Reply Noipa Agentic AI

Presented by the Trifenix:

N. Campaniello

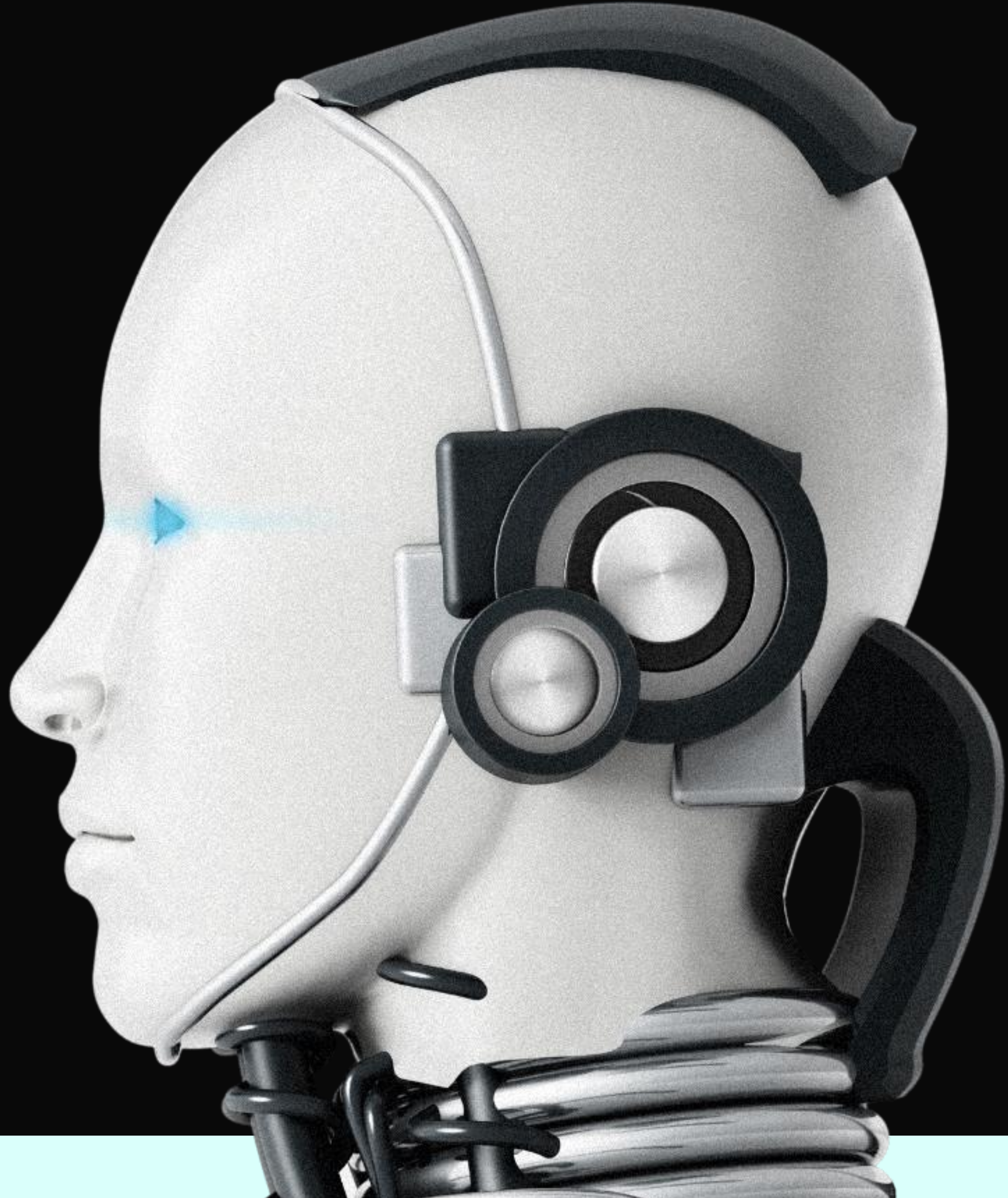
F. Cesari

A. Cappelluti

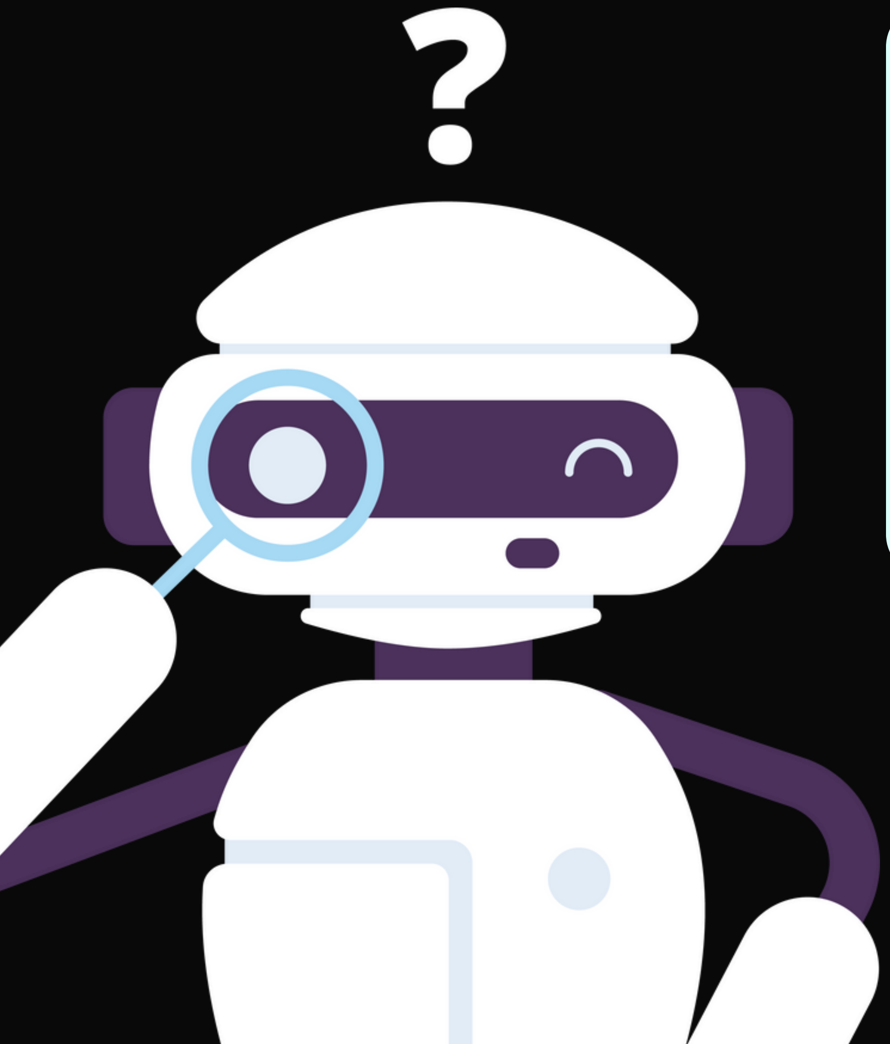
FRAMEWORK OPENAI SDK AGENT

RELEASE DATE ; March 11, 2025.

API KEY ; GPT
4.1



Explanatory Data Analysis



EntryAccessoAmministrati.csv
Users’ region of residence, affiliated administration, age group, gender & access method

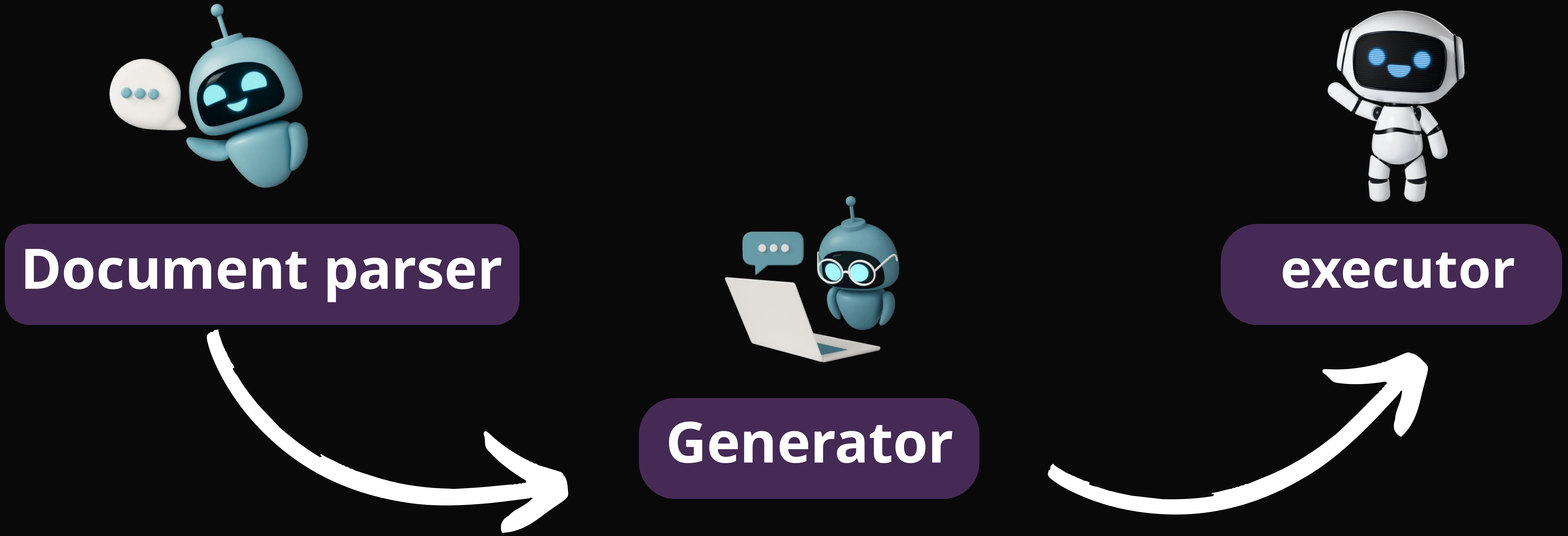
EntryAccreditoStipendi.csv
Number of payments by administration, age group, gender & payment method

Our work?
change data types
Translate column names
Add columns

EntryPendolarismo.csv
Workplace province/municipality, administration, departure vs. arrival match, distance range & user counts

EntryAmministrati.csv
Personnel distribution by municipal unit, age group & gender for the reference month

OVERVIEW



Document Parser

Initially
Key-word mapping approach

Problem
this method was static and unable to understand the broader context of the query.

