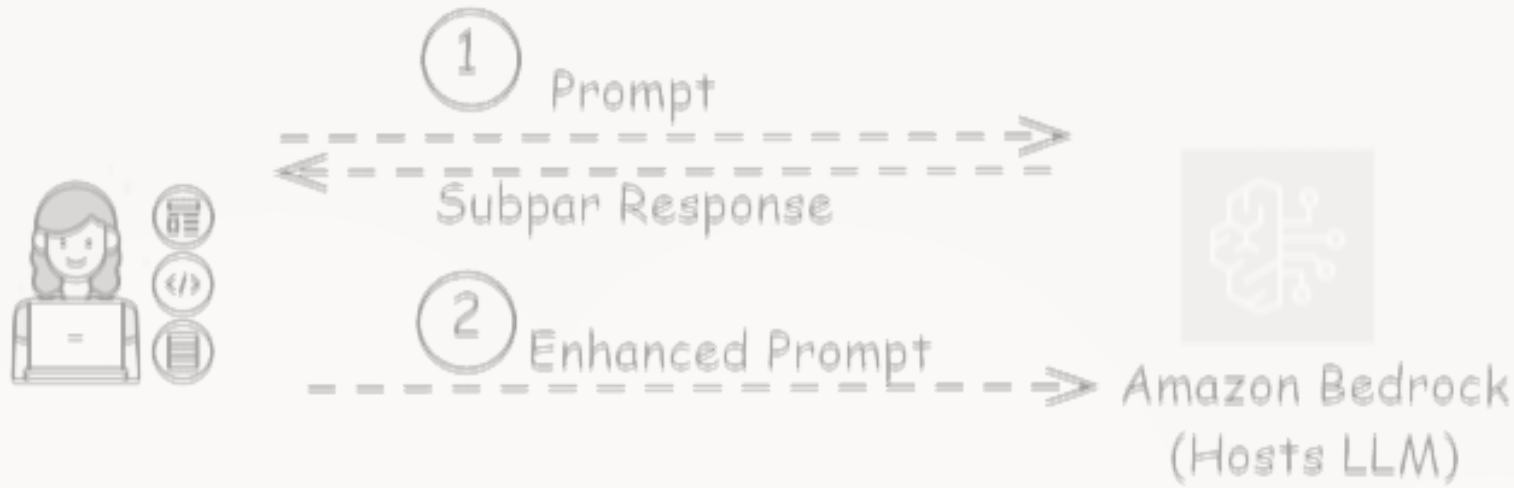
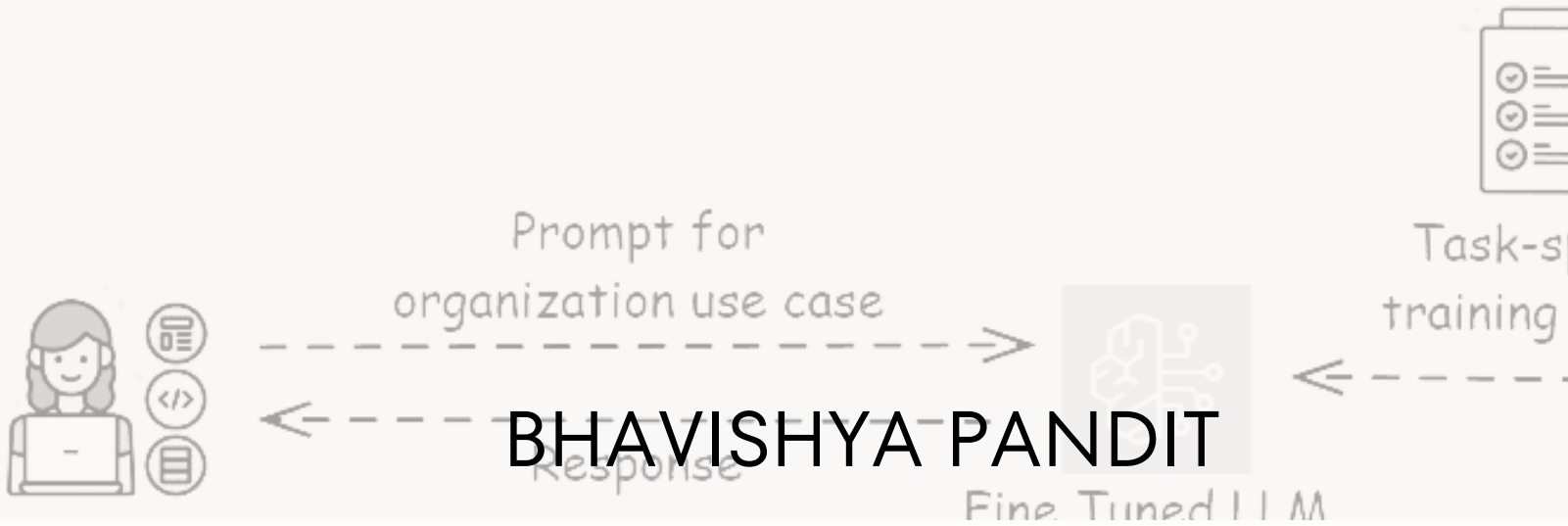


