

Backend Intern Take-Home Assignment: HumanChain-AI-Safety-incident-Log-API

Name : D. Sai Hemanth Varma

Registration Number: 12215860

College: Lovely Professional University

Github Link:

<https://github.com/danthulurisaihemanth/HumanChain-AI-Safety-incident-Log-API>

README.md Link :

<https://github.com/danthulurisaihemanth/HumanChain-AI-Safety-incident-Log-API/blob/master/README.md>

PROJECT DEMO VIDEO LINK:

https://drive.google.com/file/d/1sDFmNQykfFnHce_RDeDeSk0VfJ2-fmmg/view?usp=sharing

AI Safety Incident Log API

A simple RESTful API service to log and manage hypothetical AI safety incidents, with a web-based user interface.

Technology Stack

- **Language**: Python
- **Framework**: Flask
- **Database**: MySQL
- **Backend**: Flask(python)
- **Api's**: RESTful APIs using Flask
- **Frontend**: HTML, CSS

🧪 Tools Used

- Postman (for testing API endpoints)

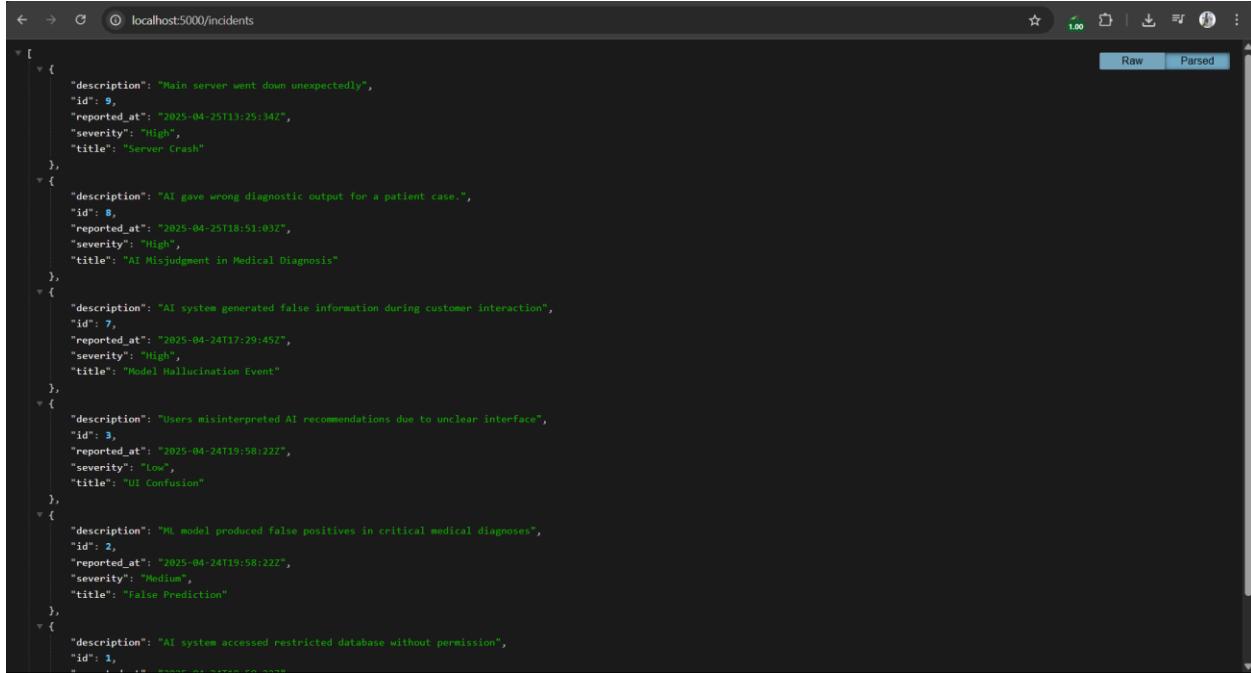
Screen Shots:

login_page:

The screenshot shows a web browser window with multiple tabs open. The active tab is titled "AI Safety Incident Log". The page has a dark header with the title "AI Safety Incident Log" and the subtitle "HumanChain - Building safer AI systems". On the left, there is a form titled "Report New Incident" with fields for "Title", "Description", "Severity" (a dropdown menu), and a "Submit Incident" button. On the right, there is a table titled "Incident Log" showing four entries:

