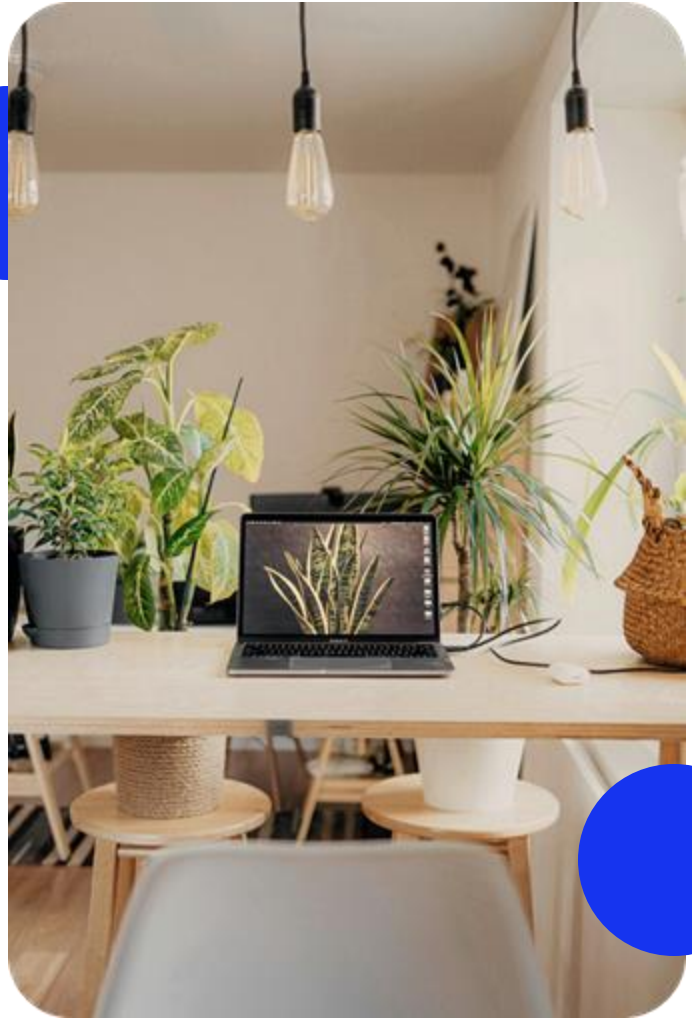
AI Masterclass

Technical Generative AI Concepts Explained Simply



github.com/Esteban-D-Lopez/TheAIWalkthrough

Learning Journey Roadmap

01	Technical Generative AI Foundations	Introduce foundational technical knowledge about AI and large language models (LLMs), laying the groundwork for understanding Generative AI.
02	GenAI Optimization Techniques 1	An overview of key LLM optimization techniques, with a deep dive into Fine-Tuning and Prompt Engineering.
03	GenAI Optimization Techniques 2	An overview of key LLM optimization techniques, with a deep dive into Retrieval Augmented Generation (RAG) and Agentic AI.
04	Generative AI Monitoring and Evaluation	An overview of key implications and practical considerations of bringing Generative AI products to life safely and efficiently.

Goals

- Understand how GenAI technology works
- Feel comfortable exploring with GenAI tools
- Start applying GenAI technology safely and responsibly

Presented by

Esteban D. Lopez

Location	New York, NY
Hometown	Quito, Ecuador
Education	University of New Orleans – BS. Accounting Louisiana State University – MS. Accounting Columbia University – FinTech Bootcamp New York University – MS. AI and Machine Learning (Current)
Professional Career	Hedge Funds Assurance Transformation & Economic Consulting AI & Data Technology Consulting AI Product Management

Learning Journey Roadmap

01 **Technical Generative AI Foundations**

Introduce foundational technical knowledge about AI and large language models (LLMs), laying the groundwork for understanding Generative AI.

02 **GenAI Optimization Techniques 1**

An overview of key LLM optimization techniques, with a deep dive into Fine-Tuning and Prompt Engineering.

03 **GenAI Optimization Techniques 2**

An overview of key LLM optimization techniques, with a deep dive into Retrieval Augmented Generation (RAG) and Agentic AI.

