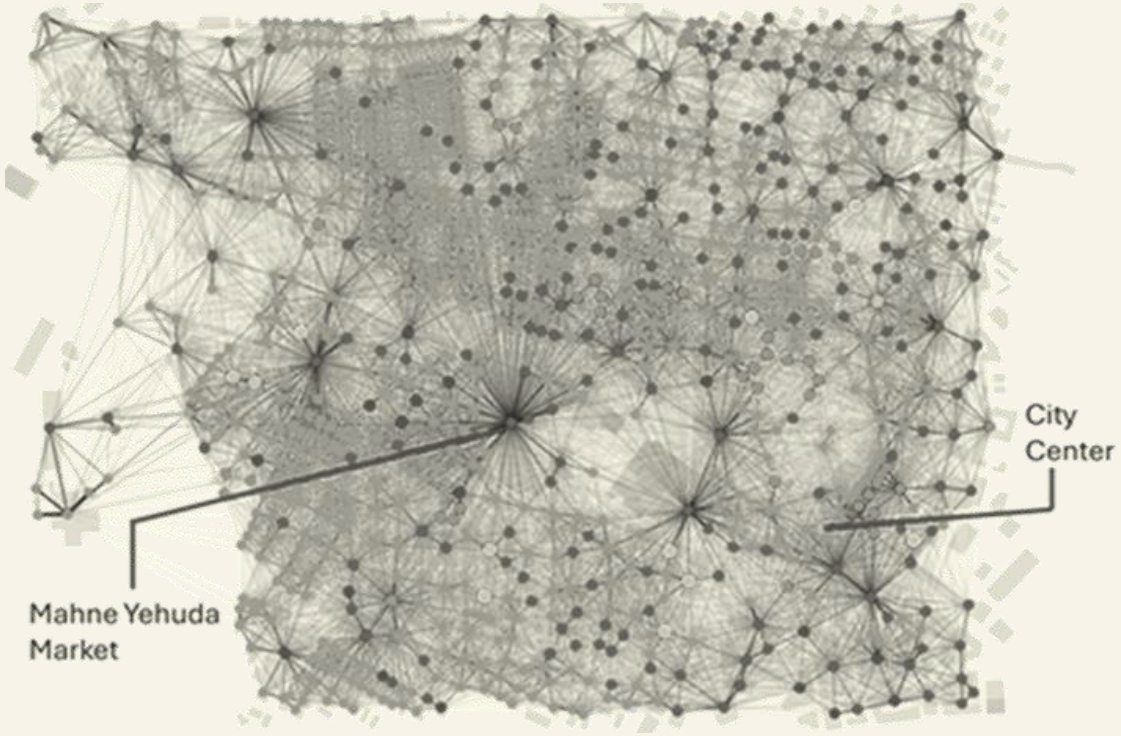
Trajectory Generation using Profile-Specific Transition Matrices

Zi Hen Lin · A. Yair Grinberger · Daniel Felsenstein

Technical University of Munich

THE HEBREW UNIVERSITY OF JERUSALEM



Generating synthetic mobility trajectories and transition matrices

Why synthetic data?

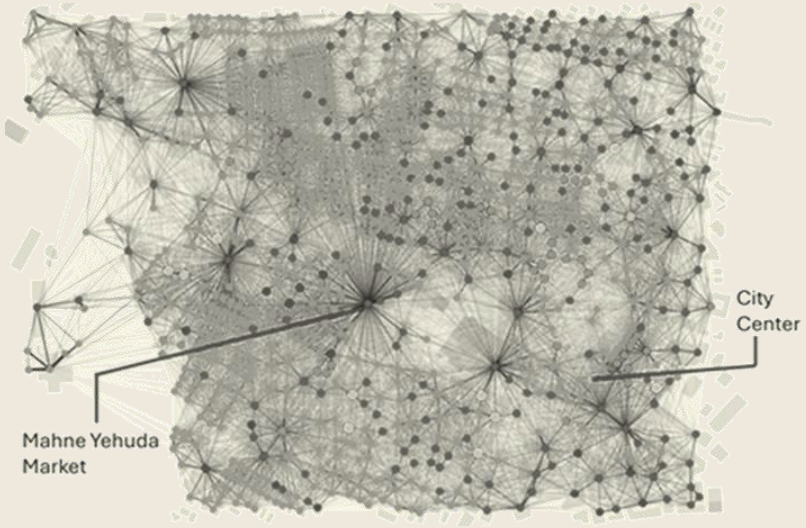
Observed mobility is costly, sparse, noisy, and privacy-sensitive.

