

NATURAL LANGUAGE PROCESSING
Alexandre Gazagnes

BILL BUSTERS

Group Andromeda



MEET OUR TEAM



NIKITA

Supreme leader



JUAN

Manager of Managing



AYELENE

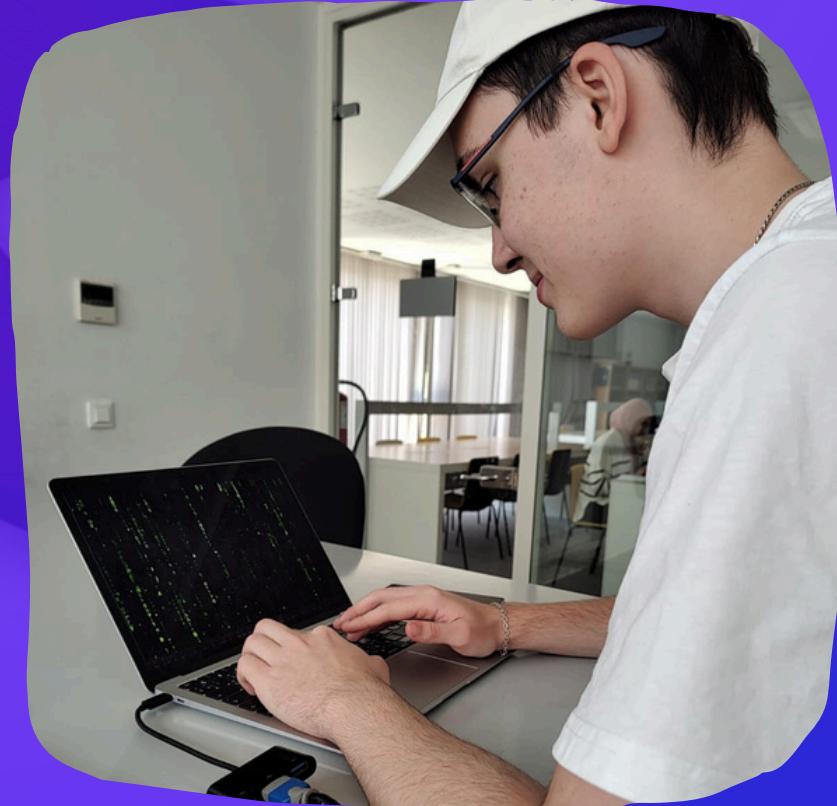
Full Time Indian



AUDE

Busy Mom

OUR JOURNEY



PROJECT OVERVIEW



Web scraping

We **track a website** that stays up-to-date with the latest legislative bills in the US, and among the ones that have passed Congress and were signed by the president, we **check for any news**, and if there indeed are new bills, we **extract the text** from them to give it to the AI Agent later