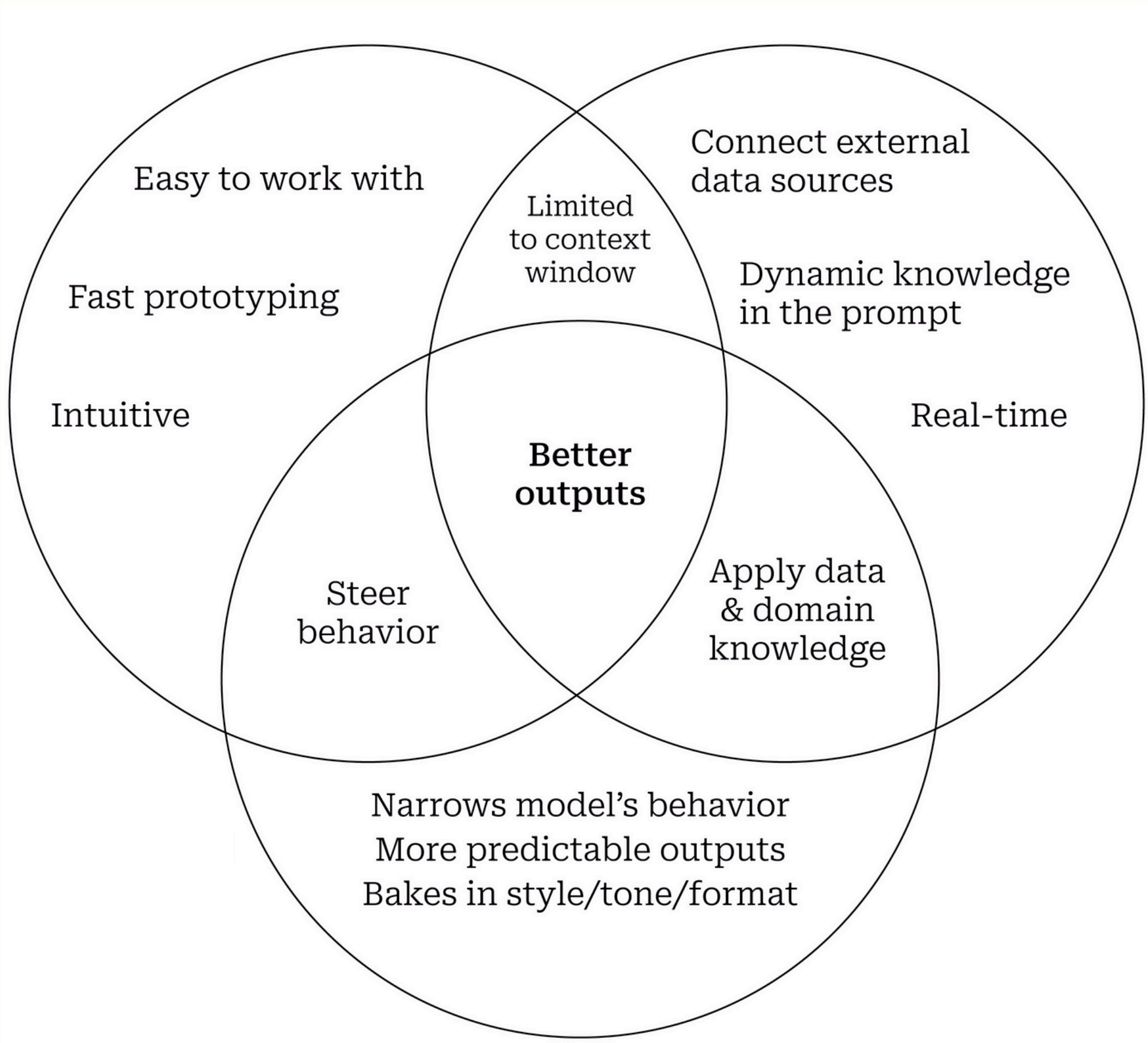
Fine Tuning vs Prompting vs RAG



BHAVISHYA PANDIT

Which technique to choose?

Prompt Engineering **RAG**



Fine Tuning

BHAVISHYA PANDIT

