Agentic AI

M2: Reflection Design Pattern

DeepLearning.AI

Reflection Design Pattern

Reflection to improve outputs of a task

DeepLearning.AI

Reflection - humans

Write an email

Didn't sign my name

Hey Tommy,

I'll be in New York next month, let me know if you'll be fre for dinner one night.

Email V1

Specific dates

Typo

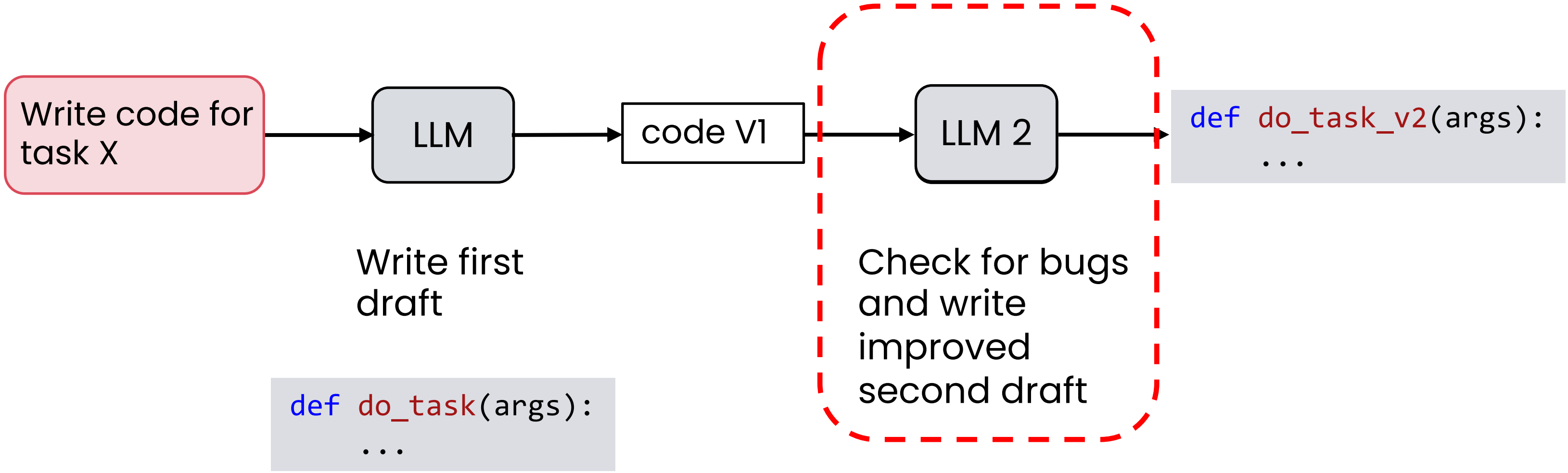
Hey Tommy,

I'll be in New York next month from the 5th–7th. Let me know if you'll be free for dinner one night.

Andrew

Email V2

Reflection to improve code



Write code for task X

LLM

code V1

LLM 2

```
def do_task_v2(args):
    ...
```

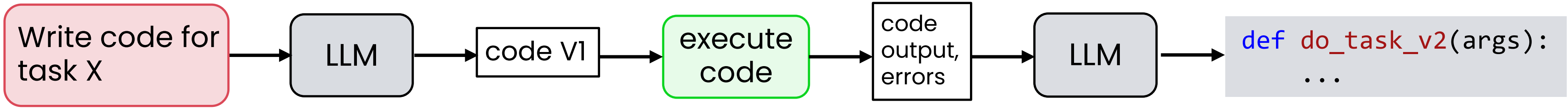
Write first draft

```
def do_task(args):
    ...
```

Check for bugs and write improved second draft

Can be a different LLM, e.g. a reasoning model

Reflection with external feedback



```
def do_task_v2(args):
    ...
```

Write first draft

```
def do_task(args):
    ...
```

```
SyntaxError:
unterminated string
literal (detected at
line 1)
```