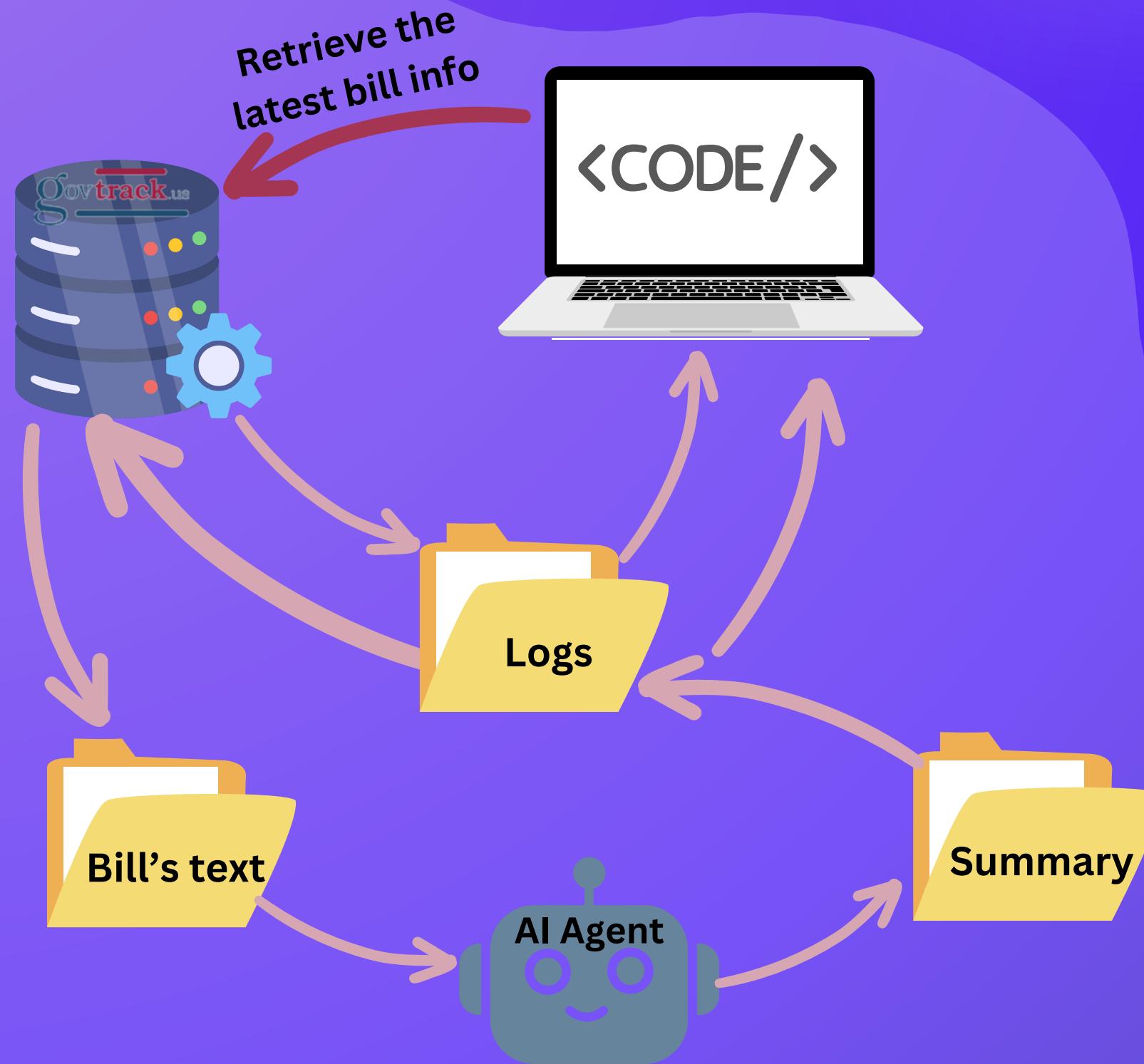
AI Explanator

Using OpenAI's API to access their latest models, we **fine-tuned it** into being able to **summarize the raw text** into an easy-to-understand piece of content



SCHEMA

SCHEMA



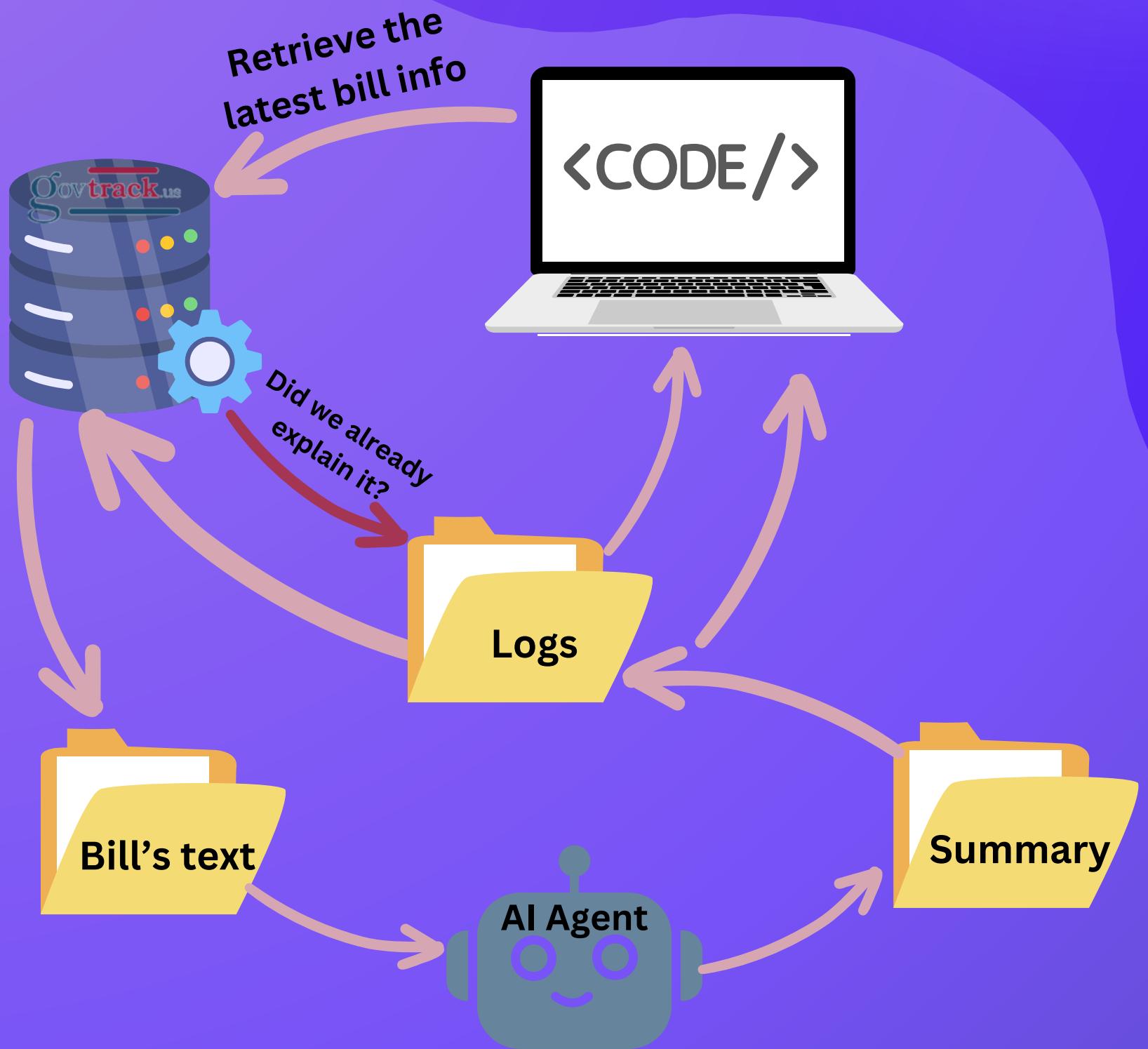
STEP 1

Applying **filters** to only show the relevant bills (passed congress and signed by the President), retrieve the title and the link for the latest bill