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.

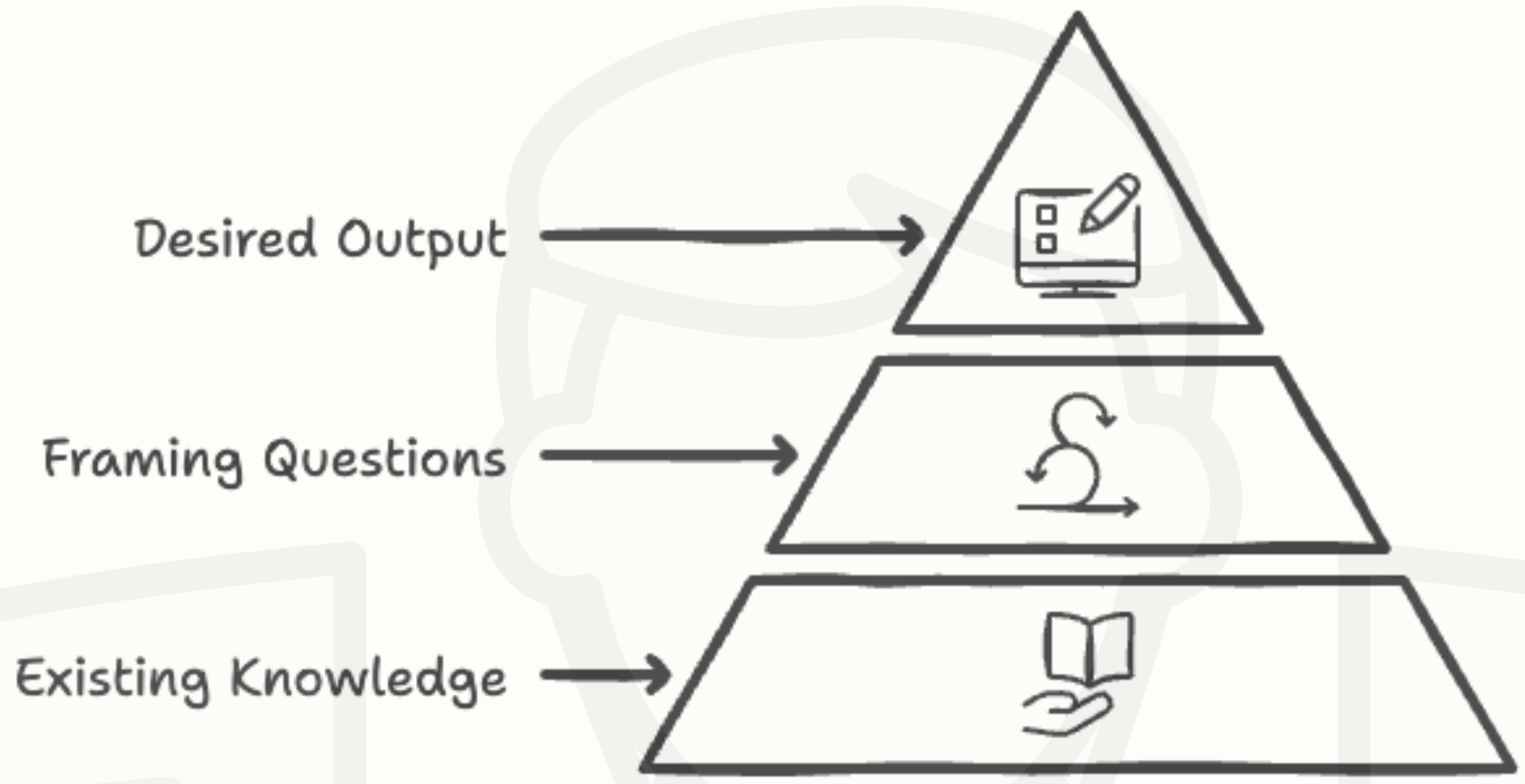
Domain Knowledge

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.

