

RAG

Enhances responses by retrieving relevant information from external knowledge bases in response to user queries.

# RAG

vs

# Prompt

# Engineering

vs

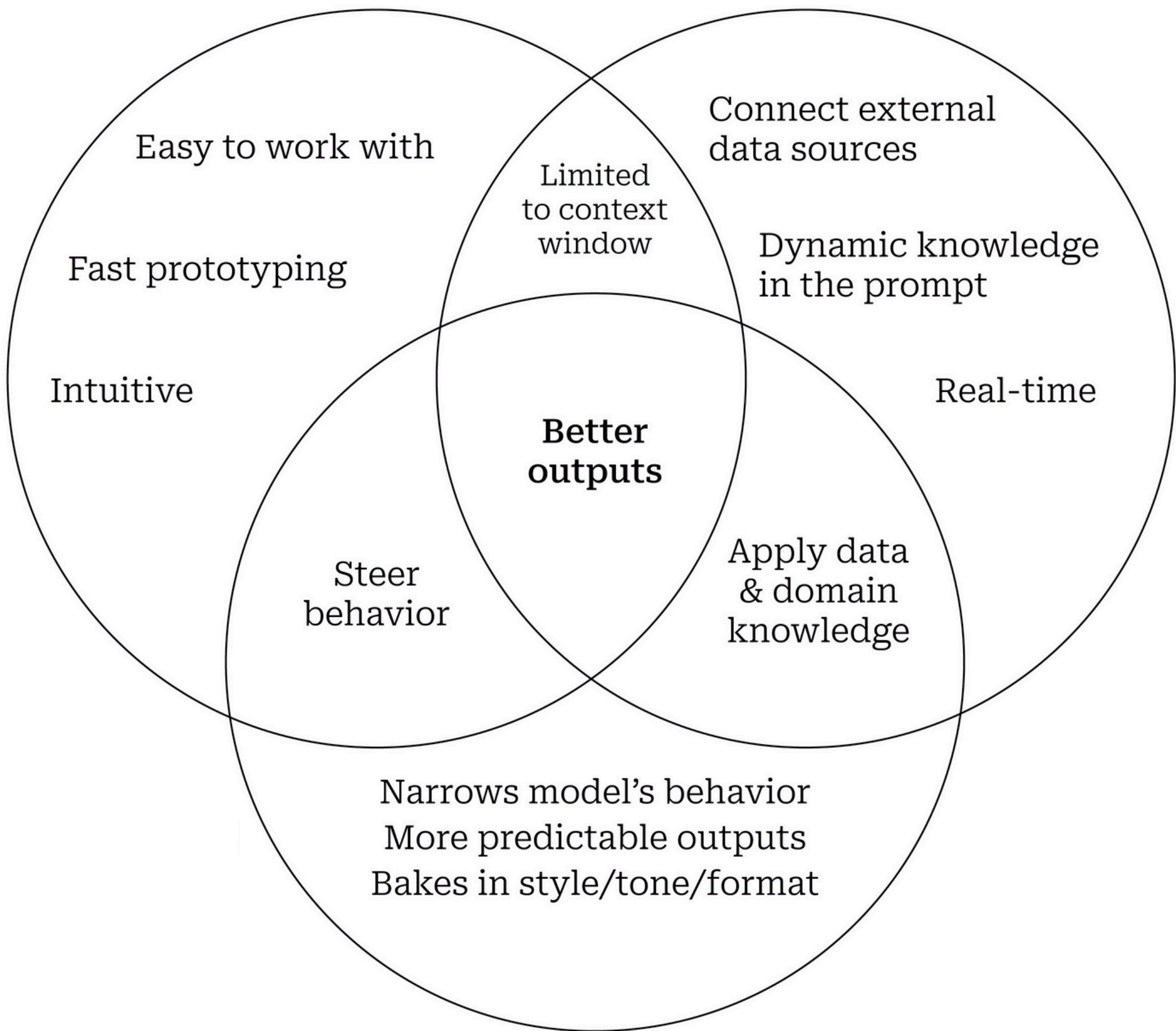
# Fine-tuning

# LLM

# Which to choose?

Prompt Engineering

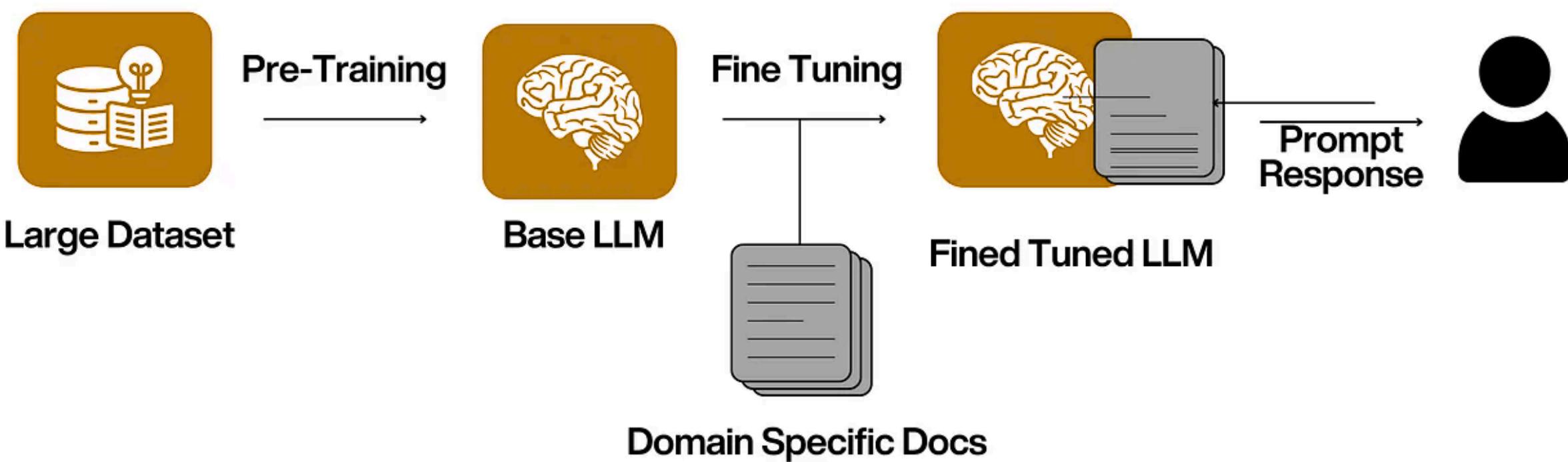
RAG



**Fine Tuning**

# What is Fine-tuning?

Fine-tuning involves adapting a pre-trained language model to a specific task by training it on a new, relevant dataset. This allows the model to learn the nuances of the target domain and improve its accuracy in generating text, translating languages, and other tasks.



## Adaptation

Fine-tuning modifies a pre-trained language model for a specific task, like writing summaries or translating languages.

## Domain Knowledge

The new dataset trains the model on specific vocabulary, syntax, and styles related to the target domain, increasing its accuracy.