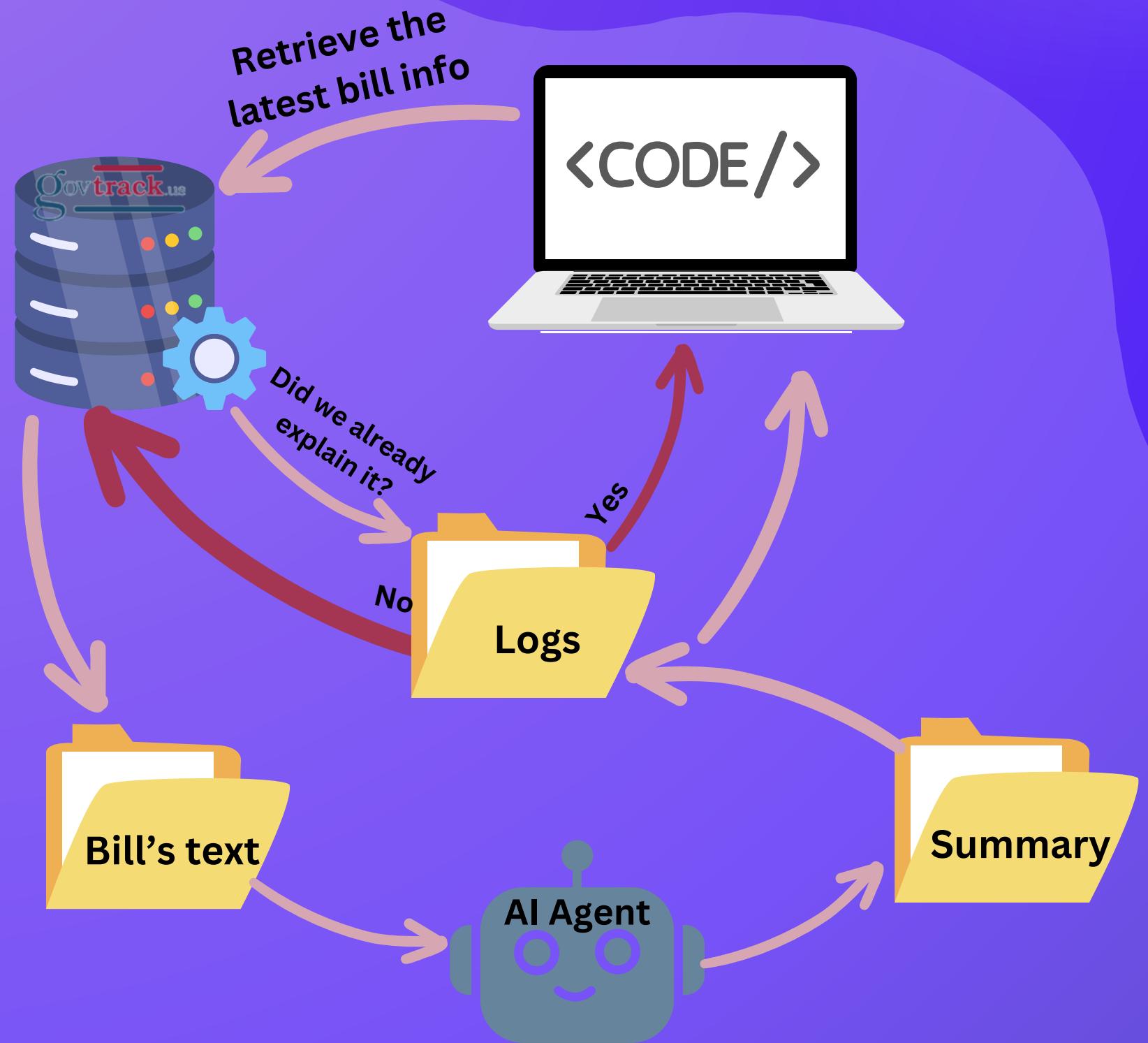
SCHEMA

STEP 2

Check our **logs** for the last bill that we completed



SCHEMA

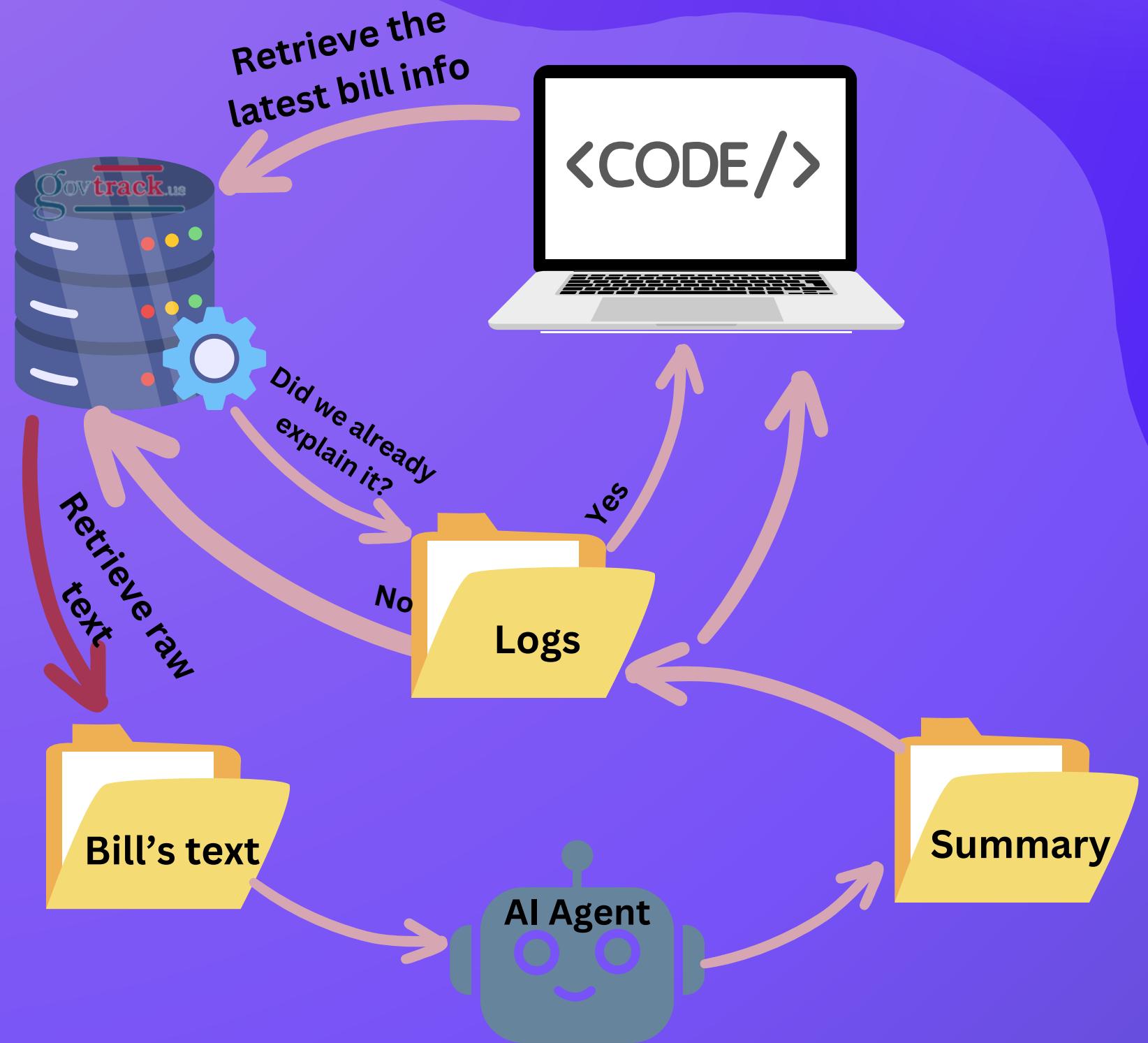


STEP 3

If we already explained the bill, quit the program