Event Type	Severity	Description	Date	Action
Model Hallucination Event	High	AI system generated false information during customer interaction	4/24/2025, 10:59:45 PM	Delete
Model Hallucination	Low	AI generated factually incorrect information about historical events	4/24/2025, 9:10:10 PM	Delete
UI Confusion	Low	Users misinterpreted AI recommendations due to unclear interface	4/25/2025, 1:28:22 AM	Delete
False Prediction	Medium	ML model produced false positives in critical medical diagnoses	4/25/2025, 1:28:22 AM	Delete

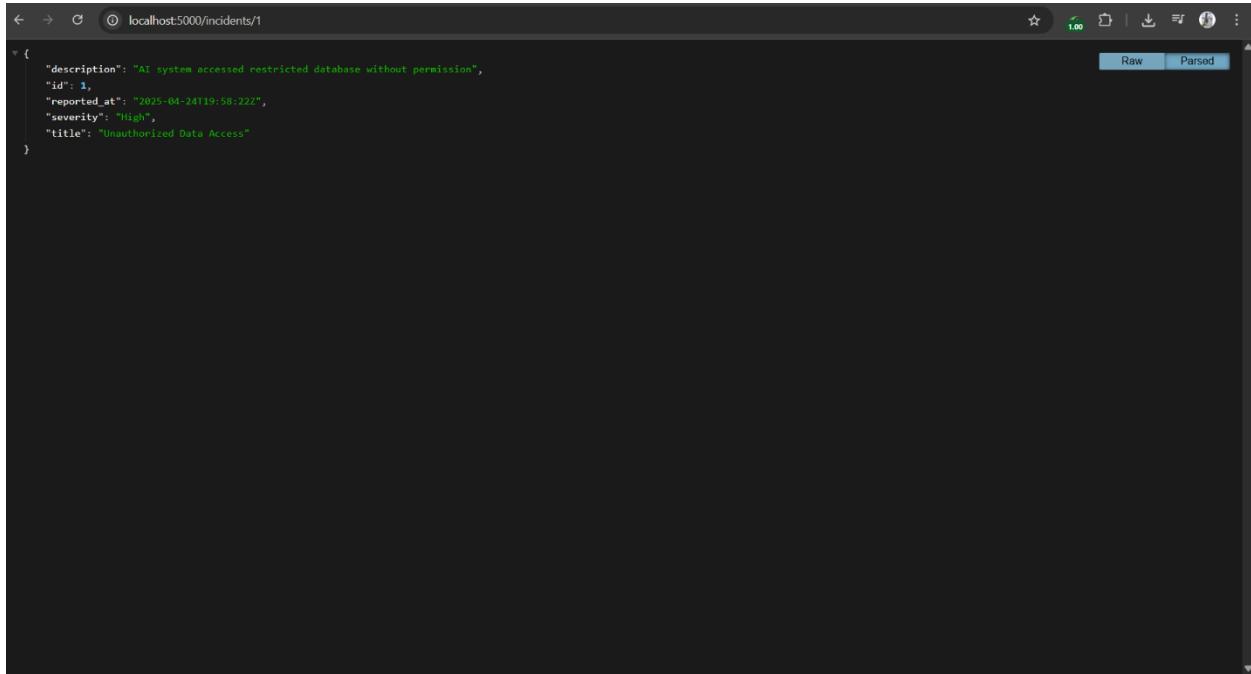
incidents_api_response



A screenshot of a web browser displaying the JSON response for all incidents. The URL is `localhost:5000/incidents`. The page shows a list of six incidents, each represented by a JSON object. The objects have properties: `description`, `id`, `reported_at`, `severity`, and `title`. The `Raw` and `Parsed` buttons are visible at the top right.

```
[{"id": 9, "reported_at": "2025-04-25T13:25:34Z", "severity": "High", "title": "Server Crash", "description": "Main server went down unexpectedly."}, {"id": 8, "reported_at": "2025-04-25T18:51:03Z", "severity": "High", "title": "AI Misjudgment in Medical Diagnosis", "description": "AI gave wrong diagnostic output for a patient case."}, {"id": 7, "reported_at": "2025-04-24T17:29:45Z", "severity": "High", "title": "Model Hallucination Event", "description": "AI system generated false information during customer interaction."}, {"id": 3, "reported_at": "2025-04-24T19:58:22Z", "severity": "Low", "title": "UI Confusion", "description": "Users misinterpreted AI recommendations due to unclear interface."}, {"id": 2, "reported_at": "2025-04-24T19:58:22Z", "severity": "Medium", "title": "False Prediction", "description": "ML model produced false positives in critical medical diagnoses."}, {"id": 1, "reported_at": "2025-04-24T19:58:22Z", "severity": "High", "title": "Unauthorized Data Access", "description": "AI system accessed restricted database without permission."}]
```