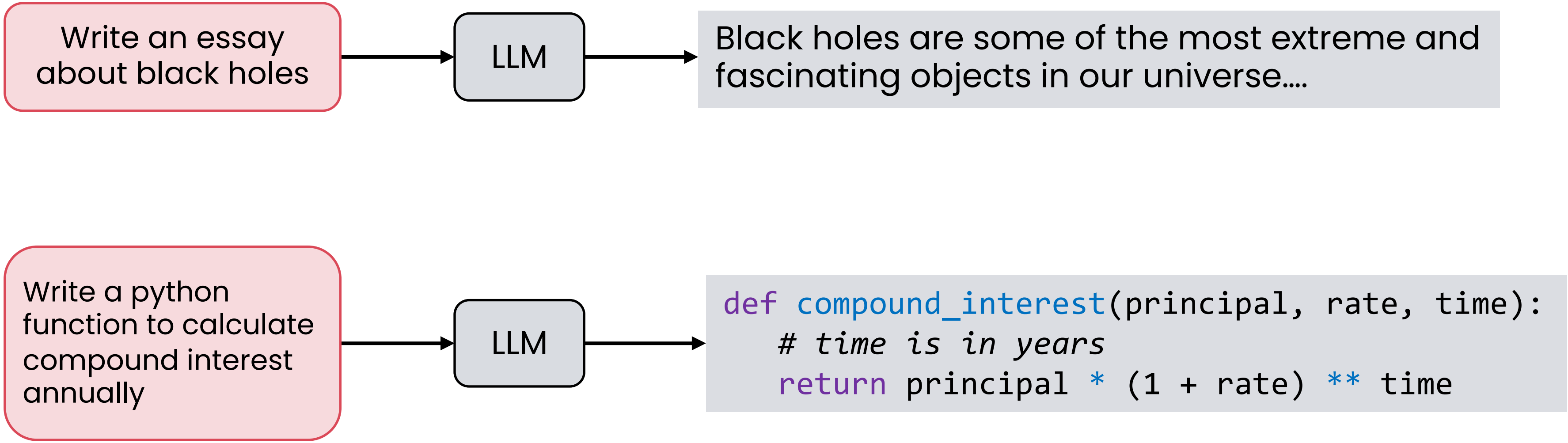
Reflect on code output or errors and write improved second draft

Reflection Design Pattern

Why not just direct generation?

DeepLearning.AI

Direct generation



Write an essay about black holes

LLM

Black holes are some of the most extreme and fascinating objects in our universe....

Write a python function to calculate compound interest annually

LLM

```python
def compound_interest(principal, rate, time):
    # time is in years
    return principal * (1 + rate) ** time
```

Zero, one, and few-shot prompting

```
Convert to
MM/DD/YYYY format

Input:
{input_date}
```

Zero-shot (no examples)

```
Convert to
MM/DD/YYYY format

Input: Jan 1st, 2025
Output: 01/01/2025

Input:
{input_date}
```

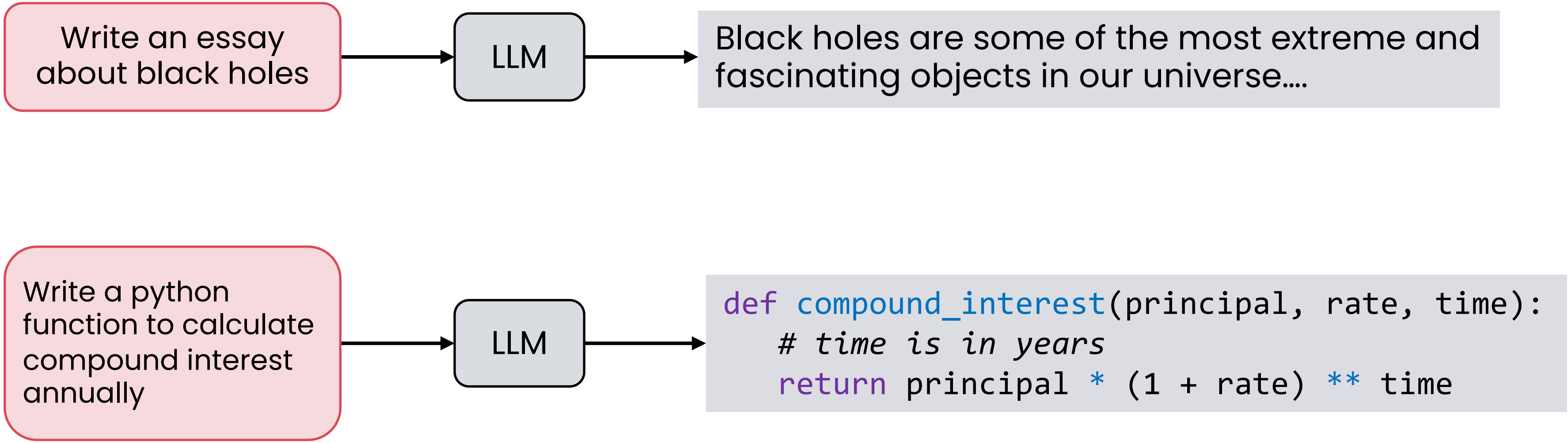
One-shot (single example)

```
Convert to
MM/DD/YYYY format

Input: Jan 1st, 2025
Output: 01/01/2025

Input: 21st June, 2025
Output: 06/21/2025

Input:
{input_date}
```