If there is indeed a new bill passed, we move on

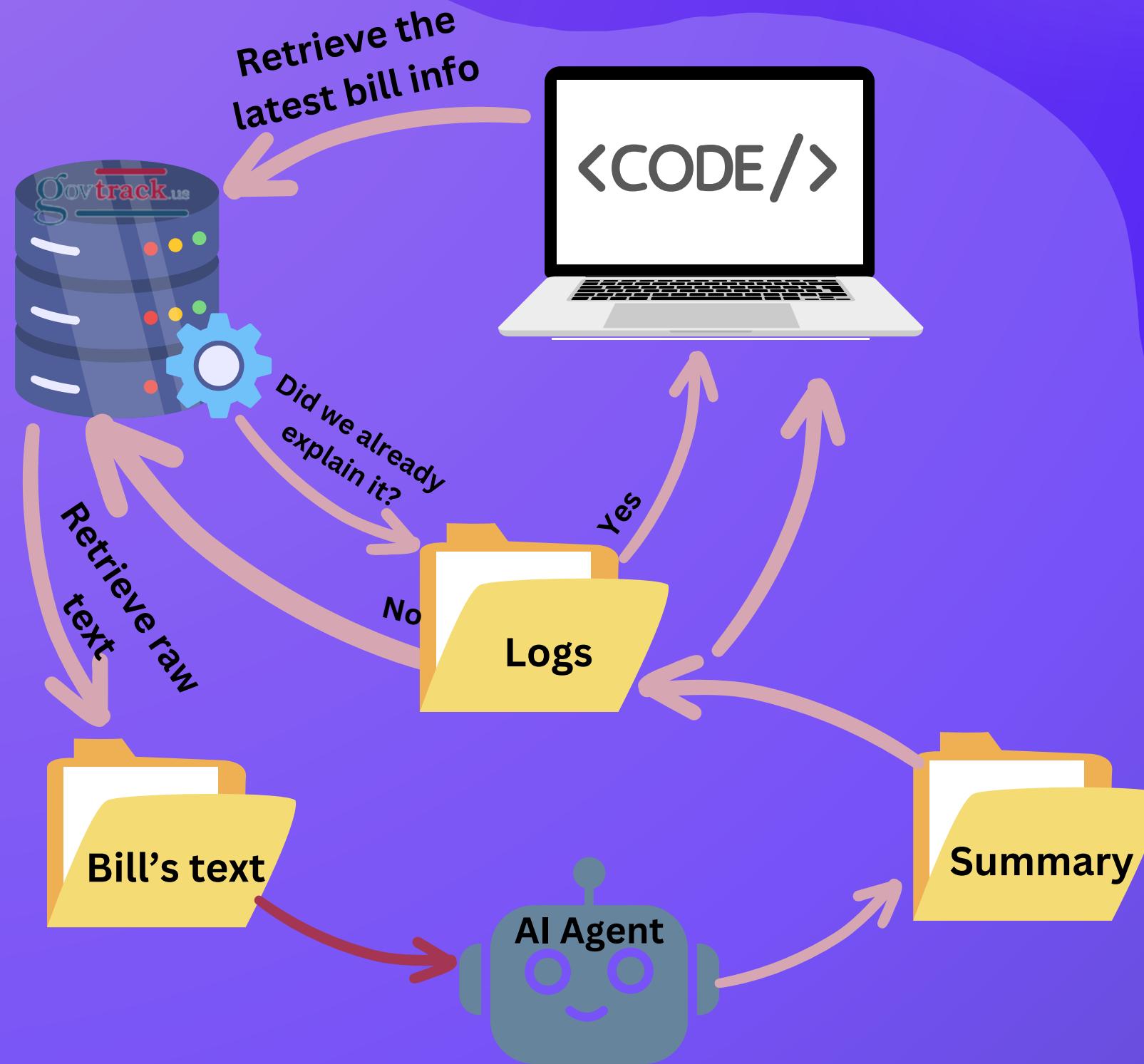
SCHEMA



STEP 4

Using the link for the bill we got earlier, we parse it for the full text of the bill and store it in a separate