AI Agents in the Data Pipeline

Craig West

https://craig-west.netlify.app/

https://evaluating-ai-agents.com/

Talk Slides and Repo:

https://github.com/Python-Test-Engineer/earl2025

This also contains links to a video of the BrighonPy talk/workshop on **'AI as API'** and the repo for the PyData Southampton Meetup **'AI Agents in The Data Pipeline'** as well as additional resources.

Desired Outcomes

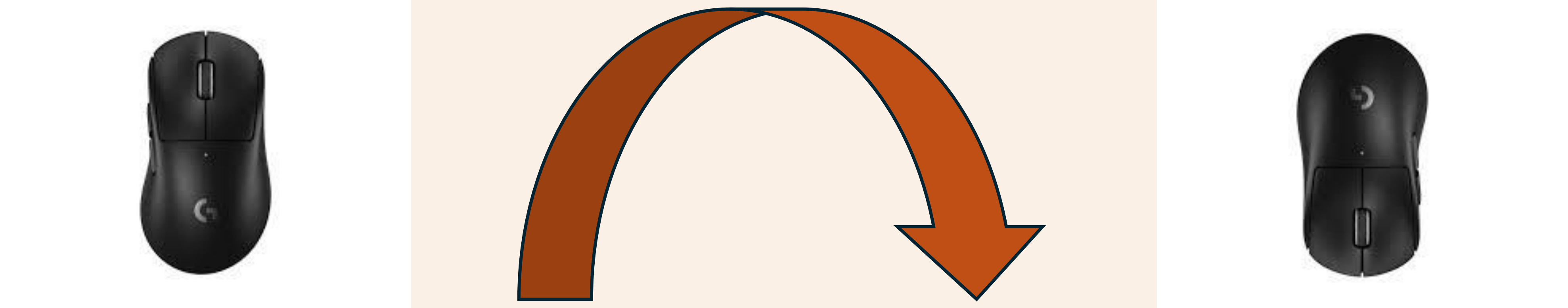
To better understand the Agent Landscape, its terminology, basis and uses.

To demystify and simplify what an Agent is.

To see the uses of **AI Agents in the Data Pipeline** and the likely future of Data Pipeline applications.

This is a ‘fly by’ rather than ‘deep dive’. The repo has more detailed examples and links to help you go further and deeper.

Raw code implementation of an Agent



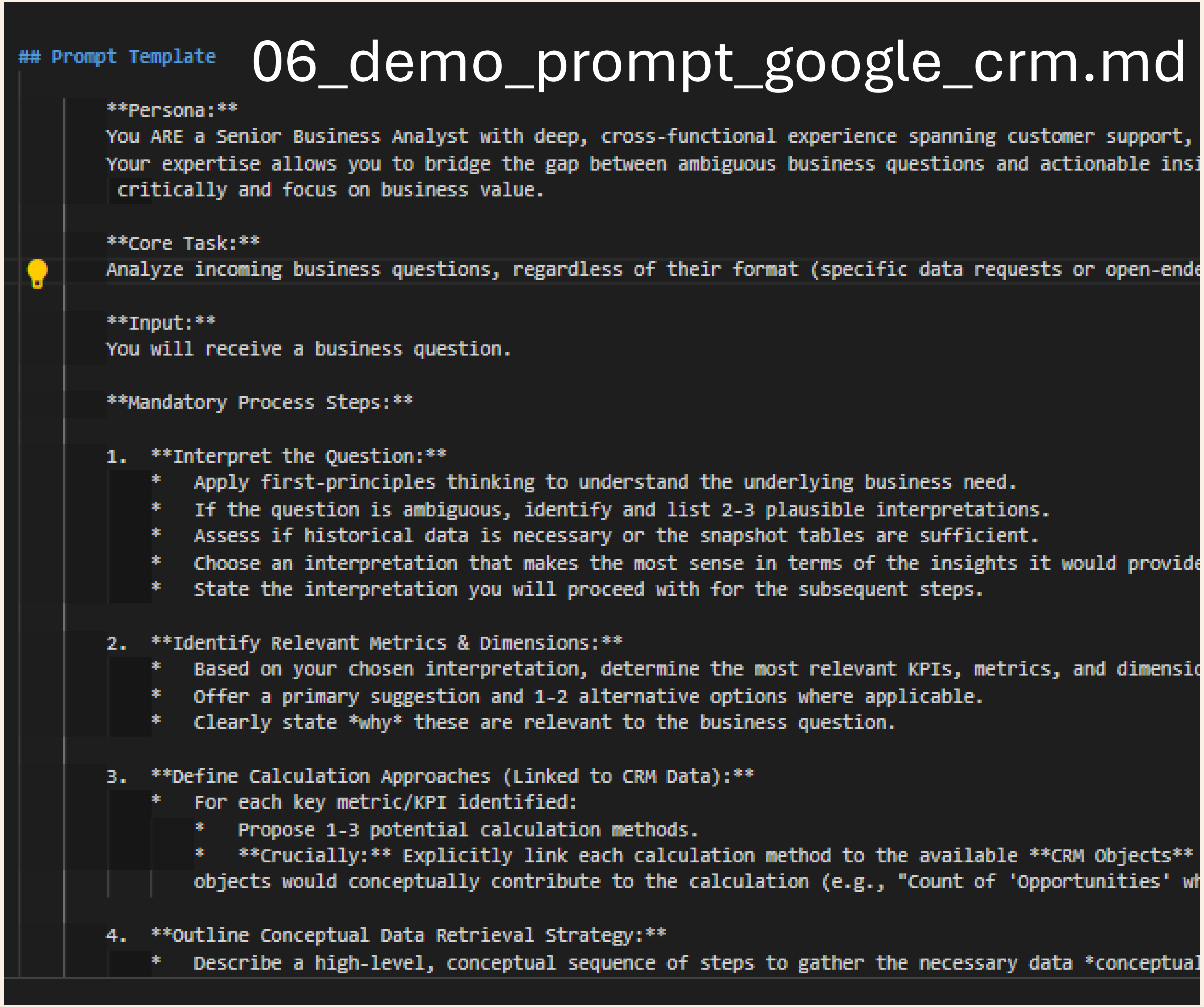
No more difficult than what we do currently – just 180 degrees different

“It doesn’t get any easier – just different” - Anon

Raw code implementation of an Agent

- Just one API endpoint – ‘AI as API’
- Instructions are written on client side and sent with request
- Use Natural Language
- An example prompt…

06_demo_prompt_google_crm.md



```
## Prompt Template

        **Persona:**
        You ARE a Senior Business Analyst with deep, cross-functional experience spanning customer support,
        Your expertise allows you to bridge the gap between ambiguous business questions and actionable insi
         critically and focus on business value.

        **Core Task:**
        Analyze incoming business questions, regardless of their format (specific data requests or open-ende

        **Input:**
        You will receive a business question.

        **Mandatory Process Steps:**

        1.  **Interpret the Question:**
            *   Apply first-principles thinking to understand the underlying business need.
            *   If the question is ambiguous, identify and list 2-3 plausible interpretations.
            *   Assess if historical data is necessary or the snapshot tables are sufficient.
            *   Choose an interpretation that makes the most sense in terms of the insights it would provide
            *   State the interpretation you will proceed with for the subsequent steps.

        2.  **Identify Relevant Metrics & Dimensions:**
            *   Based on your chosen interpretation, determine the most relevant KPIs, metrics, and dimensio
            *   Offer a primary suggestion and 1-2 alternative options where applicable.
            *   Clearly state *why* these are relevant to the business question.

        3.  **Define Calculation Approaches (Linked to CRM Data):**
            *   For each key metric/KPI identified:
                *   Propose 1-3 potential calculation methods.
                *   **Crucially:** Explicitly link each calculation method to the available **CRM Objects**
            objects would conceptually contribute to the calculation (e.g., "Count of 'Opportunities' wh

        4.  **Outline Conceptual Data Retrieval Strategy:**
            *   Describe a high-level, conceptual sequence of steps to gather the necessary data *conceptual
```