Two-shot (two examples)
Few-shot (multiple examples)

Direct generation

Zero-shot prompts



Write an essay about black holes

LLM

Black holes are some of the most extreme and fascinating objects in our universe....

Write a python function to calculate compound interest annually

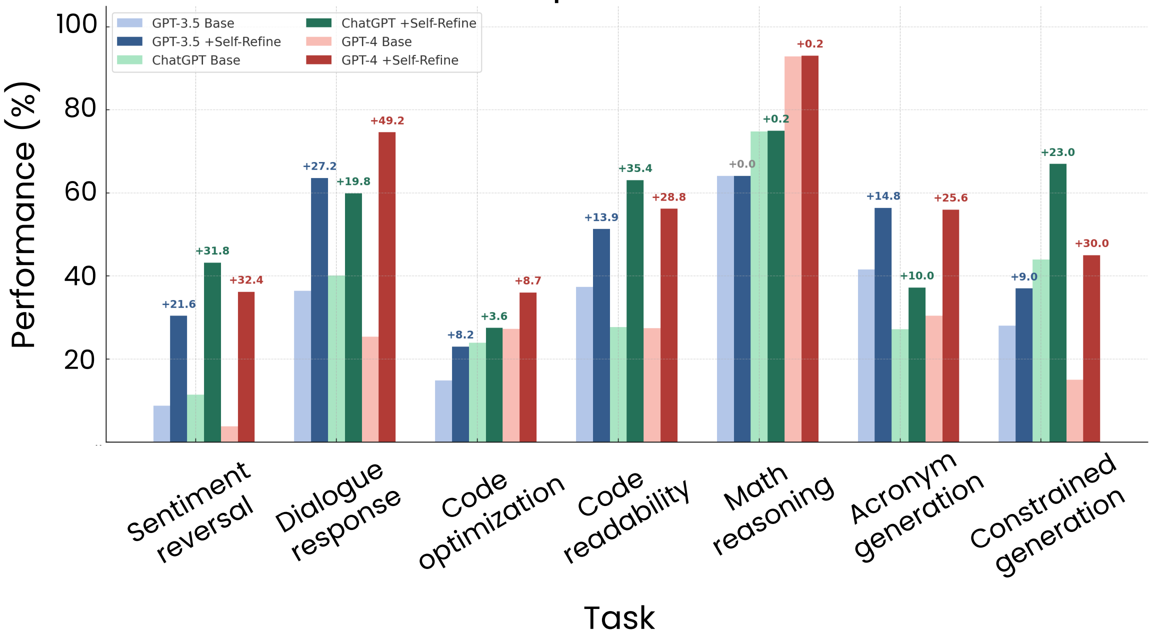
LLM

```
def compound_interest(principal, rate, time):
    # time is in years
    return principal * (1 + rate) ** time
```

Reflection has been tested

Reflection consistently outperforms direct generation on a variety of tasks.

Performance improvement with reflection



[Adapted from Madaan, A. et al. (2023) "Self-refine: Iterative refinement with self-feedback"]

DeepLearning.AI

Andrew Ng

Tasks where reflection works better

Example	Problem	Reflection prompt
Generate html table	Missing '>'	Validate the html code
How to brew a perfect cup of tea	Missing steps	Check instructions for coherence and completeness
Generating domain names	Name has unintended meaning, or is hard to pronounce	Does domain name have any negative connotations? Is the domain name hard to pronounce?

