



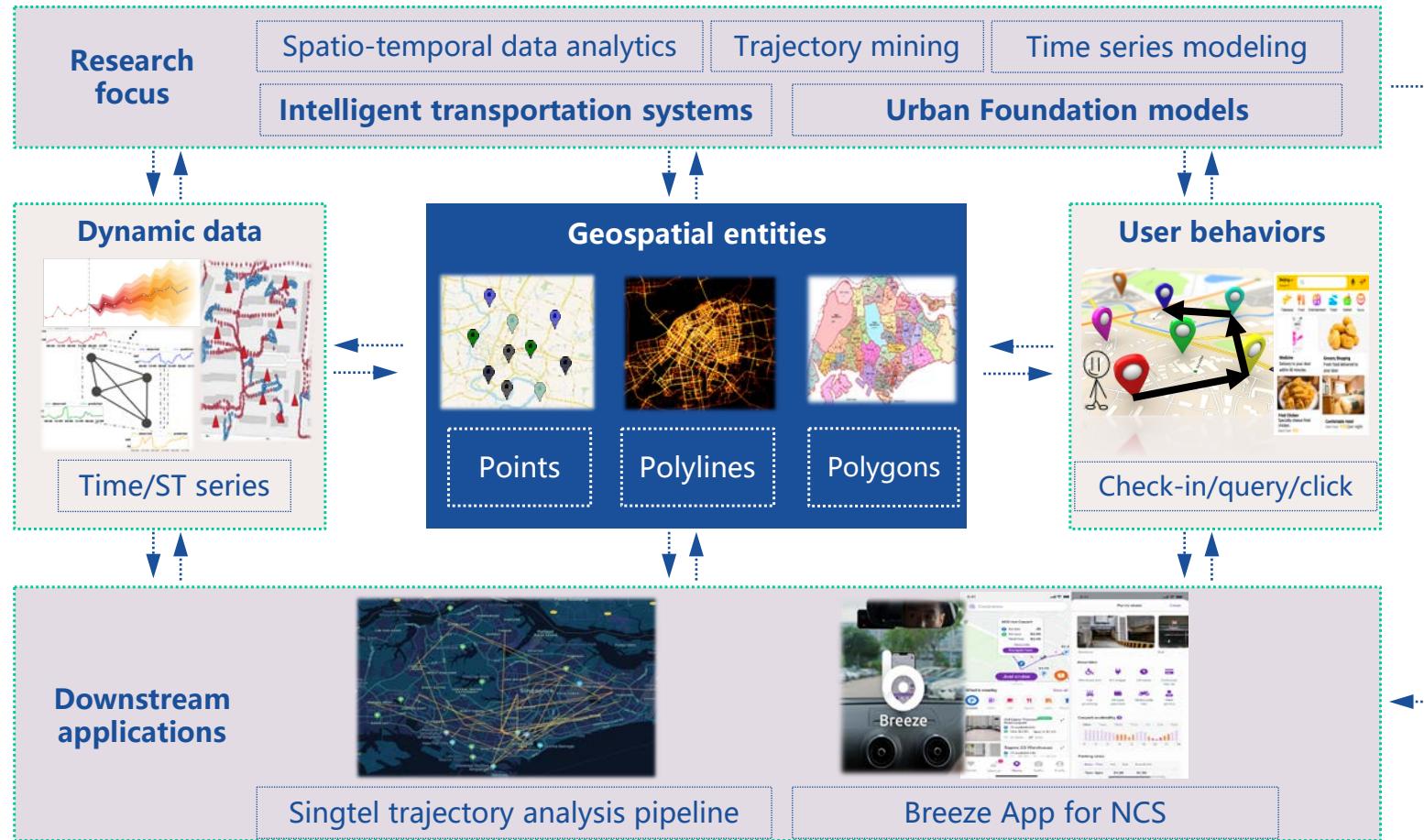
Geospatial Entity Representation: A Step Towards City Foundation Models

Gao Cong
Nanyang Technological University
Singapore



Research Overview

- Urban intelligence
 - Spatio-temporal data mining and analytics, smart city, user modeling



Outline

- Target on a specific problem on **point spatial entity**
 - Geospatial IR or Spatial Keyword Search (VLDB'09--SIGMOD'23)
 - POI recommendations (SIGIR'13 --)
 - Spatial relationship extraction (SIGMOD'23)
- Self-supervised learning for geospatial entity representation
 - Road Network Representation for Road Network Applications (CIKM'21)
 - Region Representation for Region-Level Applications (KDD'23)
 - Application of Foundation Models for Geospatial Applications
 - Efforts toward City Foundation Models.

