

AI Engineer Test

Candidate: Ngoc-Tram Nguyen

Source code: <https://github.com/ntrami/HERMES-LOGISTICS-AI-CHATBOT.git>

Video demo: <https://www.youtube.com/watch?v=dNkY6ezW46o>

1. Context:

Hermes is a lightweight AI Logistics Assistant designed to support operations managers in analyzing shipment data and responding to natural language queries. Built as a full-stack application, it features a FastAPI backend for efficient data processing and a user-friendly frontend powered by Bootstrap for styling and Plotly for interactive visualizations. The system incorporates multiple NLP strategies—including rule-based keyword matching, similarity-based techniques (using TF-IDF and cosine similarity), and a logistic regression classifier—to handle query understanding effectively. It utilizes mock shipment data with 5,000 entries to simulate real-world logistics scenarios, ensuring robust testing and demonstration.

2. Features:

Interactive Dashboard for Visualizations:



Chatbot for Logistics Queries

Chat

Which route had the most delays last week?

Route B has the highest average delay at 60.8 minutes.

Average delay minutes by route

Route	Average Delay (minutes)
Route A	58.5
Route B	60.8
Route C	57.2
Route D	56.8
Route E	54.5
Route F	53.8
Route G	53.2
Route H	51.5
Route I	49.5
Route J	30.5

Chat

Show total delayed shipments by delay reason

Total delay reasons breakdown: 3800 shipments across 3 categories. Top reason: Mechanical with 1309 occurrences (34.4%).

Delay reasons distribution

Delay Reason	Percentage
Mechanical	34.4%
Weather	33%
Traffic	32%

Ask about delays, warehouses, predictions...

Chat

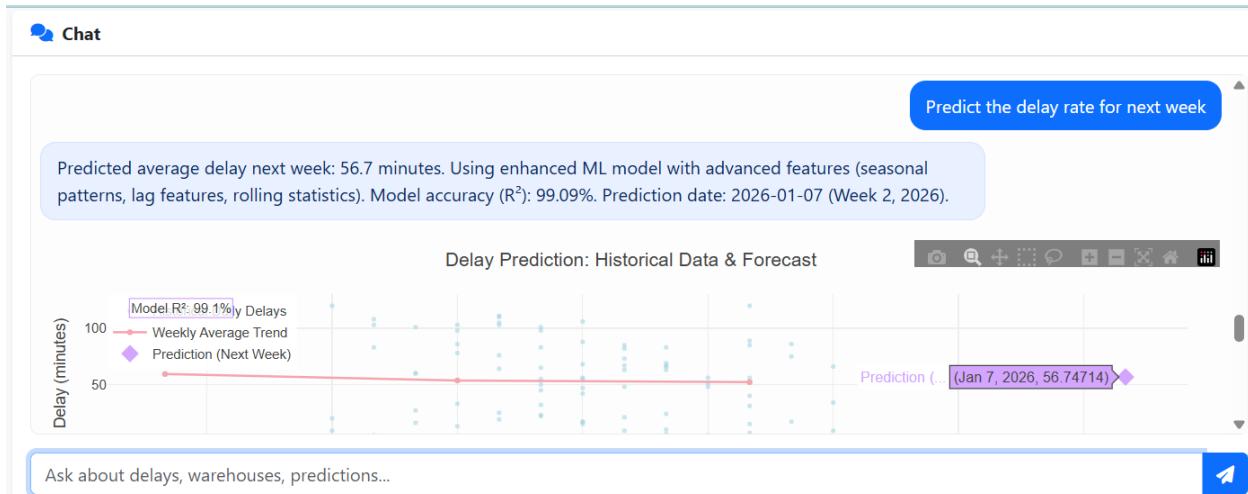
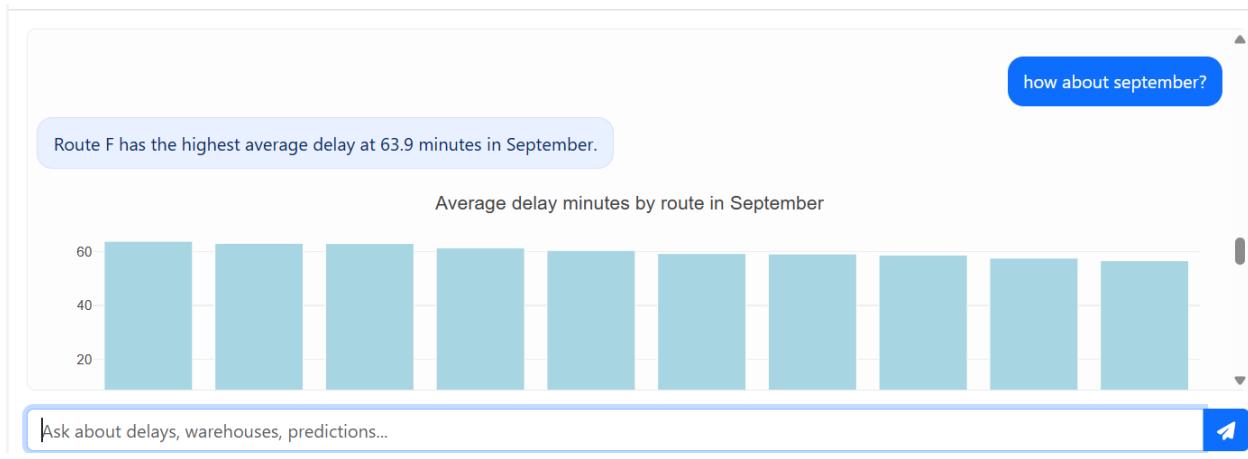
List warehouses with average delivery time above 5 days

Top 3 warehouses by fastest delivery: WH1, WH2, WH3.

Average delivery time by warehouse (days)

Warehouse	Average Delivery Time (days)
WH1	3.8
WH2	3.5
WH3	3.2

Ask about delays, warehouses, predictions...



3. Results:

Evaluating 3 methods on 50 queries...

- Method 1 (Rule-based): F1 1.0, Avg Time 0.0276s
- Method 2 (Similarity): F1 1.0, Avg Time 0.0262s
- Method 3 (ML Classifier): F1 0.822, Avg Time 0.018s

```
● PS O:\Hermes> docker-compose exec backend python eval.py
time="2025-12-10T22:36:30+07:00" level=warning msg="O:\Hermes\docker-compose.yml: `version` is obsolete"
Evaluating 3 methods on 50 queries...
Method 1 (Rule-based): F1 1.0, Avg Time 0.0276s
Method 2 (Similarity): F1 1.0, Avg Time 0.0262s
Method 3 (ML Classifier): F1 0.822, Avg Time 0.018s
```