04 **Generative AI Monitoring and Evaluation**

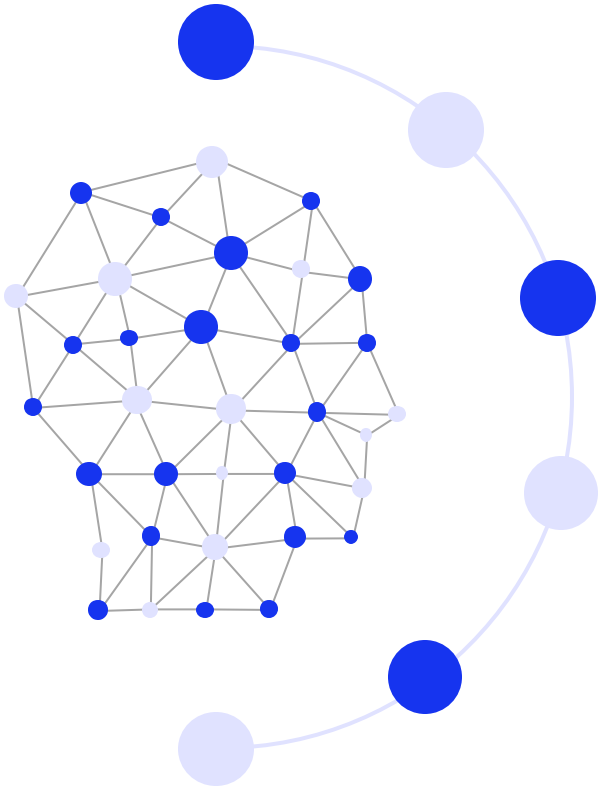
An overview of key implications and practical considerations of bringing Generative AI products to life safely and efficiently.

Goals

- Understand how GenAI technology works
- Feel comfortable exploring with GenAI tools
- Start applying GenAI technology safely and responsibly

GenAI Optimization Techniques Part I

Overview of Generative AI Principles



Generative AI

AI that creates new content—such as text, code, images, or music—rather than just analyzing data.

Neural Networks

Layers of connected artificial neurons that process data and learn complex patterns in large sets of data through training.

Large Language Models

Very large neural networks trained on massive text datasets to generate human-like language.

Transformer Architecture

The neural network design that powers modern LLMs using self-attention to understand word relationships.

Natural Language Processing

The field of enabling computers to understand, interpret, and generate human language..

Tokenization

The process of breaking text into small units (tokens) that an LLM can understand and process.,

Optimization Techniques

01 **Fine-tuning**
Training an LLM on custom data to specialize its behavior.

02 **Prompt Engineering**
Designing effective prompts to guide model outputs.

03 **Retrieval-Augmented Generation (RAG)**
Grounds LLMs with external knowledge sources.

04 **Agentic AI**
Orchestrates LLMs as multi-step, tool-using agents with memory and reasoning,

Cost
The amount of resource (data, compute, and engineering effort) needed to implement and maintain each technique.

Implementation Efficiency
How quickly and easily the technique can be deployed or iterated on in real-world workflows.

Performance
The degree of improvement the technique delivers in output quality, accuracy, and reliability.

Optimization Techniques

Technique	Description	Cost	Implementation
Fine-tuning	Further training enhances model performance.		
Retrieval Augmented Generation (RAG)	Connects model to external data sources.		
Prompt Engineering	Refines questions to maximize response quality.		
Agentic AI	Intelligent agents automate decisions and tasks.		

Part 1

Fine-tuning

Highlights how to deeply specialize a model on proprietary data or tasks, trading higher cost and effort for maximum performance gains.

Prompt Engineering

Shows how much model behavior can improve without retraining, making it the fastest, lowest-cost lever to boost performance.

LLM Fine-tuning

High-level overview of fine-tuning process for LLMs.

