



MediCar – How vehicles can navigate around using LLMs

Agenda

Milestone 1 Remote

Milestone 2 Onsite Cluj

Milestone 3 Onsite Mannheim

Overview Project Result

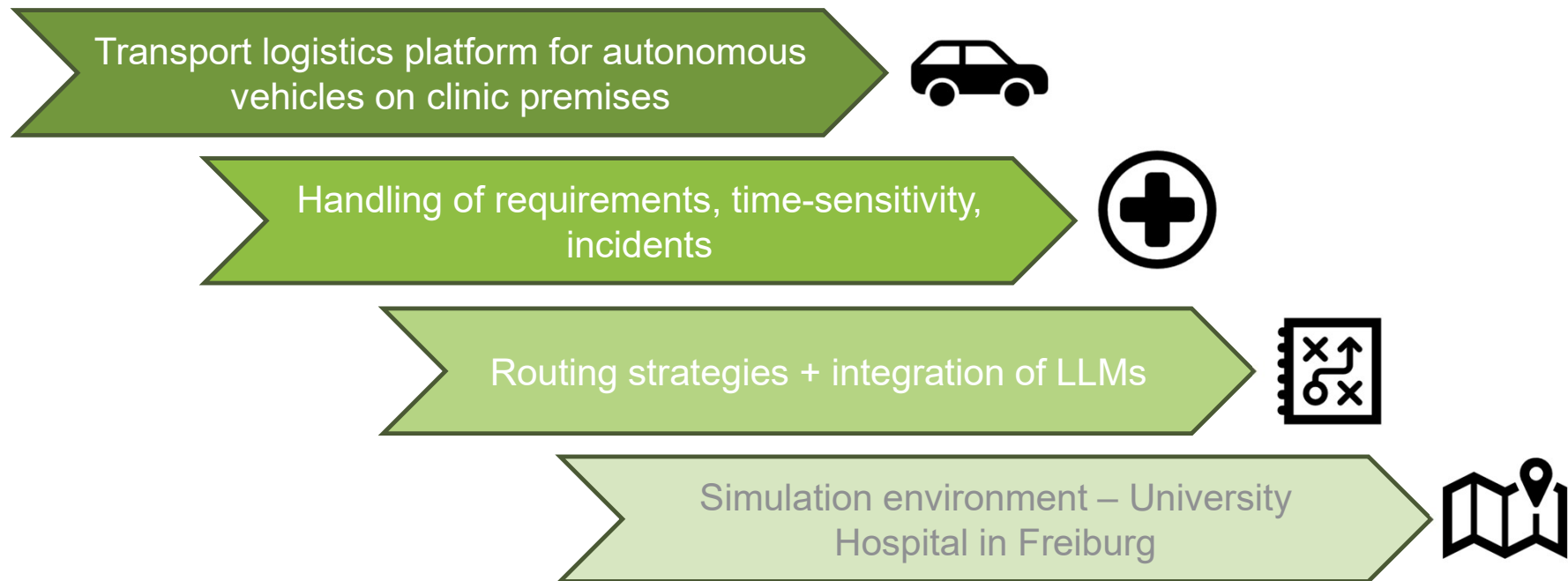
Next Steps

Research Goal 1: **Handle Incidents**

Can LLMs **simulate** the **impact** of **incidents** on routing graphs?

Research Goal 2: **Prioritize Goods**

Can LLMs be used for solving **routing** problems across
heterogeneous goods?



