoritative tone.***

Tone: Editorial

Management

Prompt 1

> ***You are a leadership coach. A team lead is struggling with prioritization. Share three frameworks they can use to manage competing requests.***

Tone: Supportive, instructive

Prompt 2

> ***You are a department head. Draft a message to your team acknowledging recent tension and outlining next steps to restore alignment.***

Tone: Empathetic, confident

Prompt 3

> ***You are an executive preparing for a board update. Summarize your top three metrics, growth levers, and open risks in under 150 words.***

Format: Bulleted or block-style summary

Content Development

Prompt 1

> ***You are a senior editor. Improve the clarity, tone, and structure of this article. Make it more engaging without losing substance.***

Tone: Crisp, confident

Prompt 2

> ***You are a content repurposing expert. Turn this long-form article into one tweet thread, one newsletter excerpt, and one TikTok script.***

Format: Multi-output summary

Prompt 3

You are a product copywriter. Write three microcopy options for a call-to-action button encouraging users to upgrade.

Tone: Clear, value-driven

Data Analysis

Prompt 1

You are a business analyst. Interpret the following product usage report and highlight three meaningful trends.

Tone: Insightful, plain-language

Prompt 2

You are a data storyteller. Write a 5-sentence summary of this chart for an executive audience with no technical background.

Tone: Accessible, high-level

Prompt 3

You are a customer insights lead. Based on this NPS feedback dataset, extract two recurring themes and suggest one recommendation.

Format: Bullet points or numbered summary

Learning & Development

Prompt 1

> ***You are an instructional designer. Turn this onboarding checklist into a simple 5-step training module for new hires.***

Format: Step-by-step breakdown

Prompt 2

> ***You are a corporate trainer. Draft a Slack message inviting employees to a new AI literacy workshop. Make it short, clear, and inviting.***

Tone: Friendly, motivating

Prompt 3

> ***You are a learning coach. Based on this employee feedback, recommend two personalized learning paths and one potential mentor pairing.***

Tone: Supportive, customized

Section 7: Prompt Scorecard & Worksheets

Prompt engineering is an iterative process. The more you test, adjust, and document, the more consistent your results become. Use the following scorecard and worksheet to improve prompt quality, troubleshoot issues, and build a reusable prompt library.

Prompt Quality Scorecard

Use this 7-point checklist to evaluate the strength of a prompt. For each item, give yourself a score from 1 (no) to 5 (excellent).

Question	Score (1–5)
1. Is the task clearly defined?	
2. Did I assign a clear role or persona to the model?	
3. Did I provide the right context or background information?	
4. Did I specify a desired output format (list, paragraph, JSON)?	
5. Did I include instructions for tone, length, or constraints?	
6. If applicable, did I request reasoning or step-by-step thinking?	
7. Is the prompt easy to understand and free of ambiguity?	

Scoring guide:

- 30–35 = Strong, high-confidence prompt
- 20–29 = Decent, may benefit from revisions
- Below 20 = Likely to produce inconsistent or unclear outputs

Prompt Refinement Worksheet (Updated for GPT-5)

Use this template to document and refine your most important prompts over time.

Field	Notes
Prompt Name	e.g. Weekly Report Summary
Goal	What you want the output to achieve — e.g. “Summarize usage data into executive-facing bullet points”
Model + Mode	e.g. GPT-5 in Thinking Mode, Claude 3 Opus, Gemini Pro (note “quick” or “deep” response mode if used)
Memory Active?	Yes / No (If the model is remembering preferences across sessions)
Temperature	e.g. 0.2 = precise/factual, 0.8 = creative/varied
Other Settings	e.g. Max Tokens = 1000, Top-p = 0.9, Multimodal = Yes/No
Output Format	e.g. JSON, table, Markdown bullets, short paragraph
Prompt (Initial)	[Paste full original prompt]
Input Context	[Paste sample input — text, data, user persona, etc.]
Output Sample	[Paste or summarize initial result]

What Worked	[What was good — tone, logic, structure, etc.]
What Needs Revision	[Note gaps — hallucination, vagueness, tone mismatch, missing logic, etc.]
Refinement Strategy	e.g. Add role, clarify format, chain steps, reduce complexity, change temperature
Final Version	[Paste refined prompt]
Final Output Sample	[Paste or describe improved result]
Result Rating	✅ Strong / 🟡 Okay / 🔴 Needs Work
Date Updated	[Track versioning over time]

Tips for Testing Prompts

- Change one variable at a time (e.g. role, tone, output format)
- Compare multiple outputs with the same prompt using different models
- Keep a running prompt library to reuse and adapt over time
- If a prompt fails, isolate why: unclear instruction, missing input, poor formatting?