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.

1

Input Prompt

A prompt is a set of instructions or context given to a language model.

2

Model Processing

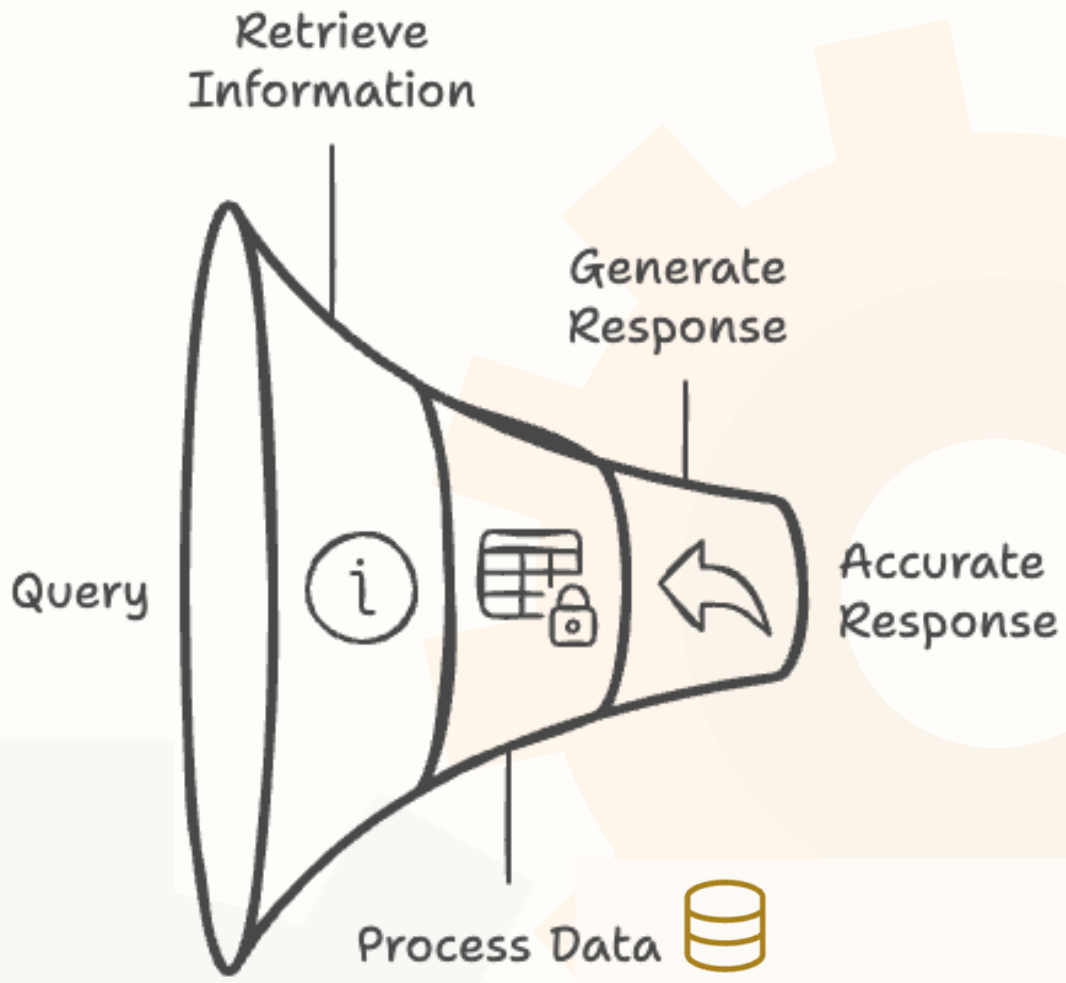
The model processes the prompt and understands the desired outcome.

3

Output Generation

The model generates text based on the prompt, fulfilling the specified instructions.

What is Retrieval Augmented Generation (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.

Knowledge Base

A RAG system utilizes a database of information, such as articles, documents, or websites, to augment the language model's understanding.