Milestones

Milestone 1

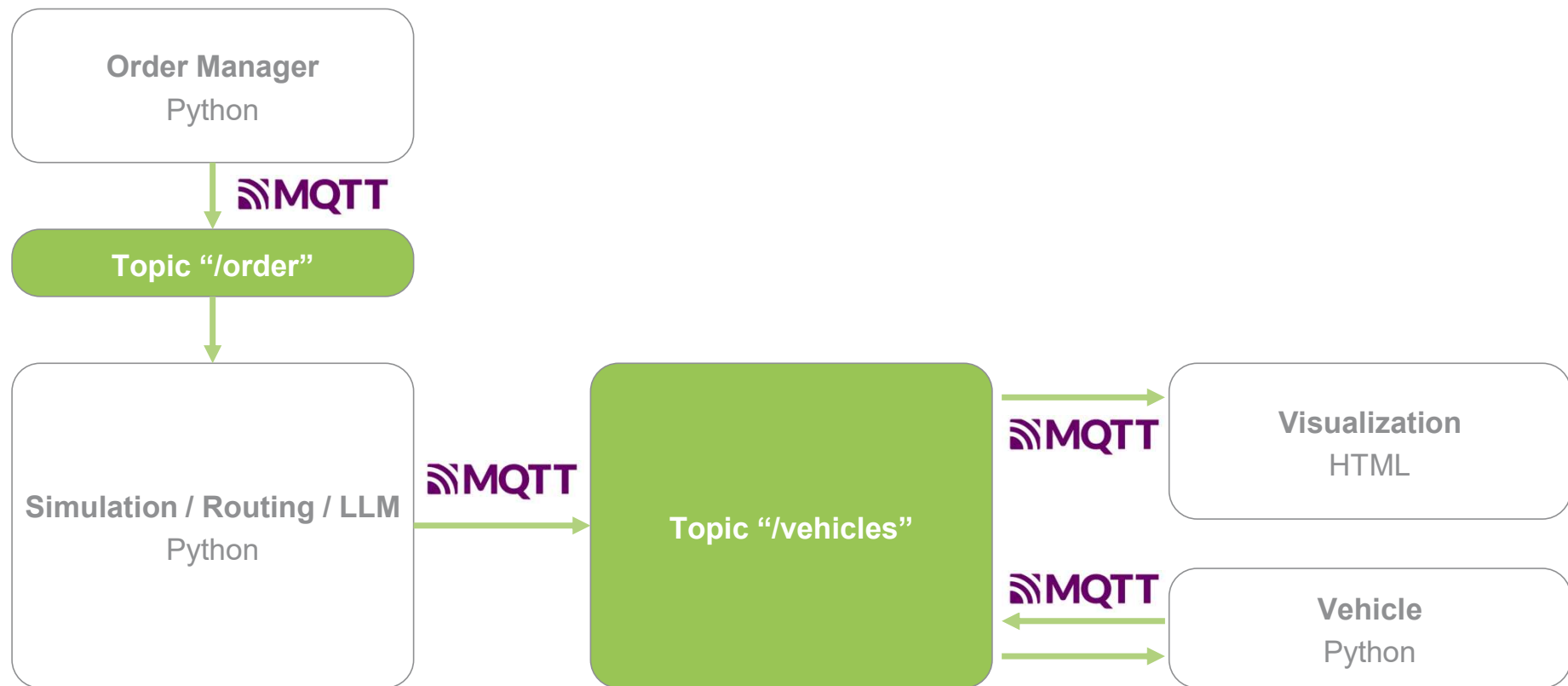
Software architecture



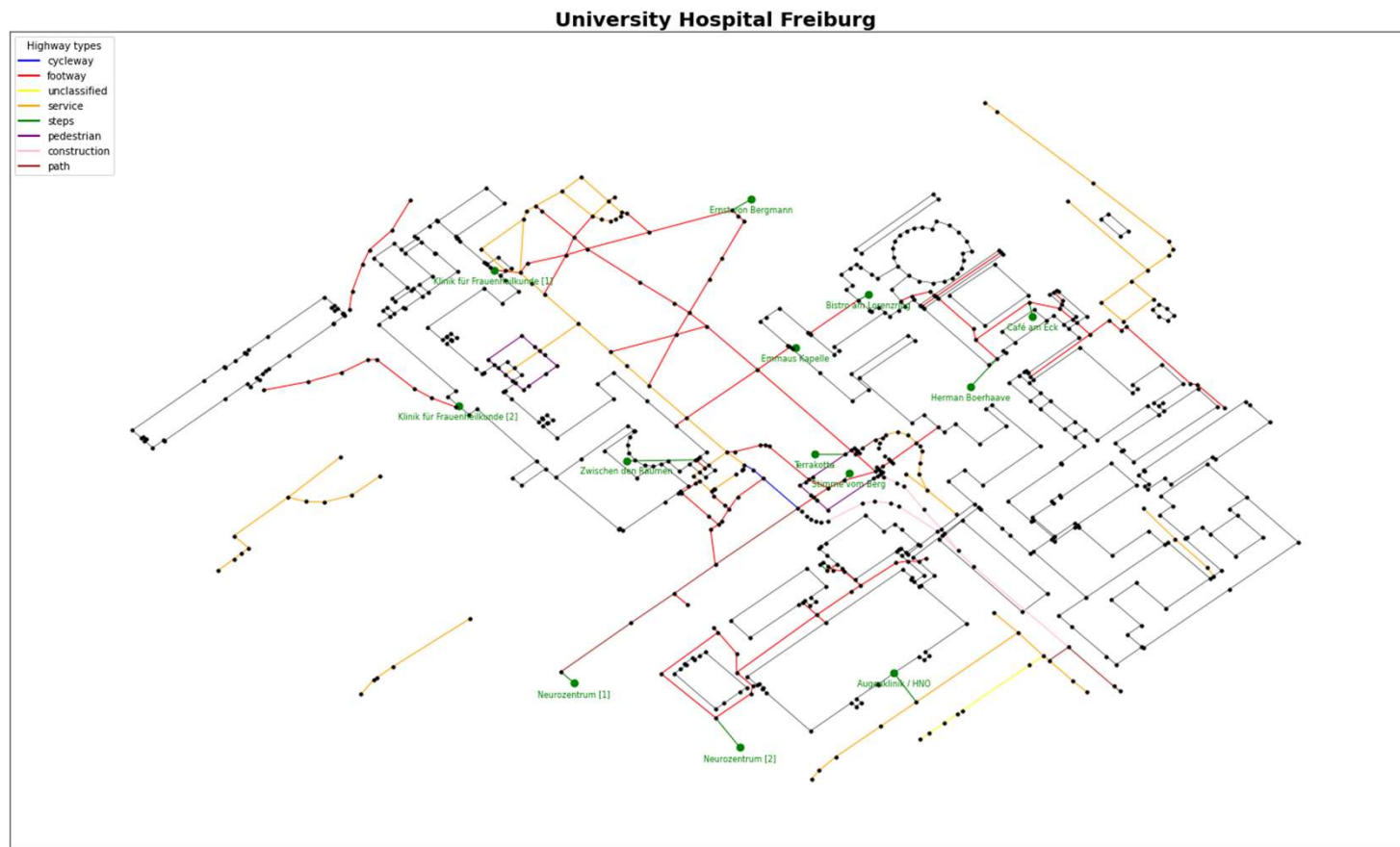
NetworkX Graph



Software Architecture



NetworkX Graph

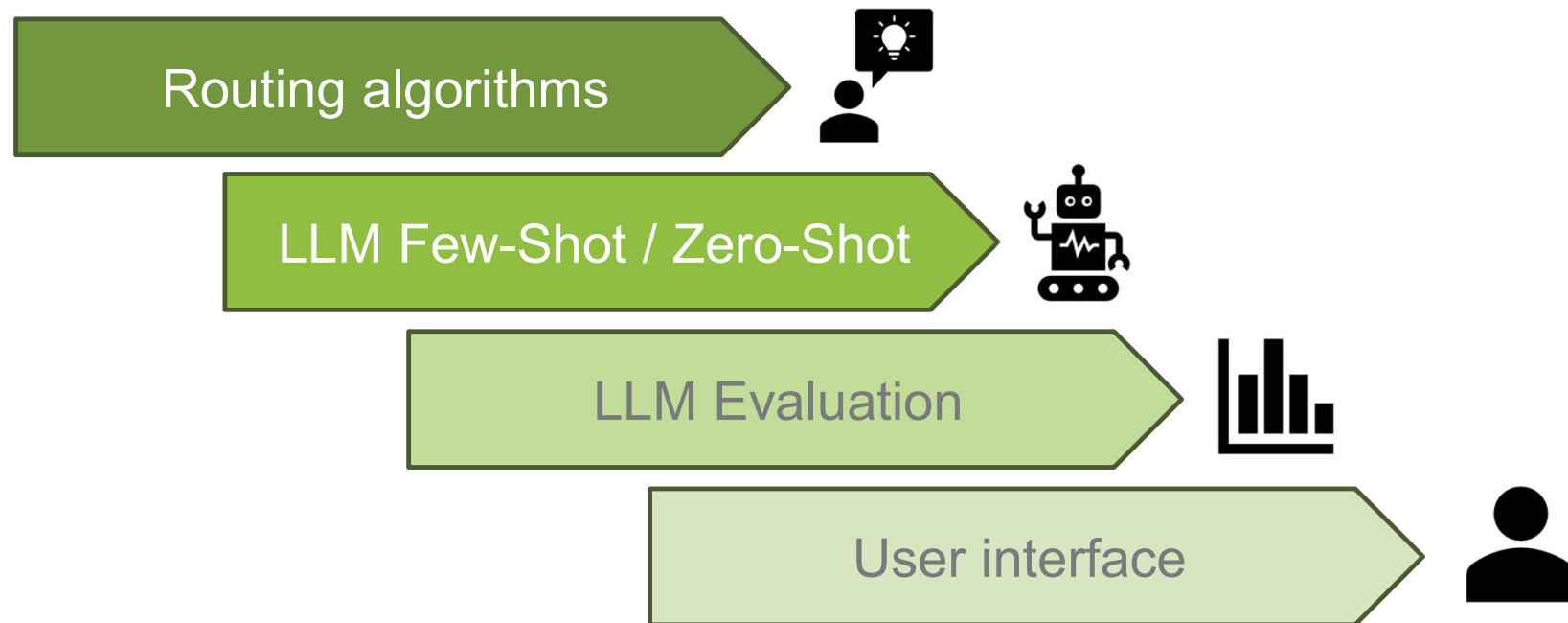


OpenStreetMap



NetworkX
Network Analysis in Python

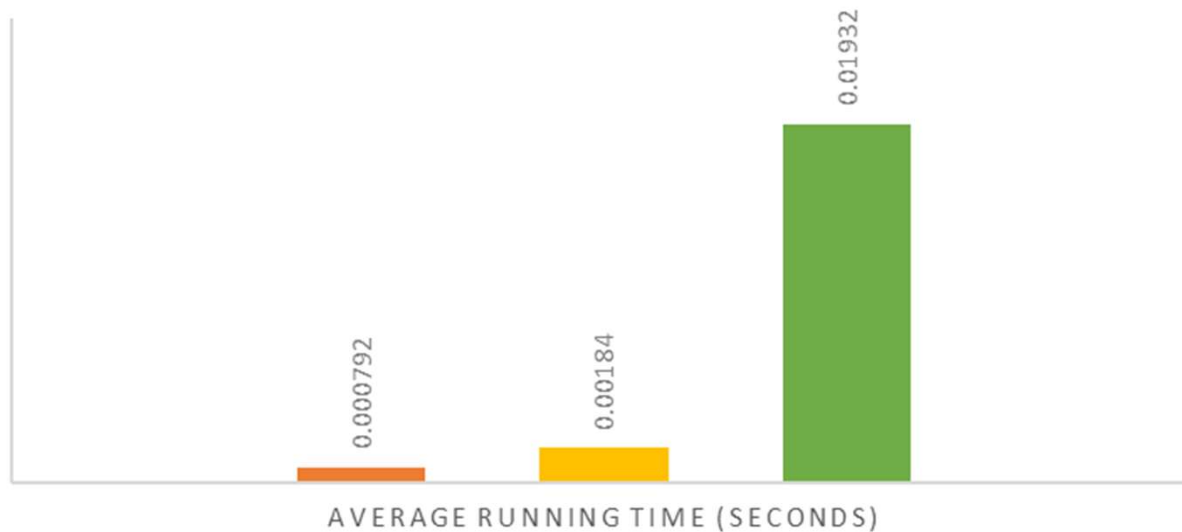
Milestone 2



Routing Algorithms

AVERAGE RUNNING TIME

■ A* ■ Dijkstra ■ Bellman-Ford



Ranking	Algorithm	Total Cost
1	A*	27,461.84
2	Dijkstra	27,461.84
3	Bellman-Ford	27,670.57

A* is the most efficient algorithm.

LLM – Few-Shot / Zero-Shot

Llama @ DWS

- Meta-model
- Open-source
- Lower latency on uni-server
- Improvement in performance

OpenAI Zero-Shot

- OpenAI model
- Relies on pre-existing knowledge and the given context
- Task-diverse
- A small number of tokens is used

OpenAI Few-Shot

- OpenAI model
- Relies on the given context and a few examples of expected output to certain inputs
- More fine-tuned for certain tasks
- Larger number of tokens is used

Evaluation metrics for LLM

Goal: Increase Accuracy!



