

# Introduction to the Reflection Pattern and its Core Components

**Created by:**

Eleni Verteouri

Gen AI Tech Lead @ UBS

**Created & Narrated by:**

Dipanjana Sarkar

Head of Community & Principal AI Scientist @ Analytics Vidhya

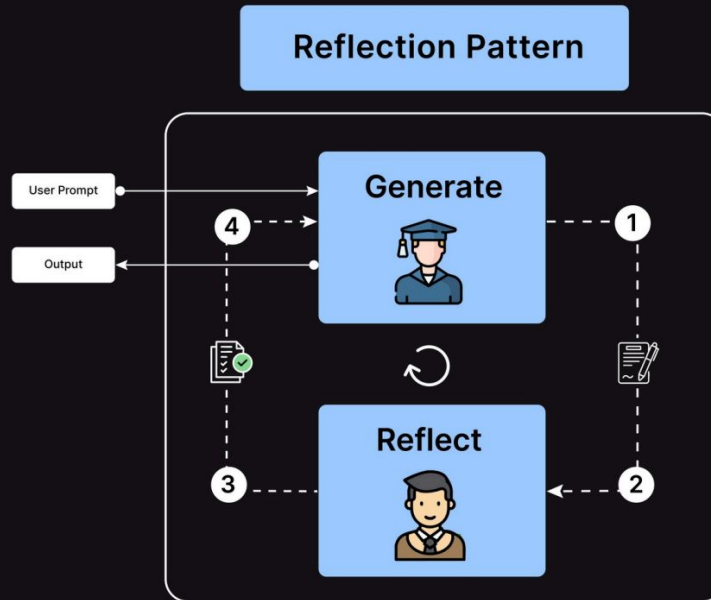
Google Developer Expert - ML & Cloud Champion Innovator

Published Author



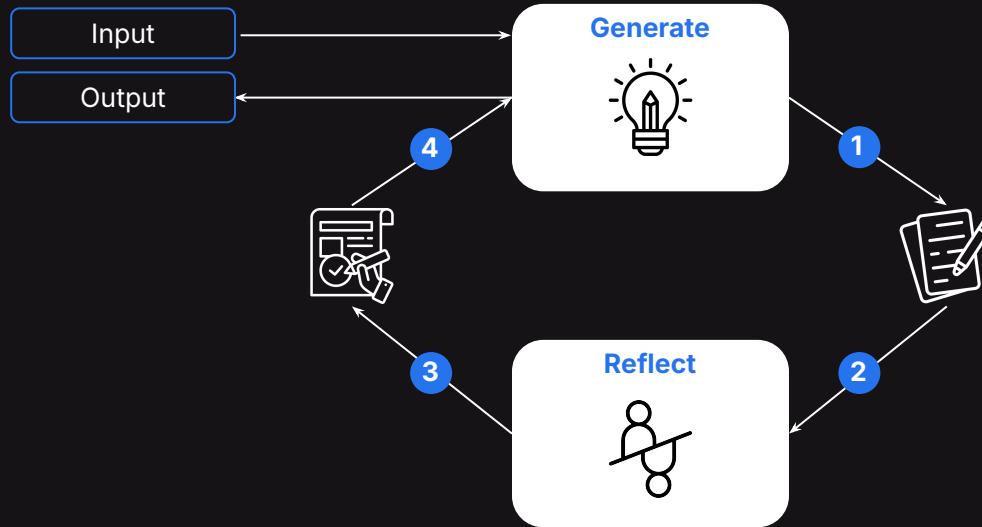
# What is the Reflection Pattern?