Tips for writing reflection prompts

Brainstorming domain names

Review the domain names you suggested.

Check if each name is easy to pronounce and thus easy to spread via word of mouth.

Consider whether each name might mean something negative in other languages.

Then output a shortlist of only the names that meet these criteria.

Improving email

Review the email first draft.

Check that the tone is professional and look for phrases that could be considered rude or insensitive.

Verify all facts, dates, and promises are accurate.

Then write the next draft of the email.

- Clearly indicate the reflection action
- Specify criteria to check

Reflection Design Pattern

Chart generation workflow

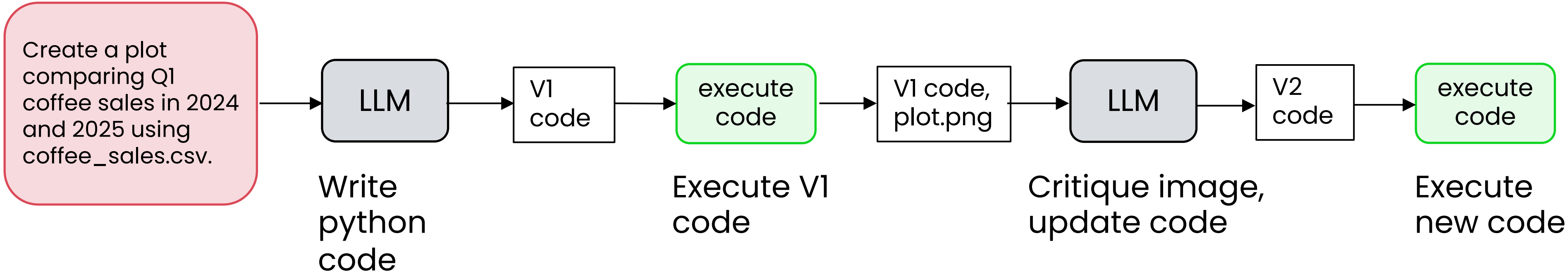
DeepLearning.AI

Visualizing coffee sales

date	price	coffee_name
2024-01-12	3.87	Latte
2024-01-28	3.87	Hot Chocolate
2024-02-09	3.87	Hot Chocolate
2024-03-01	2.89	Cappuccino
2024-03-04	3.87	Latte
…	…	…
2025-03-23	3.57	Latte

Create a plot comparing Q1 coffee sales in 2024 and 2025

Chart generation agentic workflow



Create a plot comparing Q1 coffee sales in 2024 and 2025 using coffee_sales.csv.

LLM

V1 code

execute code

V1 code, plot.png

LLM

V2 code

execute code

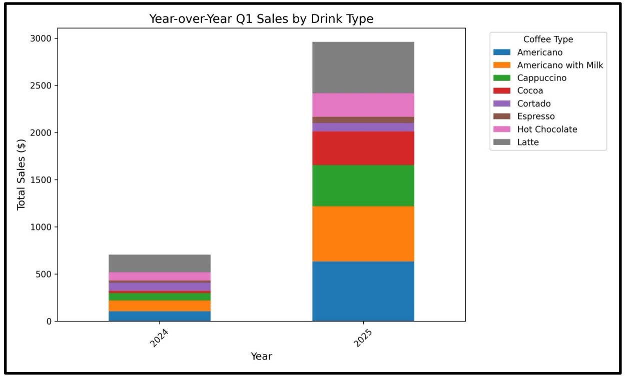
Write python code

Execute V1 code

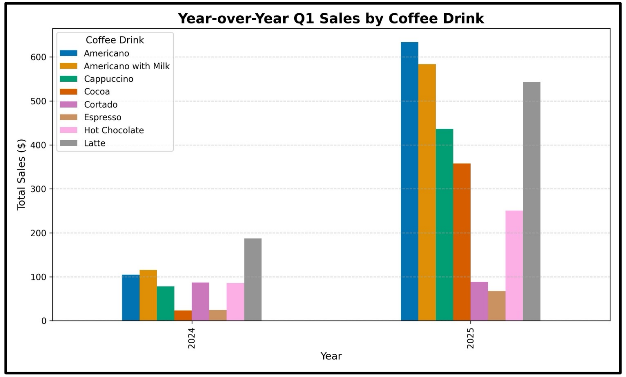
Critique image, update code

Execute new code

```
import matplotlib.pyplot as plt
import pandas as pd

# Filter for Q1 sales
q1_sales = df[df['quarter'] == 1]
....
```



plot.png



plot_v2.png

Reflection with a different LLM

LLM

Code generation

Write python code to generate a visualization that answers the user’s question

{user prompt}

LLM 2

Reflection

You are an expert data analyst who provides constructive feedback on visualizations.