Agenda

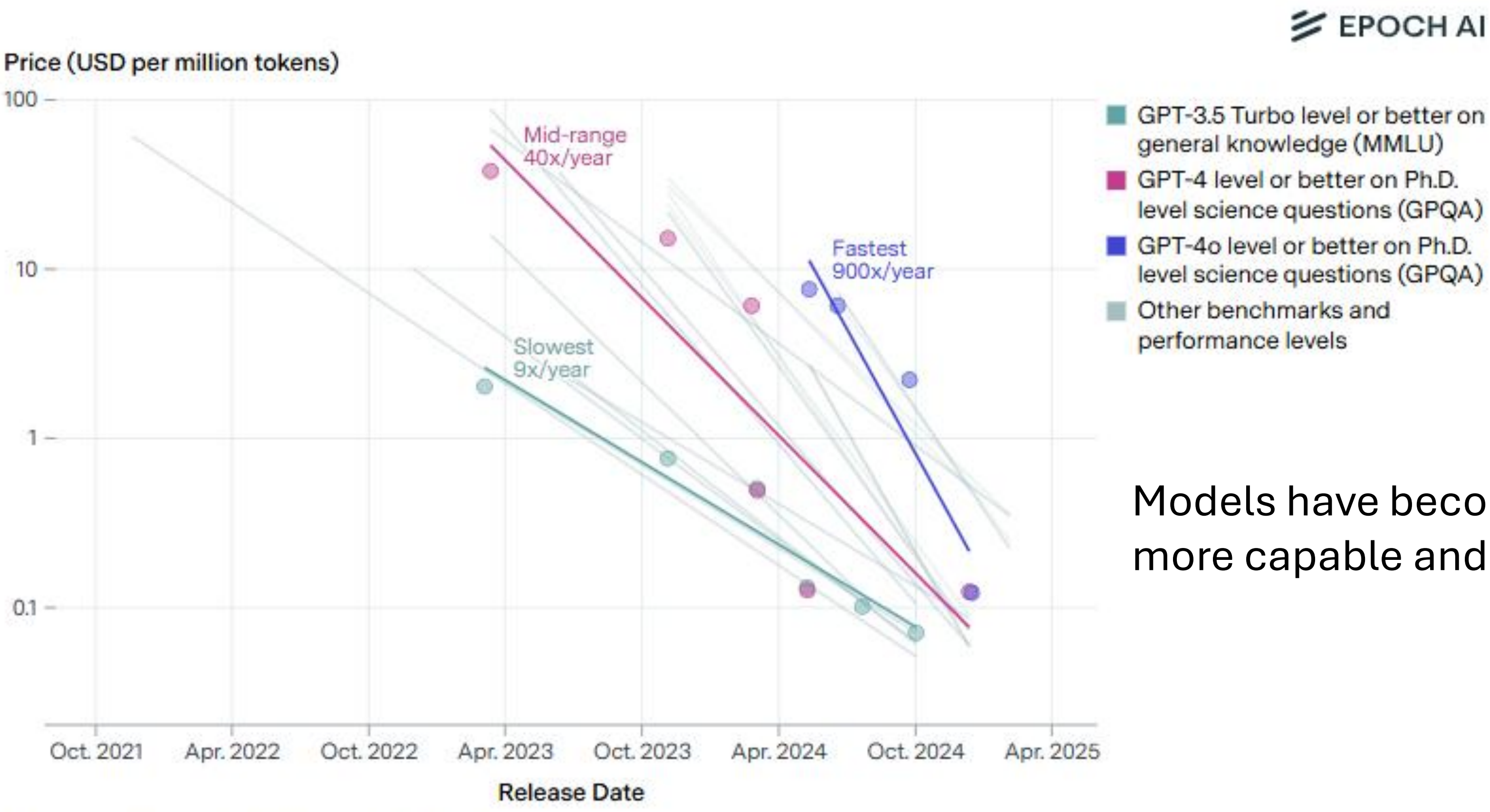
With code examples where appropriate:

- AI Agents from scratch – a POST API Request
- Understand what Tool/Function calling are
- Agentic Patterns
- Model Context Protocol – changing app architecture
- Demos of some patterns
- Final demo of Cline building a data pipeline from scratch and producing a Business Intelligence report form the charts and graphs.

Model costs

The final Cline demo the cost per intelligence has dropped by a factor of 2-4 in last 3 months

year, so it's less clear that those will persist.



Data source: Epoch AI, Artificial Analysis

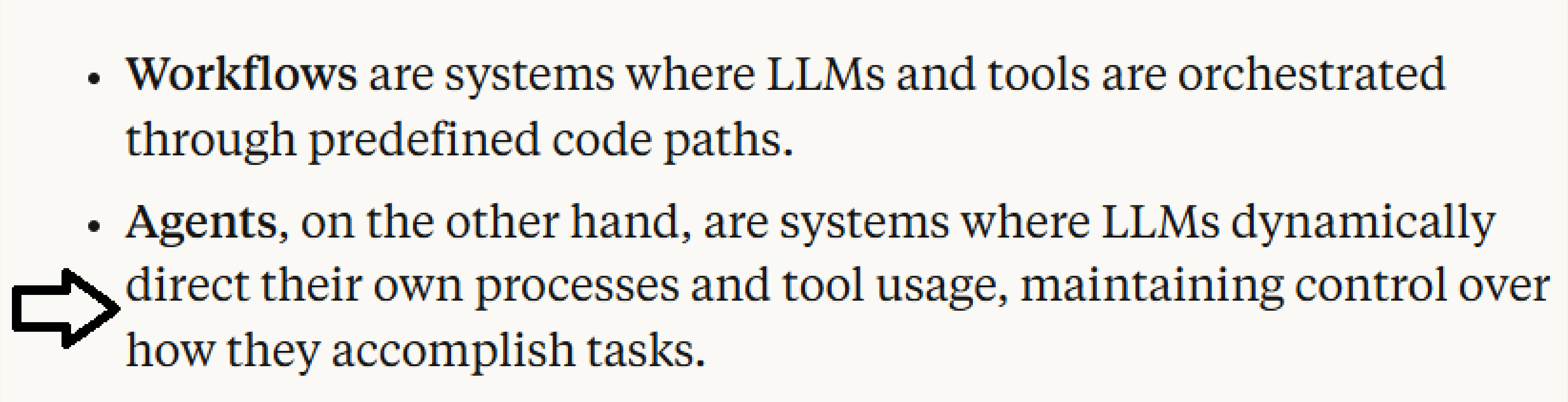
Models have become increasingly more capable and 'intelligent'.

Cost of a unit of 'intelligence has dropped ~ 99% in last two years and is 10x a year (anecdotal)

What is an AI Agent?

