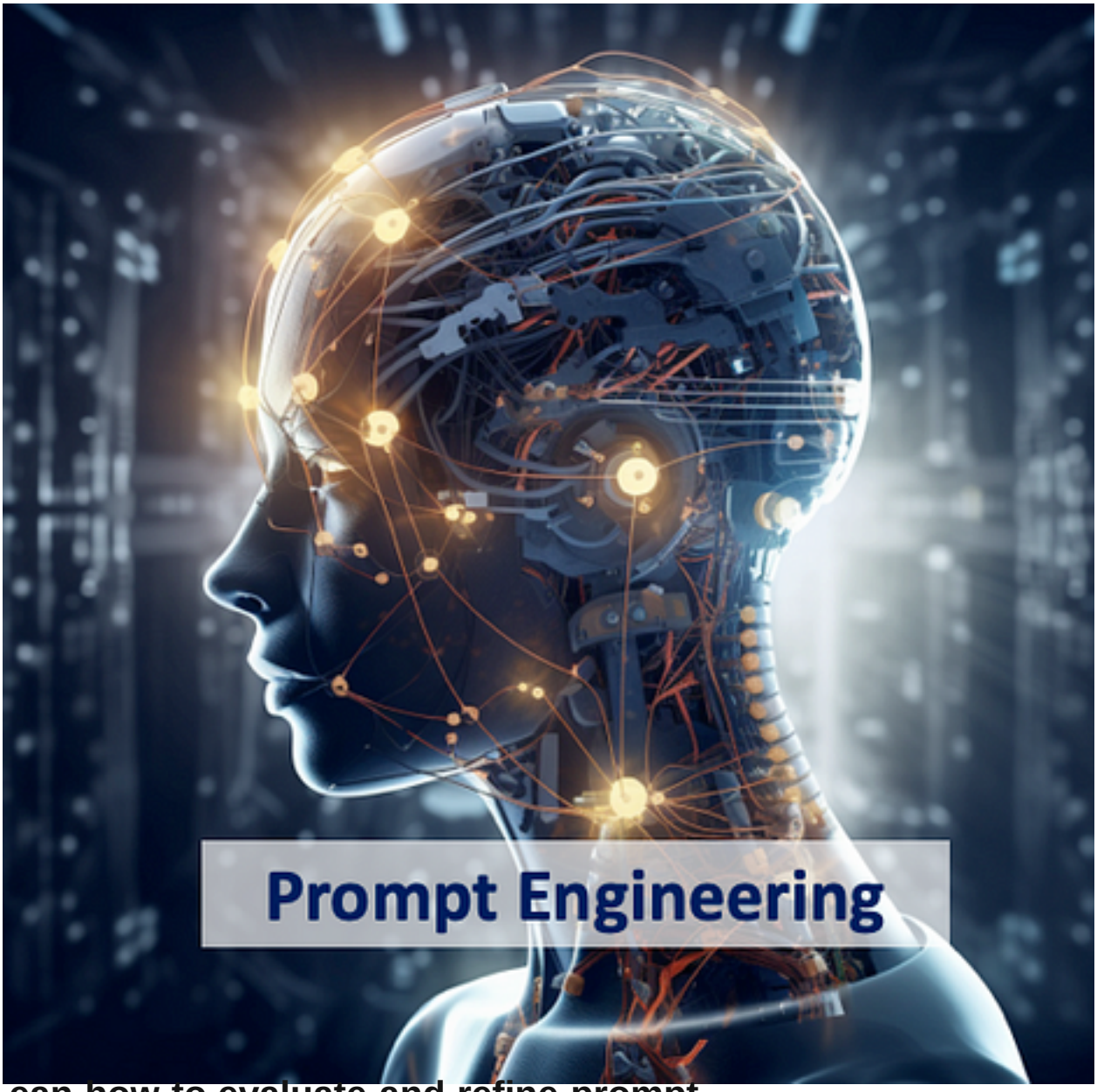
Prompt Engineering — Evaluating and Refining Your Prompt

Lean how to evaluate and refine prompt

Tony • 5 min read • 2023-06-01

https://tonylixu.medium.com/prompt-engineering-evaluating-and-refining-your-prompt-6baf455ef451



Lean how to evaluate and refine prompt

> Note, full "Prompt Engineering" mind map is available at: "**Prompt Engineering Mind Map**"

What is Prompt Evaluation and Refining?