Trajectory generation approaches

- **Interpretable structures**
- **Learning-based generators**
- **LLMs**

Here, we focus on transition matrices

Individual trajectories may look fine, but are the visiting the right places?

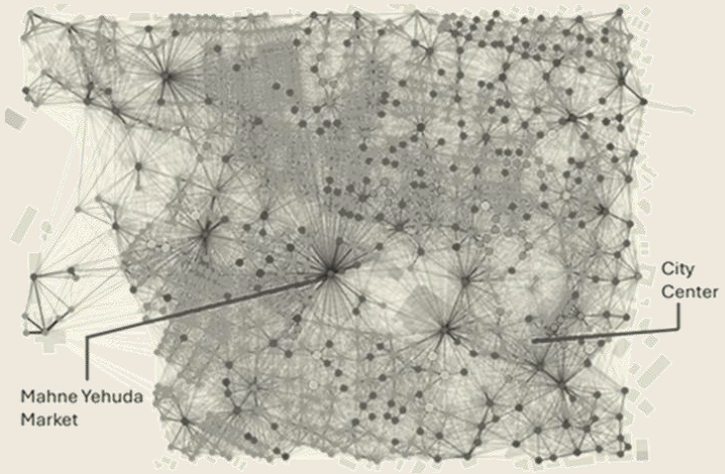


CONCEPT

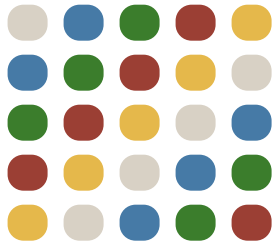
Transition matrix as a graph encoding spatial and behavior contexts; LLMs as a complex behavioral context

