PROMPT
Engineering

✳ master great prompts ✳

Lets start with the Prompt Engineering Mindset

Definition

At its core, Prompt Engineering is the art and science of crafting efficient inputs (prompts) to get the best possible output from Generative AI models (like ChatGPT, Gemini, and Claude).

It involves selecting the right words, phrasing, structure, and examples to guide the AI toward a specific, accurate, and high-quality result.

How to think about it

Prompt engineering is not about memorising tricks—it's about communicating with intent.

Treat the AI like an incredibly capable but extremely literal intern who has read everything on the internet yet has no idea what you personally care about unless you spell it out.

Core mindset

The core mindset that ties every technique together:

1. **You are the director not the requester:** The AI doesn't "know" what you want—it mirrors the world you paint in your prompt. The clearer and more deliberate your instructions, the better it performs.
2. **Clarity > Cleverness:** A straightforward, slightly longer prompt almost always beats a short "clever" one. Being explicit is a superpower.
3. **The AI has no context unless you give it:** It can't read your mind, your project brief, your deadlines, or your audience's background. Everything that matters must be in the prompt (or the conversation history).
4. **Iterate without ego:** First responses are rarely perfect. Great results come from refining, not from nailing it in one shot. Think of prompting as sculpting—start with a rough block and chip away.
5. **Specificity compounds:** Every detail you add (role + audience + format + constraints + examples) multiplies the quality of the output. Stack techniques ruthlessly.
6. **Test like a scientist:** Change one variable at a time (add a role, then examples, then Chain of Thought) so you learn what actually moves the needle for your use case.

General tips to apply to every prompt you'll ever write

- **Start with the end in mind:** What exact deliverable do you need? (a 200-word email, a JSON object, a 5-bullet summary, etc.)

- **Front load the important stuff:** Put role, task, and format in the first 1–2 sentences whenever possible.

- **When in doubt, over-explain on the first try**; you can always shorten later.

- **Combine techniques:** The most powerful prompts use 3–5 techniques at once (e.g., Role + Few-Shot + CoT + Strict JSON Output + Constraints).

- **Remember context length trade-offs:** For creative tasks, allow some freedom; for precision tasks (code, data extraction, analysis), lock it down with strict instructions and low temperature if your interface allows.

- Keep a personal "prompt library" of your best-performing prompts—you'll reuse the patterns constantly.

Understanding the context window

- **The context window is the hidden resource that controls everything in Prompt Engineering**

- Explaining the **context window:**

  - The context window is the total amount of information (measured in tokens) that an AI model can 'see' and process in a single interaction.

  - Think of it as AI's short term memory:

    - (Your prompt + any uploaded files/data + the conversation history + the output its generating) must fit inside this fixed window size.

  - 1 token ≈ 3–4 characters or roughly ¾ of a word in English.

- **Example:**

  - The sentence "Prompt engineering is incredibly powerful" is about 8 tokens.

- **Every modern LLM has a maximum context window:** Once you hit that limit, the model literally forgets the earliest parts of the conversation or data. This is where hallucination happens.

Why the context window matters to prompt engineers

You are not just paying for intelligence—you are paying for memory.

Everything competes for space in that window:

Component	Typical token cost	What happens if its too big
System prompt / role	50–500 tokens	Reduces room for data/output
Conversation history	Can be thousands of tokens	Older messages get dropped
Uploaded files / data	A 10-page PDF ≈ 4,000–8,000 tokens	May truncate or fail entirely
Your actual prompt	100–2,000+ tokens (with examples)	Leaves less room for reasoning
Chain of Thought reasoning	Often 1,000–5,000+ tokens	Improves quality but eats space
Generated output	Whatever you ask for (e.g., 2,000-word report = ~2,500 tokens)	If output would exceed remaining space, model stops mid-sentence (or refuses)

The aim: Total tokens used = Input tokens (prompt + history + data) + Output tokens ≤ Model's max context

Major LLM Context Windows (as of November 18, 2025)