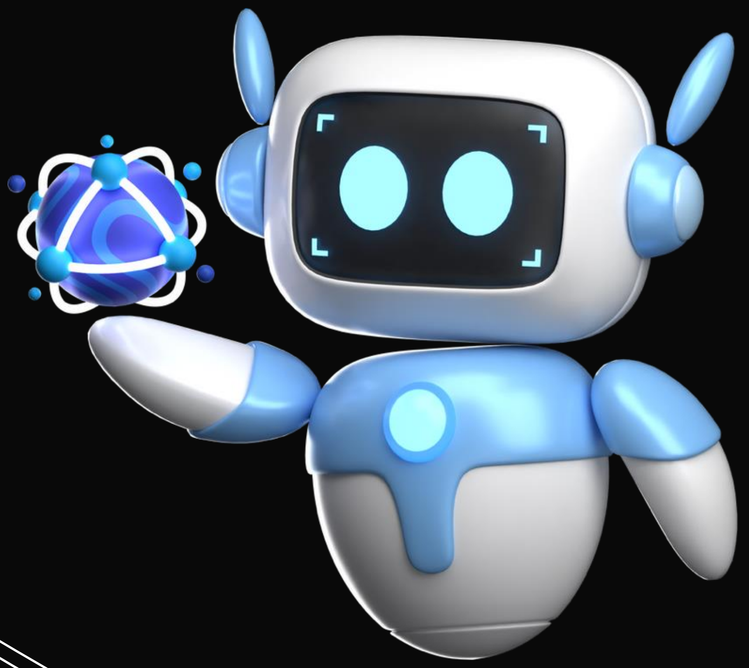
Hugging Face

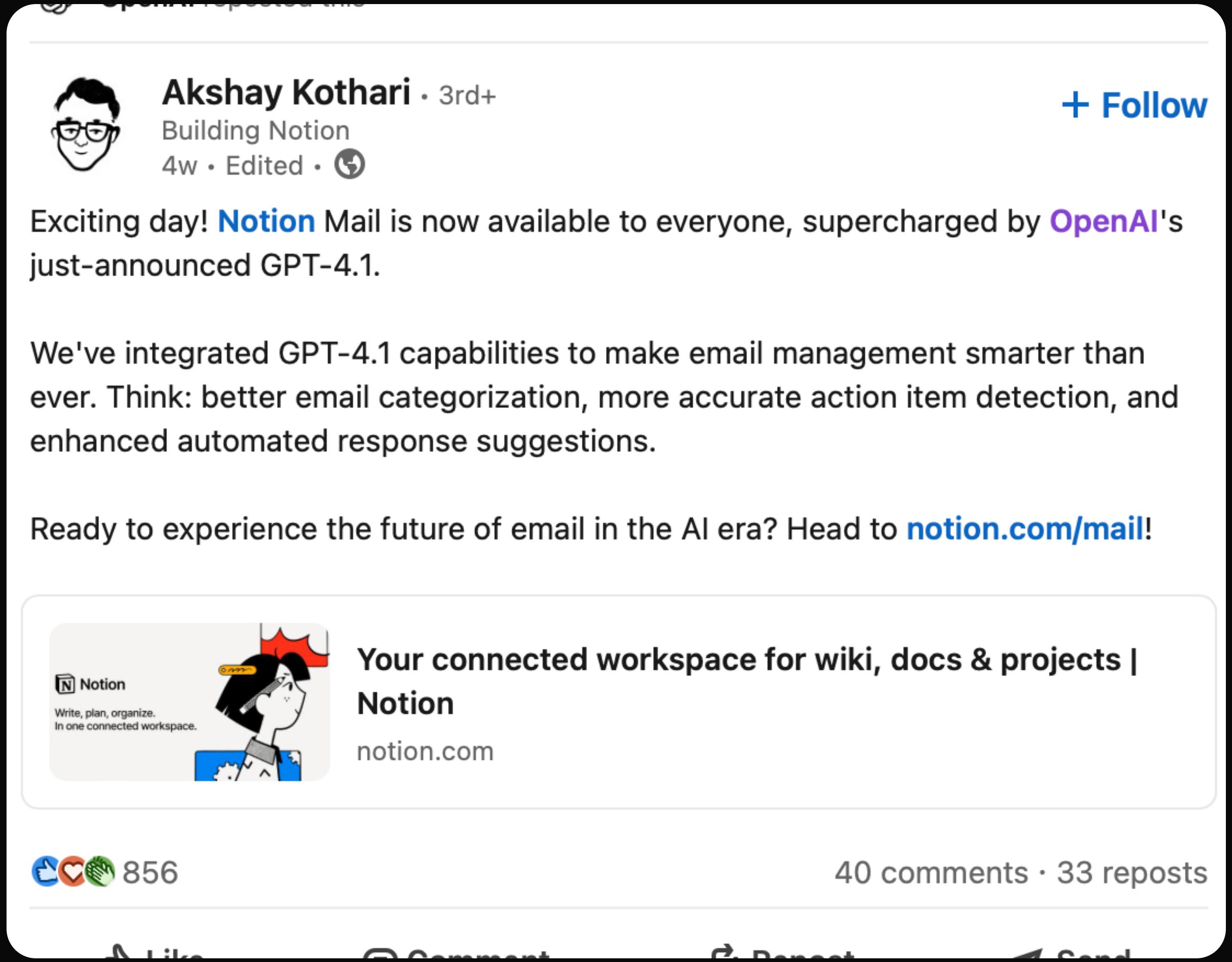
Sentence transformer ┈┈┈> **Llama index**