Evaluating and refining prompts involve assessing the quality and effectiveness of prompts you

provide to an AI model and making necessary improvements to achieve desired outputs.

Evaluating Prompts

This refers to the process of testing and assessing how well a given prompt works in instructing an AI model to generate desired responses. This typically involves:

• Implementing the prompt and observing the output

• Assessing the accuracy and relevance of the response

• Evaluating if the AI model adhered to the prompt's intent

• Comparing the output to your expected or desired result

Refining Prompts

Based on the evaluation, you might find that a prompt needs to be improved or refined. This could mean:

• Making instructions more explicit

• Adding more context or detail

• Changing the phrasing or structure of the prompt

• Modifying the prompt to address any issues observed during evaluation

Example one

```
Initial Prompt: "Tell me a joke."
Evaluation: The AI might respond with a joke, but it might not be appropriate
for all audiences.

Refined Prompt: "Tell me a family-friendly joke."
```

Example two

```
Initial Prompt: "Summarize the news."
Evaluation: The AI might not know which news article you're referring to,
leading to a vague or generic response.

Refined Prompt: "Summarize the key points of the following news article: [insert
article text here]"
```

This iterative process of evaluation and refinement is vital for getting the most out of AI models. It helps tailor the model's output to your specific needs and improves the overall effectiveness and accuracy of the system.

Why Evaluation and Refining is Crucial?

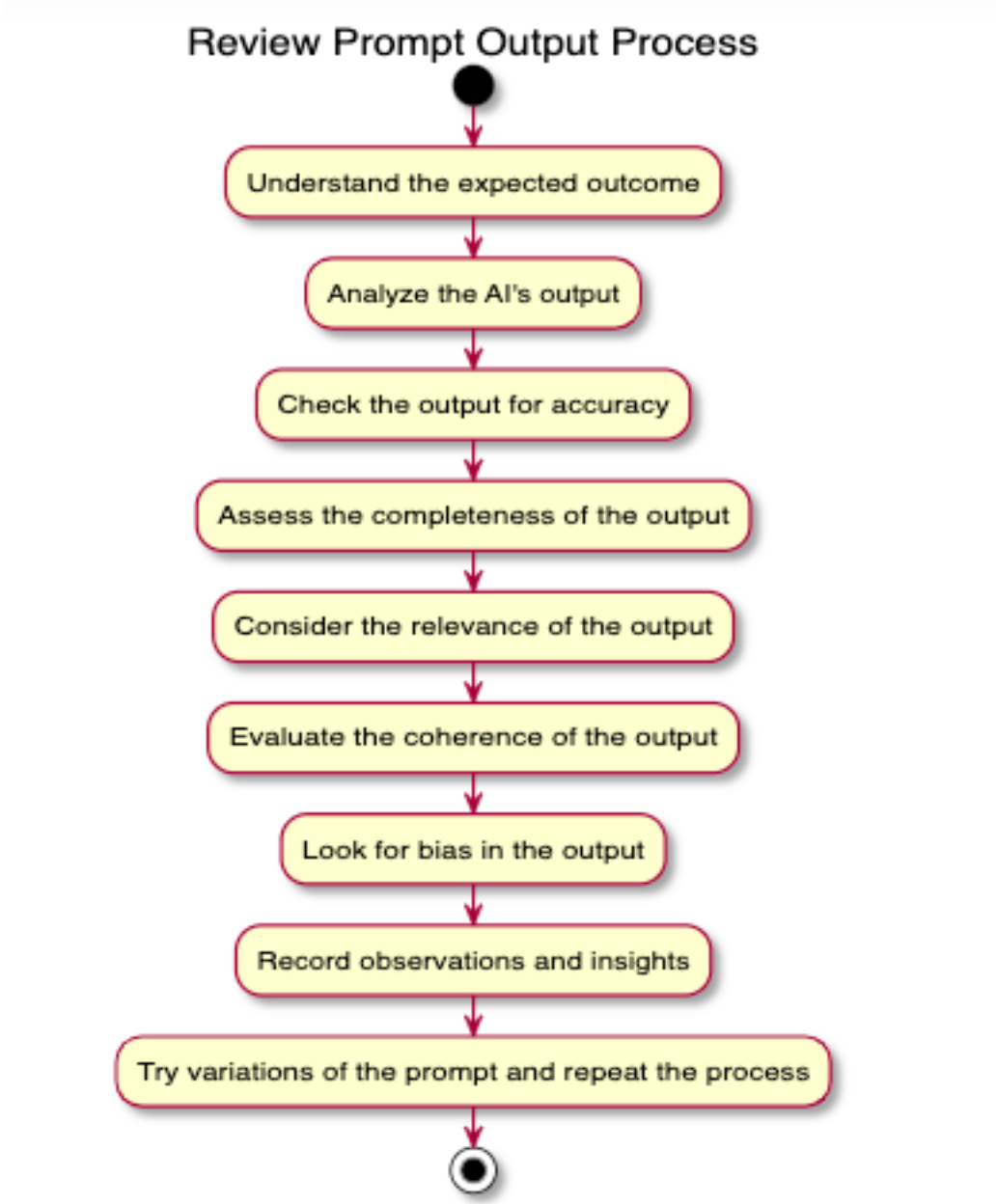
By constantly evaluating and refining prompts, you ensure that your application remains effective, efficient, and user-friendly over time. This also helps you gain deeper insights into the model's