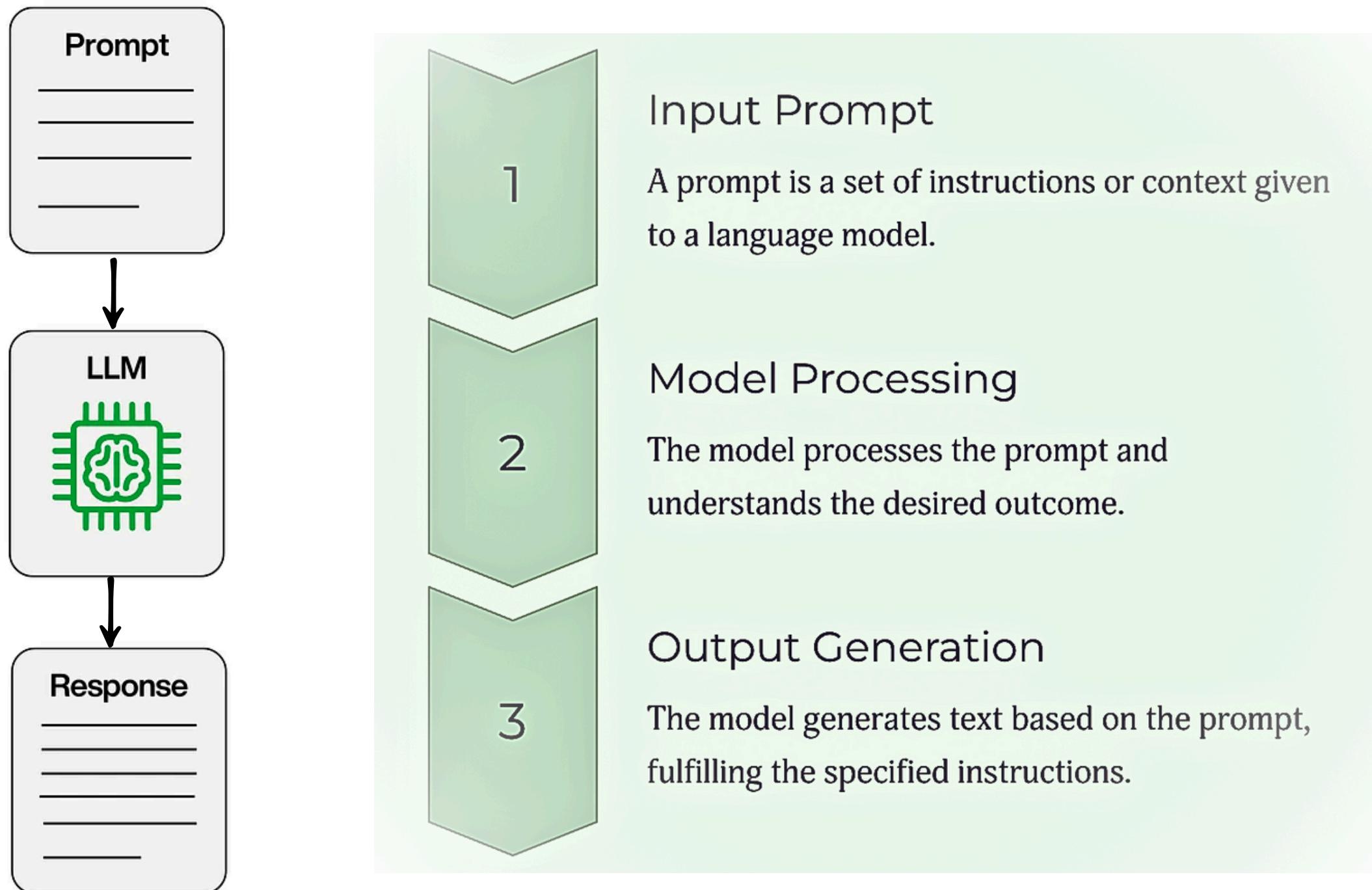
## Customization

Fine-tuning provides a way to tailor language models to suit unique needs and requirements, improving their performance on specific tasks.

- **Analogy:** Instead of asking your chef friend for pasta, you send them to cooking school for a month to master your family's pasta recipe. Now they'll always cook it your way.
- **Pros:** Very accurate for niche tasks, captures tone/style/domain.
- **Cons:** Costly, slower, needs lots of quality data.