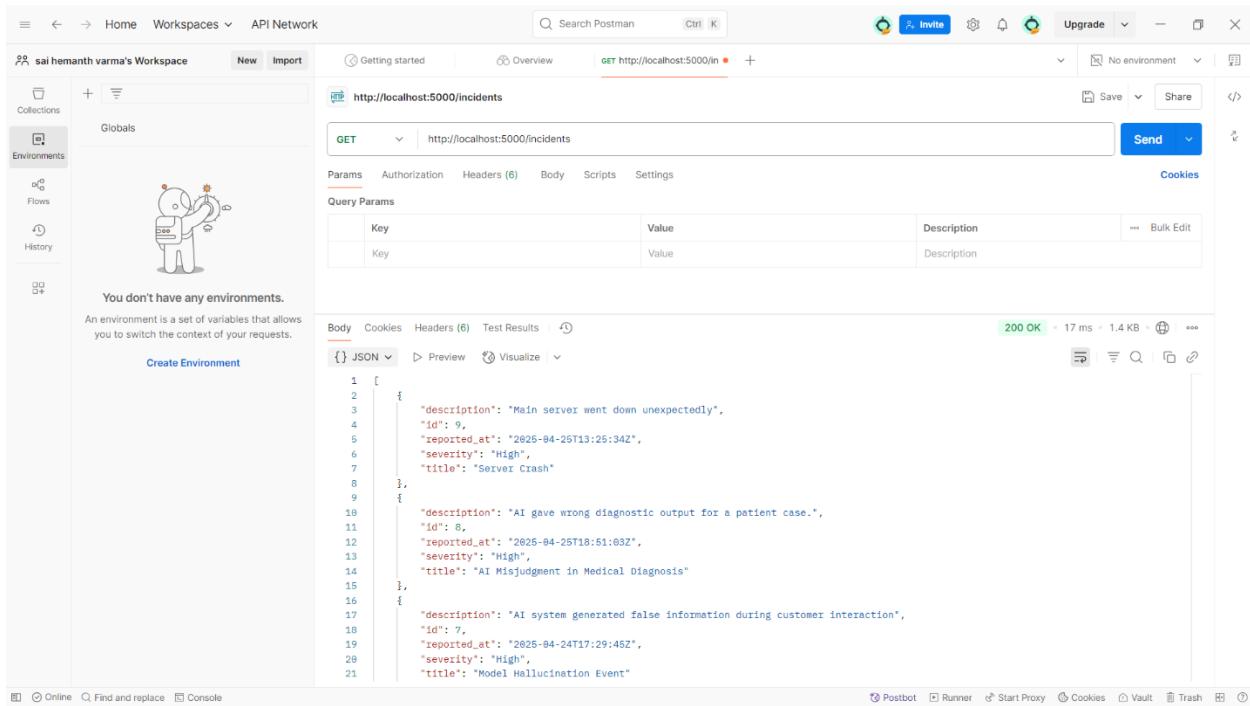
single_incident_api_response



A screenshot of a web browser displaying the JSON response for a single incident. The URL is `localhost:5000/incidents/1`. The page shows one incident, which is the third item from the list in the previous screenshot. The `Raw` and `Parsed` buttons are visible at the top right.

```
{"id": 1, "reported_at": "2025-04-24T19:58:22Z", "severity": "High", "title": "Unauthorized Data Access", "description": "AI system accessed restricted database without permission."}]
```