Our Research on Point Spatial Entity

Geospatial IR
(Spatial keyword query)
(VLDB'09 --)



POI / User recommendation
(SIGIR'13 --)

POI data exploration
(SIGMOD'18)

Integrating POI with tweets & photos
(CIKM'17)



Constructing knowledge graphs

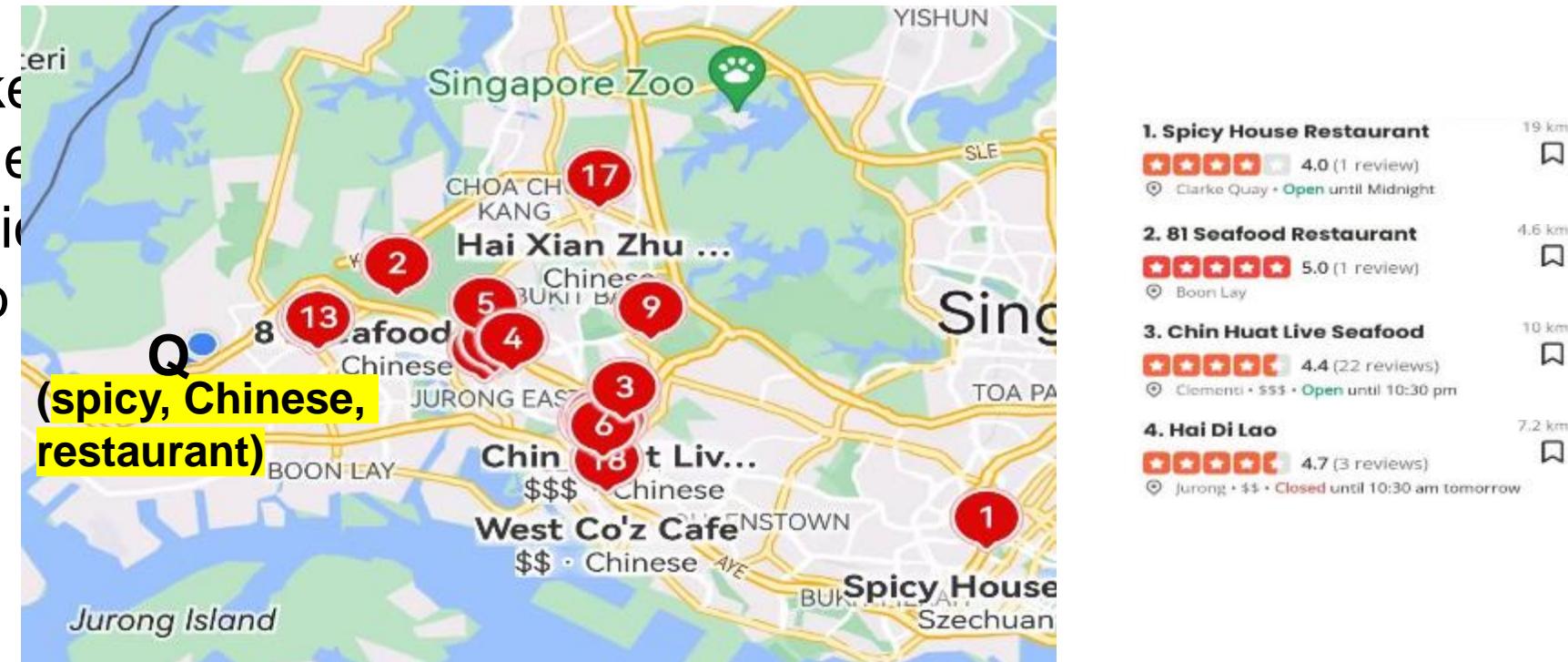
Semantic relationship extraction:
competitiveness/complementary
(VLDB'22)