Generate Python Code

Python written tool was static ⟶ Let’s adopt the prompt based system

- Dynamic code generation
- Flexible output code
- Accuracy blow up





Akshay Kothari • 3rd+
Building Notion
4w • Edited •

Exciting day! Notion Mail is now available to everyone, supercharged by OpenAI's just-announced GPT-4.1.

We've integrated GPT-4.1 capabilities to make email management smarter than ever. Think: better email categorization, more accurate action item detection, and enhanced automated response suggestions.

Ready to experience the future of email in the AI era? Head to notion.com/mail!

Your connected workspace for wiki, docs & projects | Notion
notion.com

856 · 40 comments · 33 reposts

Code Executor & Memory

Code Executor:

Main reporter’s Tool

Takes in input the generated python code and automatically run it

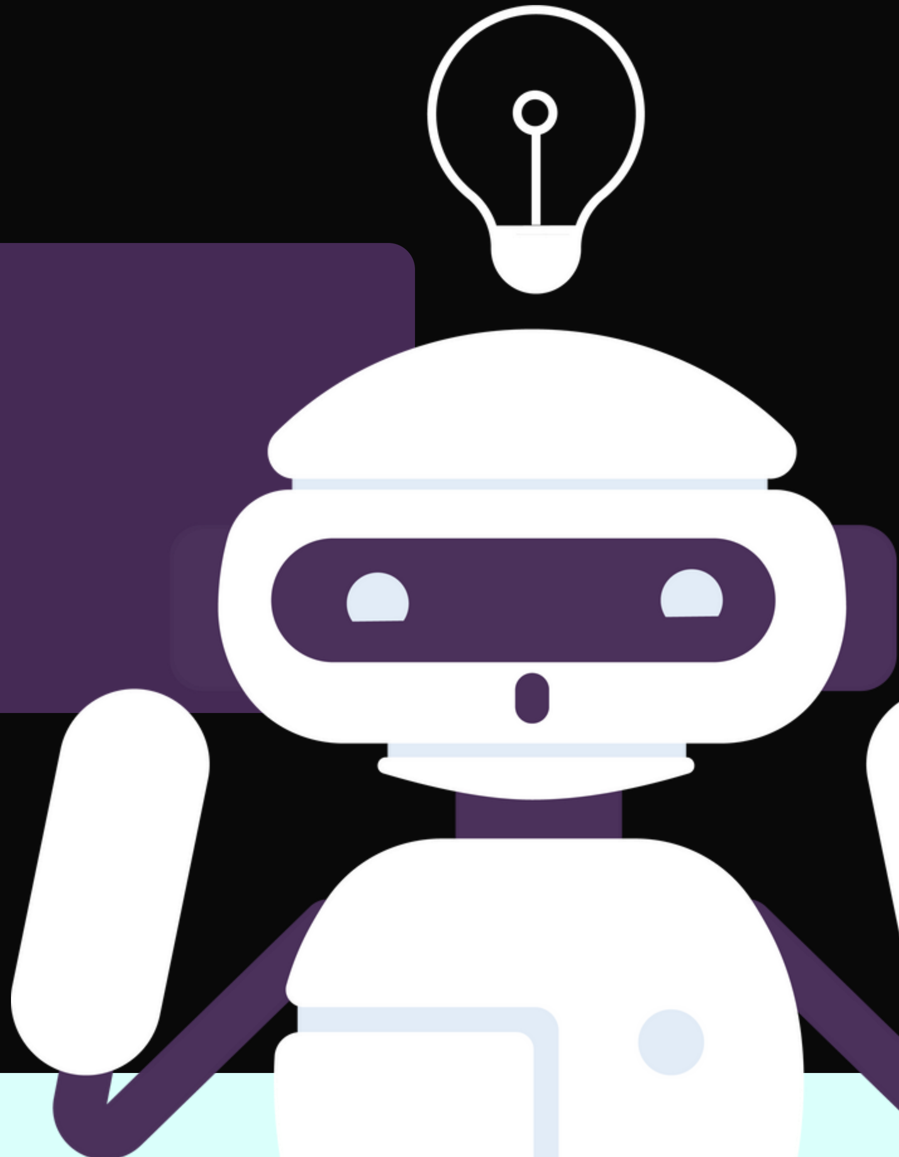
in this way is possible directly display the answers of the user questions.