post_request_image

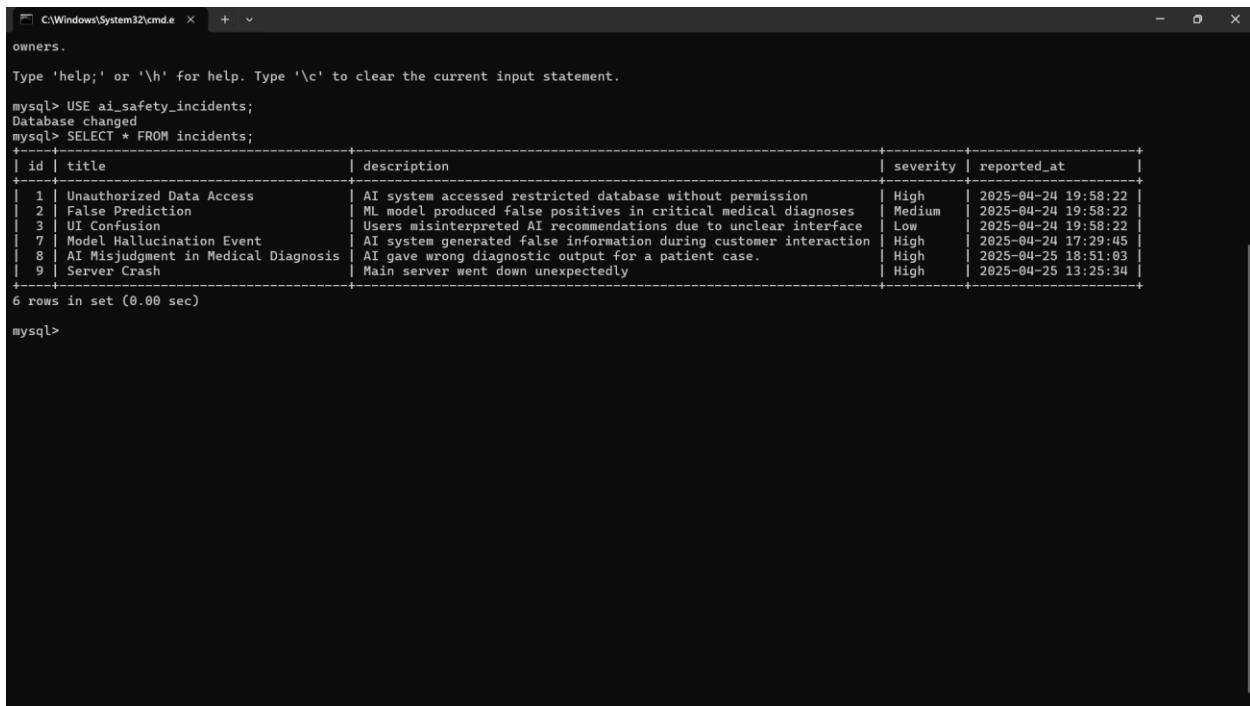


The screenshot shows the Postman application interface. On the left, there's a sidebar with 'Collections' (selected), 'Flows', and 'History'. The main area has tabs for 'Getting started', 'Overview', and a selected 'GET http://localhost:5000/incidents'. Below this, there are sections for 'Params', 'Authorization', 'Headers (6)', 'Body', 'Scripts', and 'Settings'. Under 'Query Params', there's a table with columns 'Key', 'Value', and 'Description'. The 'Body' tab is selected, showing a JSON response with 21 numbered lines. The response is as follows:

```
1 [
2   {
3     "description": "Main server went down unexpectedly",
4     "id": 9,
5     "reported_at": "2025-04-25T13:26:34Z",
6     "severity": "High",
7     "title": "Server Crash"
8   },
9   {
10     "description": "AI gave wrong diagnostic output for a patient case.",
11     "id": 8,
12     "reported_at": "2025-04-25T18:51:03Z",
13     "severity": "High",
14     "title": "AI Misjudgment in Medical Diagnosis"
15   },
16   {
17     "description": "AI system generated false information during customer interaction",
18     "id": 7,
19     "reported_at": "2025-04-24T17:29:45Z",
20     "severity": "High",
21     "title": "Model Hallucination Event"
22 }
```

The status bar at the bottom shows '200 OK' with a green icon, '17 ms', '1.4 KB', and a timestamp of '4:04'. Other icons include 'Postbot', 'Runner', 'Start Proxy', 'Cookies', 'Vault', 'Trash', and a refresh symbol.