01 Start from a Pretrained Base Model

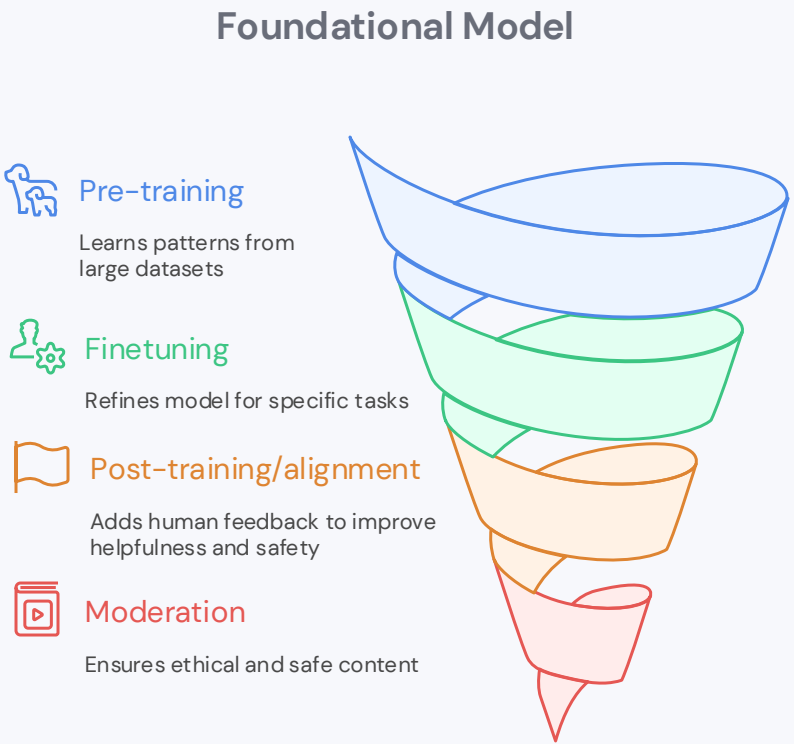
- Use an existing LLM already trained on massive general text datasets
- The model already knows grammar, facts, and reasoning patterns

02 Optimized Performance

- Feed it curated examples from your target domain
- Adjusts model weights so it learns your style, terminology, and goals

03 Validate and Deploy

- Test on held-out data to measure accuracy and prevent overfitting
- Deploy the fine-tuned model as a specialized version of the original



LLM Fine-tuning – Analysis

Benefits	Fine-tuning deeply customizes an LLM's behavior, improving accuracy, tone, and efficiency for specific tasks while reducing the need for complex prompts.
Drawbacks	Fine-tuning is resource-intensive, requiring high-quality labeled data, significant compute, and periodic retraining as domain knowledge evolves.

Sample Use Cases

Customer Support Bots

Fine-tuning on past chat transcripts ensures consistent tone and domain-specific answers without long prompts.

Legal Document Drafting

Fine-tuning on contracts and filings gives the model precise terminology and formatting, reducing editing time.

Medical Question Answering

Fine-tuning on vetted clinical guidelines improves accuracy and safety, minimizing the risk of hallucinated advice.

Prompt Engineering

High-level overview of prompt engineering process for LLMs.

01 Design Effective Prompts

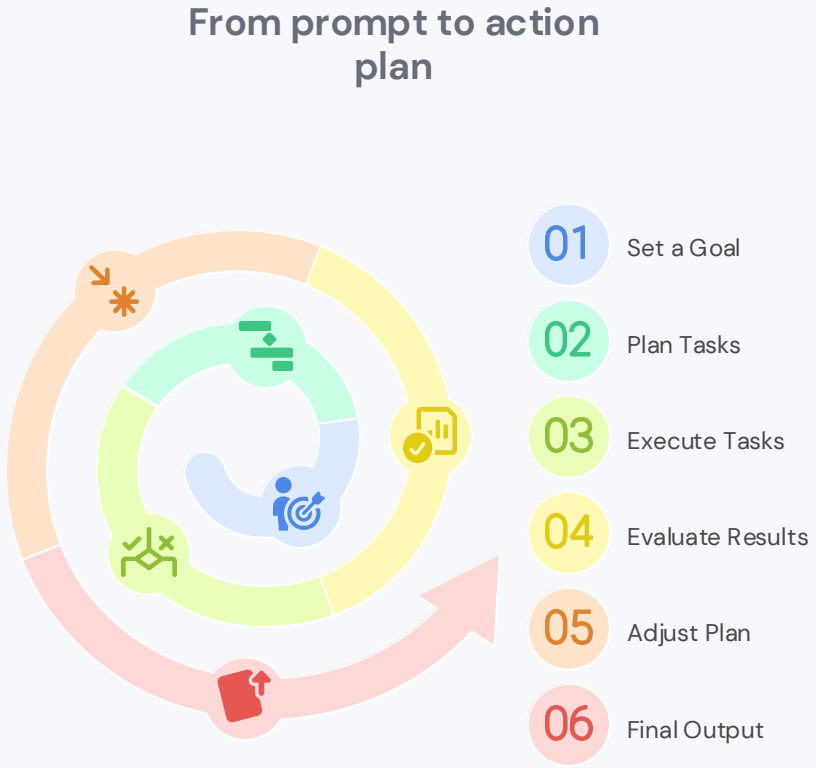
- Write clear, structured instructions that guide the model's behavior
- Include context, format, tone, role, and constraints using proper technique for the use case