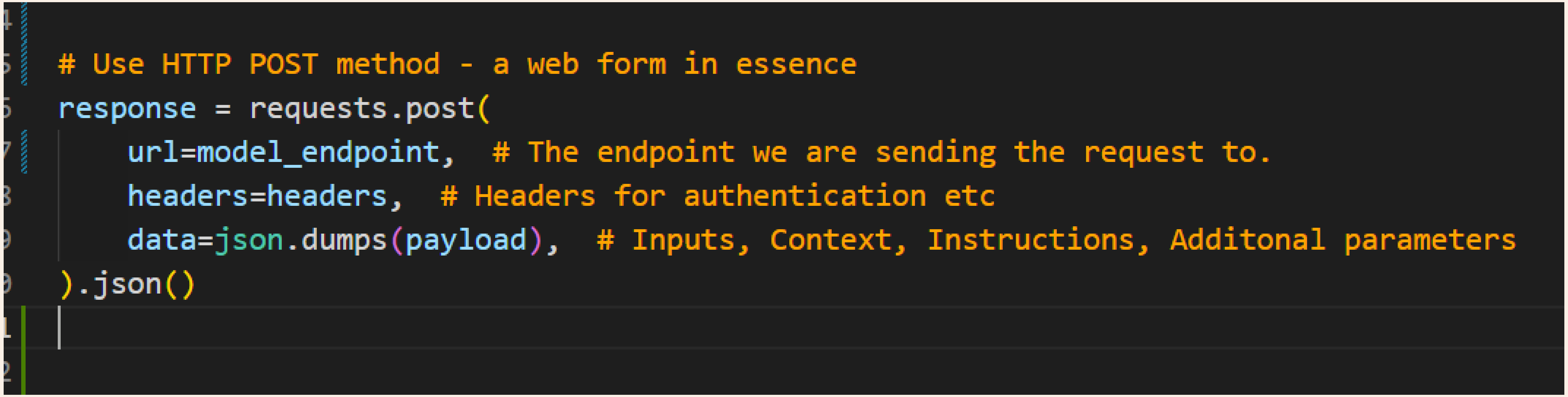
Many definitions but people opt for ‘Agentic Apps’

Anthropic



- **Workflows** are systems where LLMs and tools are orchestrated through predefined code paths.
- **Agents**, on the other hand, are systems where LLMs dynamically direct their own processes and tool usage, maintaining control over how they accomplish tasks.

Raw code implementation of an Agent

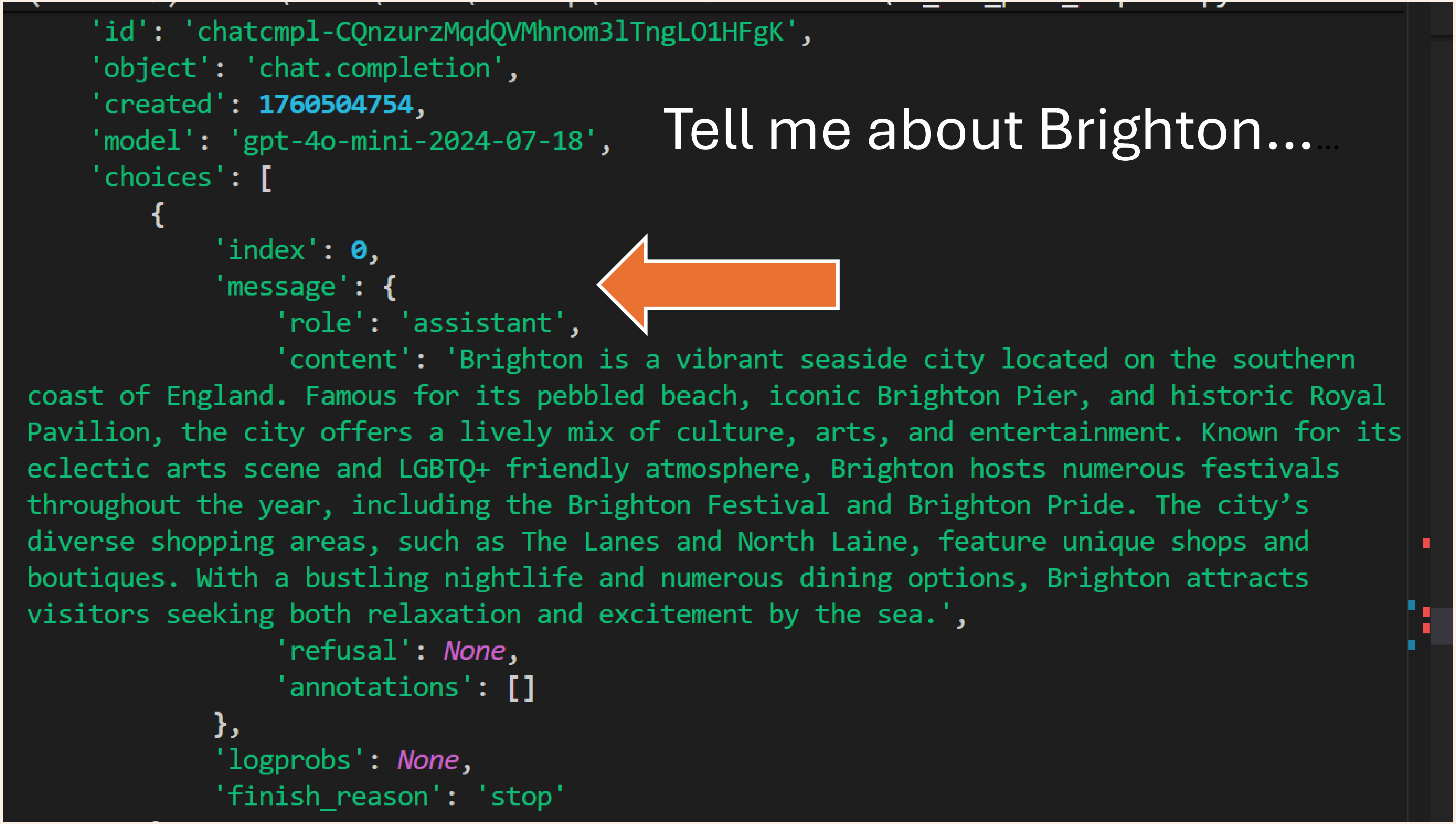


```python
# Use HTTP POST method - a web form in essence
response = requests.post(
    url=model_endpoint,  # The endpoint we are sending the request to.
    headers=headers,  # Headers for authentication etc
    data=json.dumps(payload),  # Inputs, Context, Instructions, Additonal parameters
).json()
```

We saw an example of a prompt earlier – ‘payload’

Let’s look at `01_raw_post_request.py` in the repo…

Tell me about Brighton...