Spatial context

buildings · land use · floorspace



Transition matrix



Edges encode behavior; nodes encode urban space

Behavioral context

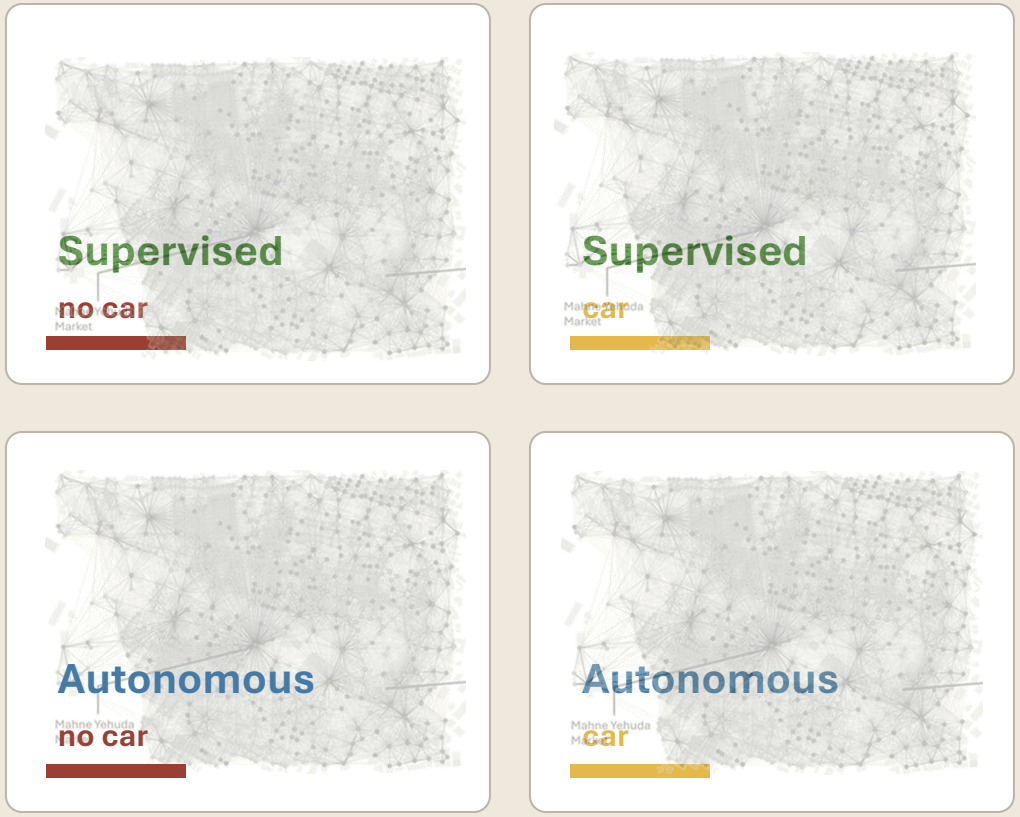
- socio-demographics
- movement history
- activity need
- **autonomy regime**

EXPERIMENT

Crossing autonomy regimes with mobility profiles

gravity-derived choice set

unconstrained



“You have no car”

“You have a car”

Prompt (GPT-5.4)

You are an employee aged {age}... Your home is {home_id}

You should generate a trajectory with **arbitrary length**

The trajectory should always **start from and return to home**... **no building will be visited twice** except for home.

For each iteration, output time, uniq_id, behavioral reasoning in json format.

1,014
downtown Jerusalem buildings

4,000
generated trajectories

Violations (1)

TEMPORAL VIOLATION

Stop 4: output

{'time': '07:35', 'uniq_id': 53369867, 'behavioral_reasoning':

After spending the night at home 52682478, I start a new day by walking to this nearby mixed-use building for a quick breakfast and coffee. Since I do not have a car, a very short errand close to home is practical before continuing elsewhere, and this stop would likely be brief rather than a long stay.}

Stop 3: reasoning

I need to figure out how to proceed because **the developer states that home is 52684489**, but it's not listed in the user's provided data. Since I can't choose a building not in the list, I'm stuck but it's not listed in the user's provided data. Since I can't choose a building not in the list, I'm stuck. **The current list includes 52682478, which could be a mistake for home.**

NUMERICAL CHOICE DRIFT

Stop 8: Output

{'time': '19:50', 'uniq_id': 52719107, …}

Stop 8: Reasoning

53317277 is about 145m southeast and categorized as mixed-use…

I could pick 53317277 for a quick dinner or coffee after running errands…

Stopping at a small mixed-use building would make sense... So, I think 52719107 is the best choice…

I am considering selecting 53317277, which is southeast of my current location... about 145 meters away.

Violations (2)