Usable Edges

Dataset:

- 112 events on edges:
52 true, 60 false cases

Metrics:

- Accuracy
- Precision, Recall
- F1 score
- ROC - AUC



Dynamic Edge Weights

Dataset (length dependent):

- 112 events on edges:
43 true, 69 false cases

Metrics:

- Accuracy
- Precision, Recall
- F1 score
- ROC - AUC



Geographic Position

Dataset:

- 10 events on nodes

Metrics:

- LLM Answers
/ Explainable Results
- Accuracy
- Hallucination
Coincidences



Function Calling

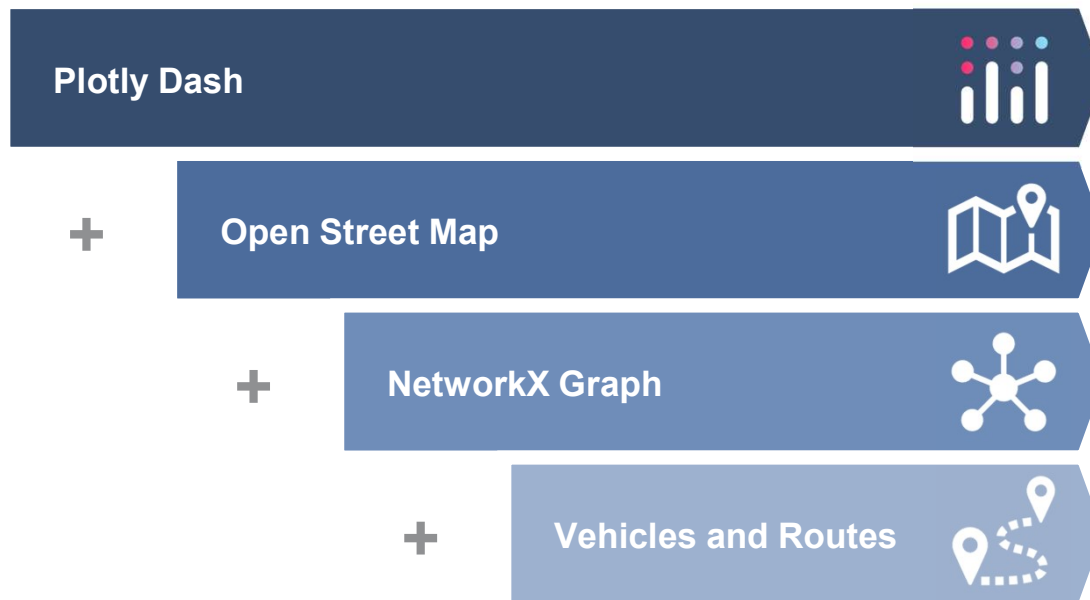
Dataset (1 node vs multiple nodes):

- 20 events on nodes:
10 true, 10 false cases

Metrics:

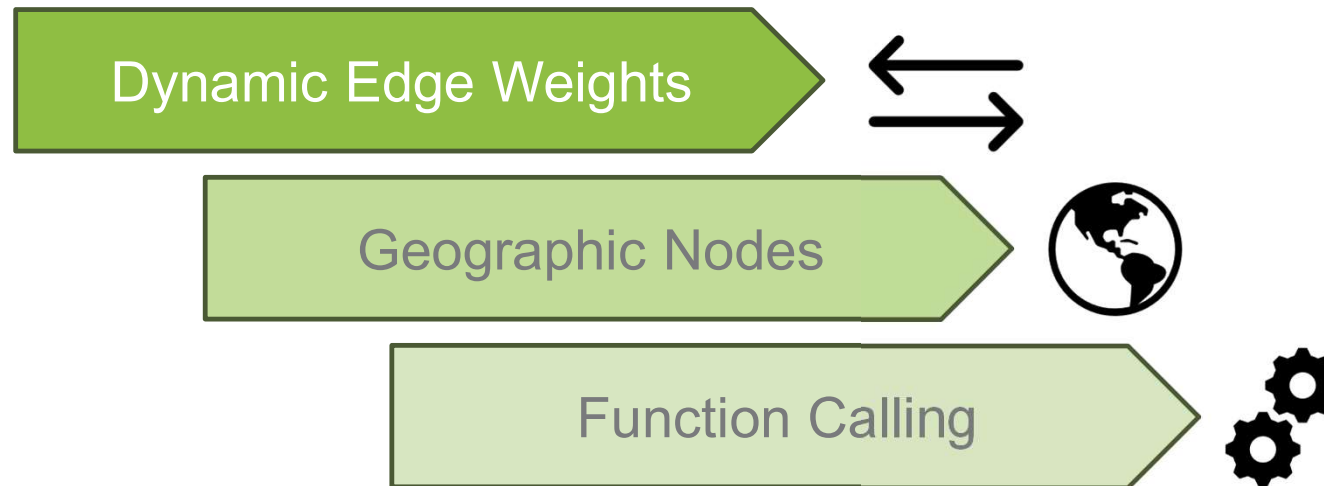
- Accuracy

Plotly Dash User Interface

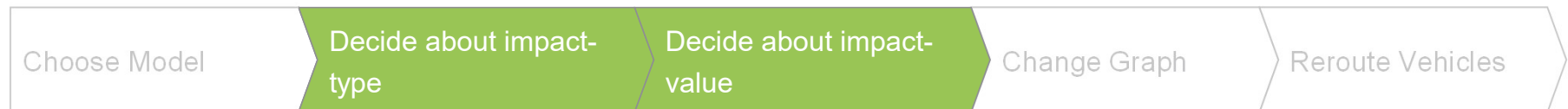


MQTT Event Channel

Milestone 3

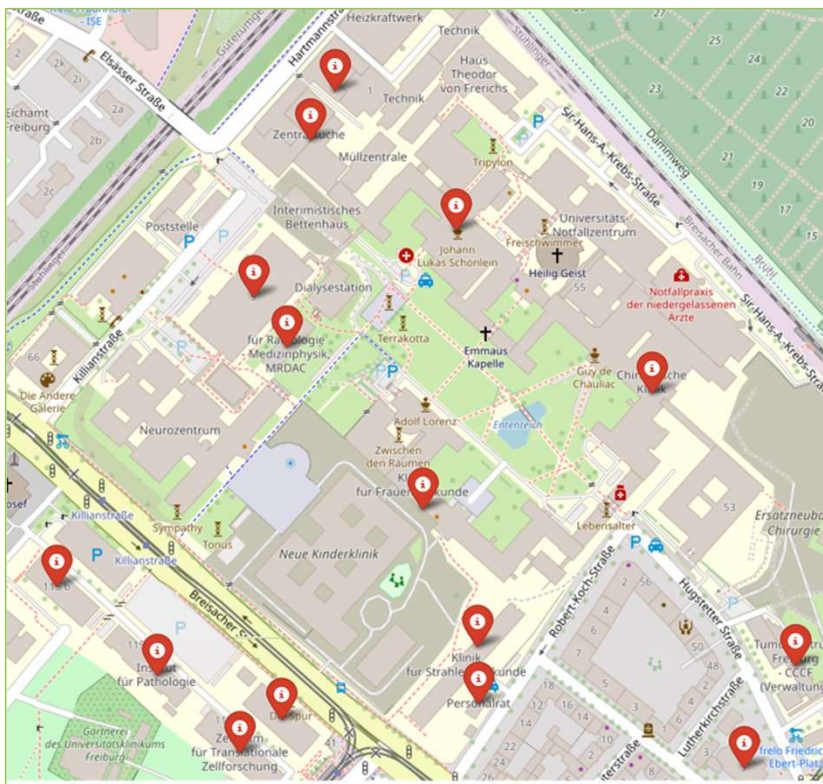


Dynamic Edge Weights



Component	User	LLM #1	LLM #2	NetworkX	A*
Action	Choose model that supports dynamic edges (e.g., GPT-3.5 Few-shot Dynamic).	Let LLM decide whether the incident impacts the entire length of the edge or a point in between.	Let another LLM decide about the severity of the impact. This is a single number on a scale.	Update weights from impacted edges in the graph. Publish changes.	Identify vehicles that need to be rerouted and broadcast a new updated route .