Memory:

This allow the agent to mantain a conversational buffer

Buffer goes up to 6 answers-questions

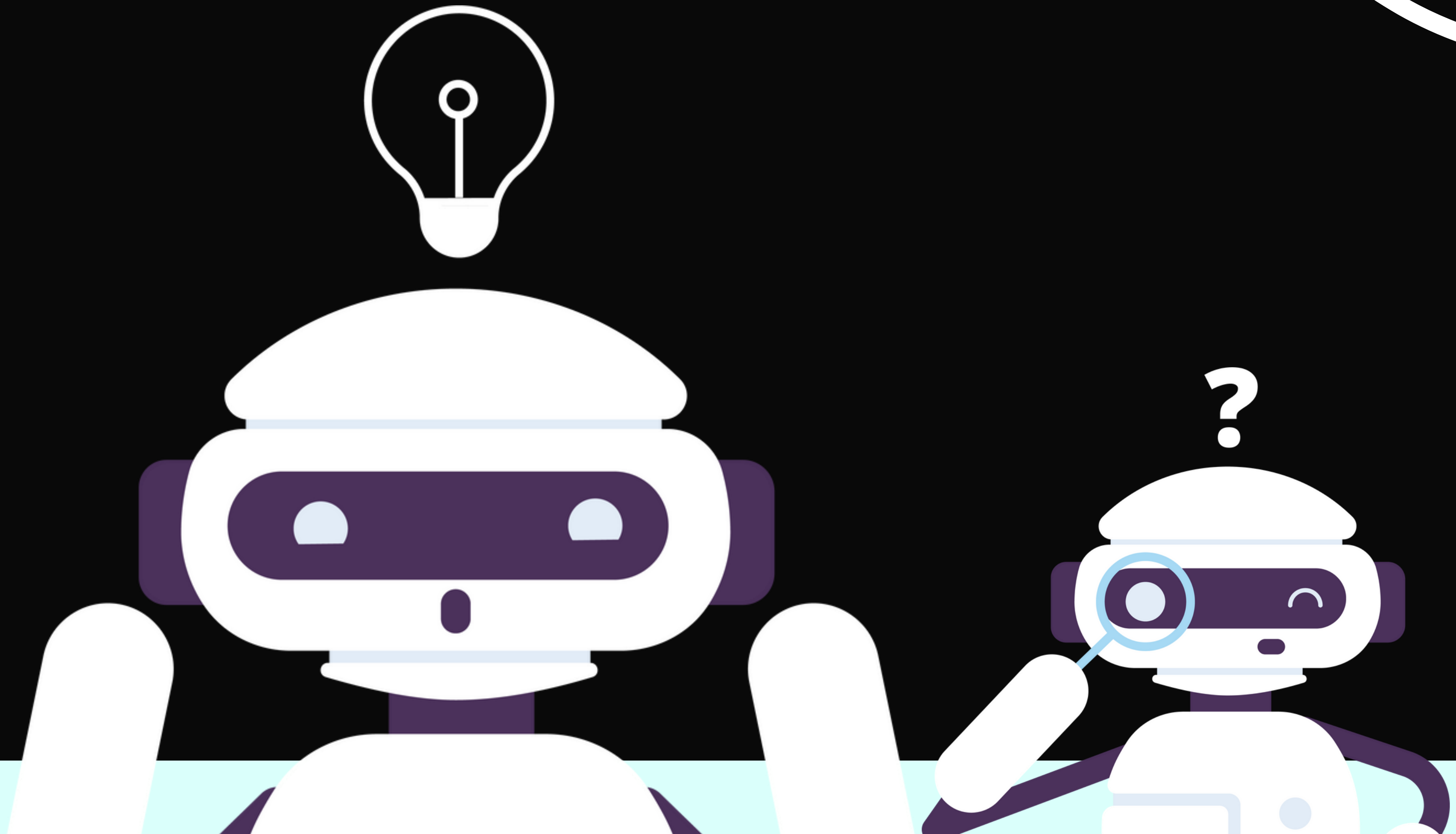
coherent system that avoid repetition and help to respond better



Memory was introduced for the agents.