# What is Prompting?

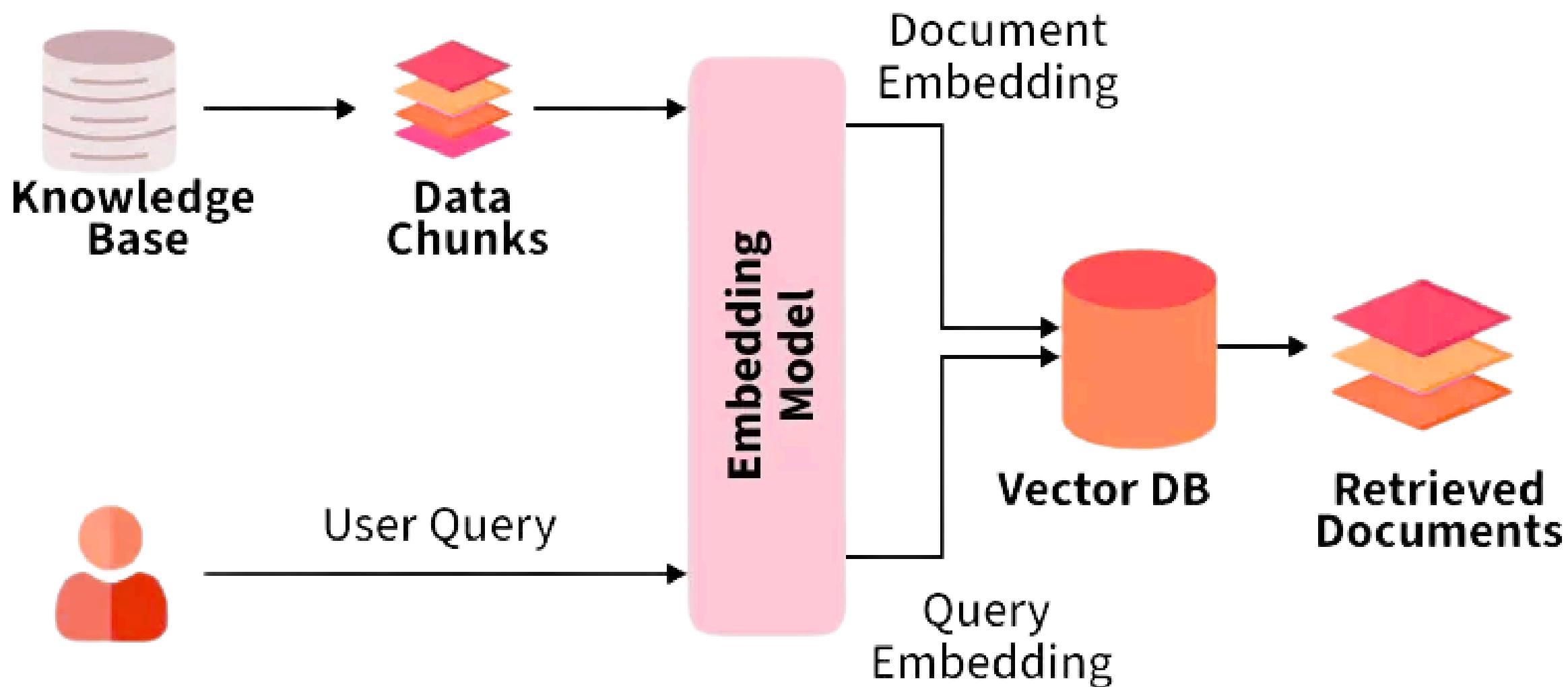
Prompting involves providing a language model with specific instructions or context in the form of text prompts. This helps guide the model towards desired outputs and allows for more control over the generated text.



- **Analogy:** Like asking your super-talented chef friend to cook pasta. The way you describe the dish (extra cheese, no garlic, spicy) decides how it turns out.
- **Pros:** Cheap, fast, zero training.
- **Cons:** Limited, the chef only knows what's already in their head.