02 Test and Refine Outputs

- Experiment with variations, examples, and technique
- Observe how the model responds and adjust wording accordingly

03 Standardize and Automate

- Save successful prompts as reusable templates or chains
- Use them consistently across workflows, tools, or APIs



Prompt Engineering – Analysis

Benefits	Prompt engineering is fast, low-cost, and flexible, letting you shape model behavior without retraining or additional infrastructure.
Drawbacks	It can be inconsistent and hard to scale, as outputs depend heavily on wording and may drift with model updates.

Sample Use Cases

Content Generation

Carefully designed prompts can produce consistent tone and structure for marketing copy or blog drafts.

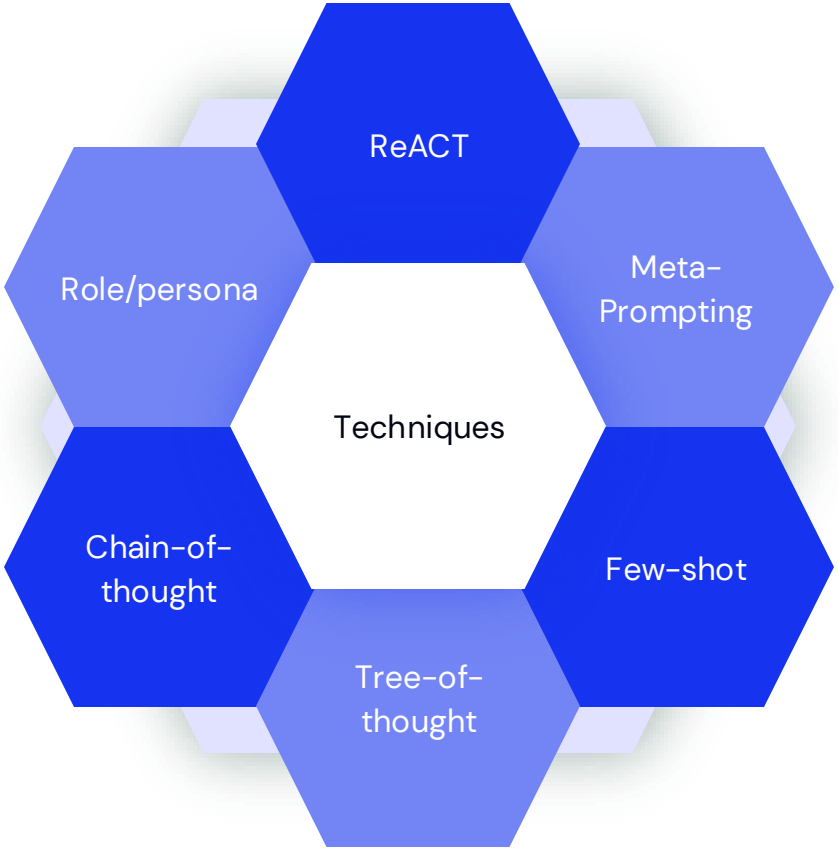
Customer Support Responses

Templates help models follow brand voice and answer FAQs reliably.