81.4%
segments include next-stop ID in reasoning

225
supervised trajectories violated transition dynamics

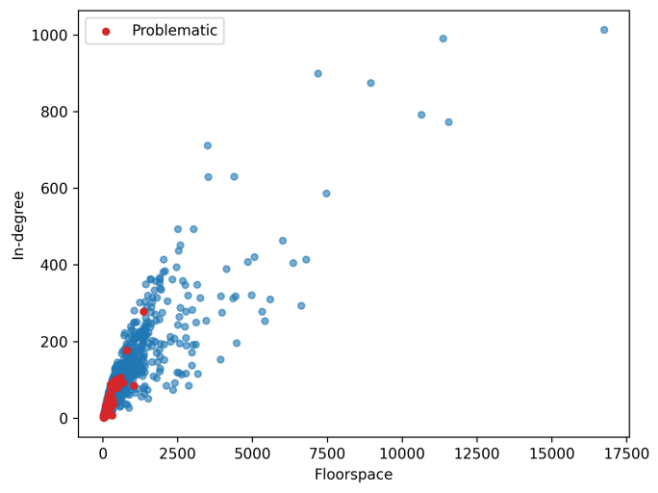
134
did not reach home by iteration 11

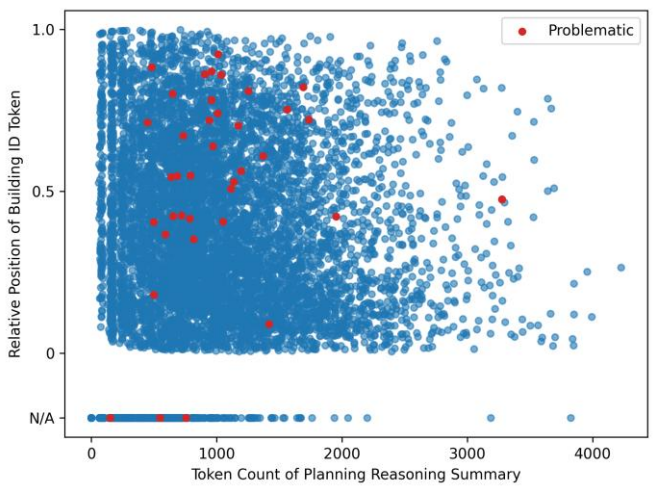
91
output home while ignoring adjacency

Two causes

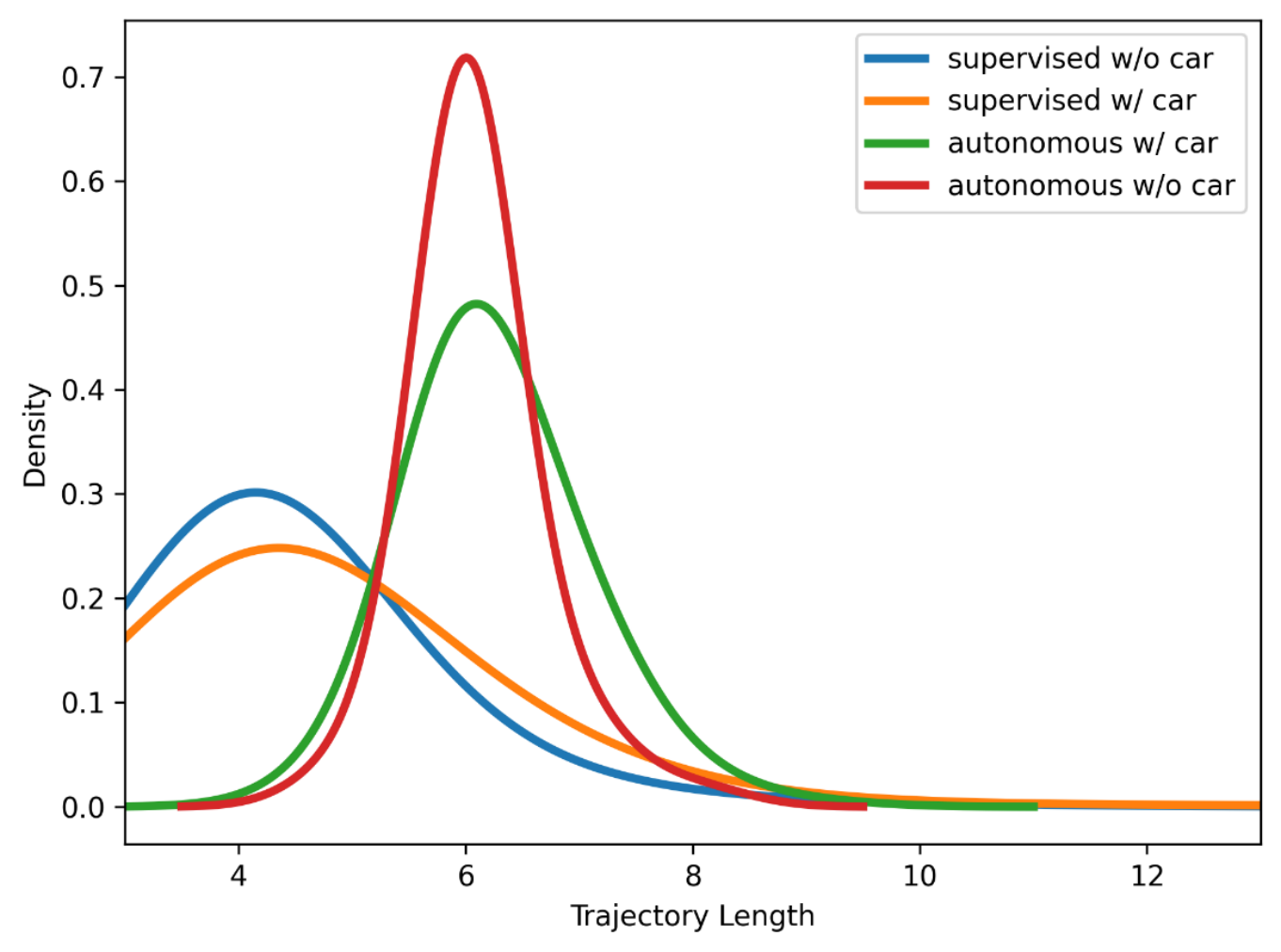
Conflating previous building lists with the current adjacency list.