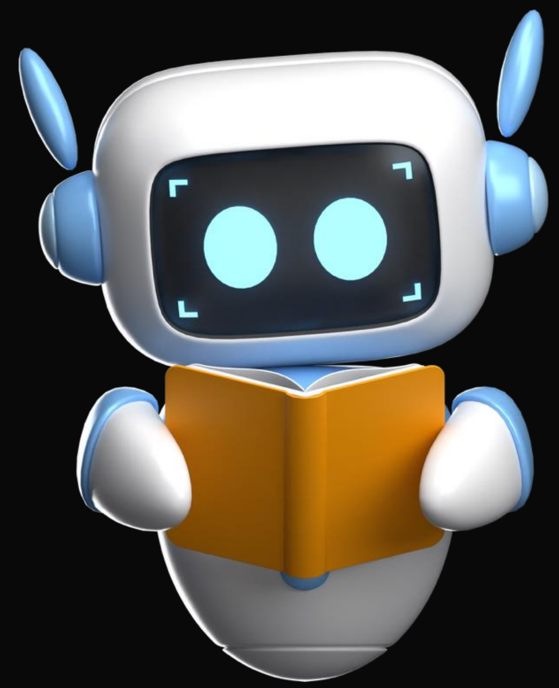
- Parts of the code written with AI assistance were:
  The Python code executor.
  The memory module.

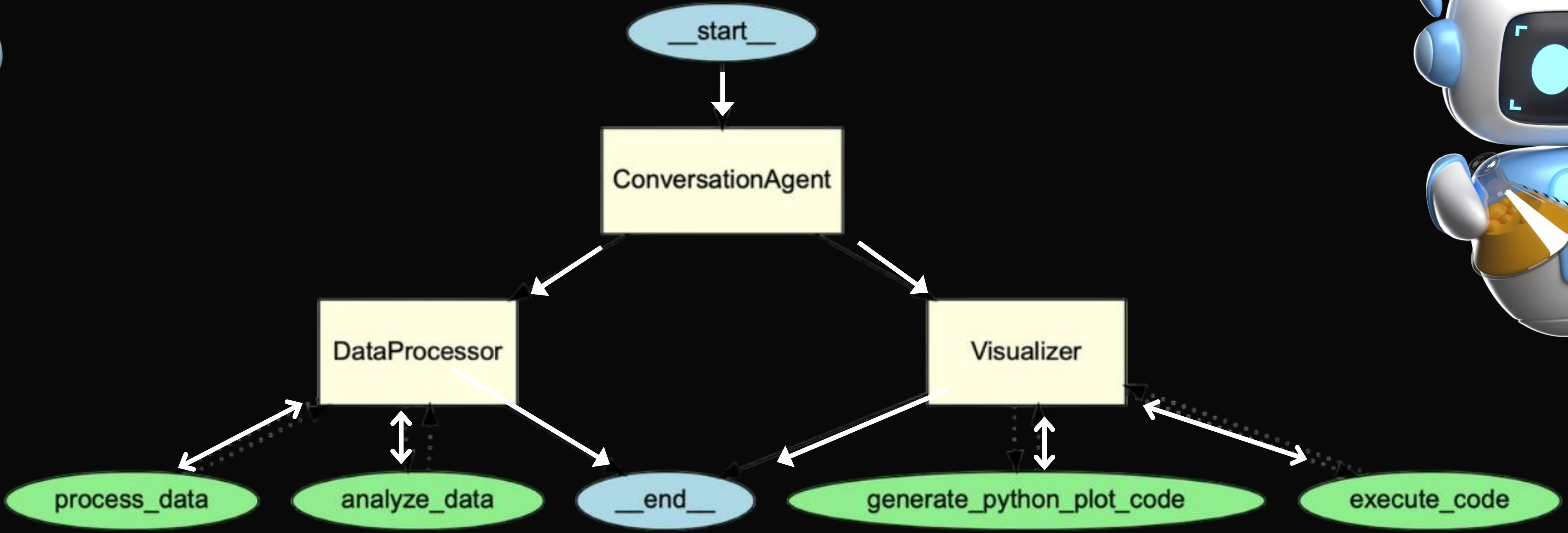
This was necessary because the framework used, OpenAI Agent SDK, did not include built-in functionalities like those provided by LangChain.