The reflection pattern helps the LLM refine its outputs in a continuous feedback loop, as shown:



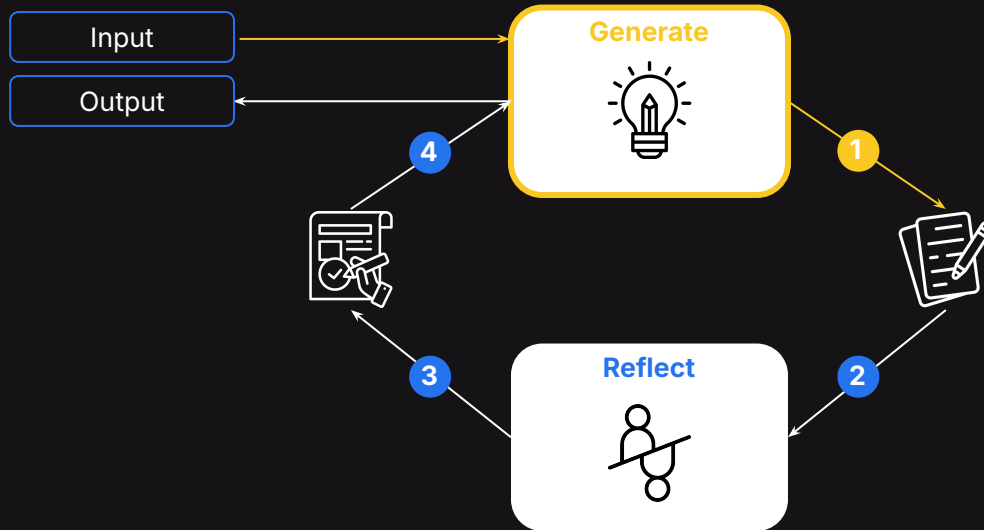
# A Deep Dive into the Basic Components

## Architectural Overview:



# Basic Components of the Reflection Pattern

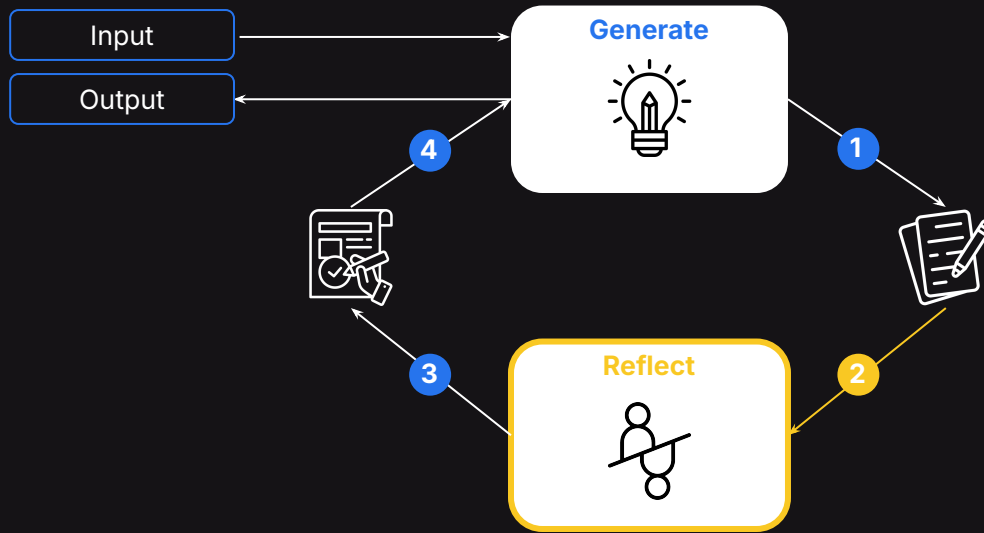
## The Generation Phase:



**Generate:** The LLM creates an initial response based on the user prompt.

# Basic Components of the Reflection Pattern

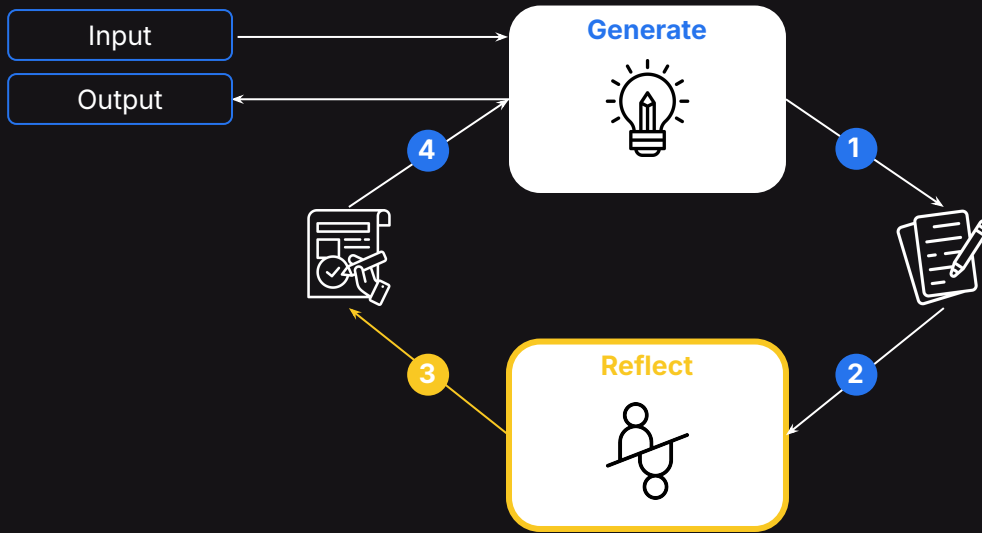
## The Reflection Phase:



**Reflect:** The LLM reviews its output, identifying improvement areas, such as content, relevance, clarity, or style.

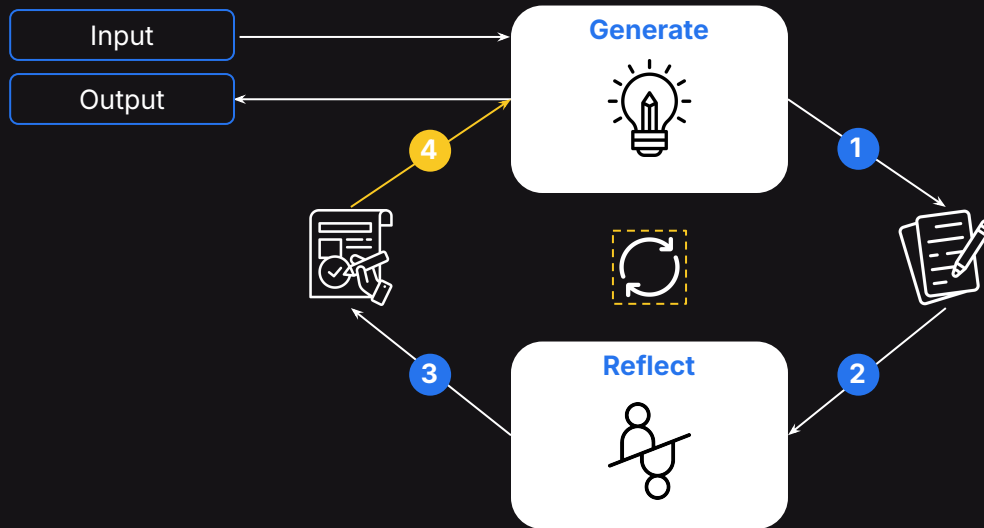
# Basic Components of the Reflection Pattern