{V1 code} {plot.png} {conversation history}

Step 1: Critique the attached chart for readability, clarity, and completeness.

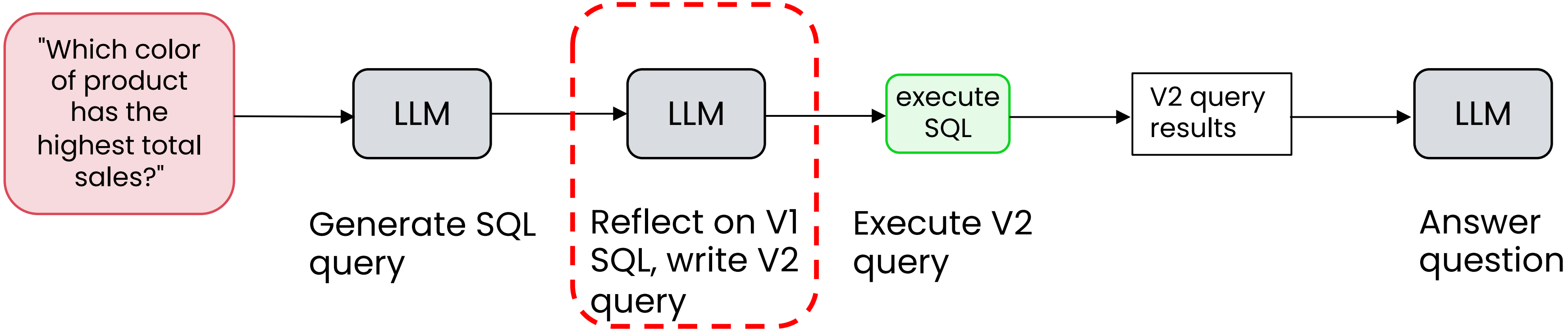
Step 2: Write new code to implement your improvements.

Reflection Design Pattern

Evaluating the impact of reflection

DeepLearning.AI

Create a dataset of prompts and answers



Generate SQL query

Reflect on V1 SQL, write V2 query

Execute V2 query

Answer question

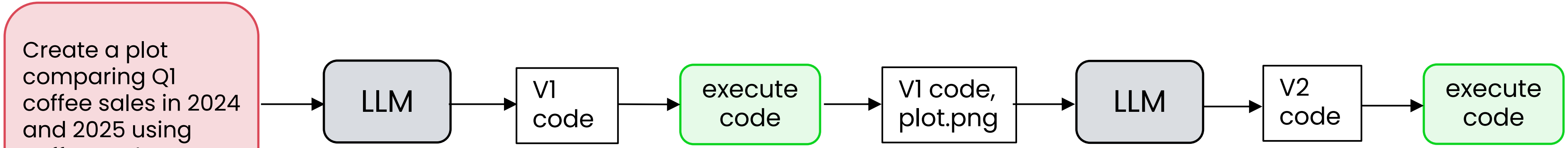
Prompts	Ground truth answer	No reflection	With reflection
Number of items sold in May 2025?	1201	980	1201
Most expensive item?	Airflow sneaker	Airflow sneaker	Airflow sneaker
How many styles carried?	14	14	14
		87% correct	95% correct

Run each time you change reflection prompt

DeepLearning.AI

Andrew Ng

What about subjective tasks?



Create a plot comparing Q1 coffee sales in 2024 and 2025 using coffee_sales.csv.