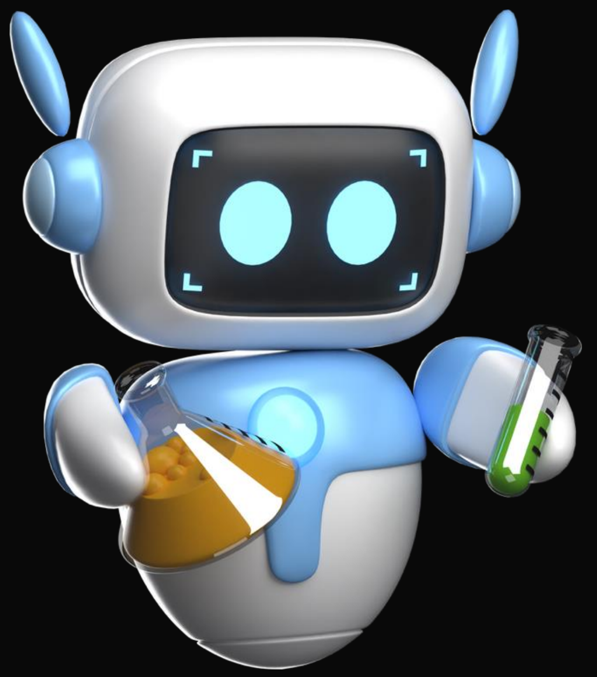


TECHNIQUE IMPLEMENTATION

Hierarchical Multi-Agent

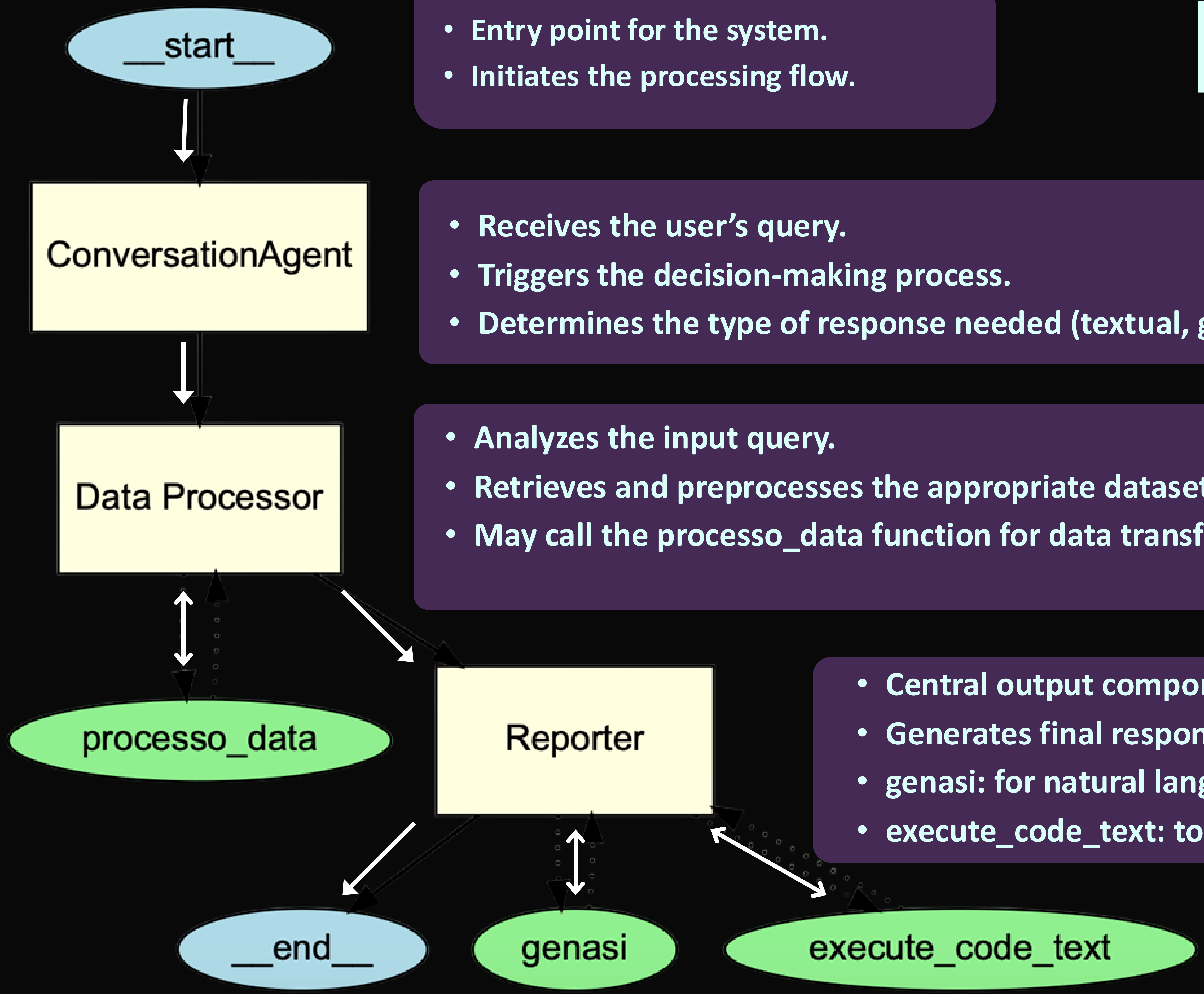






- **Redundant Tooling: Both the DataProcessor and Visualizer wrap nearly identical helper tools, leading to overlapping responsibilities and wasted implementation effort.**
- **Maintenance Overhead: Updating or fixing similar tool logic in multiple agents doubles the work and increases the risk of inconsistent behavior.**
- **Performance Inefficiency: Routing closely related tasks through separate agents adds unnecessary communication steps and slows down end-to-end execution.**

PIPELINE



- Entry point for the system.
- Initiates the processing flow.