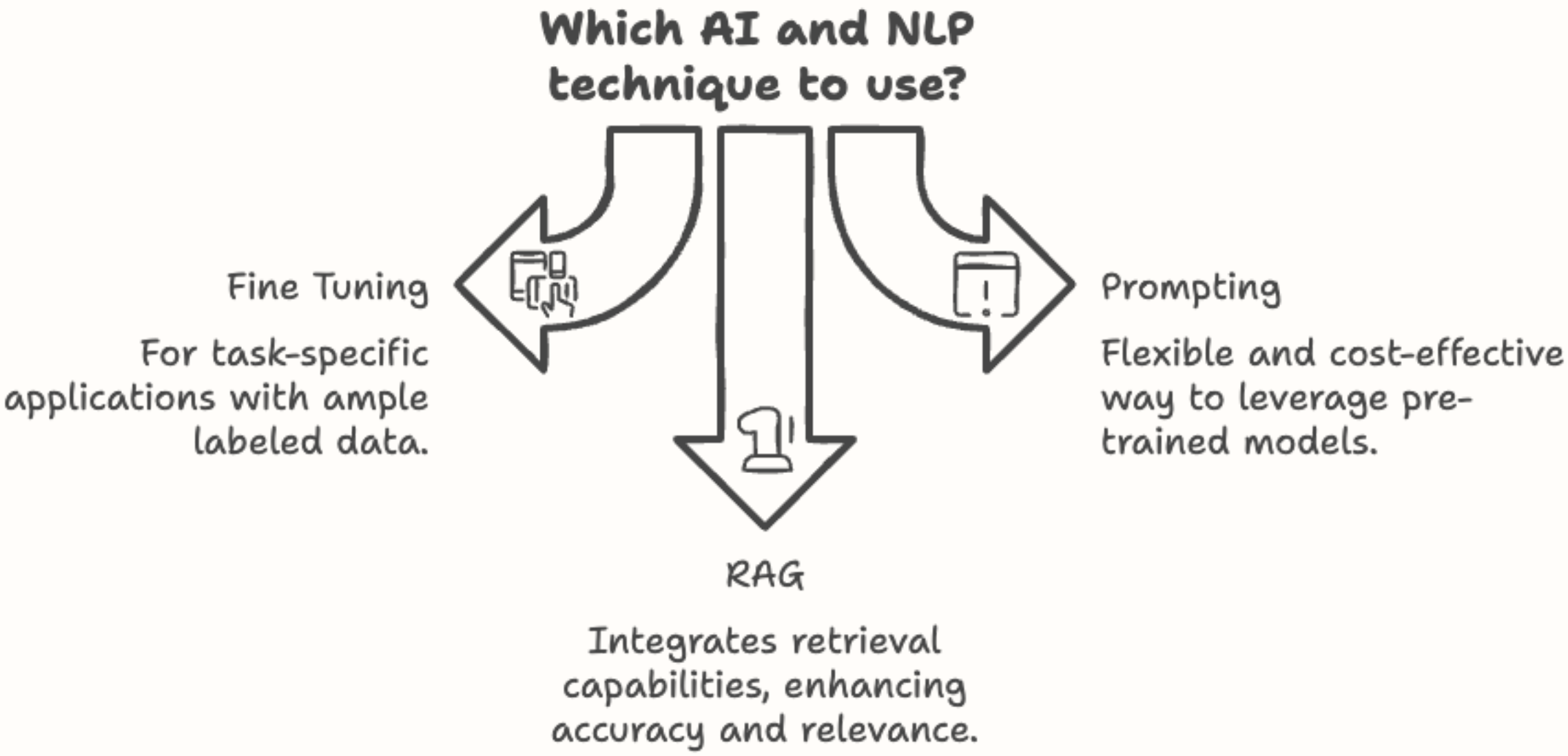
Information Retrieval

The model retrieves relevant information from the database based on the input prompt or query.

Generation

The model integrates the retrieved knowledge into its response, providing more comprehensive and contextually relevant outputs.

Conclusion: Choosing the Right Approach



Fine Tuning, Prompting, and RAG are three distinct techniques that serve different purposes in the realm of AI and natural language processing.

- 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 FOR MORE AI/ML POSTS

BHAVISHYA PANDIT