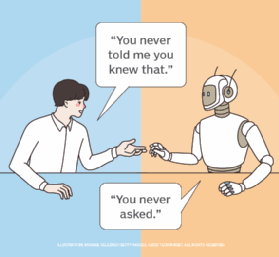
What is Prompting?

Prompting is the process of creating instructions, questions, or commands (prompts) to guide an AI model to generate a specific, desired output. It's how you communicate with an AI to get it to perform a task, such as writing text, generating code, or creating an image. Effective prompts are often detailed and clear, providing the necessary context for the AI to produce a relevant and accurate response.

* **How it works:** When you provide a prompt, the AI model uses its training data to interpret your request and generate a response.
* **Why it's important:** The quality of the prompt directly influences the quality of the AI's output. A well-crafted prompt leads to more accurate, relevant, and useful results, while a vague one may lead to a poor or unhelpful response.



* **Examples:**
  + **Basic prompt:** "Write an email inviting colleagues to a webinar."
  + **Detailed prompt:** "Write a formal email inviting my team to a webinar on AI in marketing, which starts on November 15th at 10 AM PST. Please include a link to the registration page and mention that the guest speaker is Dr. Jane Doe."

### **Common types of prompts**

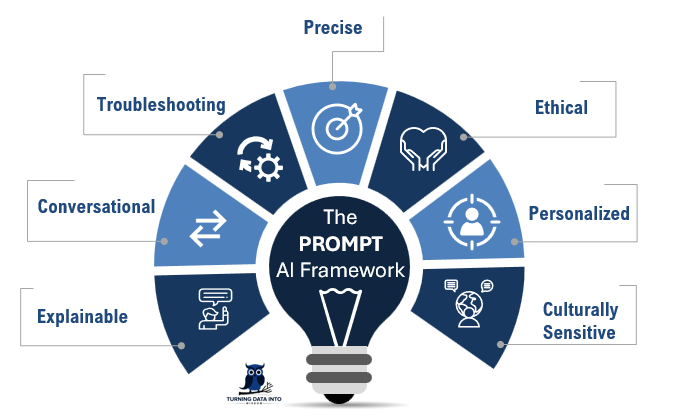
Prompting can range from simple commands to complex instructions, with several common techniques:

* Zero-shot prompting: Giving a direct command or question without examples, like asking for a five-bullet-point summary of an article.
* Few-shot prompting: Providing the AI with examples of desired input-output pairs to guide its response format or style. An example would be showing the AI a few product review summaries before asking it to summarize a new one.
* Chain-of-thought (CoT) prompting: Encouraging the AI to think step-by-step to solve complex problems more accurately. This could involve asking the AI to solve a math problem by showing each step.
* Role-based prompting: Instructing the AI to adopt a specific persona to tailor its response, such as asking it to act as a financial advisor.
* Instructional prompting: Using clear verbs and commands, often with formatting or length constraints. An example is asking the AI to write a professional email under 100 words.

Key components of a prompt framework

Effective frameworks break down a user's request into essential parts that provide clear direction and context to the AI model. Typical components include:

* **Role/Persona:** Instructing the AI to adopt a specific identity, such as "Act as a financial analyst" or "You are a friendly chatbot".
* **Task/Objective:** Clearly stating the desired goal, such as "Summarize this document" or "Generate a marketing strategy".
* **Context:** Providing relevant background information and constraints to inform the AI's response.
* **Format:** Specifying the desired output structure, such as "in bullet points," "as a table," or "a 3-section landing page".
* **Audience:** Defining the intended reader so the AI can tailor the tone and language accordingly.