Geographic Node Representation



Action: "There is a radiation leak at *Klinik für Strahlenheilkunde*."
Context: List of nodes with (x,y)-coordinates

Output ▾

The radiation leak at Klinik für Strahlenheilkunde will directly affect the following nodes:

1. Klinik für Frauenheilkunde

Explanation: Close proximity to Klinik für Strahlenheilkunde, likely to experience increased traffic due to emergency response or evacuation procedures.

2. Dialysestation

Explanation: Close proximity to Klinik für Strahlenheilkunde, likely to experience increased traffic due to emergency response or evacuation procedures.

Geographic Node Representation

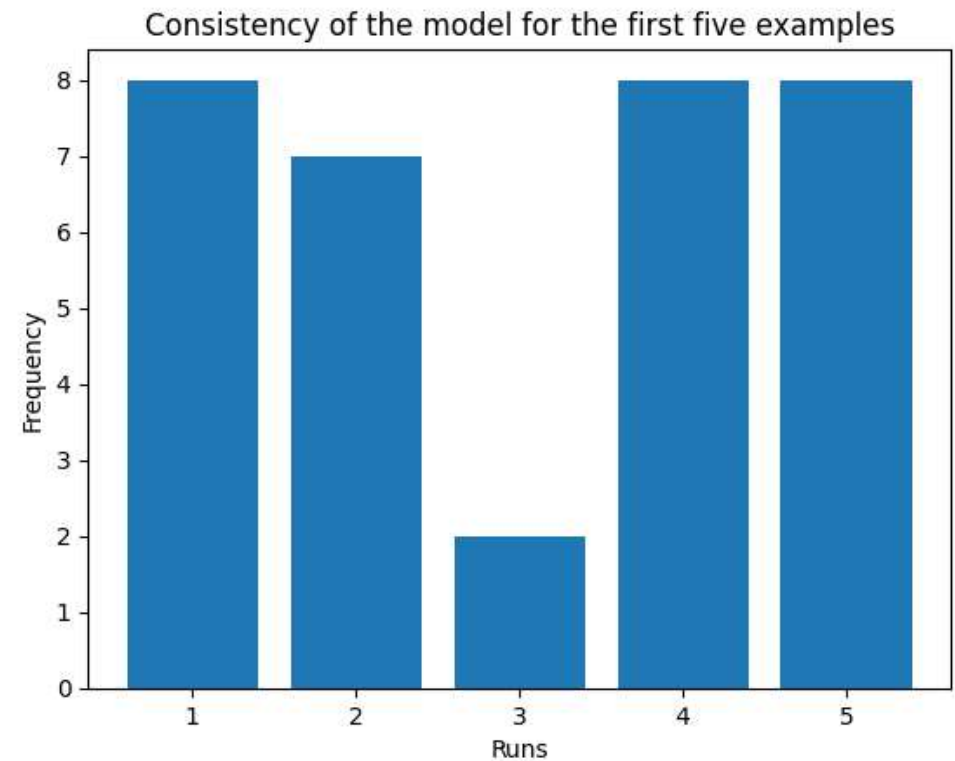
1) Named Nodes & Temp = 0

Metric	LLM / Ground Truth	Accuracy
Mean	2.64	0.4533
Min	0.6	0.2
Max	5	0.7333

2) Renamed Nodes & Temp = 0

Metric	LLM / Ground Truth	Accuracy
Mean	0.99	0.6667
Min	0.2	0.2
Max	2	0.8667

3) Hallucination Test



Function Calling

1 ➤ Instruct the LLM to determine if the event impacts multiple nodes or only the given one.

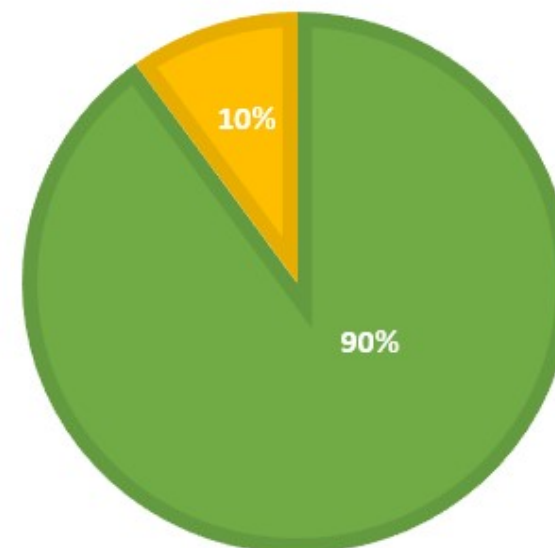
2 ➤ Define function to assess the impact.