directory, with a file name = Bill's ID

SCHEMA



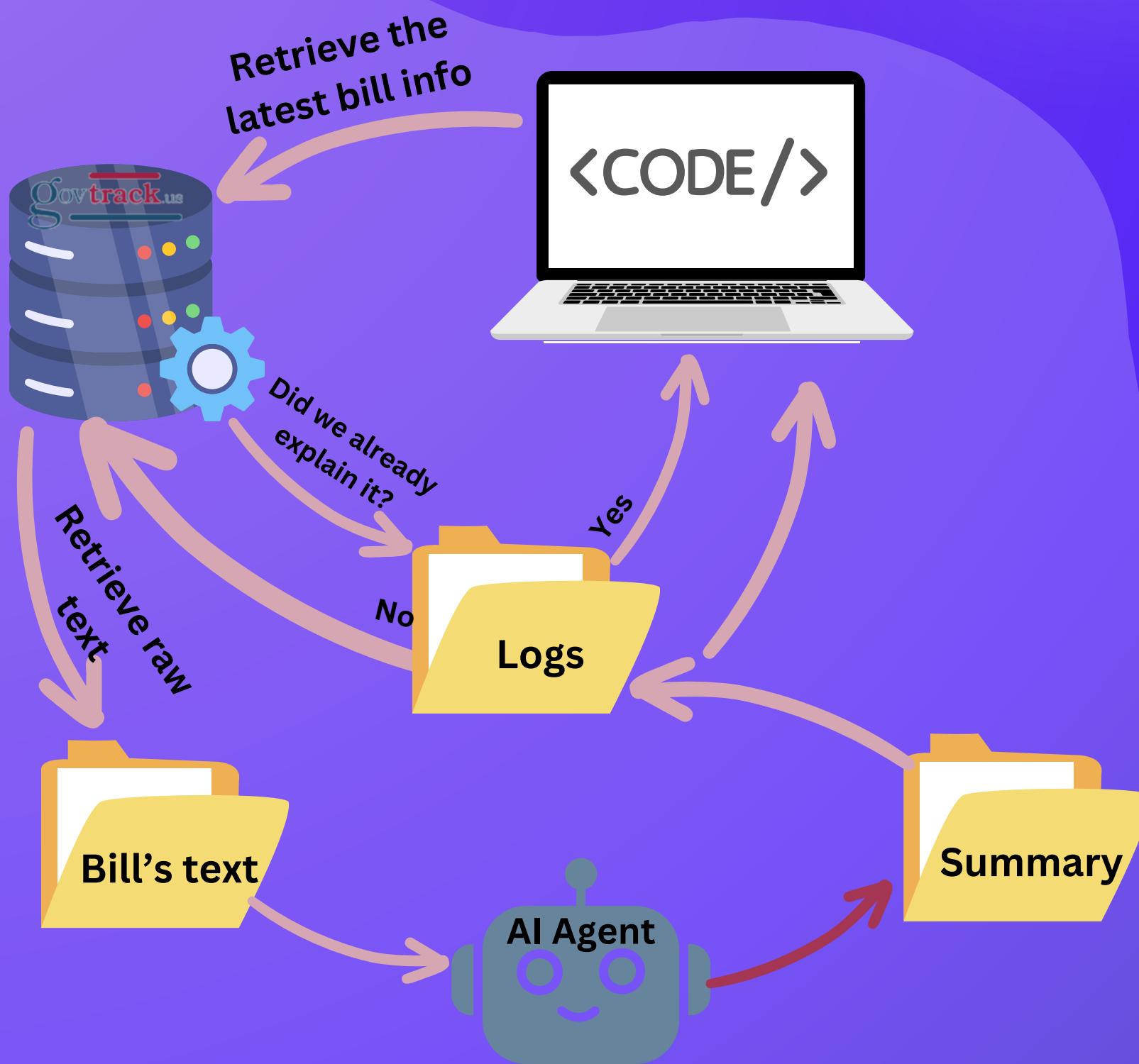
STEP 5

We send the text to a custom-made AI Agent, using OpenAI's API (For this project we used gpt-3.5-16k as it's cheap to test on)