Code Assistance

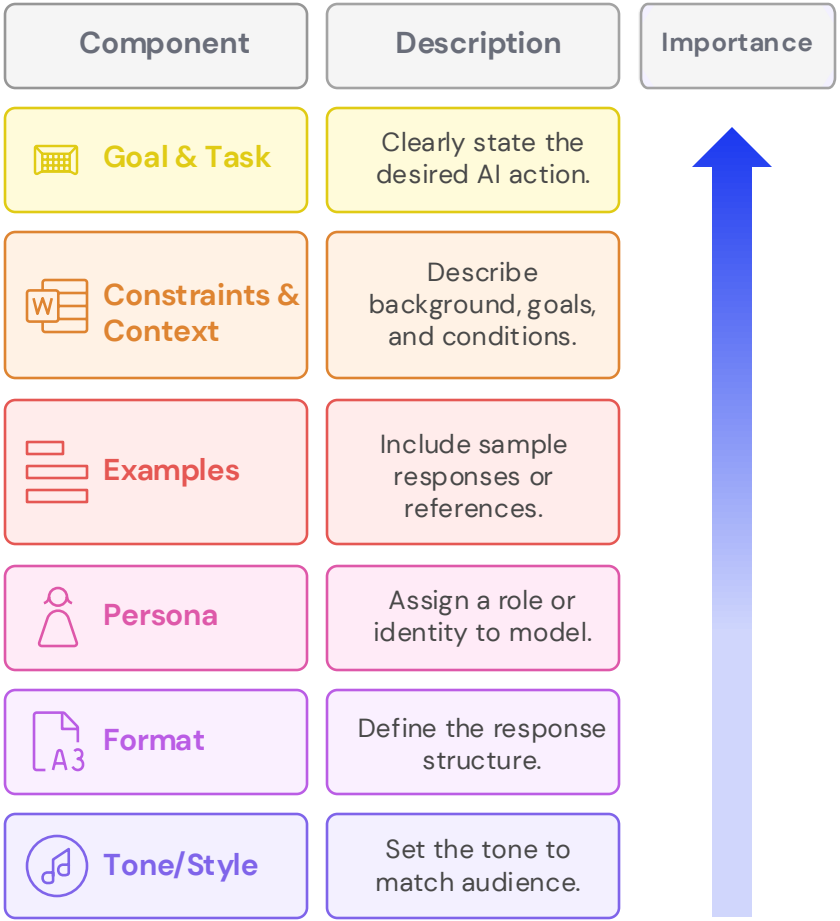
Structured prompts guide the model to generate bug-free snippets and explain logic clearly.

Prompt Engineering Workshop



Basic Best Practices

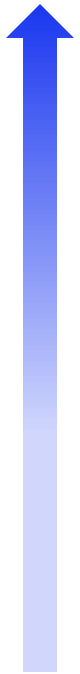
Prompt Engineering



Best Practices – Sample: Meeting Summary

Component	Description	Top Prompt	Mid Prompt	Min Prompt
Goal & Task	Clearly state the desired AI action.	Summarize the attached meeting transcript.	Summarize this meeting transcript.	Give me a brief summary of the main decisions from this meeting.
Constraints & Context	Describe background, goals, and conditions.	Focus only on key decisions, action items, and owners. Ignore small talk.	Focus on decisions and action items.	
Examples	Include sample responses or references.	Example summary: “Decisions: Approved Q3 budget… Actions: John to update deck by Friday…”		
Persona	Assign a role or identity to model.	Act as a corporate communications specialist.		
Format	Define the response structure.	Use bullet points with bold section headers.	Use bullets and keep it concise.	
Tone/Style	Set the tone to match audience.	Keep it concise and professional.		

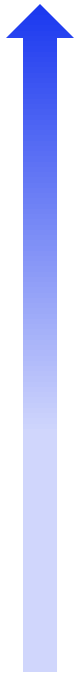
Importance



Best Practices – Sample: Analyze Sales Data

Component	Description	Top Prompt	Mid Prompt	Min Prompt
Goal & Task	Clearly state the desired AI action.	Analyze this sales dataset and identify trends.	Analyze this sales data for monthly revenue trends by region.	What trends do you see in this sales data?
Constraints & Context	Describe background, goals, and conditions.	Focus on monthly revenue growth by region. Ignore other columns.		
Examples	Include sample responses or references.	For example: “Revenue grew +12% MoM in LATAM, flat in EMEA.”		
Persona	Assign a role or identity to model.	Act as a business analyst.	Analyze these as if you were a CFA.	
Format	Define the response structure.	Return a table with Region, Trend, and % Change columns.	Show results in a simple table.	
Tone/Style	Set the tone to match audience.	Clear, concise, no jargon.		