Final Thought

Prompting isn't about tricks. It's about clarity. The better you guide the model, the more valuable it becomes. Once you learn to speak its language, you can do almost anything, from strategy to storytelling to automation.

This guide is your starting point. Now go make the AI work *for you*.

Section 8: Glossary of Common Prompt Engineering Terms

This glossary breaks down key terms you've seen throughout the guide.

Term	Definition	Example / Context
Prompt	The input or instruction you give to a language model to generate a response.	“Summarize this article in 3 bullet points.”
LLM (Large Language Model)	A type of AI trained on vast amounts of text to predict and generate language. Examples include ChatGPT, Claude, and Gemini.	ChatGPT is an LLM developed by OpenAI.
Token	A chunk of text the model processes — often a word or part of a word. Models like GPT-5 can now handle hundreds of thousands or even millions of tokens in one session.	“summarize” = 1 token; “artificial intelligence” = 2–3 tokens
Zero-Shot Prompting	Giving the model a task without any examples.	“List 3 benefits of this product.”
Few-Shot Prompting	Giving the model examples to help it understand how to respond.	Example: “Hi, I’m Alex.” → “Hello Alex, nice to meet you!”
System Prompt	A behind-the-scenes instruction that sets the model’s behavior, tone, formatting, or constraints. In GPT-5, system prompts can persist across longer chats with better consistency.	“Always respond in JSON.”
Role Prompting	Telling the model to take on a specific persona or point of view. GPT-5 is especially responsive to specific, niche roles (e.g., "You are a UX researcher at a fintech company…").	“You are a product manager…”
Contextual Prompting	Including relevant background or data to help the model generate better results. GPT-5’s expanded context window allows full documents, transcripts, and multi-part examples to be included without loss of structure.	“Given our brand voice guidelines, write a tweet…”
Step-Back Prompting	Asking the model to reflect or solve a broader question before executing the task.	“What does the user want?” → Then write the message

Chain-of-Thought (CoT)	Instructing the model to think through reasoning steps before giving a final answer.	"Think step by step…"
Self-Consistency	Running the same prompt multiple times and selecting the most common or consistent result. GPT-5 shows improved consistency on its own, but this technique is still useful for creative tasks or when resolving ambiguity.	Used for ambiguous or high-stakes tasks.
Tree of Thoughts (ToT)	Exploring multiple reasoning paths or options before converging on a solution.	"Explore 3 campaign strategies and evaluate pros and cons."
ReAct	A prompting framework where the model reasons and then acts — often used with tools like web search or code interpreters. GPT-5 supports advanced ReAct-style prompting in agentic workflows, especially with tools like search, browser, or code interpreter.	"Search for X → Summarize → Recommend Y"
Automatic Prompt Engineering (APE)	Using AI to generate, test, and refine multiple versions of a prompt.	"Write 5 variations of this onboarding prompt and rate them for clarity."
Temperature	A setting that controls randomness. Lower = more focused, higher = more creative.	0.2 = precise, 0.8 = more varied output
Top-p / Top-k	Sampling settings that determine which words the model can choose from next — affects diversity in output.	Useful for tuning creativity vs. precision.
Max Tokens	The maximum length (in tokens) that the model can generate.	Prevents overly long or rambling answers.
Hallucination	When the model generates false or made-up information that sounds correct.	"This company raised $50M" — when it actually didn't.
Format Instruction	A prompt that specifies how the output should be structured (e.g., list, table, paragraph).	"Respond in JSON" or "Use bullet points."
Persona	A defined role or identity assigned to the AI in a prompt.	"You are a friendly customer support agent."

Memory	GPT-5 (in ChatGPT) can remember facts, preferences, and behavior across sessions — allowing it to personalize responses and build continuity over time.	"The model remembered my product's name from last week's conversation."
Thinking Mode / Quick Response Mode	In GPT-5-powered products, you can often toggle between "thinking mode" (deep reasoning) and "quick response" (faster, shallower output).	Click "Get Quick Response" to skip reasoning and generate fast drafts, or leave it off for thoughtful analysis.
Test-Time Compute	A technique where the model dynamically allocates more processing power ("compute") based on how complex the task is.	GPT-5 may use more compute when writing a business strategy than summarizing a tweet.
Multimodal Input	The ability for a model to understand and respond to more than just text — including images, audio, and video.	"Upload this screenshot and write alt text that's brand-appropriate."

FORWARDFUTURE

Helping everyone benefit from AI with timely news & accessible education.

Website: forwardfuture.ai

YouTube: @matthew_berman

X: @forward_future_

Thanks for reading!

- Matt, Nick, and the Forward Future Team