## The Modification/Iteration Phase:



**Modify:** The LLM revises the output based on its reflections, enhancing the response.

# Basic Components of the Reflection Pattern



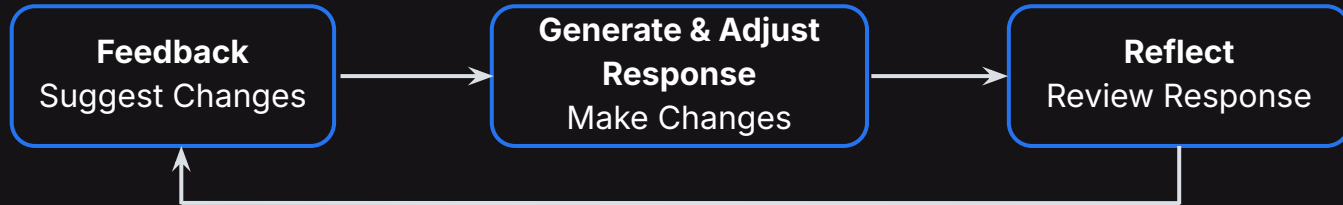
**Repeat:** The process loops as needed, allowing iterative refinements for generating the most appropriate final response.

# Why does Adaptability Matter?

- **Dynamic Environments:** AI systems operate in contexts where data, goals, or conditions evolve rapidly.

**Example:** Customer preferences in automated service systems change frequently.

- **Sustainability:** Adaptability ensures AI remains effective over time, avoiding obsolescence.