behavior, enabling you to make better use of the technology.

Several reasons that you need to evaluating and refining:

- **Improve Accuracy**: The accuracy of the AI's responses largely depends on the quality and clarity of the prompts. Evaluation and refinement ensure that prompts guide the model to produce accurate and useful outputs.
- **Performance Optimization**: Different prompts can result in varying levels of performance, even when achieving the same goal.
- **Context Sensitivity**: AI models often need precise and contextual information to generate accurate responses. Refining prompts helps in incorporating relevant context and fine-tuning the amount of detail required.
- **Avoiding Misinterpretations**: AI models can sometimes misinterpret prompts due to vagueness or ambiguity. Evaluating and refining prompts help in making instructions explicit and reducing the chance of misinterpretation.
- **Cost-Efficiency**: More effective prompts can lead to correct results in less time, thus reducing computational cost and making your application more efficient.

Step one: Review Output



Suppose we give an AI model the following prompt:

```
“Translate the following English text to Spanish: ‘Hello, how are you?’”
```

The AI model produces this output:

```
“Hola, ¿cómo estás?”
```

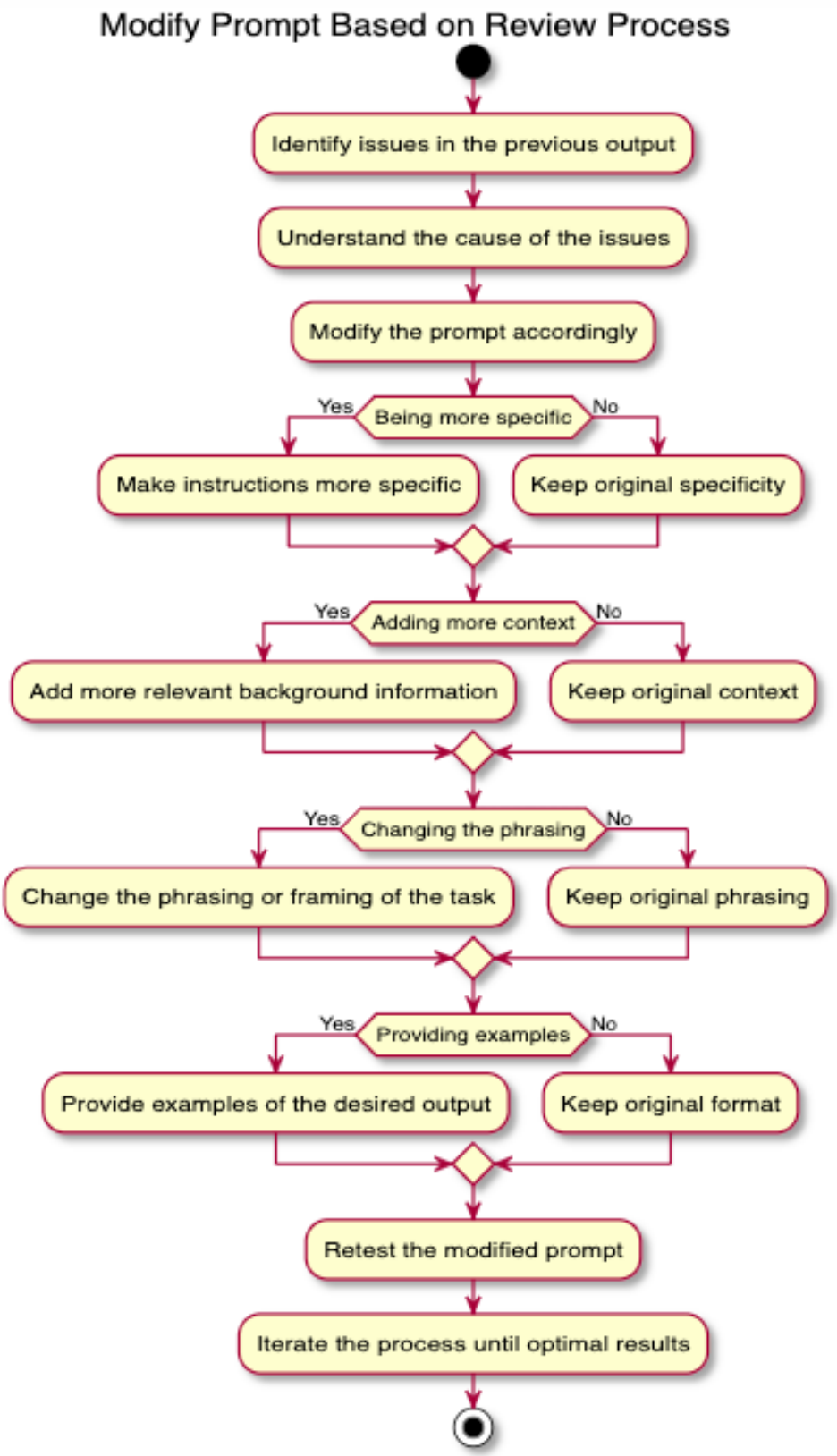
We then review this output as follows:

1. **Understand the Expected Outcome**: The expected outcome here is a correct translation of the English phrase into Spanish.
2. **Analyze the Output**: The AI model has provided a response in Spanish.
3. **Check for Accuracy**: The translation is accurate, as “Hola, ¿cómo estás?” is the correct

Spanish translation for “Hello, how are you?”