Reasoning about one numeric candidate but outputting another building ID.

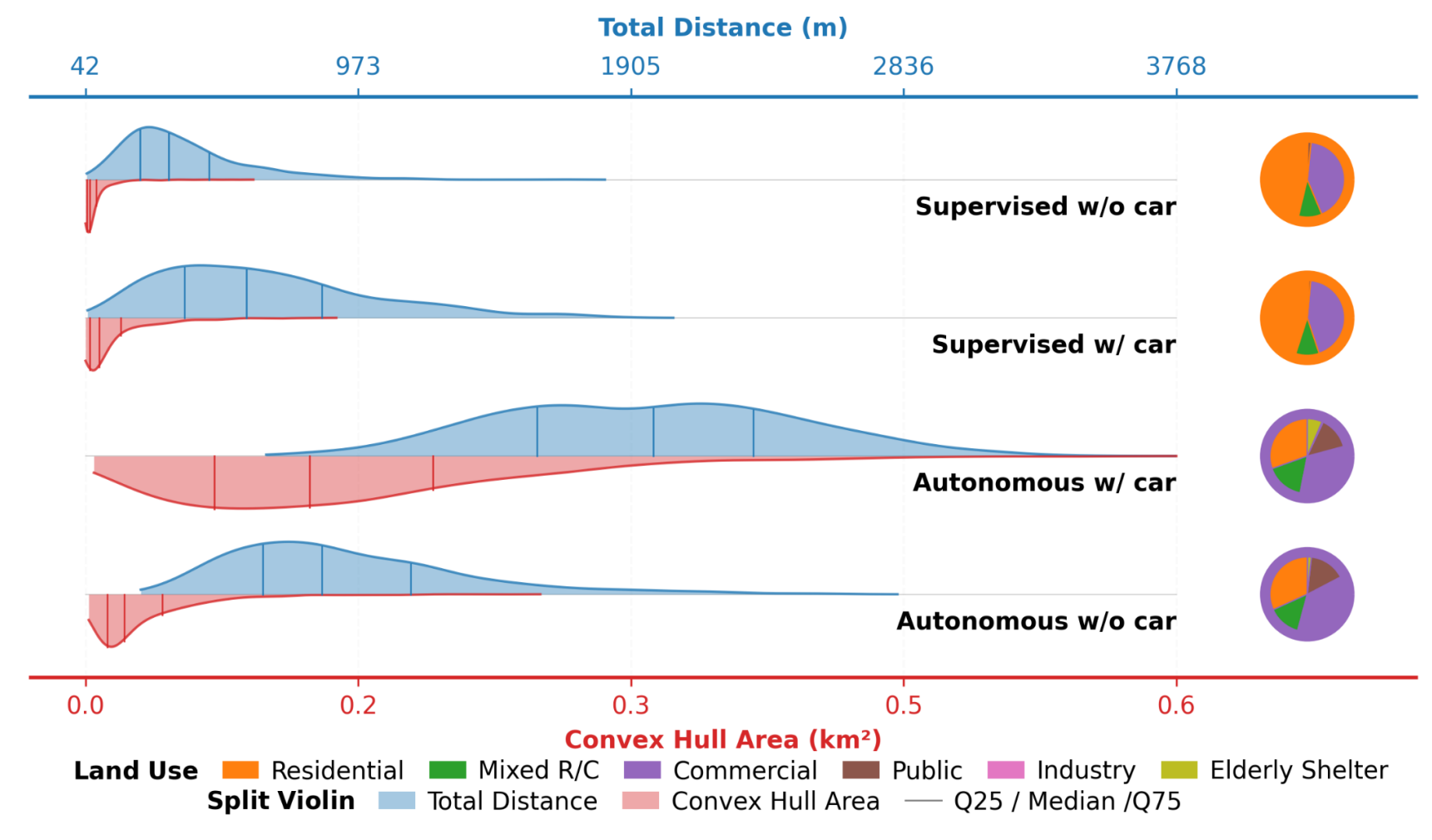




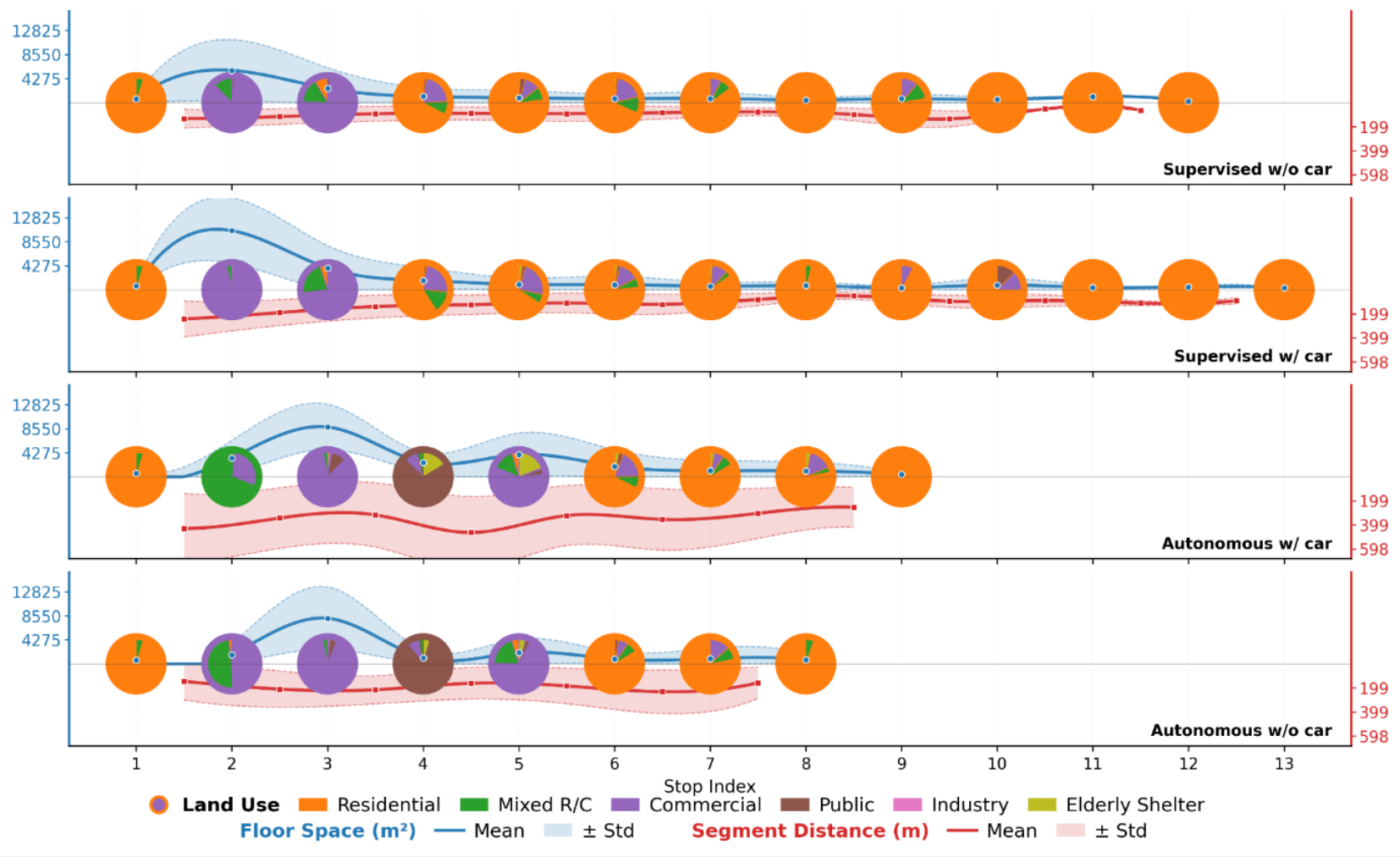
Trajectory lengths



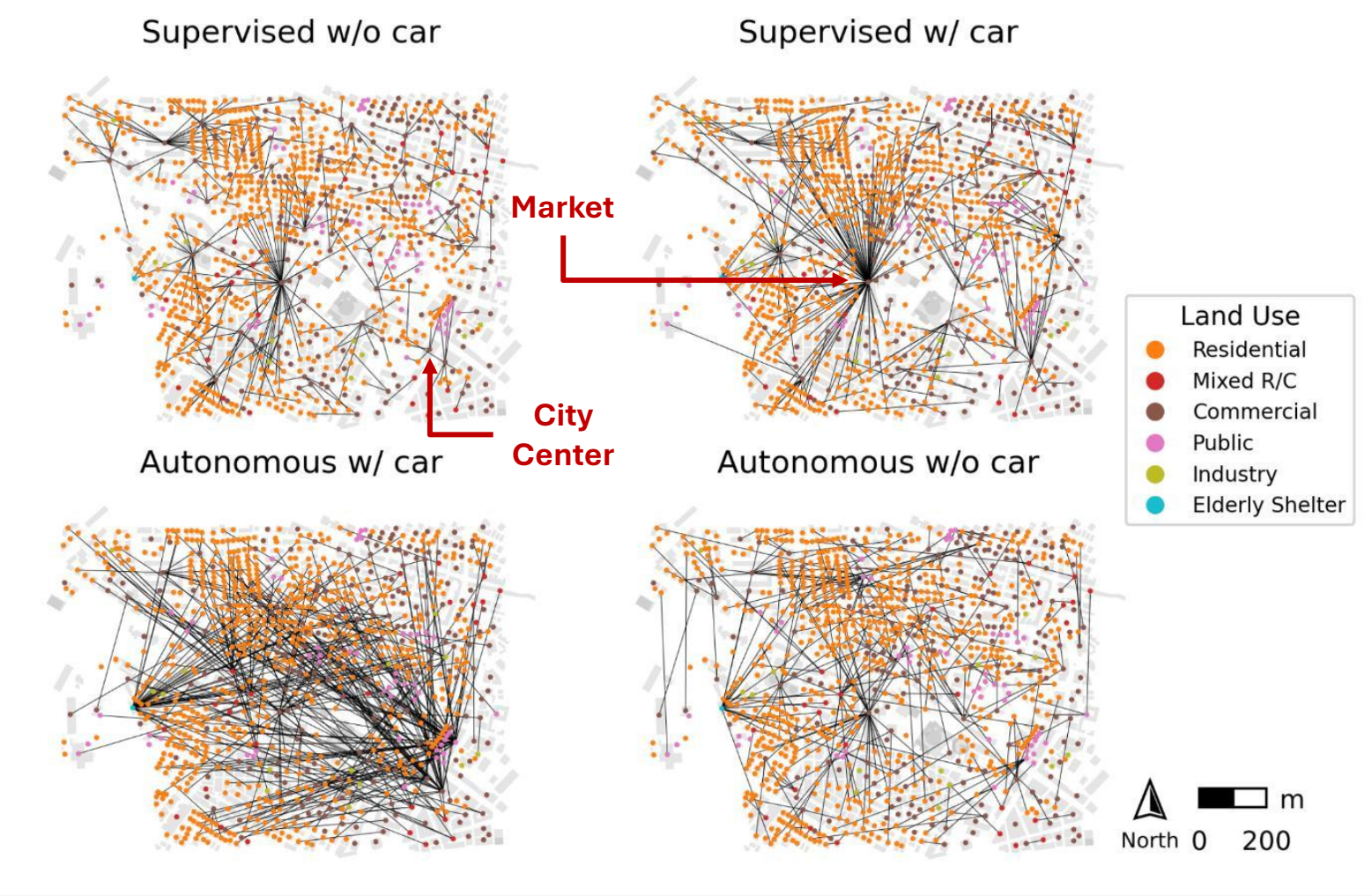
Trajectory characteristics



Stop-level profiles



Transition matrices



Conclusion