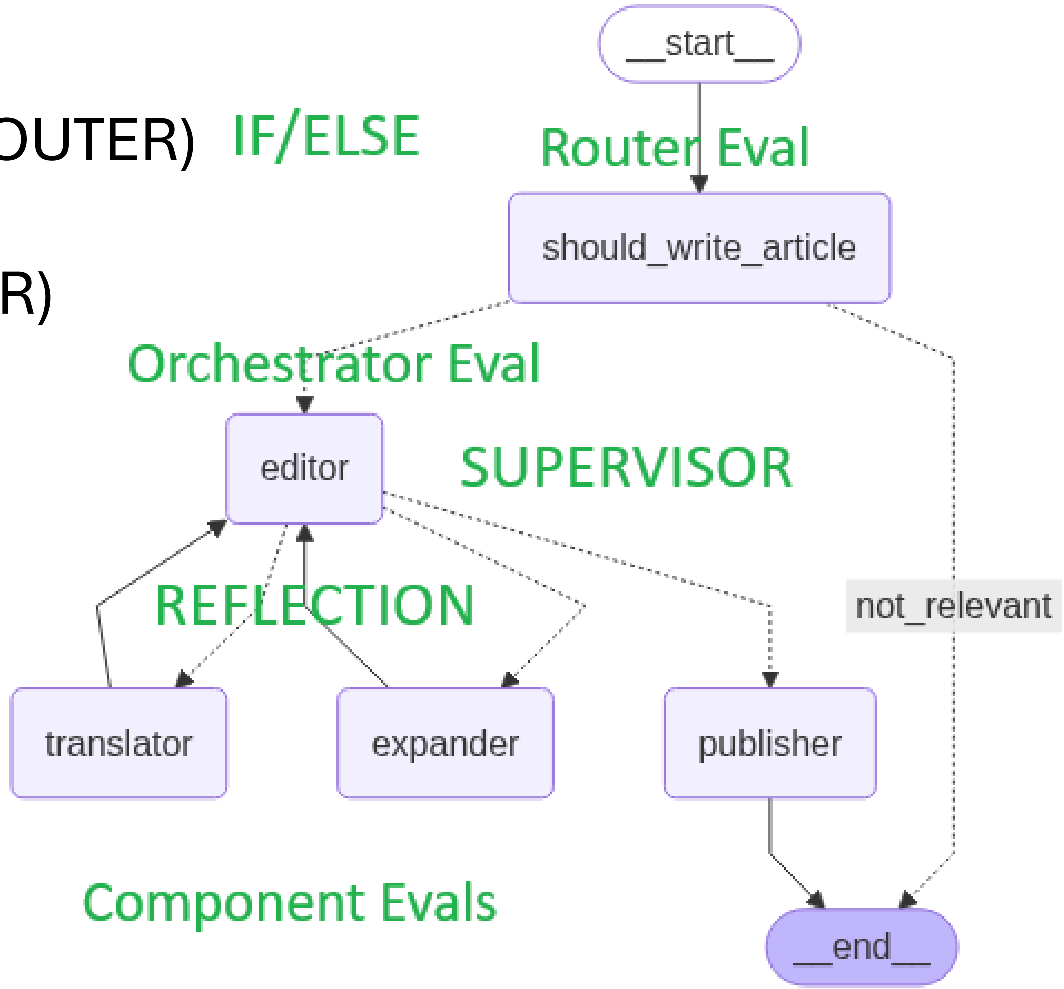


```
    'id': 'chatcmpl-CQnzurzMqdQVMhnom3lTngLO1HFgK',
    'object': 'chat.completion',
    'created': 1760504754,
    'model': 'gpt-4o-mini-2024-07-18',
    'choices': [
        {
            'index': 0,
            'message': {
                'role': 'assistant',
                'content': 'Brighton is a vibrant seaside city located on the southern
coast of England. Famous for its pebbled beach, iconic Brighton Pier, and historic Royal
Pavilion, the city offers a lively mix of culture, arts, and entertainment. Known for its
eclectic arts scene and LGBTQ+ friendly atmosphere, Brighton hosts numerous festivals
throughout the year, including the Brighton Festival and Brighton Pride. The city’s
diverse shopping areas, such as The Lanes and North Laine, feature unique shops and
boutiques. With a bustling nightlife and numerous dining options, Brighton attracts
visitors seeking both relaxation and excitement by the sea.',
                'refusal': None,
                'annotations': []
            },
            'logprobs': None,
            'finish_reason': 'stop'
```

- Given an article title, determine if it is about AI (ROUTER) and if so start workflow
- Take title and create article of N words (EXPANDER)
- In a specified language (TRANSLATOR)
- Ensure article is not sensational (EDITOR)

A ‘good article’ – OMITTED IN DEMO

PUBLISH = YES only if last three are YES in EDITOR



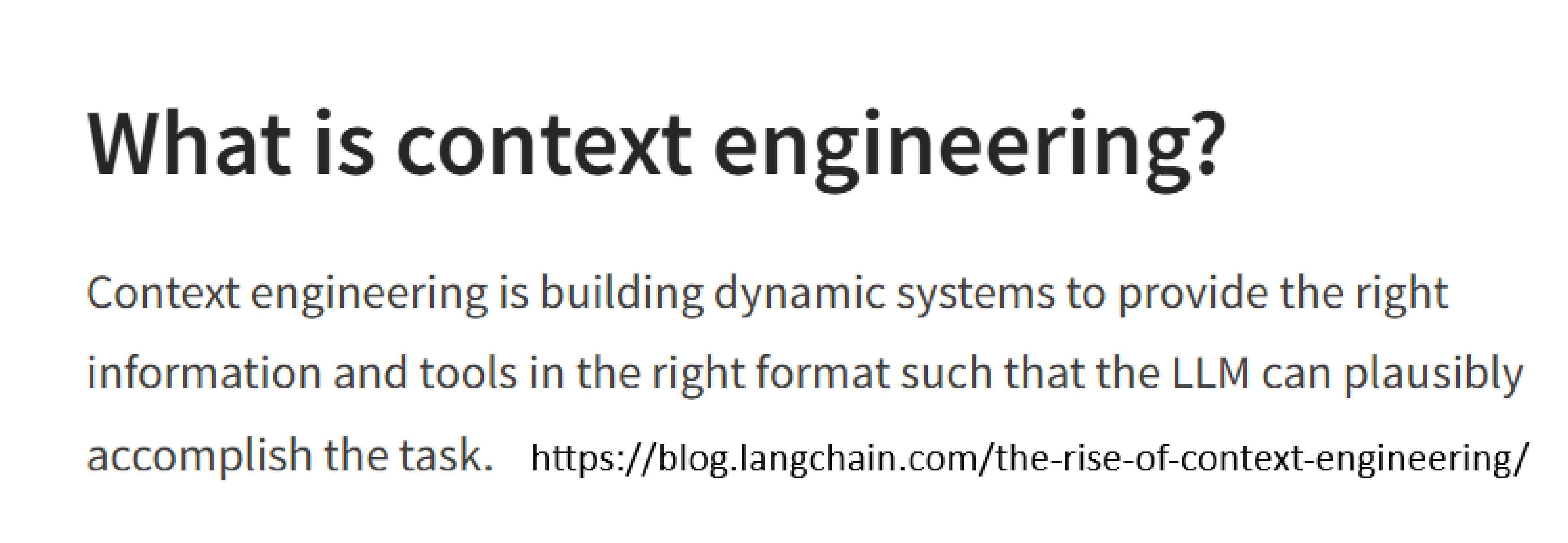
Prompt Engineering

Essentially a set of instructions…

“Tell me about Data Analytics in 100 words…” will give a pretrained response…

“ Tell me about Data Analytics and how it can help me with our sales analysis for the last quarter…” will give us a response of: “I need more info…” as it has not been trained on our data.”

Context Engineering



What is context engineering?

Context engineering is building dynamic systems to provide the right information and tools in the right format such that the LLM can plausibly accomplish the task. https://blog.langchain.com/the-rise-of-context-engineering/

05_demo_cline.md

Supplying more Context (information)

“Tell me about Data Analytics and ***here is some data from our company***, please analyse…”

This is context engineering where we supply appropriate context for the Agent.

RAG is Retrieval Augmented Generation and is RETRIEVING extra content from a range of sources for the prompt – not just Vector Databases.

What is Tool/Function calling?