Provider	Model	Max Context Window	Notes / Limitations
Anthropic	Claude 4 Opus / Sonnet	200,000 tokens	Standard for most users; 1M available in beta for Sonnet 4/4.5 via API (premium pricing for >200k)
OpenAI	GPT-4o / o3 series	128,000 tokens	Output often capped ~16k in consumer tools
OpenAI	GPT-4.1 / GPT-5 series	Up to 1,000,000 tokens	Newer reasoning-focused models; full 1M in API, sometimes limited in ChatGPT
Google	Gemini 2.5 Pro	1,000,000–2,000,000 tokens	2M available in preview/enterprise; excellent long-context recall up to ~500k in practice
Google	Gemini 2.5 Flash	1,000,000 tokens	Faster, lower-cost variant
xAI	Grok 4	256,000 tokens	Balanced speed + context (this model)
xAI	Grok 4 Fast	2,000,000 tokens	Cost-efficient variant with massive context
Meta	Llama 4 Scout	10,000,000 tokens	Industry-leading; requires significant GPU memory for full use
Meta	Llama 4 Maverick	1,000,000 tokens	General-purpose flagship
Alibaba	Qwen3-Next	256,000 native (extendable to 1M)	Strong long-context performance
Alibaba	Qwen3-Long	10,000,000 tokens	Specialized ultra-long variant
DeepSeek	DeepSeek V3 / R1	128,000 tokens	Excellent cost-performance ratio

11 techniques to remember

Now in this lesson we're going to cover 11 techniques that will cover 95% of real world use cases.

Technique	When it shines most	One liner power moves
Role prompting	You need expertise, tone, or a specific viewpoint	"You are a …" in the very first sentence
Output structure guide	Consistent format, style, or transformation needed	Take this information and output it into this format
Chain of thought reasoning	Math, logic, analysis, or multi-step decisions	"Let's think step by step"
Clear & specific instructions	You keep getting close-but-not-quite responses	Spell out length, format, tone, includes/excludes
Context setting	Generic answers that miss your actual situation	Give background, goals, constraints, audience
Output formatting	You need JSON, tables, code blocks, or exact structure	"Respond in this exact markdown/JSON format…"
Constraints & limitations	Prevent rambling, jargon, hype, or off-topic output	"Max 150 words · No emojis · Avoid claims of 'best'"
Iterative refinement	First draft is good but not perfect	Build on the previous answer instead of starting over
Audience targeting	Same info needs to land differently	"Explain to a 7-year-old / CEO / total beginner"
Task decomposition	The request is big and intimidating	Break it into numbered steps and tackle one by one
Perspective taking	You want balanced analysis or to anticipate objections	"Analyze from the viewpoint of X, Y, and Z"

Role prompting

- Role prompting is one of the most powerful techniques for shaping AI responses.
- By assigning the AI a specific role, persona, or professional identity, you fundamentally change how it approaches your question.
- The AI will adopt the knowledge base, communication style, and perspective of that role.
- Use this when you need expertise in a specific domain, want responses tailored to a particular profession, or need the AI to think from a specific viewpoint.
- This works because it primes the AI's knowledge retrieval and response generation toward that domain.

Expert guidance

Instruct the AI to act as an authority figure or specialist to gain in depth insights and explanations.

Example prompt: "You are a senior cardiovascular surgeon with 20 years of experience. Explain to a medical student the difference between Type A and Type B aortic dissections, including diagnostic criteria and treatment approaches."

Creative persona

Ask the AI to embody a creative role, like a writer, artist, or storyteller, to generate imaginative content.

Example prompt: "You are a friendly kindergarten teacher who loves making learning fun. Help me explain the water cycle to 5 year olds using a story about Droplet the water drop."

Professional perspective

Direct the AI to analyze information from a specific professional viewpoint, such as a lawyer, doctor, or teacher.

Example prompt: "You are a seasoned UX designer at a top tech company. Review this mobile app wireframe and provide detailed feedback on usability, accessibility, and design patterns."

Output structure guidance

- This teaches the AI to do what you want to do.
- For example: You might have a transcript of a meeting - and you provide the AI with an example of the minutes for your organisation and ask it to turn the transcript into the minutes.
- This is incredibly effective for tasks with specific formats, consistent transformations or unique styles

Text transformation example

Demonstrate how text should be reformatted or transformed based on given inputs.

Example prompt:
Please take the following transcript of a meeting on X date and transform it into the minutes format attached.

Educational example

Please take the following data, and answer the following questions

Example prompt:
Please take the following information, and answer the following questions

Chain of thought reasoning

- Chain of thought prompting asks the AI to show its work by thinking through problems step-by-step, just like you'd solve a math problem by writing out each step.
- This dramatically improves accuracy for complex reasoning tasks, mathematical problems, logical puzzles, and multi step processes.
- By making the AI's reasoning visible, you can spot errors, understand the logic, and get more reliable answers.
- **Use phrases like 'Let's think step by step,' 'Show your reasoning,' or 'Walk me through your thought process.'**
- This is essential for tasks requiring logic, calculation, or sequential decision-making.