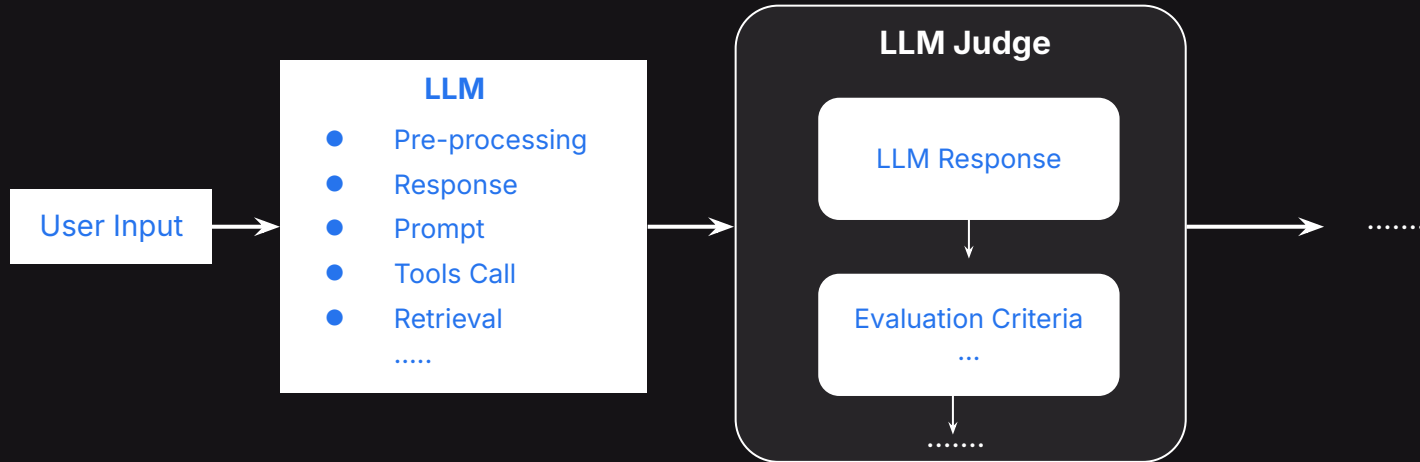




# How to Critique and Improve Generated Content?

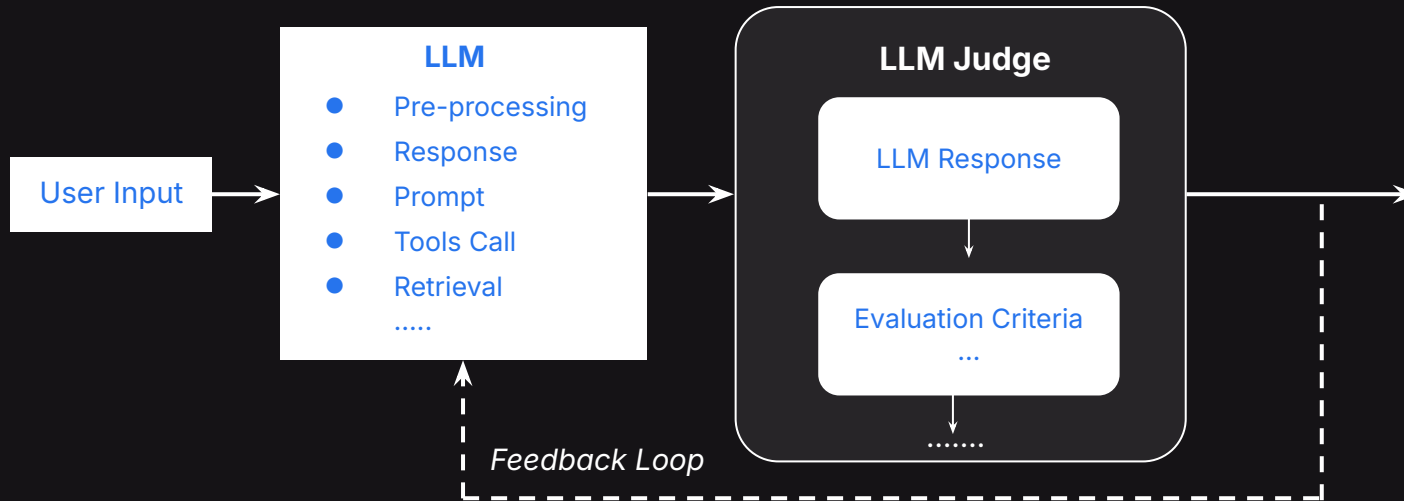
## LLMs as a Judge:

- LLM-as-a-Judge involves using an LLM to score or grade another LLM's response based on a defined prompt and criteria like relevance, specific conditions, domain guidelines and more.
- It powers evaluation methods to improve RAG and Generative AI Systems **but can also be used to give feedback in Reflection-powered Agentic AI Systems to improve responses.**



# How is LLM as a Judge used for Reflection:

- **Single Output Scoring (No Reference):** The judge LLM scores a response based on criteria like input quality or domain guidelines — no ideal answer is given.
- **Single Output Scoring (With Reference):** The judge LLM compares the response to an ideal answer for more reliable scoring.
- **The first method is the most commonly used method in the Reflection Pattern**



# Top LLM-as-a-Judge Scoring Methods

**G-Eval:** This framework uses **CoT prompting** to stabilize and make LLM judges more reliable and accurate in terms of metric score computation

## Task Introduction:

You will be given one summary written for a news article. Your task is to rate the summary on one metric .....

## Evaluation Criteria:

Factual Accuracy (1–5) – The extent to which the summary accurately reflects facts presented in the article, without introducing errors or hallucinations.

Auto CoT or User-Specified

## Evaluation Steps:

- Read the news article carefully and identify all factual details (e.g., names, dates, events).
- Compare the summary to the article. Check whether it presents the correct facts without distortions or false additions.
- **Assign a score for factual accuracy on a scale of 1 to 5, where 1 is the lowest and 5 is the highest based on the Evaluation Criteria.**

## Task Article:

Article: NASA confirmed today that its Artemis II mission will launch in May 2025. The crew of four includes Christina Koch, Victor Glover, Reid Wiseman, and Jeremy Hansen. This will be the first crewed mission to orbit the Moon since Apollo 17.