SCHEMA

STEP 6

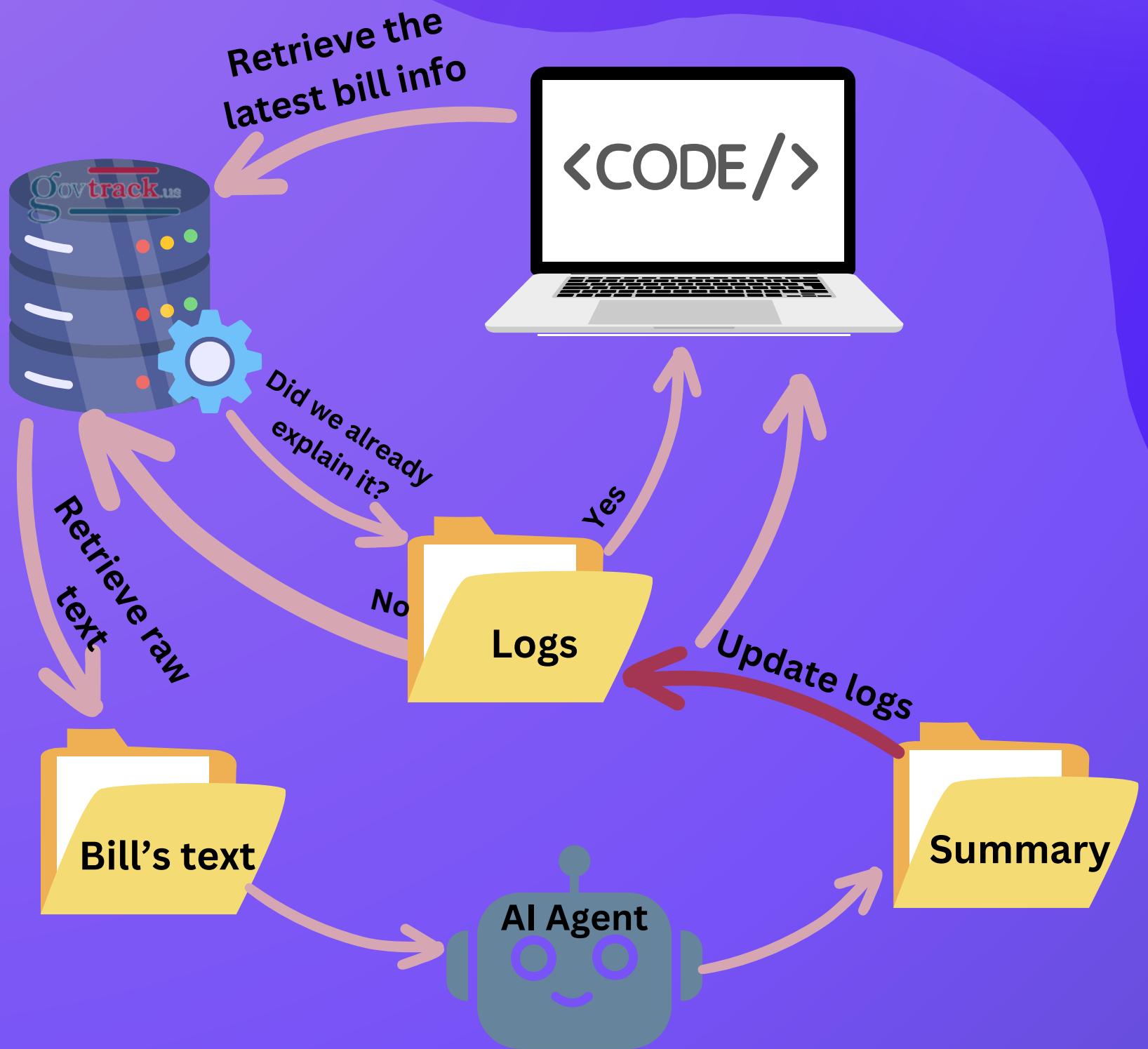
We **save the output** of the agent into a new directory with custom file name (like in previous steps)