Problem solving example

Instead of just providing a solution, the AI is prompted to list out each step it took to arrive at the solution, clarifying its approach to breaking down the problem.

Logical reasoning example

For complex logical deductions, the AI explains its inferential process, detailing how each premise leads to a specific conclusion or eliminates other possibilities.

Analysis example

When analysing data or text, the AI demonstrates its analytical steps, such as identifying key patterns, drawing comparisons, or interpreting nuances.

Clear and specific instructions

- Vague prompts get vague results.
- Being crystal clear about exactly what you want—including length, format, style, structure, and any other requirements—is fundamental to good prompting.
- Think of it like giving directions: 'Go to the store' versus 'Drive 2 miles north on Main Street, turn left at the red barn, the store is the blue building on the right.'
- The more specific you are, the more likely you'll get exactly what you need. Include word counts, number of items, formatting rules, tone requirements, and what to include or exclude. This eliminates ambiguity and reduces back-and-forth.

Content creation example

Vague: "Write a blog post about AI."

Specific: "Write a 500-word blog post on the benefits of AI in healthcare, targeting a general audience, using an informative yet enthusiastic tone. Include a clear introduction, three main benefits, and a conclusion. Ensure it's SEO-friendly with keywords like 'AI healthcare solutions' and 'medical technology innovation'."

Data processing example

Vague: "Analyse this data."

Specific: "Analyze the Q3 sales data from the attached CSV file. Identify the top 5 performing products, calculate the month-over-month growth for each product category, and summarise key trends and anomalies in a bulleted list, focusing on actionable insights for the marketing team."

Code generation example

Vague: "Write some Python code."

Specific: "Write a Python function that takes a list of integers as input and returns a new list containing only the even numbers, sorted in ascending order. Include docstrings explaining the function's purpose, arguments, and return value, and add unit tests covering edge cases like empty lists or lists with no even numbers."

Context setting

- AI doesn't know anything about your specific situation unless you tell it.
- Context setting means providing all the relevant background information, constraints, and circumstances that frame your request.
- This includes your skill level, the project you're working on, who you're creating for, time constraints, available resources, and any other situational factors.
- Good context transforms generic answers into tailored, actionable advice.

Technical help example

Vague: "How do I fix my computer?"

Specific: "My Windows 11 laptop, a Dell XPS 15, is showing a 'Blue Screen of Death' error with code 0x00000133 every time I try to open a video editing application. I've already tried updating drivers and running a disk check. What are the next diagnostic steps I should take, considering I need to finish a project by tomorrow?"

Career advice example

Vague: "What's a good career path for me?"

Specific: "I have a Bachelor's degree in Marketing with 3 years of experience in social media management for SaaS startups. I'm looking for a career path that offers strong growth potential, allows for creative problem-solving, and has remote work flexibility. I'm particularly interested in roles related to product marketing or brand strategy, but I'm open to suggestions in similar fields. What specific roles, skills, and certifications should I pursue to transition into a more senior position within the next 2-3 years?"

Business strategy example

Vague: "How can I increase sales?"

Specific: "Our small e-commerce business sells handmade jewelry, primarily targeting women aged 25-45 in the US. Our average order value is $75, and we currently rely heavily on Instagram ads. We've seen a 10% decline in sales over the last quarter due to increased ad costs. We have a monthly marketing budget of $2,000 and want to explore new channels or strategies to increase sales by 15% within the next six months, focusing on sustainable customer acquisition rather than just short term boosts."

Output formatting

- The format of the AI's response is just as important as its content.
- Output formatting means explicitly specifying the structure you want: JSON, markdown tables, bullet lists, numbered steps, code blocks, or custom formats.
- This is crucial when you need to parse the output programmatically, present information in a specific way, or integrate the response into another system.
- Well formatted output saves time on manual reformatting and ensures consistency. Be precise about structure, delimiters, field names, and organisation. This technique is essential for API responses, data pipelines, reports, and any structured information.