## Task Summary:

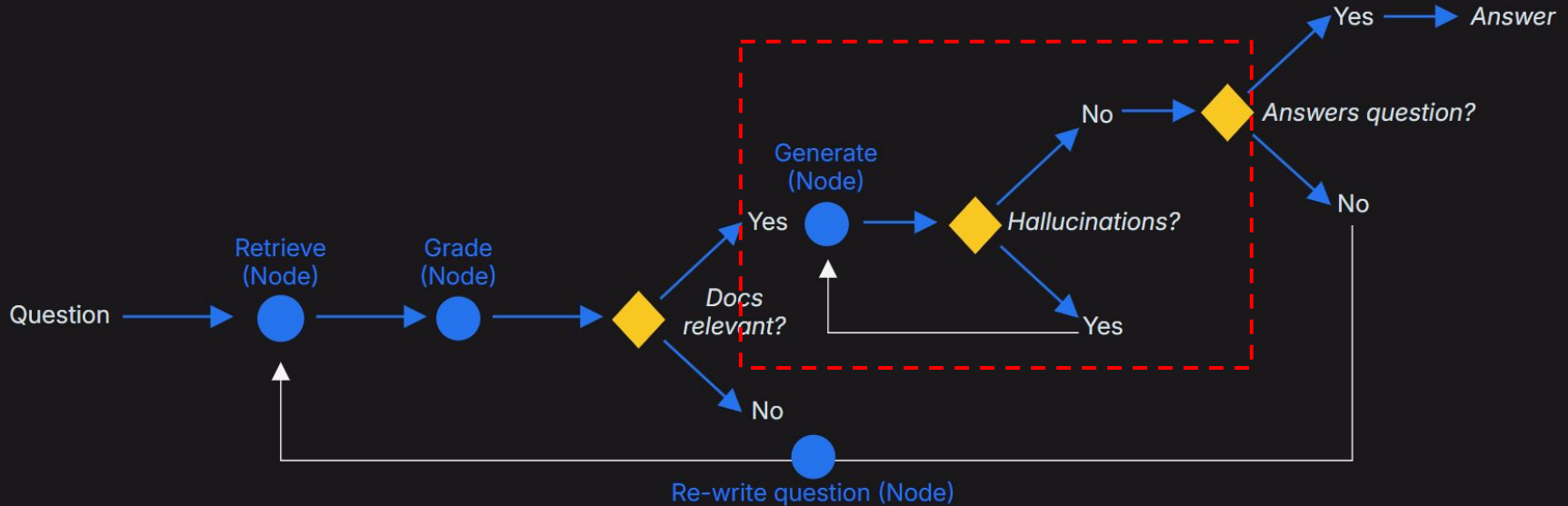
Summary: NASA's Artemis II mission is set for May 2025 and includes four astronauts: Christina Koch, Victor Glover, and two engineers from Canada. It will land on the Moon for the first time since 1972.

**LLM-as-a-Judge Evaluation(scores ONLY):** Factual Accuracy:

- Omitted Engineers in summary
- False additions of location and year
- Mission name, date, key members are correct
- **Overall Score: 3**

**Note:** The above is just a representative example

# LLM-as-a-Judge in Agentic RAG



- Use standard vector database retrieval for context based on query
- Leverages agentic reflection pattern to use an LLM to reflect on the context and check for relevancy
- Also checks for hallucinations and if the question is answered using the same pattern to make the system more accurate

# Drawbacks of LLM as a Judge



## **Non-Deterministic Scores:**

LLMs may give different scores for the same output at different times.



## **Narcissistic Bias:**

Some LLMs tend to favor their own outputs (e.g., GPT-4, Claude-3.5).



## **Verbose Preference:**

LLM judges often prefer longer responses over concise ones.



## **Low Granularity:**

Fine-grained scoring (e.g., 1-10) can be inconsistent and unreliable.



## **Position Bias:**

LLMs may favor the first option in pairwise comparisons.

# Applications of the Reflection Pattern



## AI-Assisted Writing

Generating drafts, critiquing them, and refining for clarity and tone.



## Code Generation & Debugging

Identifying bugs, critiquing generated fixes, and iterating for robust solutions.



## Design Tools

Creating prototypes, reviewing them, and refining designs based on feedback.

# Thanks!