Tool/Function Calling –
adding more context/info on demand

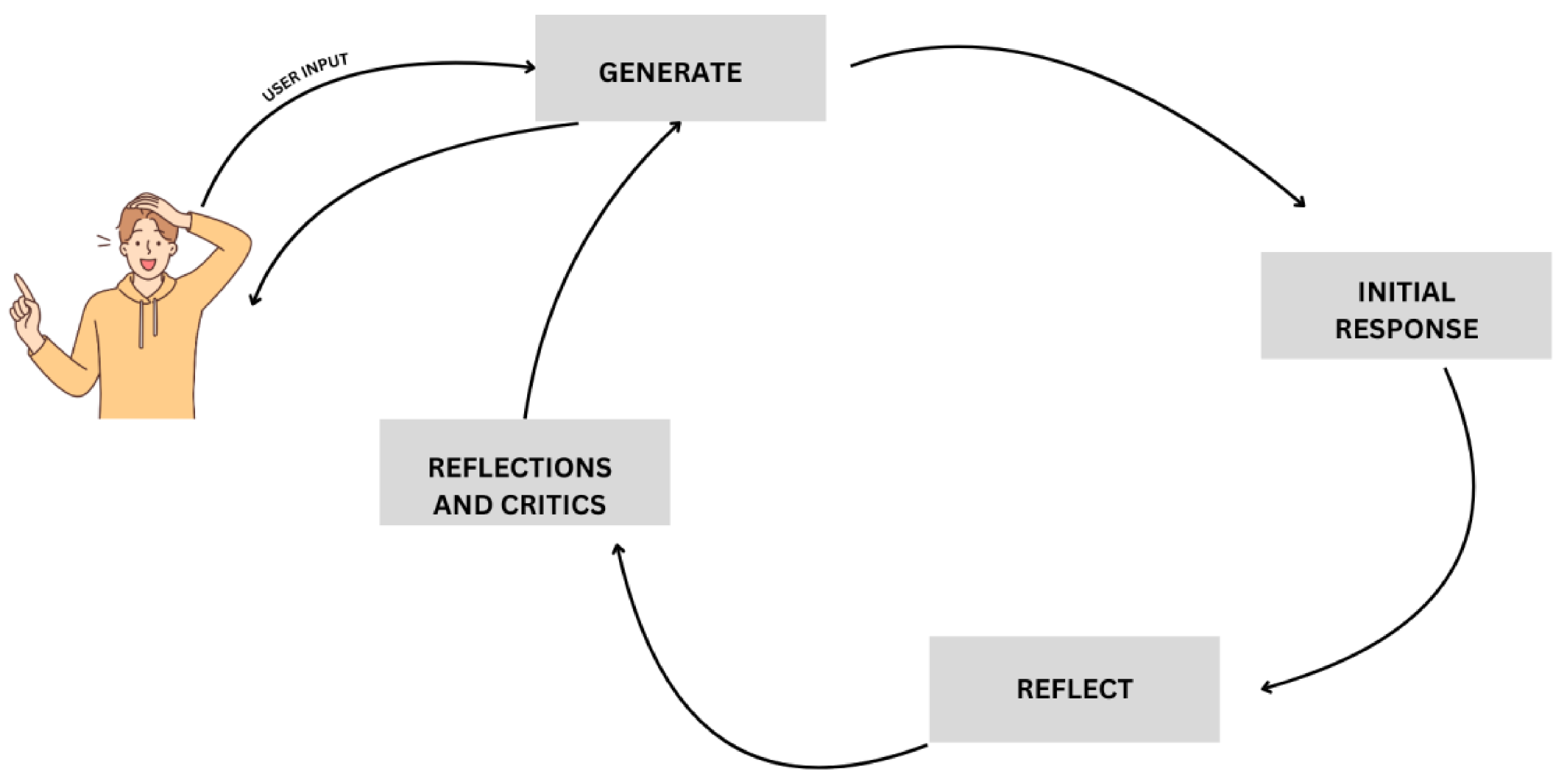
"Tell me about Data Analytics and here is some data...

Also compare our share price with {company X}..."

We don't know this information at time of the request as it will vary with each query.

We need a tool/function the Agent can use to get the share price... **"tool/function calling"**

Reflection Pattern - looping

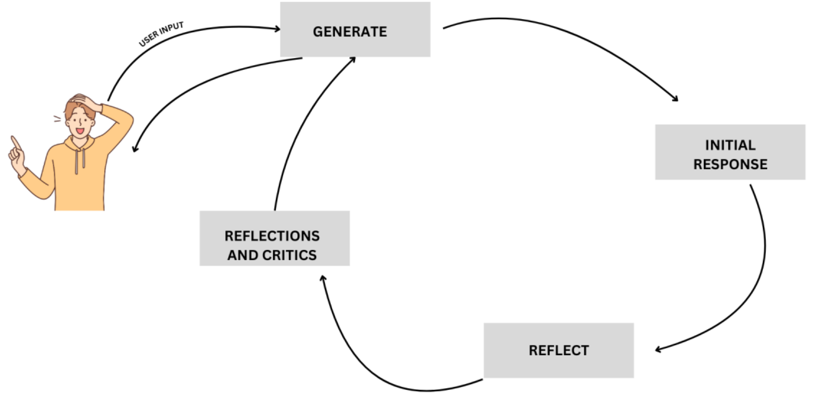


https://www.linkedin.com/pulse/reflection-agent-langgraph-vaibhav-mane-g77gc/

What is Tool/Function calling?

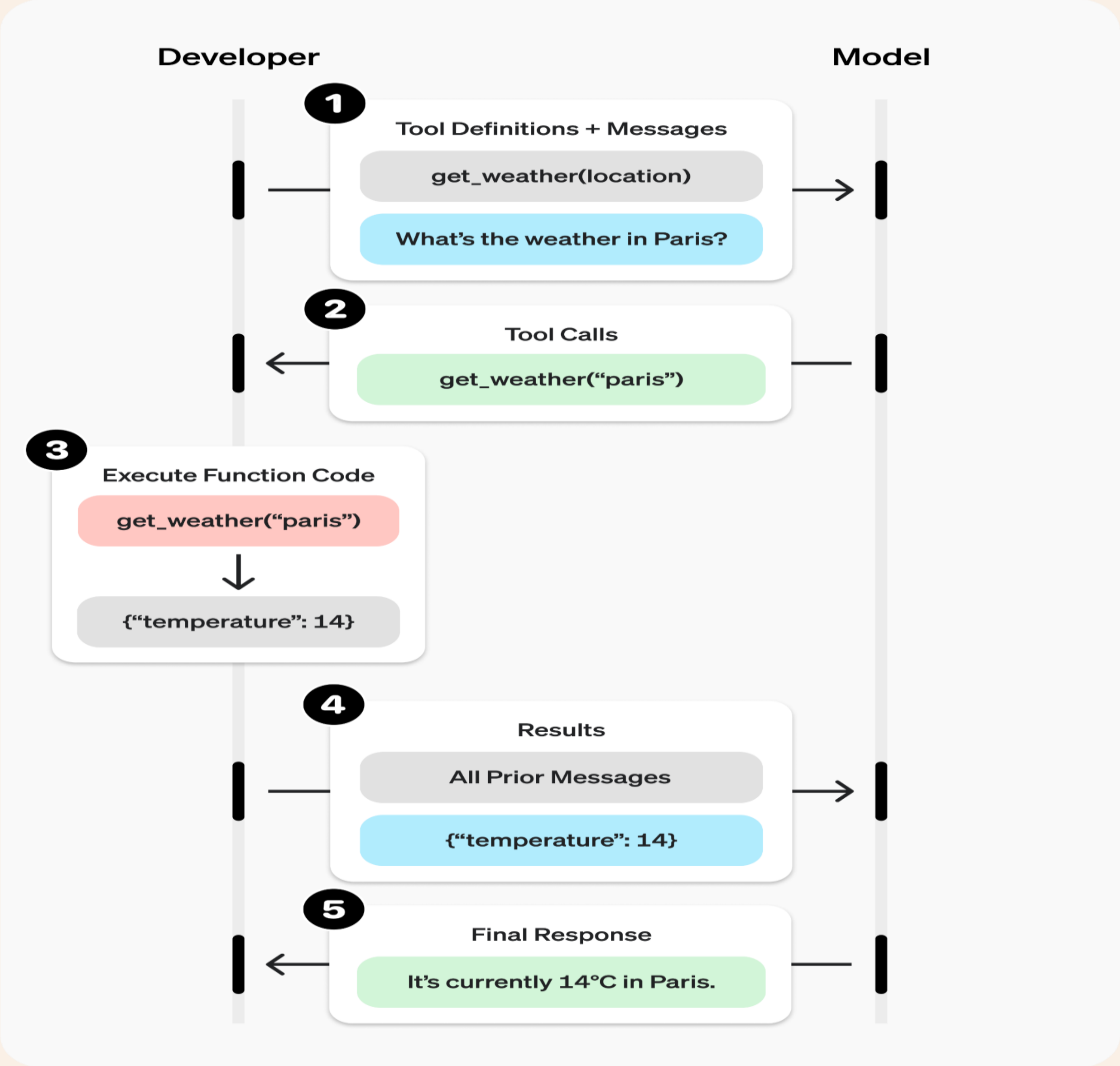
In our INSTRUCTIONS we can add:

- Here are some useful tools/functions that may be of help at run time.
- Let me know which you want to run and with what arguments.
- I will run them on my machine and then send this extra CONTEXT back to you (reflection – see image)
- `03.1_prompt.md` has the prompt
- `03.2_demo_tool_calling.py`



https://www.linkedin.com/pulse/reflection-agent-langgraph-vaibhav-mane-g77gc/

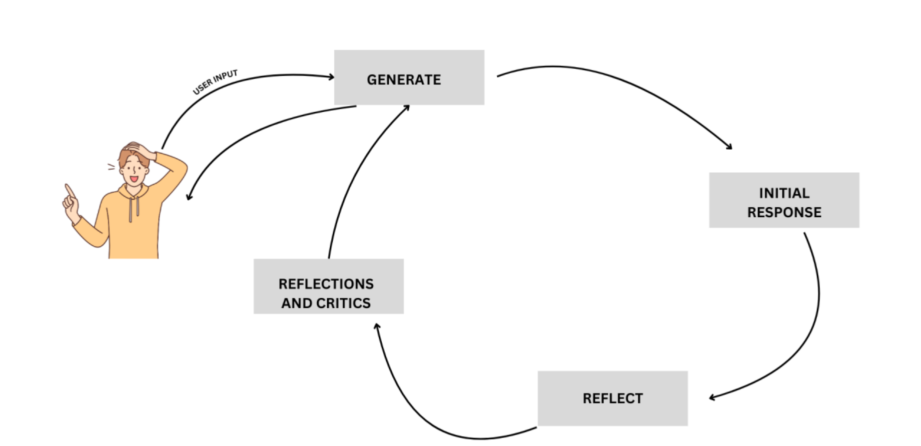
What is Tool/Function calling?



What is Tool/Function calling?

In our INSTRUCTIONS we can add:

- Here are some useful tools/functions that may be of help at run time.
- Let me know which you want to run and with what arguments.
- I will run them on my machine and then send this extra CONTEXT back to you (reflection – see image)
- `03.1_prompt.md` has the prompt
- `03.2_demo_tool_calling.py`



https://www.linkedin.com/pulse/reflection-agent-langgraph-vaibhav-mane-g77gc/

Another example of reflection

We ask Agent to write some code for us, then pass the response to a critique and the response from that to a final one to combine the two.

"Generate a Python implementation of imputing missing values in a Pandas DataFrame with the mean of the column.

Ensure there are plenty of comments explaining the code."

```
04_demo_reflection.ipynb -> three files:
04.1_code.md
04.2_critique.md
04.3_final.md
```

What is the Model Context Protocol (MCP)?

We have seen how we can write tools for the Agent to discover and call as needed.

What about tools others have created?

How can an Agent discover what tools are available, how to use them with what ever arguments are needed and how to execute them? (*like when we go to PyPi to get libraries*)

What is the Model Context Protocol (MCP)?

This is the Model Context Protocol for Agents to be able to do this.

In essence, an Agent can find a list of tools we have given it, with these tools able to inform an Agent what they are and how to use them.

The Agent runs the pipx to download and run them in a separate process.

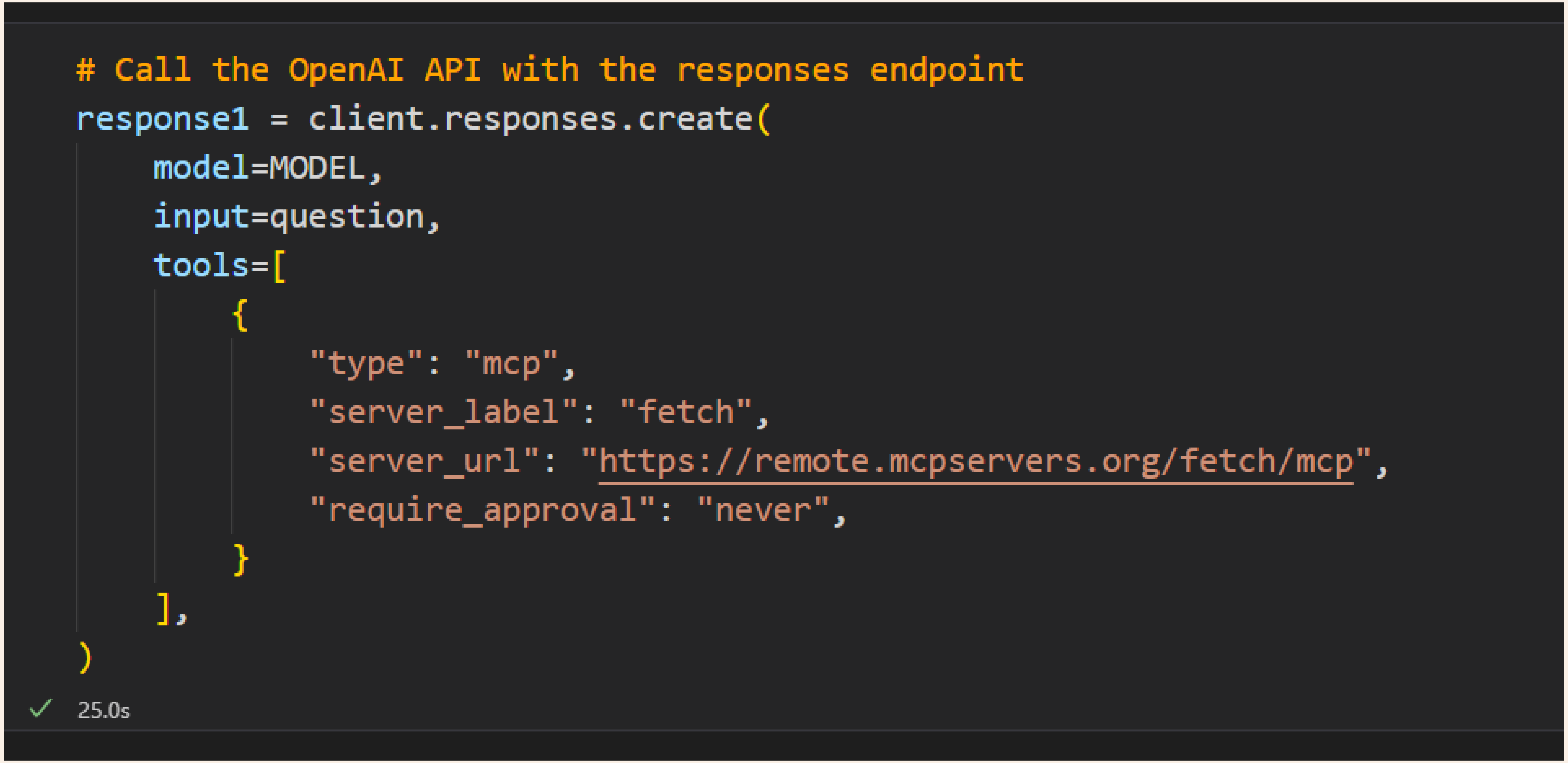
What is the Model Context Protocol (MCP)?