SCHEMA

STEP 7

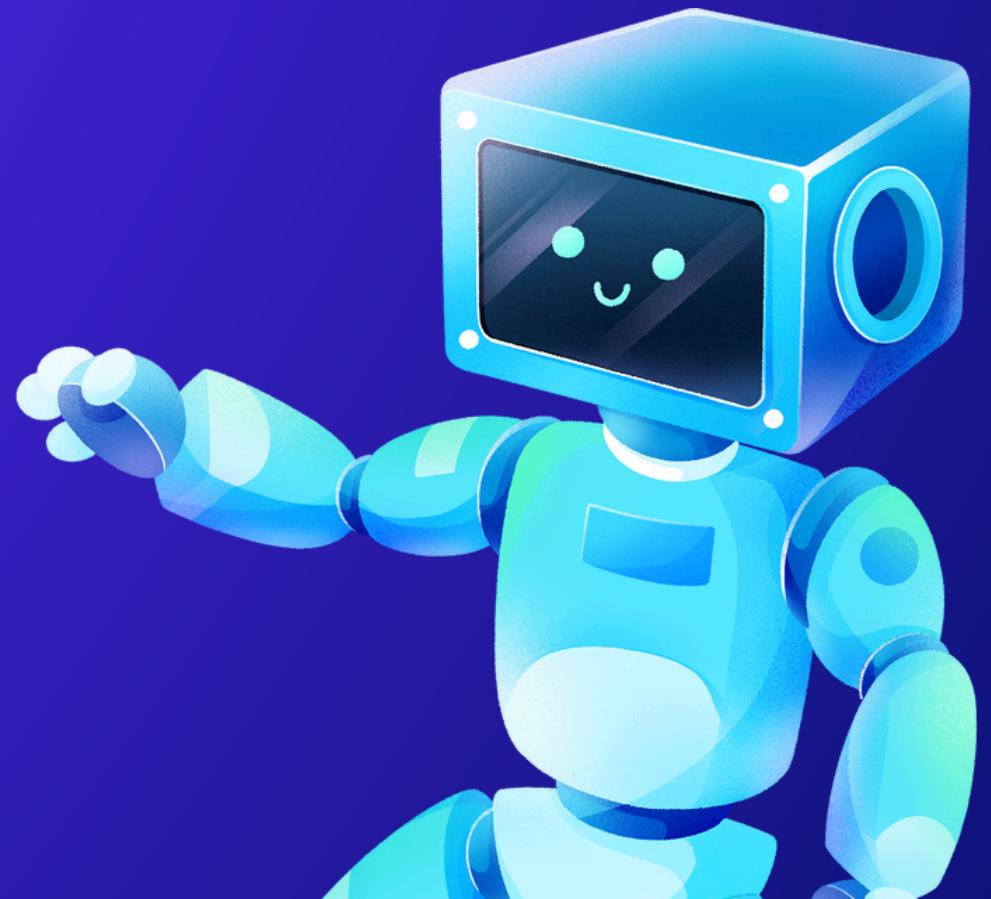
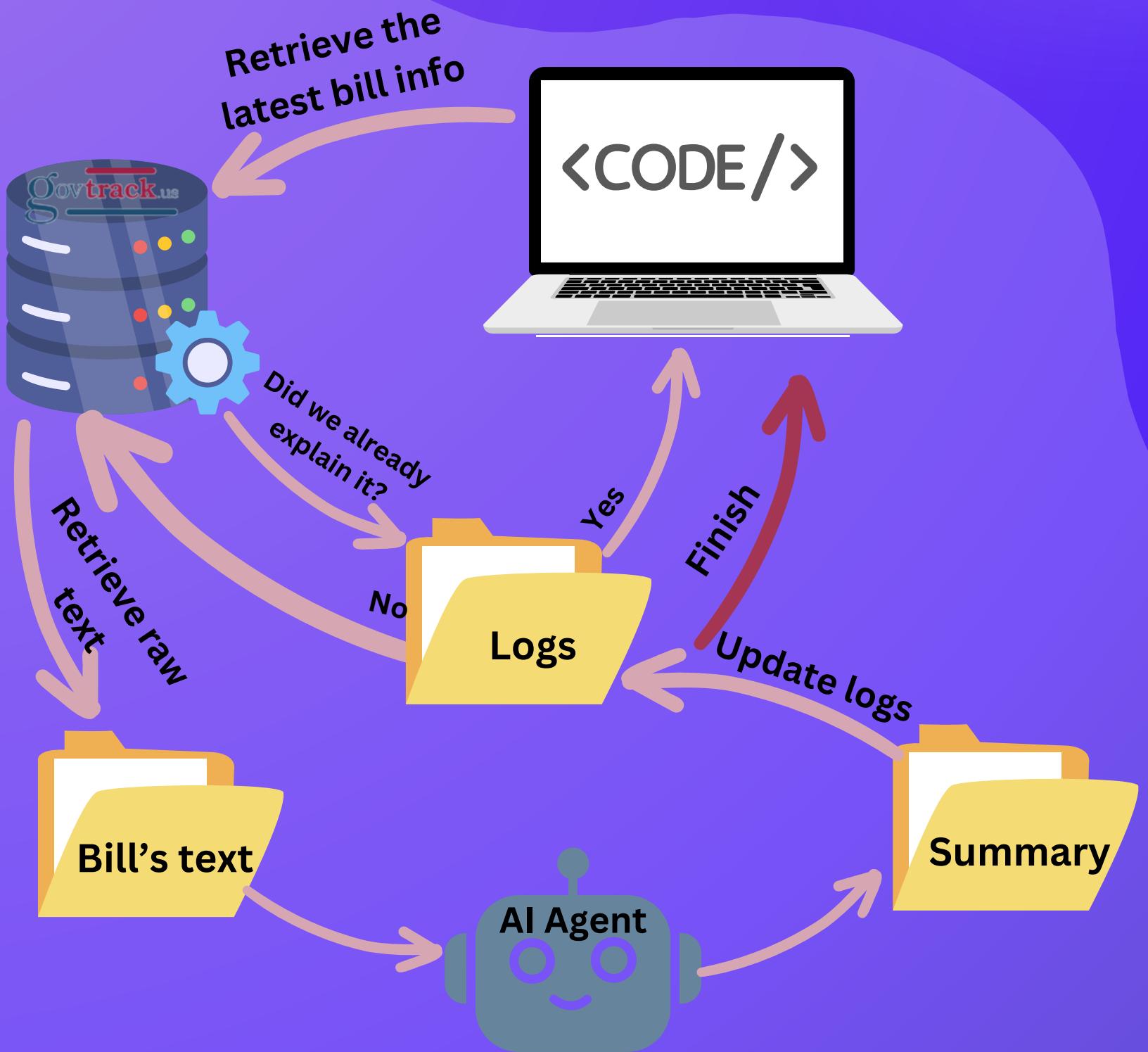
Update the logs file with
the title of the bill