database_schema



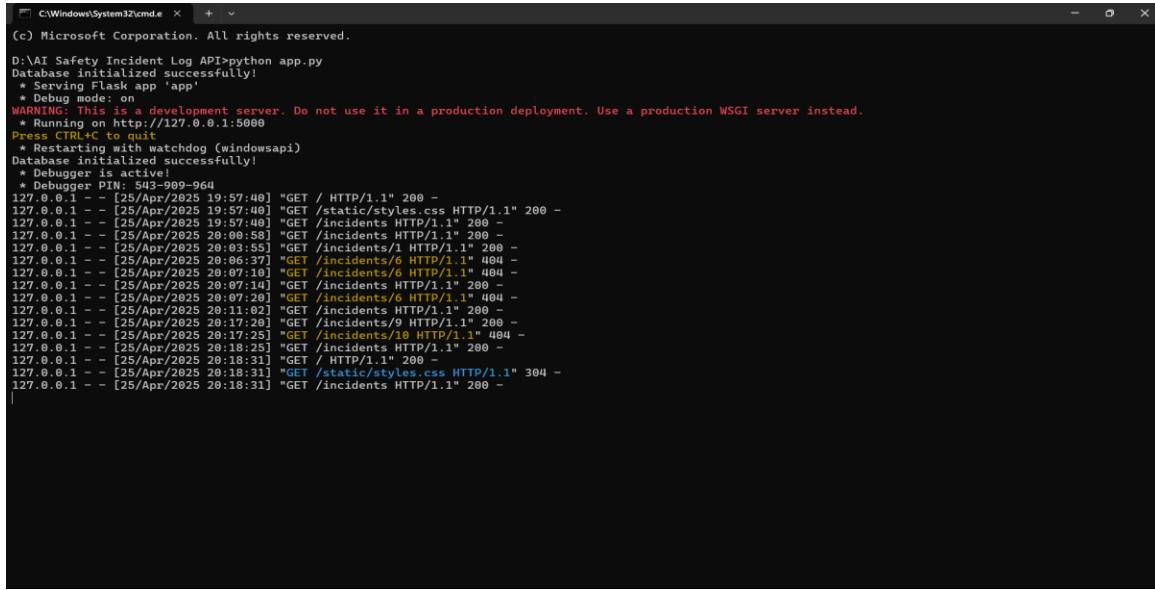
```
C:\Windows\System32\cmd.e + v
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> USE ai_safety_incidents;
Database changed
mysql> SELECT * FROM incidents;
+----+-----+-----+-----+-----+
| id | title          | description                | severity | reported_at    |
+----+-----+-----+-----+-----+
| 1  | Unauthorized Data Access | AI system accessed restricted database without permission | High    | 2025-04-24 19:58:22 |
| 2  | False Prediction   | ML model produced false positives in critical medical diagnoses | Medium  | 2025-04-24 19:58:22 |
| 3  | UI Confusion       | Users misinterpreted AI recommendations due to unclear interface | Low     | 2025-04-24 19:58:22 |
| 7  | Model Hallucination Event | AI system generated false information during customer interaction | High    | 2025-04-24 17:29:45 |
| 8  | AI Misjudgment in Medical Diagnosis | AI gave wrong diagnostic output for a patient case. | High    | 2025-04-25 18:51:03 |
| 9  | Server Crash       | Main server went down unexpectedly | High    | 2025-04-25 13:25:34 |
+----+-----+-----+-----+-----+
6 rows in set (0.00 sec)

mysql>
```