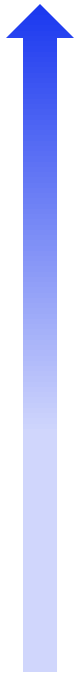
Importance

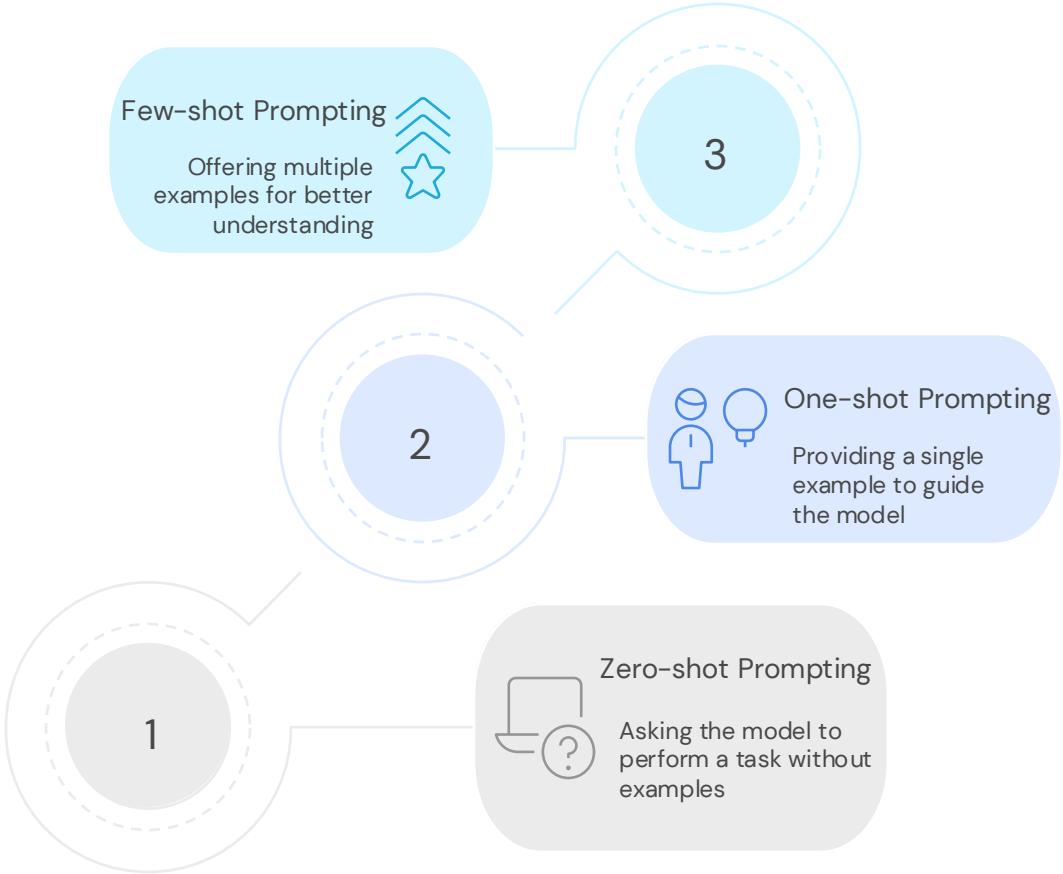


Best Practices – Sample: Writing Customer Email

Component	Description	Top Prompt	Mid Prompt	Min Prompt
Goal & Task	Clearly state the desired AI action.	Draft a customer outreach email about our new product launch.	Write a short, friendly email announcing our new product launch to existing clients.	Draft a short email about our new product launch.
Constraints & Context	Describe background, goals, and conditions.	Audience is mid-level managers at existing client companies. Keep under 150 words.	Keep it under 150 words.	
Examples	Include sample responses or references.	Here's an email I've sent before to other customers: [email].		
Persona	Assign a role or identity to model.	Act as a marketing copywriter.		
Format	Define the response structure.	Write in paragraph form with subject line and body.		
Tone/Style	Set the tone to match audience.	Friendly, confident, and concise – use my previous email.	Keep the tone warm and friendly.	

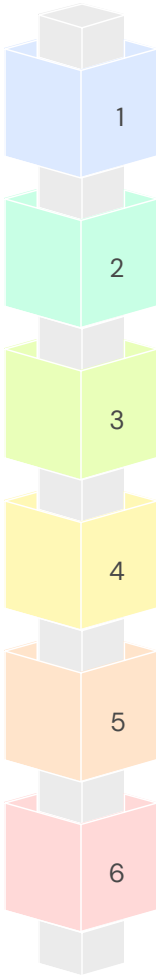
Importance





N-Shot Prompting

Provide n-samples in your prompt



Chain-of-Thought
Guide the model to explain its reasoning step-by-step.