LLM

V1 code

execute code

V1 code, plot.png

LLM

V2 code

execute code

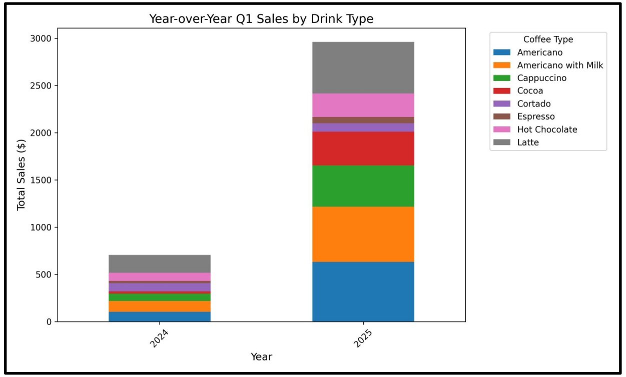
Write python code

Execute V1 code

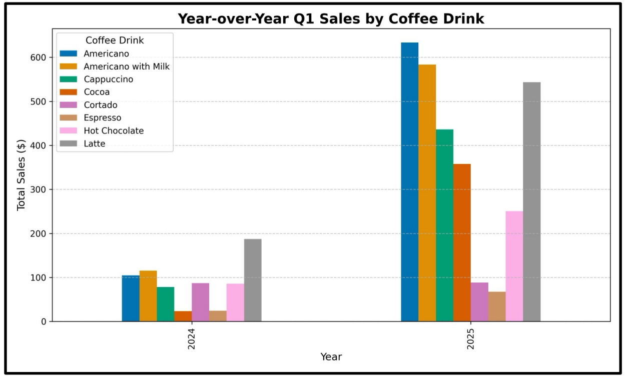
Critique image, update code

Execute new code

Can you measure which chart is better with an eval?



Before reflection

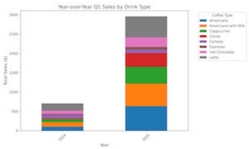


After reflection

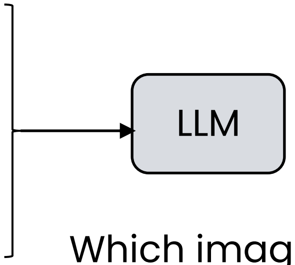
DeepLearning.AI

Andrew Ng

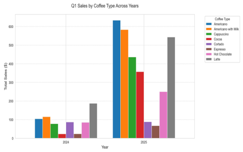
Using an LLM as a judge



plot.png



Which image is better?



plot_v2.png

Known issues with using LLMs for comparison:

- Answers often not very good
- Position bias

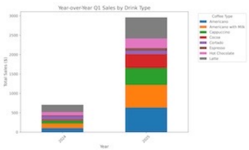
A B

LLM picks A more often

DeepLearning.AI

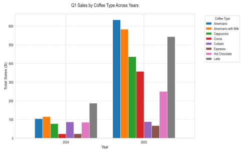
Andrew Ng

Grading with a rubric gives more consistent results



plot.png

LLM



plot_v2.png

Rubric

Assess the attached image against this quality rubric. Each item should receive a score for 1 (true) or 0 (false). Return the scores for each item as a json object

1. Has clear title
2. Axis labels present
3. Appropriate chart type
4. Axes use appropriate numerical range
5. …

Input	No reflection	With reflection
User query 1	4	6
User query 2	5	8
....	....	....
User query 10	5	7

Run each time you change reflection prompt

DeepLearning.AI

Andrew Ng

Evaluating reflection

- Objective evals
  - Code-based evals are easier
  - Build a dataset of ground truth examples