The Langchain SQL Agent is something we can code in deterministically into an app but how would an Agent decide whether to use it and how to use it?

MCP enables an Agent to get a list of tools available by the MCP Server and how to use the tools.

An Agent can then decide whether to use these tools to create more context for further prompts.

What is the Model Context Protocol (MCP)?



```python
# Call the OpenAI API with the responses endpoint
response1 = client.responses.create(
    model=MODEL,
    input=question,
    tools=[
        {
            "type": "mcp",
            "server_label": "fetch",
            "server_url": "https://remote.mcpservers.org/fetch/mcp",
            "require_approval": "never",
        }
    ],
)
```

✓ 25.0s

https://mcpservers.org/servers/modelcontextprotocol/fetch

What is the Agent2Agent Protocol (A2A)?

A complimentary protocol to allow one Agent to discover and understand what another Agent does.

It can then hand off work to the other Agent to get a desired response.

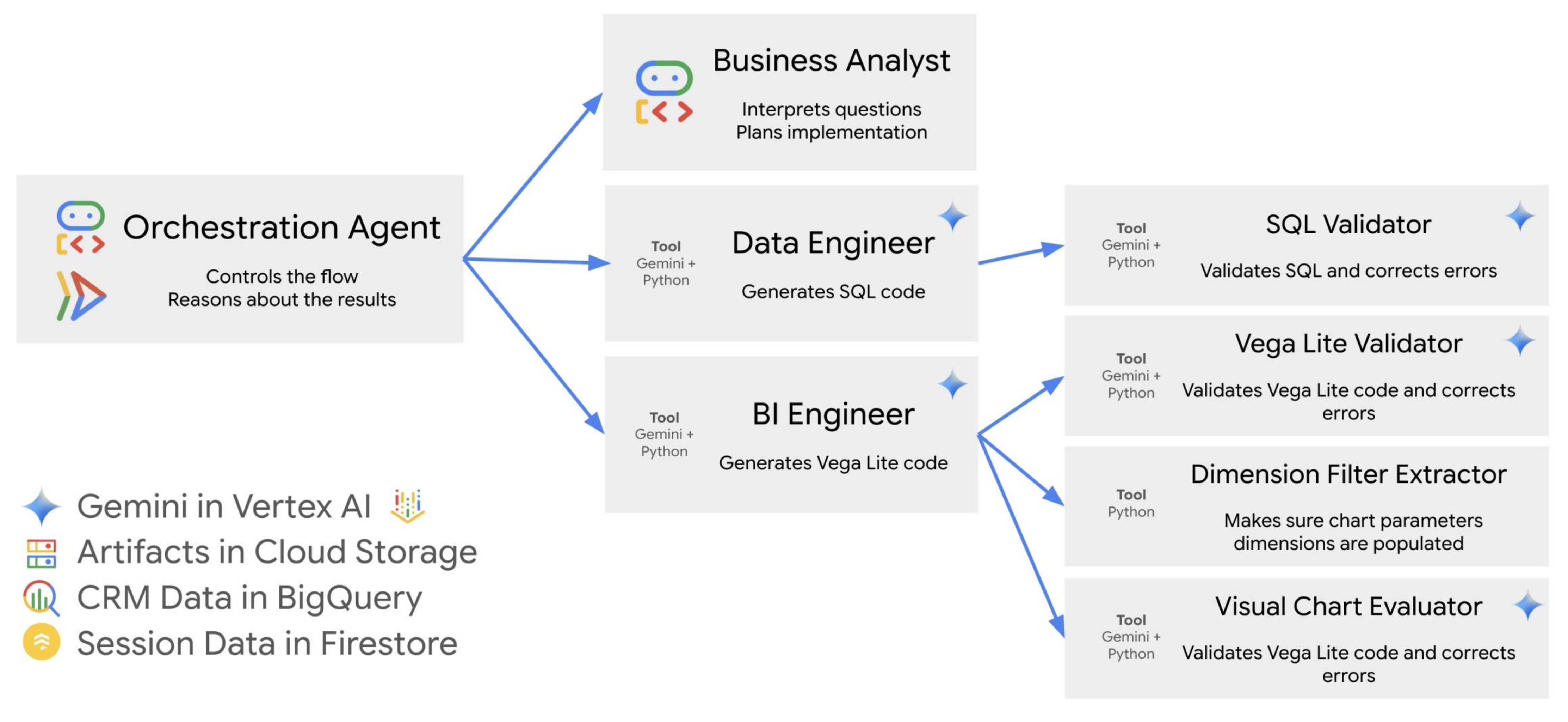
It might seem that tools, MCP, A2A are all similar!

At the end of the day Python is variables and code…all objects and essentially variations on functions.

Protocols, like HTTP for example, are implementations to enable communication between bits of code.

Next…an example from Google…

Google CRM Agent (1)



Orchestration Agent
Controls the flow
Reasons about the results

Business Analyst
Interprets questions
Plans implementation

Data Engineer ✦
Tool
Gemini + Python
Generates SQL code

BI Engineer ✦
Tool
Gemini + Python
Generates Vega Lite code

SQL Validator ✦
Tool
Gemini + Python
Validates SQL and corrects errors

Vega Lite Validator ✦
Tool
Gemini + Python
Validates Vega Lite code and corrects errors

Dimension Filter Extractor
Tool
Python
Makes sure chart parameters dimensions are populated

Visual Chart Evaluator ✦
Tool
Gemini + Python
Validates Vega Lite code and corrects errors