ReAct
Interleaves reasoning steps with actions like tool use or searches.

Self-Critique
Asks the model to evaluate or improve its own output

Role Prompting
Assign a persona to control tone and realism.

Meta-Prompting
Gives the model instructions on how to create or improve prompts.

Multi-step Prompting
Break tasks into smaller prompts for complex outputs.

Advanced Prompting Techniques

Specialized prompting strategies designed to enhance AI's ability to generate accurate, consistent, and tailored responses.

These techniques go beyond basic questioning to improve creative and analytical outputs.

Which prompting technique should be used?

Chain-of-Thought Prompting

Use for complex reasoning tasks requiring step-by-step explanations.

ReAct (Reason + Act)

Ideal for agentic AI or task planning workflows.

Self- Critique

Improves accuracy by choosing the best answer from multiple attempts.

Role Prompting

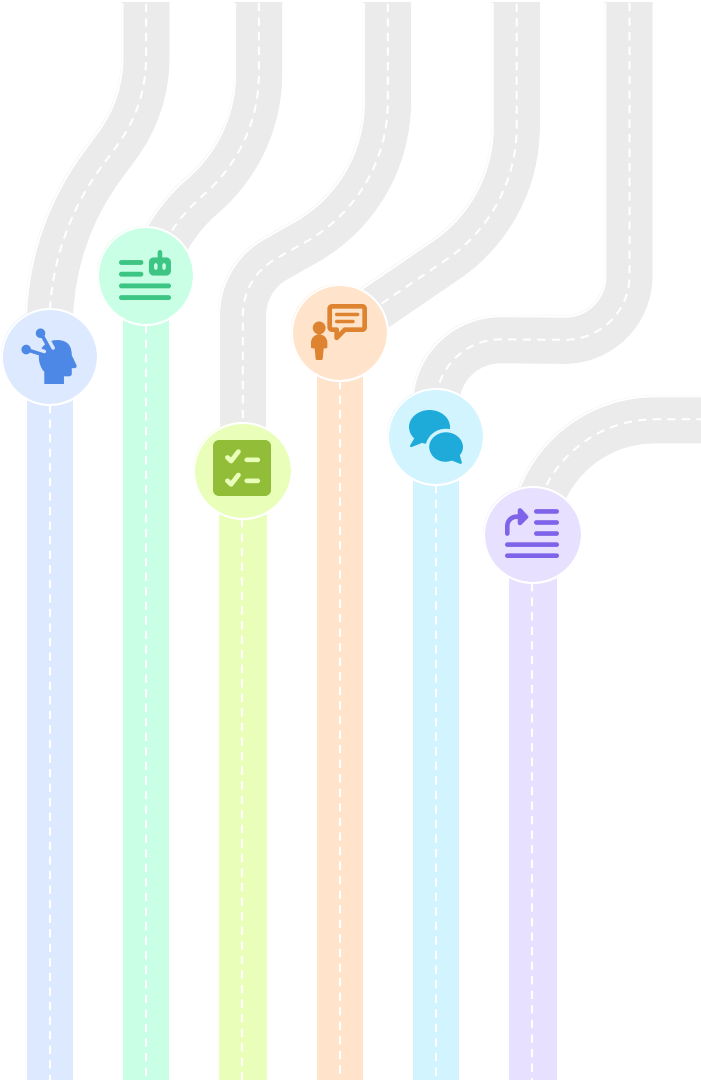
Controls tone and aligns with Specific audiences or realism.

Meta-Prompting

Useful for creating custom GPTs or tools that have a specific use case

Multi-step Prompting

Suitable for long or structured outputs like plans or emails.



Put it in practice

3 Things to Take With You

Writing

"... Follow George Orwell's 6 rules for effective writing. No emm-dashes."

Meta-prompt

"Draft a prompt so I can turn an AI large language model into a world class [role] to do [task]"

Chain-of-thought

"[Task]... Complete this step-by-step and show your reasoning."

[sources]

Thank you!

Questions?