Example 1 - JSON for programming

"Analyse this product review and return the result in this exact JSON format: { 'sentiment': 'positive' | 'neutral' | 'negative', 'rating_estimate': 1-5, 'key_topics': ['topic1', 'topic2'], 'mentioned_features': ['feature1', 'feature2'], 'customer_type': 'first-time' | 'repeat' | 'unknown', 'urgency_level': 'low' | 'medium' | 'high' }

Review: [your review text]"

Example 2 - Markdown table

"Compare these three project management tools in a markdown table with these columns: | Feature | Asana | Trello | Monday.com | Include rows for: Price, Team Size Limit, Integrations, Mobile App, Gantt Charts, Time Tracking"

Example 3 - Structured report

"Provide the analysis in this exact format:

EXECUTIVE SUMMARY[2-3 sentences]

KEY FINDINGS1. [Finding with supporting data]2. [Finding with supporting data]3. [Finding with supporting data]

RECOMMENDATIONS- **Immediate Action:** [what to do now]- **Short-term (1-3 months):** [what to do soon]- **Long-term (6-12 months):** [strategic changes]

RISKS[Bullet list of potential risks]"

Constraints and limitations

- Constraints define the boundaries of what you want—and just as importantly, what you don't want.
- This includes length limits (word counts, character limits), forbidden content (no jargon, no controversial topics), required simplicity levels, time periods to focus on, and any other boundaries.
- Constraints help focus the AI's response and prevent unwanted elements.
- They're especially useful for creating content that must fit specific requirements like social media posts, presentations, or regulated industries.
- Good constraints make responses more usable and appropriate for your exact needs.

Social media

Create 3 LinkedIn post options about our new product launch. Constraints:

- Maximum 150 words each
- Professional but conversational tone
- No emojis or hashtags
- Must mention 'innovation' and 'customer-focused'
- Avoid any technical specifications
- Include a clear call-to-action
- Do not make claims about being '#1' or 'the best'

Educational content

Explain blockchain technology. Constraints:

- Maximum 200 words
- 8th-grade reading level
- Use only everyday analogies (no computer science terms)
- No mention of cryptocurrency or Bitcoin
- Focus only on the core concept of distributed ledgers
- Use at least one real world analogy

Business writing

Draft an email declining a vendor proposal. Constraints:

- Maximum 100 words
- Maintain positive relationship for future opportunities
- Do not cite specific reasons for decline
- Do not mention competitor names
- Professional and respectful tone
- Include appreciation for their time
- Keep door open for future consideration

Iterative refinement

- Iterative refinement recognises that you rarely get the perfect answer on the first try.
- Instead of starting over, you build on previous responses by asking targeted follow up questions to clarify, expand, or adjust specific parts.
- This conversational approach lets you incrementally shape the output toward exactly what you need. Ask for elaboration on specific points, request examples for unclear concepts, or have the AI revise particular sections while keeping others.
- This is more efficient than re-prompting from scratch and leverages the context of the ongoing conversation.

Progressive detail:

First prompt: "Outline a marketing plan for a new fitness app."

[AI provides outline]

Refinement 1: "Great start. Can you expand section 3 (Social Media Strategy) with specific platform recommendations and posting frequency?"

[AI expands that section]

Refinement 2: "Perfect. Now for Instagram specifically, can you provide 5 concrete post ideas with caption templates?"

[AI provides detailed Instagram content]

Targeted improvements:

First prompt: "Write a product description for noise-cancelling headphones."

[AI provides description]

Refinement 1: "The description is good, but can you make the opening sentence more attention-grabbing? Keep everything else the same."

[AI revises opening]

Refinement 2: "Now add a paragraph specifically addressing the concerns of frequent travelers. Insert it after paragraph 2."

Clarification:

First prompt: "Explain how neural networks learn."

[AI provides explanation]

Refinement 1: "You mentioned 'backpropagation' in step 3. Can you explain that specific concept with a simple analogy?"

[AI explains backpropagation]

Refinement 2: "That makes sense. Now, using that same level of simplicity, explain gradient descent."

Audience targetting

- The same information needs to be communicated very differently depending on who's receiving it.
- Audience targeting means explicitly stating who the response is for—their age, expertise level, background knowledge, interests, or role.
- This causes the AI to automatically adjust complexity, vocabulary, examples, and depth of explanation.
- A concept explained to a child versus a PhD student versus a business executive requires completely different approaches.
- Always specify your audience to get appropriately tailored communication that resonates with and is understandable by your specific readers or listeners.