SCHEMA

STEP 8

Finish the program



WORK DISTRIBUTION & COMMITS

Aude

Main logic

Juan & Ayelene

Scraping

Nikita

Agent

Fixed issues with config, added logic to work with files Nikses committed 19 hours ago	fc9006f		
Added the main logic for the main program AudeGrd committed 20 hours ago	8a04091		
Added function to check the latest bill jjmayorgah committed yesterday	bbf06e8		
update to work in local repo, not google drive, saving in .txt, cutting short useless content from text file Nikses committed yesterday	4f76242		
updated .gitignore Nikses committed yesterday	af8052c		
-o- Commits on Apr 30, 2024			
Initial webscrape logic IleneBiju7 committed 2 days ago	5638ff3		
Main logic for agent Nikses committed 2 days ago	2ff3cf7		
Added gitignore Nikses committed 2 days ago	27342d0		
Stopped tracking .DS_Store and .idea/ Nikses committed 2 days ago	03eb91b		

30 webscrape.py

```
... ... @@ -1,4 +1,32 @@
1 - #functions to scrape the website for the latest bill
1 + # functions to scrape the website for the latest bill
2 + import requests
3 + from bs4 import BeautifulSoup
4 + from docx import Document
5
6
7 url = 'https://www.govtrack.us/congress/bills/browse?status=passed#sort=-introduced_date&current_status[]>=28'
```

IleneBiju7 1 hour ago Author Collaborator

Not being used

Nikses 1 hour ago Owner

Are you kidding me check the later commits 😂

IleneBiju7 via email 32 minutes ago Author Collaborator

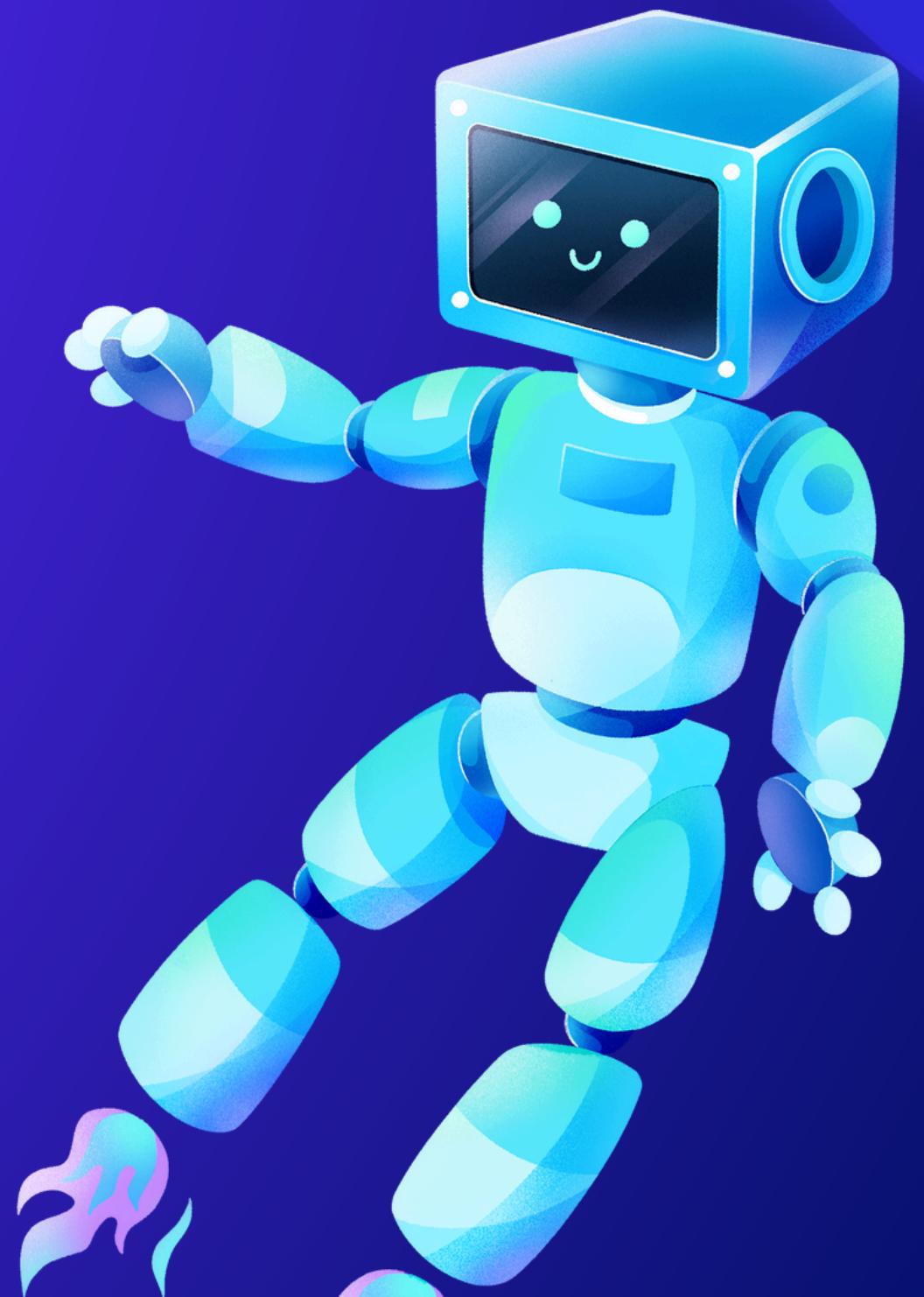
Okay smartass

Reply...



A background featuring abstract geometric shapes like triangles and hexagons in white and light blue, with small glowing teal dots at their vertices, set against a dark purple gradient.

CODE



THANK YOU!