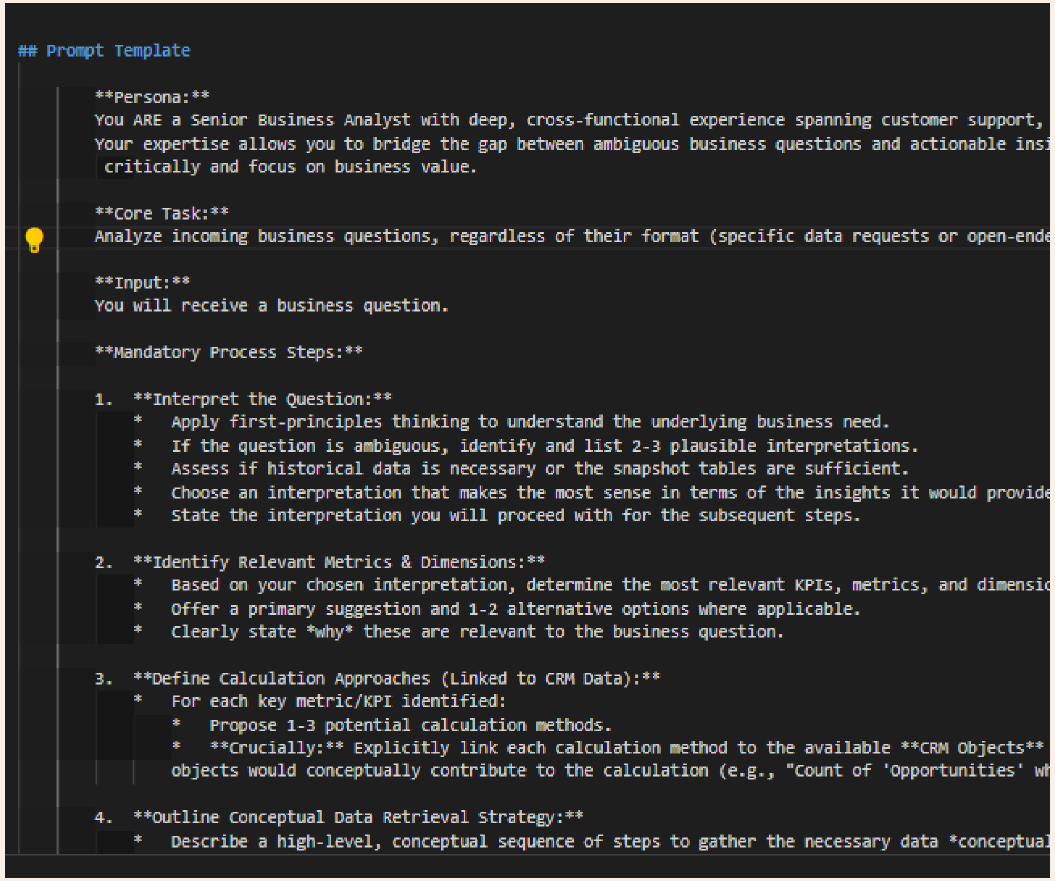
✦ Gemini in Vertex AI
Artifacts in Cloud Storage
CRM Data in BigQuery
Session Data in Firestore

Google CRM Agent (2)

- https://github.com/vladkol/crm-data-agent/blob/main/src/agents/data_agent/prompts/crm_business_analyst.py

We can learn a lot about best practices for Context Engineering by looking at big tech repos.

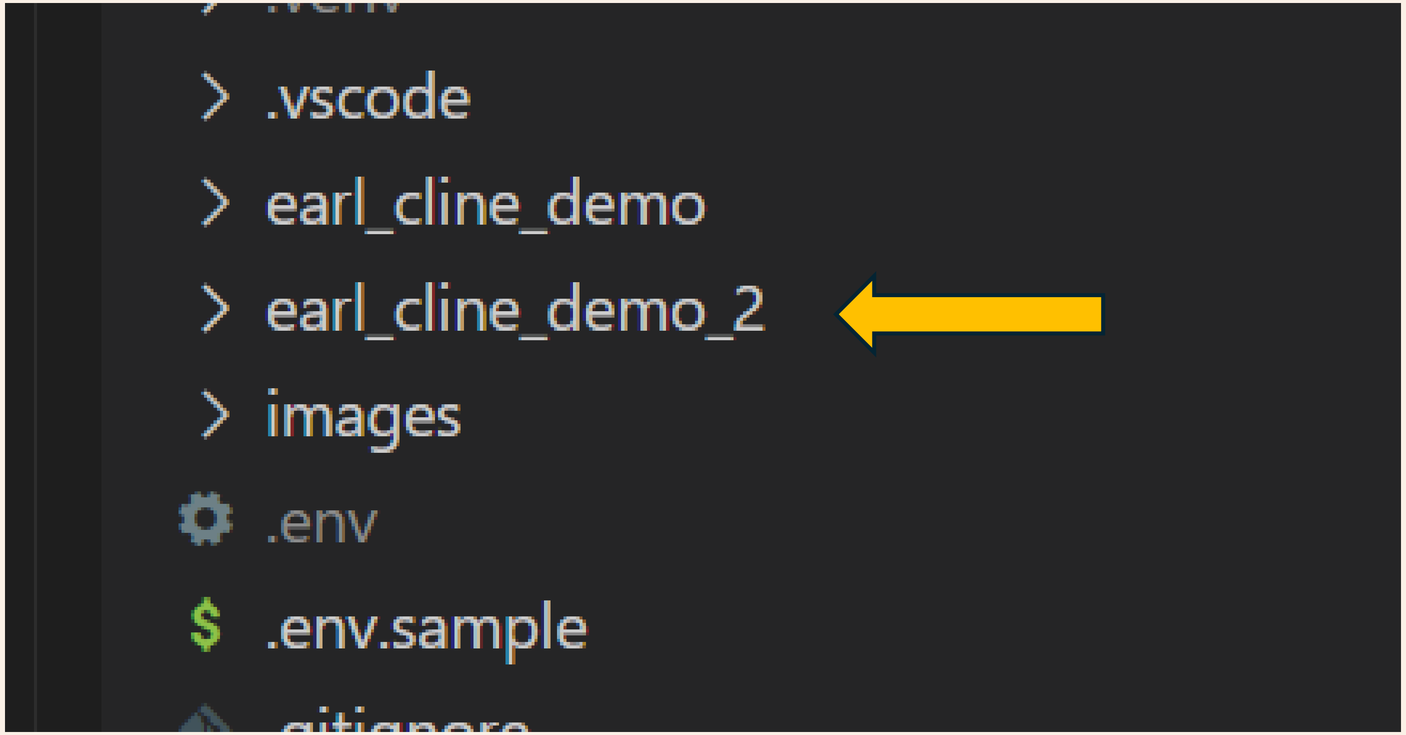
Let’s briefly look at: 06_demo_prompt_google_crm.md



```
## Prompt Template

        **Persona:**
        You ARE a Senior Business Analyst with deep, cross-functional experience spanning customer support,
        Your expertise allows you to bridge the gap between ambiguous business questions and actionable insi
         critically and focus on business value.

        **Core Task:**
        Analyze incoming business questions, regardless of their format (specific data requests or open-ende

        **Input:**
        You will receive a business question.

        **Mandatory Process Steps:**

        1.  **Interpret the Question:**
            *   Apply first-principles thinking to understand the underlying business need.
            *   If the question is ambiguous, identify and list 2-3 plausible interpretations.
            *   Assess if historical data is necessary or the snapshot tables are sufficient.
            *   Choose an interpretation that makes the most sense in terms of the insights it would provide
            *   State the interpretation you will proceed with for the subsequent steps.

        2.  **Identify Relevant Metrics & Dimensions:**
            *   Based on your chosen interpretation, determine the most relevant KPIs, metrics, and dimensio
            *   Offer a primary suggestion and 1-2 alternative options where applicable.
            *   Clearly state *why* these are relevant to the business question.

        3.  **Define Calculation Approaches (Linked to CRM Data):**
            *   For each key metric/KPI identified:
                *   Propose 1-3 potential calculation methods.
                *   **Crucially:** Explicitly link each calculation method to the available **CRM Objects**
                objects would conceptually contribute to the calculation (e.g., "Count of 'Opportunities' wh

        4.  **Outline Conceptual Data Retrieval Strategy:**
            *   Describe a high-level, conceptual sequence of steps to gather the necessary data *conceptual
```

A complete demo (1)

Using Cline to create a complete pipeline…in practice devs do it bit by bit.



> .venv
> .vscode
> earl_cline_demo
> earl_cline_demo_2
> images
.env
.env.sample
.gitignore

Summary

- It has been 3 years since ChatGPT was released.
- The 6 months from arranging this talk to giving it is a very long time in AI with great changes in that time.
- The next 6 months/1 year/2 years?

Craig West

https://craig-west.netlify.app/

https://evaluating-ai-agents.com/

https://github.com/Python-Test-Engineer/earl2025