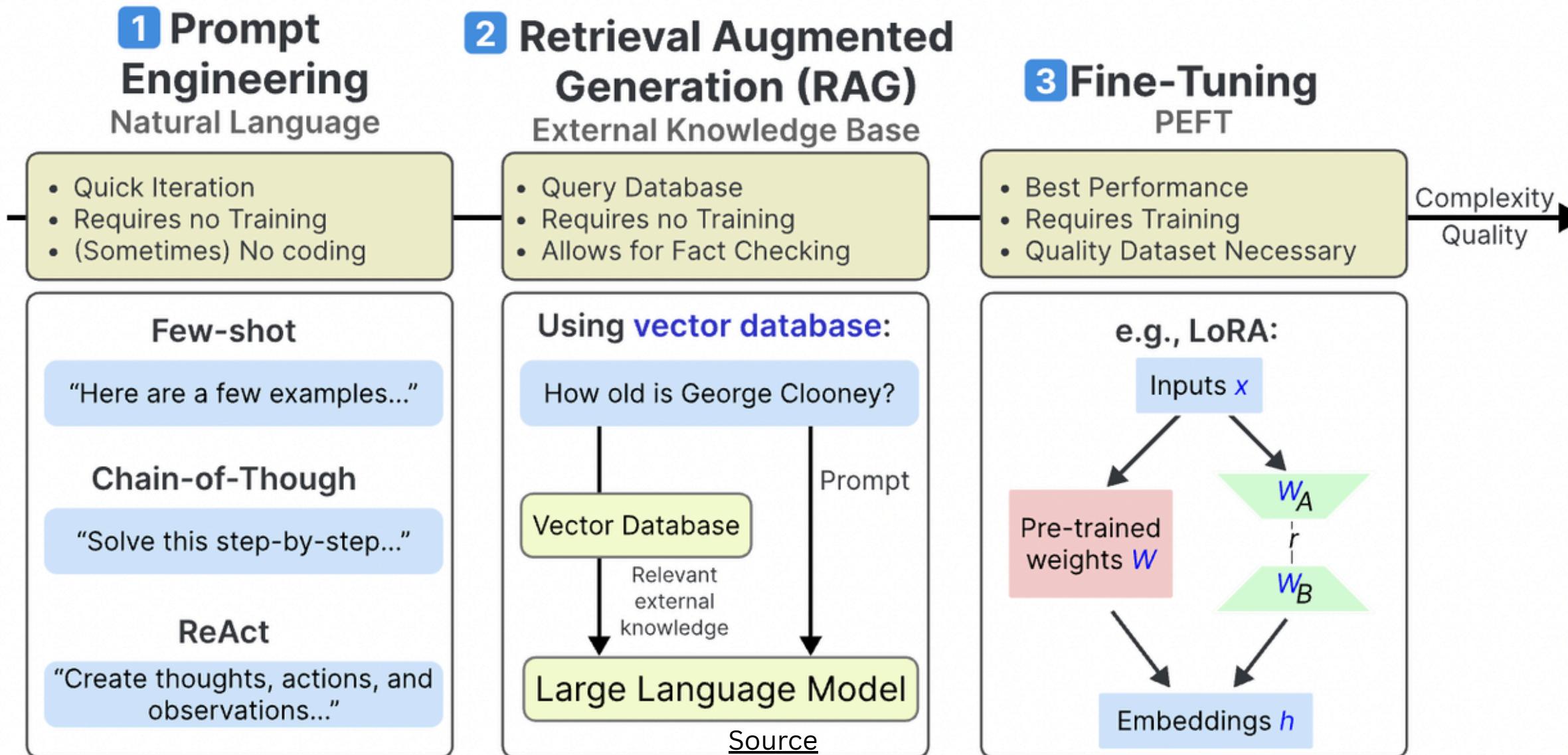
# What is RAG?

RAG combines the capabilities of language models with a knowledge base, allowing them to access and incorporate external information into their responses. This enables the model to generate more accurate and comprehensive outputs.



- **Analogy:** You don't change the chef. You just give them access to a giant recipe book every time you ask for a dish. They'll look up what's needed while cooking.
- **Pros:** Keeps answers up-to-date, no need to retrain, great for dynamic knowledge.
- **Cons:** Needs infra setup (databases, embeddings), retrieval quality is crucial.

# Choosing the Right Approach



- Fine tuning is ideal for task-specific applications with ample labeled data.
- Prompting offers a flexible and cost-effective way to leverage pre-trained models.
- RAG stands out by integrating retrieval capabilities, enhancing the accuracy and relevance of generated responses. Understanding these methods allows practitioners to choose the most suitable approach for their specific needs and objectives.



**Follow to stay updated on  
Generative AI**



**SAVE**



**LIKE**



**REPOST**