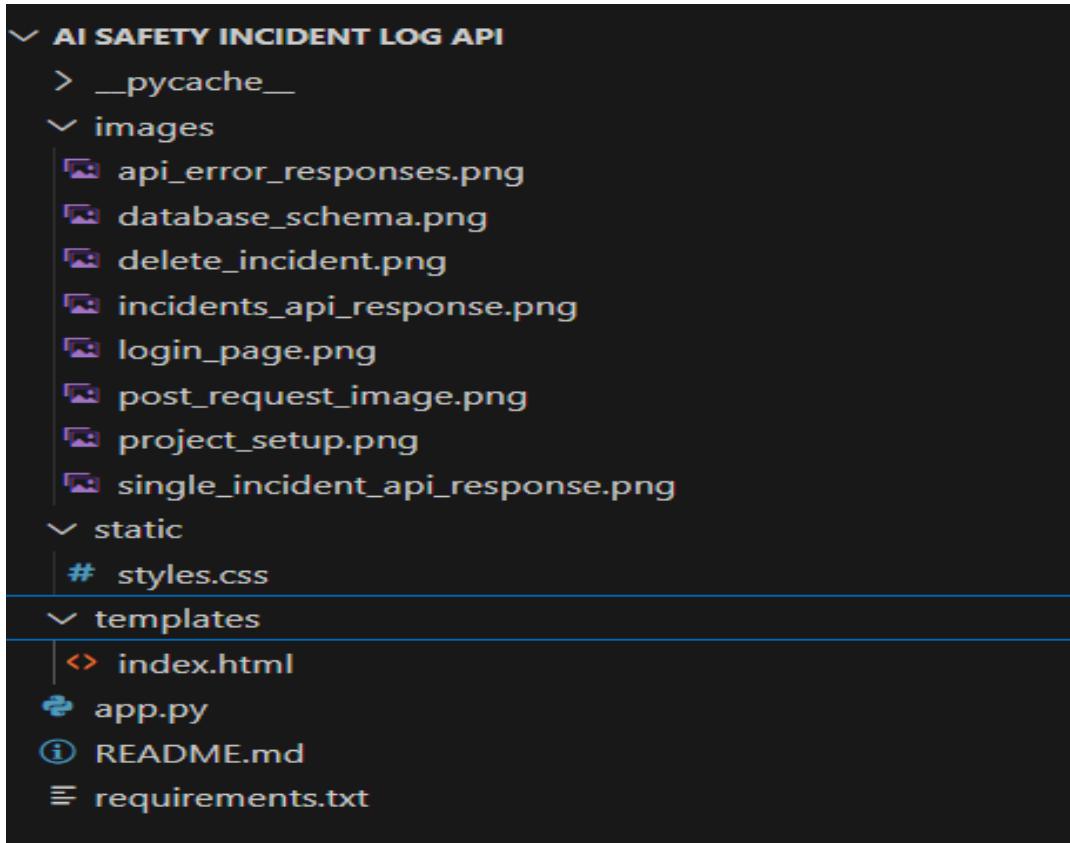
project setup



```
C:\Windows\System32\cmd.exe x + v
(c) Microsoft Corporation. All rights reserved.

D:\AI Safety Incident Log API>python app.py
Database initialized successfully!
* Serving Flask app 'app'
* Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
* Restarting with watchdog (windowsapi)
Database initialized successfully!
* Debugger is active
* Debugger PIN: 123-989-964
127.0.0.1 - - [25/Apr/2025 19:57:48] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [25/Apr/2025 19:57:48] "GET /static/styles.css HTTP/1.1" 200 -
127.0.0.1 - - [25/Apr/2025 19:57:48] "GET /incidents HTTP/1.1" 200 -
127.0.0.1 - - [25/Apr/2025 20:00:58] "GET /incidents HTTP/1.1" 200 -
127.0.0.1 - - [25/Apr/2025 20:03:55] "GET /incidents/1 HTTP/1.1" 200 -
127.0.0.1 - - [25/Apr/2025 20:06:37] "GET /incidents/6 HTTP/1.1" 404 -
127.0.0.1 - - [25/Apr/2025 20:07:10] "GET /incidents/6 HTTP/1.1" 404 -
127.0.0.1 - - [25/Apr/2025 20:07:13] "GET /incidents/10 HTTP/1.1" 200 -
127.0.0.1 - - [25/Apr/2025 20:07:13] "GET /incidents/6 HTTP/1.1" 200 -
127.0.0.1 - - [25/Apr/2025 20:11:02] "GET /incidents/9 HTTP/1.1" 200 -
127.0.0.1 - - [25/Apr/2025 20:11:02] "GET /incidents/9 HTTP/1.1" 200 -
127.0.0.1 - - [25/Apr/2025 20:17:25] "GET /incidents/10 HTTP/1.1" 404 -
127.0.0.1 - - [25/Apr/2025 20:17:25] "GET /incidents HTTP/1.1" 200 -
127.0.0.1 - - [25/Apr/2025 20:18:31] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [25/Apr/2025 20:18:31] "GET /static/styles.css HTTP/1.1" 304 -
127.0.0.1 - - [25/Apr/2025 20:18:31] "GET /incidents HTTP/1.1" 200 -
```

project structure



Thank You Note

I would like to express my sincere gratitude to SparkleHood for providing me with this opportunity to demonstrate my skills through this project assignment. The challenge of creating an AI Safety Incident Log API has been both educational and engaging, allowing me to apply my knowledge of backend development in a practical context.

I appreciate the time and consideration given to my application. This project has further reinforced my interest in contributing to SparkleHood's mission and becoming part of your team.

Thank you for reviewing my submission. I look forward to the possibility of discussing it further and learning more about how I could contribute to SparkleHood's innovative work in the future.