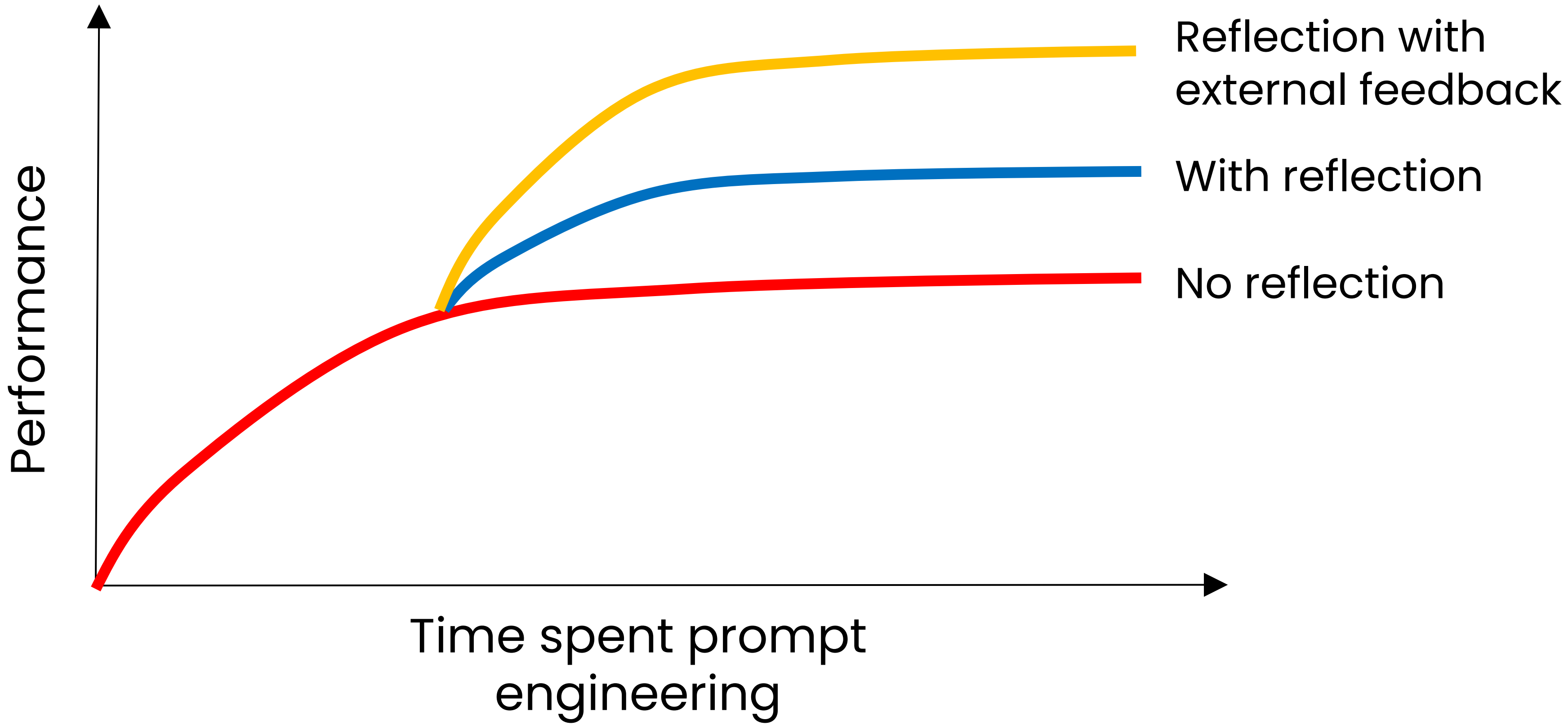
- Subjective evals
  - Use LLM as a judge
  - Rubric-based grading is better

Reflection Design Pattern

Using external feedback

DeepLearning.AI

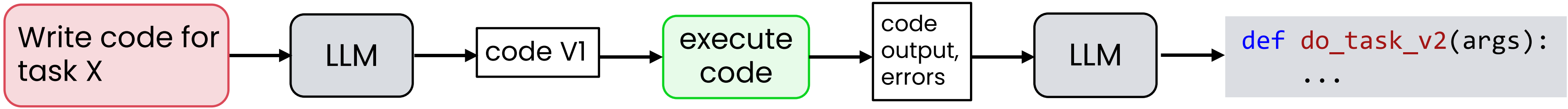
Return on investment on prompt engineering



DeepLearning.AI

Andrew Ng

Reflection with external feedback



```
def do_task_v2(args):
    ...
```

Write first draft

```
def do_task(args):
    ...
```

```
SyntaxError:
unterminated string
literal (detected at
line 1)
```

Reflect on code output or errors and write improved second draft

Other examples of tools to help reflection

Challenge	Example	Source of feedback
Mentioning competitors	Our company's shoes are better than RivalCo	Pattern matching for competitor names
Fact checking an essay	The Taj Mahal was built in 1648	Web search results
LLM won't follow output length guidelines	Essay is over word limit	Word count tool

DeepLearning.AI

End of M2