Same topic - different audience

Prompt for CEO: "Explain our database migration project to our CEO who cares about business impact, not technical details.
Focus on: timeline, cost, risks, and business benefits.
Maximum 5 sentences."
Result: Business-focused, ROI-driven, executive summary.
Prompt for engineers: "Explain our database migration project to our engineering team. Include: technical architecture, migration steps, potential challenges, rollback plan, and performance expectations."
Result: Technical, detailed, implementation-focused.

Age specific competence

"Explain how vaccines work:
- Version 1: For a curious 6-year-old
- Version 2: For a high school biology student
- Version 3: For a medical resident
Use appropriate vocabulary and depth for each."

Role specific outputs

"Explain this data breach incident in three versions:
- For affected customers (focus: what they should do, how we're protecting them)
- For our board of directors (focus: liability, financial impact, remediation)
- For our technical security team (focus: attack vector, vulnerabilities, prevention)"

17

Task decomposition...

- Complex tasks often overwhelm AI (and humans!).
- Task decomposition breaks large, complicated requests into smaller, manageable sub-tasks that can be tackled sequentially or in parallel.
- Instead of asking for everything at once, you create a clear roadmap of steps and either ask the AI to handle them one-by-one or work through the entire workflow systematically.
- This improves quality, makes debugging easier, and prevents the AI from getting lost in complexity. Use this for projects, multi-step processes, comprehensive analyses, or anything with multiple distinct components.

Task decomposition continued

Content creation

❌ Overwhelming: "Create a complete content marketing strategy for my business."

✅ Decomposed:

"Let's build this step-by-step:

Step 1: First, help me identify my target audience personas (demographics, pain points, goals)

[Complete step 1, then...]

Step 2: Based on these personas, what content topics would resonate with each group?

[Complete step 2, then...]

Step 3: For the top 3 topics, what content formats would work best (blog, video, infographic, etc.)?

[Complete step 3, then...]

Step 4: Create a 3-month content calendar with these topics and formats

[Complete step 4, then...]

Step 5: For each content piece, outline the promotion strategy"

Software project example

"I need to build a task management app. Let's break this down:

Phase 1: Define core features and user stories

Phase 2: Design the database schema

Phase 3: Plan the API endpoints

Phase 4: Sketch the UI wireframes

Phase 5: Identify the technology stack

Phase 6: Create a development timeline

Let's start with Phase 1. What are the essential features for an MVP task management app?"

Business analysis example

"Analyse whether we should expand to the European market.

Break this into:

Market size and opportunity analysis

1. Competitive landscape assessment
2. Regulatory and compliance requirements
3. Financial projections and investment needed
4. Risk analysis
5. Go/no-go recommendation
6. Complete each analysis separately, then provide a final integrated recommendation."

Perspective prompting

- Sometimes you need the AI to analyse something from multiple viewpoints or argue different sides of an issue.
- Perspective taking asks the AI to explicitly consider different stakeholders, opposing viewpoints, or alternative interpretations.
- This is invaluable for decision making, understanding controversies, anticipating objections, or creating balanced content. By forcing examination from multiple angles, you get more nuanced, comprehensive analysis and avoid blind spots in your thinking.

Decision analysis

"Analyse the decision to implement a 4 day work week from three perspectives:
- **Perspective 1** - Employee viewpoint: [Benefits, concerns, what they value most]
- **Perspective 2** - Management viewpoint: [Operational challenges, productivity concerns, cost implications]
- **Perspective 3** - Customer viewpoint: [Service availability, response times, quality impact]

Then provide a balanced recommendation considering all three."

Stakeholder analysis

"Our city is considering banning petrol powered leaf blowers. Analyse this from:
- Homeowners' perspective
- Landscaping companies' perspective
- Environmental activists' perspective
- City government perspective

For each: What are their main arguments? What do they stand to gain or lose?"

Devils advocate

"I believe we should switch from office based to remote first work. Present:
- Argument FOR (strengthen my position with evidence)
- Argument AGAINST (steelman the opposing view)
- Rebuttal to the counter arguments
- Final assessment of which position is stronger"

Now lets have a practice

https://askbeevs-prompt-tutor.replit.app/

I built this app with a prompt.

And for particular tasks - go find world class prompt libraries

https://askbeevs.com

Has a whole library of them!

And finally - AI knows how to write prompts!

Hack out what you think you want,
and ask AI to write you a prompt!

PRACTICE! ITERATE! DON'T STRESS!

YOU CANT GET IT WRONG!