3 ➤ `function_call = auto`

4 ➤ Evaluate the answers.

RESULT

■ LLM Answer = Ground Truth ■ LLM Answer != Ground Truth



Project Results

Dash
127.0.0.1:8050

Intelligent Hospital Logistics

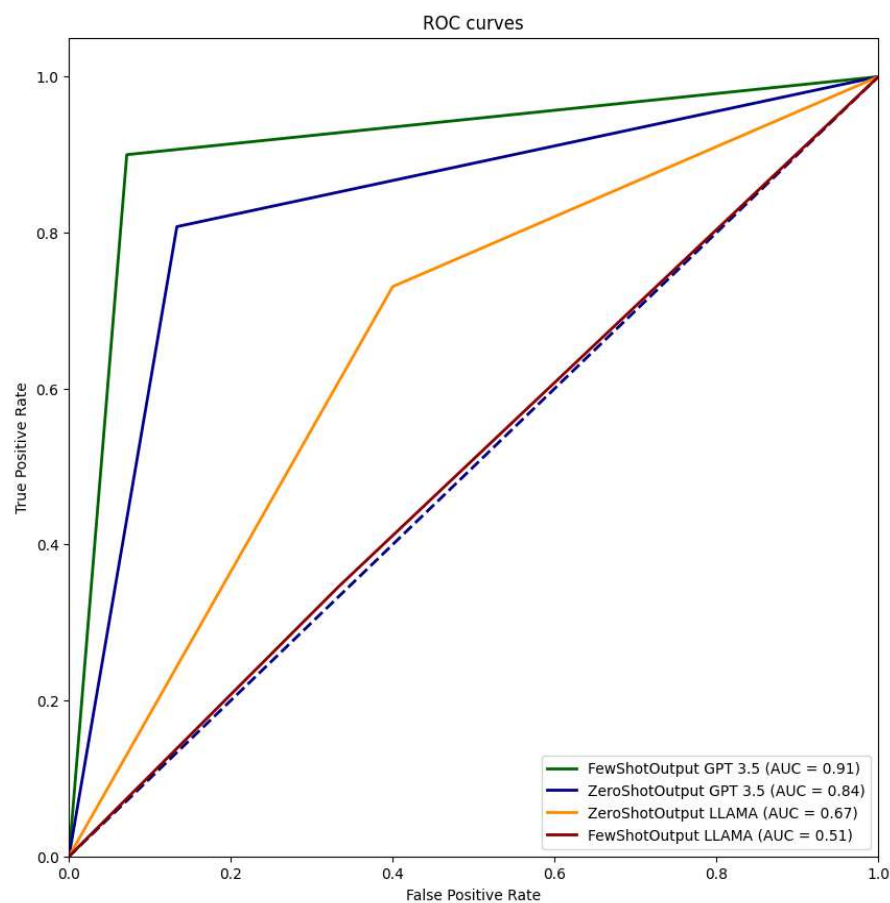
Incidents
Status

LLM
Prompt...
GPT-3.5-Few
Submit

Incident Generation
Seed used by all vehicles: 42
Update Seed
Randomize Seed
Generate Incidents from vehicles:
☐ off
☒ on

Ongoing Incidents

Can Large Language Models **simulate** the **impact** of **incidents** on routing graphs?



1) Usable Edges

Ranking	Model	Accuracy
1	GPT 3.5 Few Shot	0.9151
2	GPT 3.5 Zero Shot	0.8393
3	Llama 2 Zero Shot	0.6607
4	Llama 2 Few Shot	0.5179

2) Dynamic Edge Weights

Ranking	Model	Accuracy
1	GPT 3.5 Few Shot	0.8428

3) Geographic Position

Ranking	Model	Accuracy
1	GPT 3.5 Zero Shot	0.6667

4) Function Calling

Ranking	Model	Accuracy
1	GPT 3.5 Zero Shot	0.9000

NEXT STEPS

Our path towards the final presentation



Thank You For Your Attention!



And This International Team Project!