- Receives the user’s query.
- Triggers the decision-making process.
- Determines the type of response needed (textual, graphical, etc.).

- Analyzes the input query.
- Retrieves and preprocesses the appropriate dataset.
- May call the processo_data function for data transformation.

- Central output component.
- Generates final responses using:
- genasi: for natural language generation.
- execute_code_text: to run Python code when needed.

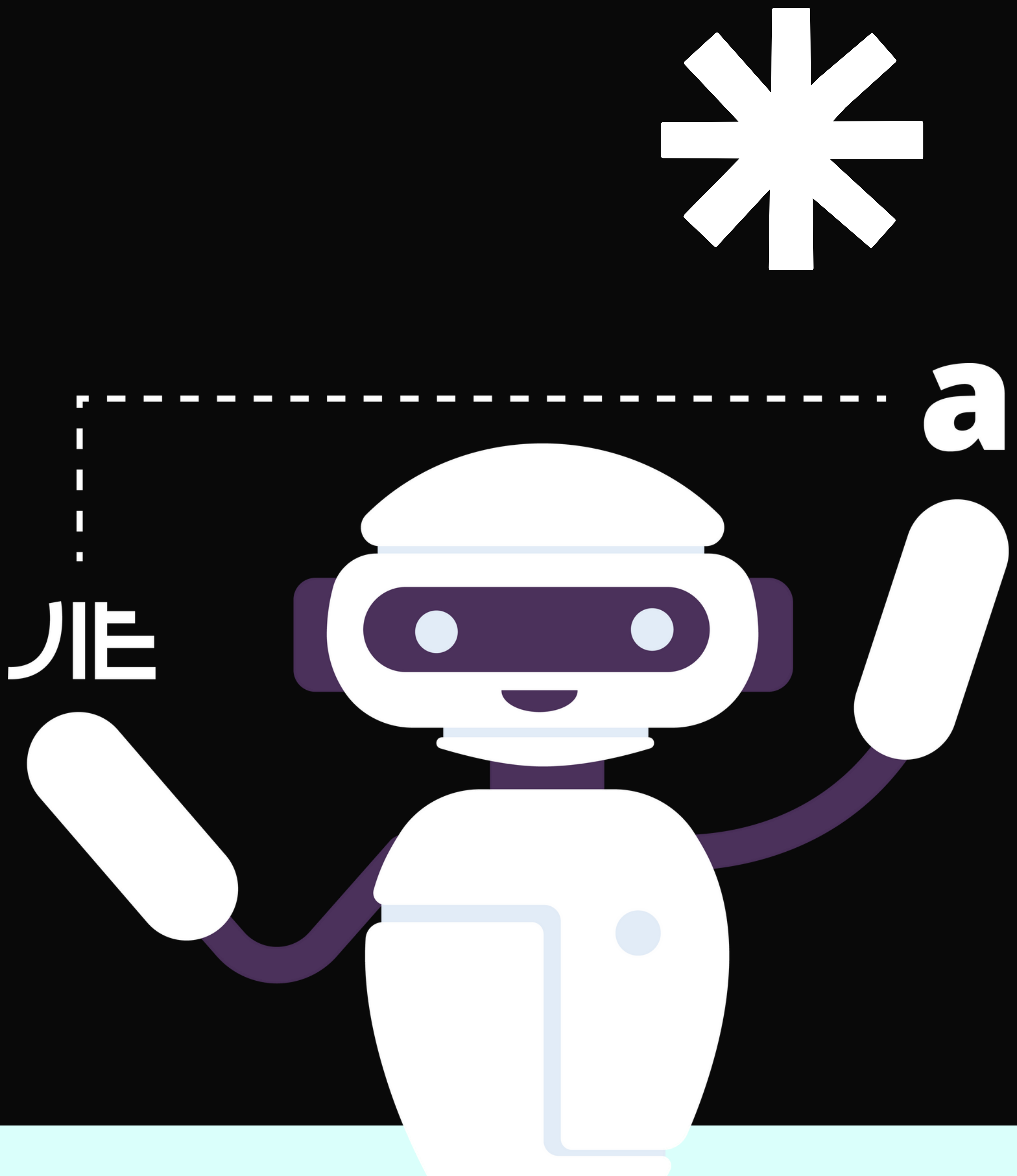
Response Evaluation

- To evaluate the agents' answers:
  - **DeepEval** was used to evaluate specific answers provided by the agent but it lacks context
  - **Giskard** was also tested, but it is currently unable to evaluate plots/graphs.

↓

ChatGPT-o3 was used with full context, resulting in a score of 91 out of 100.

so we have built an excel for query evaluation



Limitations

Hallucinations & Inaccuracies:

May generate plausible-sounding but incorrect or fabricated information

Latency & Cost:

High-quality responses can incur noticeable delays and significant per-token API expenses.

**

Thankyou

@reallygreatsite