4. **Assess Completeness**: The AI has translated the entire sentence, so the response is complete.
5. **Consider Relevance**: The output is relevant as it’s a direct response to the prompt.
6. **Evaluate Coherence**: The translated sentence is coherent and makes sense in the target language.
7. **Look for Bias**: In this particular case, there’s no bias involved as it’s a simple translation task.
8. **Record Observations**: Given that the AI model has produced an accurate, relevant, and complete response, we record that the prompt worked well in this case.
9. **Repeat with Variations**: For a comprehensive evaluation, we could try the same prompt with different sentences and observe the results.

Step two: Modify Prompt



Guidelines:

- **Identify Issues**: Based on the output review, identify the specific issues or areas that need improvement. Did the AI misunderstand the prompt? Did it produce an inaccurate result, or did it miss some information?
- **Understand the Cause**: Try to understand why the AI might have produced the undesired

result. Is the prompt too vague or ambiguous? Was the context insufficient, or was the instruction unclear?

- **Modify the Prompt**: Make adjustments to your prompt to address the identified issues. This could involve:

```
1. Being more specific: If the prompt was too vague, try to be
more specific in your instruction.
2. Adding more context: If the AI lacked the necessary context to
understand the prompt, provide more background information.
3. Changing the phrasing: If the AI misunderstood the prompt,
consider changing the phrasing or the way you're framing the task.
4. Providing examples: In some cases, providing an example of the
desired output can help guide the AI.
```