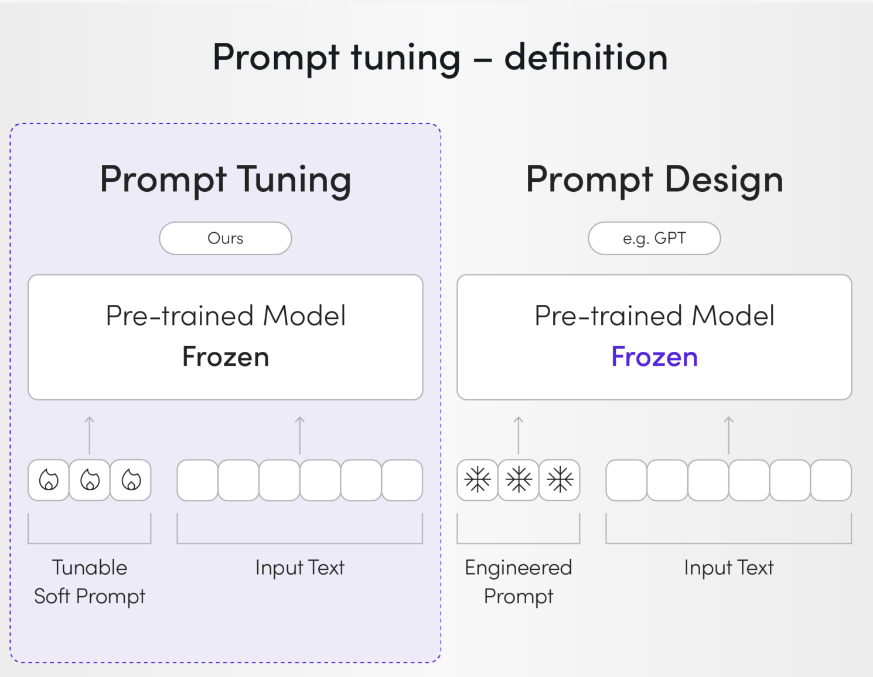


## **Prompt Tuning**

Prompt tuning is an efficient method for adapting a large language model (LLM) to a new task without the costly and time-consuming process of retraining the entire model. It works by adding and training a small set of "soft prompts"—learnable input parameters—that subtly guide the pre-trained model's output towards the desired behavior. The vast majority of the model's parameters remain unchanged, or "frozen".

#### **How Does Prompt-Tuning Work?**

* Prompt-tuning involves crafting specific input prompts – either manually or using AI-designed embeddings – that contextualize the task for the model. These prompts act as instructions, or cues, that steer the model to generate task-specific predictions or decisions. By giving the AI model a specific persona or context, it can enhance its relevance and efficiency.
* **Manual Prompts (Hard Prompts):** Originally, prompt-tuning relied on manually written instructions. For example, to adapt a language model for translation, a simple input like “Translate English to French: cheese” would prompt the model to respond with “fromage.” While effective for some tasks, manual prompts often require significant trial and error to achieve optimal results.
* **AI-Designed Prompts (Soft Prompts):** More advanced prompt-tuning methods replace hand-crafted prompts with embeddings—strings of numbers that are trained by AI. These embeddings, injected into the model’s input layer, are task-specific but don’t alter the model itself. This technique leverages the capabilities of large language models to handle complex tasks efficiently. The optimization of these prompts can be achieved without changing the model parameters, making the process more efficient.



* **Key Benefits of Prompt-Tuning:**
  + **Efficiency:** Since the model remains “frozen” (its parameters are unchanged), prompt-tuning eliminates the need for retraining. This reduces computing and energy usage by up to 1,000 times compared to fine-tuning, saving thousands of dollars. This approach is known as parameter efficient fine tuning, which optimizes the training process with reduced computational costs.
  + **Low Data Requirements:** Prompt-tuning is particularly useful for tasks with limited labeled data. Prompts essentially act as substitutes for additional training data by distilling the model’s knowledge for specific use cases.
  + **Scalability Across Domains:** Prompts can be tailored for diverse tasks, such as analyzing legal contracts, detecting fraud, or summarizing documents, without needing separate models for each task. This allows the model to excel at specific tasks by creating tailored prompts for each application.