Taking the best of both worlds?

Distribution shifts

Gravity-based and car-based logics internalized by the model

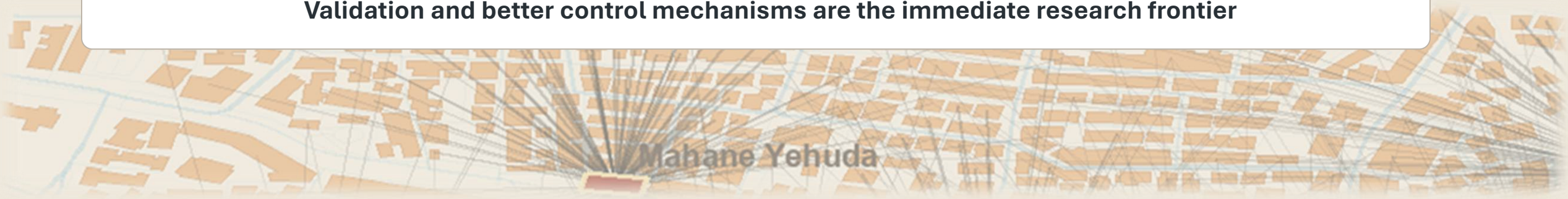
Supervision tradeoff

Gravity constraints improve coherence but amplify return-home and adjacency violations

Next steps

Structured guidance
Explicit coordinates
Controlled rule violations

LLMs can estimate semantic, profile-specific transition structures;
Validation and better control mechanisms are the immediate research frontier



Thank you!

Zi Hen Lin
zihen.lin@mail.huji.ac.il

Yair Grinberger
yair.grinberger@mail.huji.ac.il

Daniel Felsenstein
daniel.felsenstein@mail.huji.ac.il