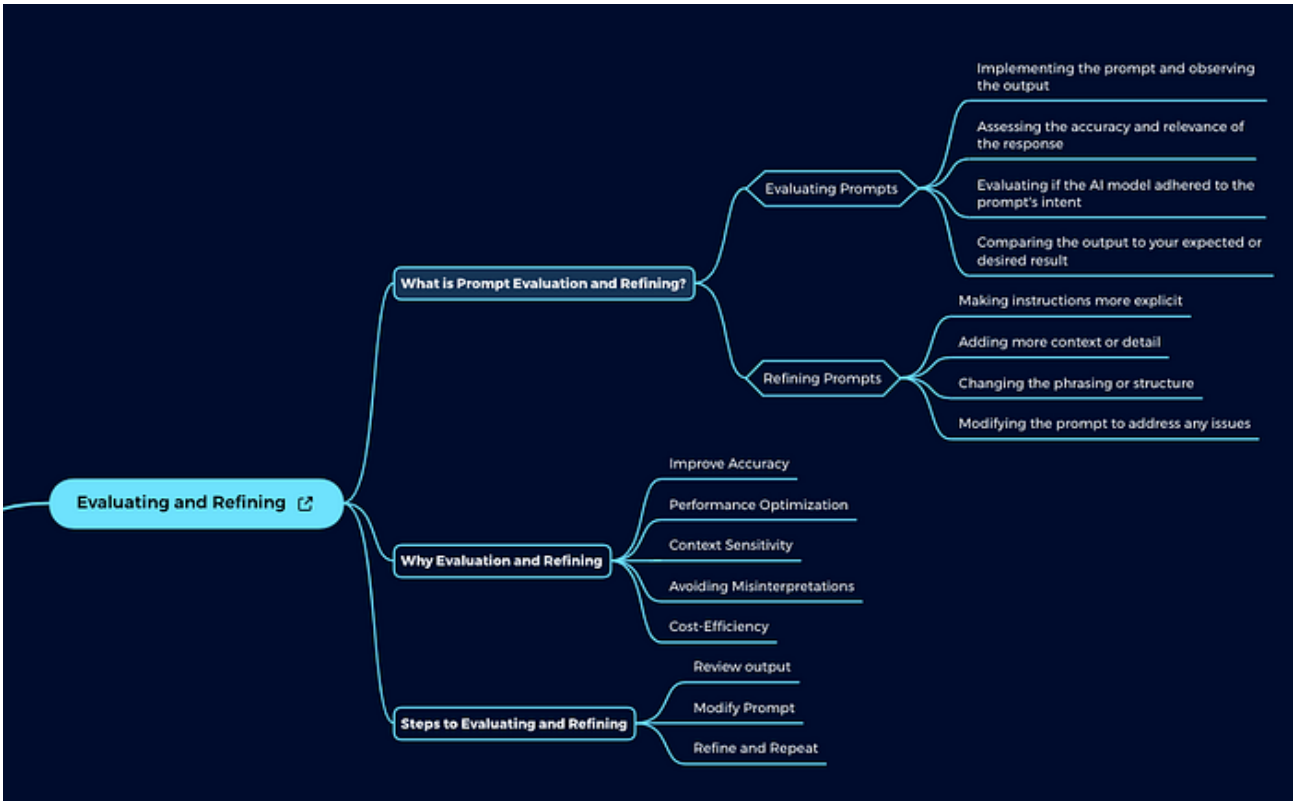
- **Retest the Prompt**: Once you've modified the prompt, use it to generate a new output from the AI. Repeat the review process to evaluate if the modifications have led to improvements.
- **Iterate**: This process might need to be repeated several times before you find the most effective prompt. Keep iterating and refining based on your observations and learnings.

For example, if you originally prompted a model with "Summarize the following text," and found the summaries were too brief, you might adjust the prompt to be "Provide a detailed summary of the following text," or even, "Summarize the following text in 5 sentences."

Step three: Refine and Repeat if necessary

If needed, refine and repeat the above two process. For enhanced control, it's advisable to incorporate slight adjustments during each iteration. The more experience you gain in crafting prompts for a specific model, the fewer the iterations required to obtain optimal results.

Conclusion


Evaluating and Refining
What is Prompt Evaluation and Refining?
Evaluating Prompts
Implementing the prompt and observing the output
Assessing the accuracy and relevance of the response
Evaluating if the AI model adhered to the prompt's intent
Comparing the output to your expected or desired result
Refining Prompts
Making instructions more explicit
Adding more context or detail
Changing the phrasing or structure
Modifying the prompt to address any issues
Why Evaluation and Refining
Improve Accuracy
Performance Optimization
Context Sensitivity
Avoiding Misinterpretations
Cost-Efficiency
Steps to Evaluating and Refining
Review output
Modify Prompt
Refine and Repeat