Entity resolution/ **geospatial relationship extraction** (WWW'22, SIGMOD'23)

Extracting locations and entity linkage
(WWW'18)

Spatial Keyword Query (Geographic IR)

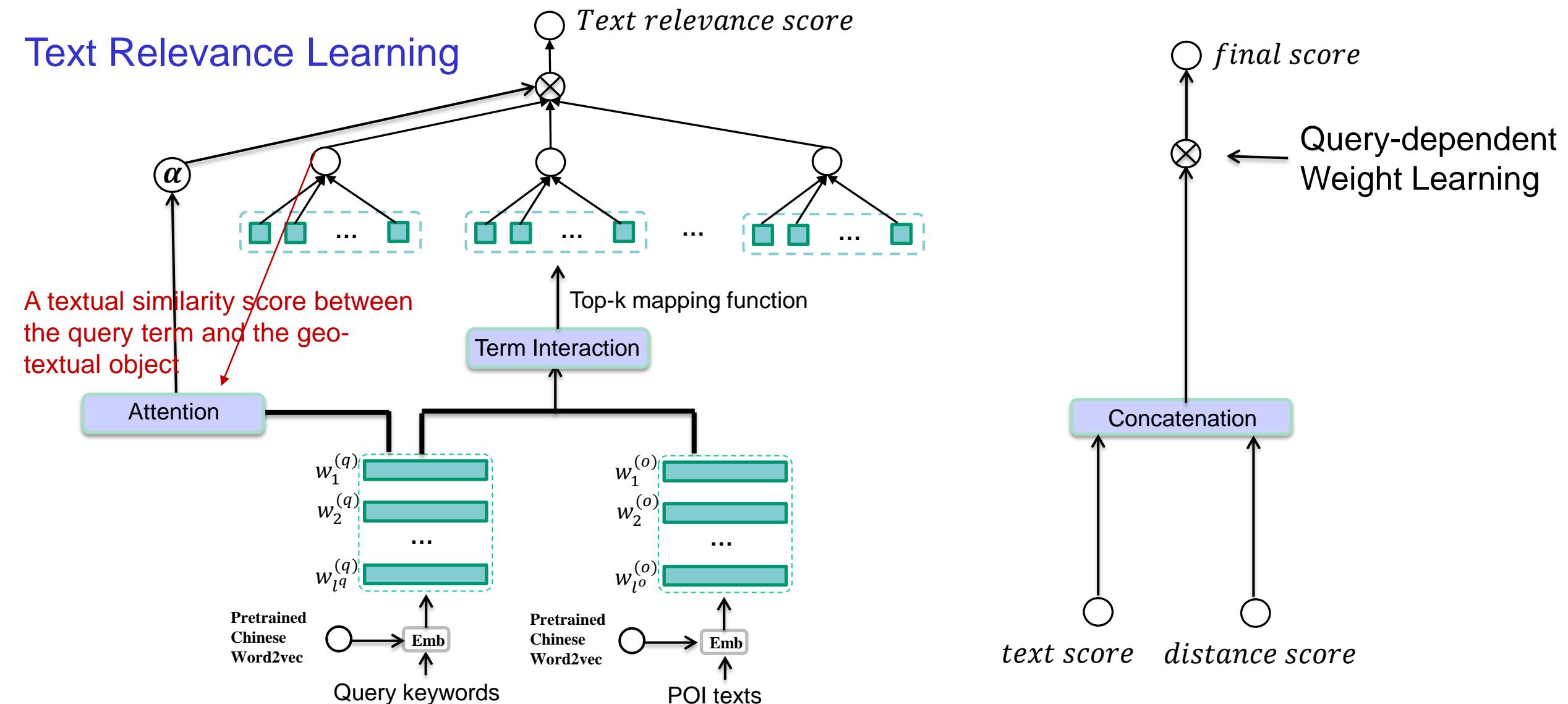
- Take query keywords and location as input and output retrieved objects/documents
- Applications of spatial keyword queries
 - Geographic search engines
 - location-based services
 - locally targeted web advertising



Spatial Keyword Query Example on Yelp (or Meituan)

Geospatial entity representation learning

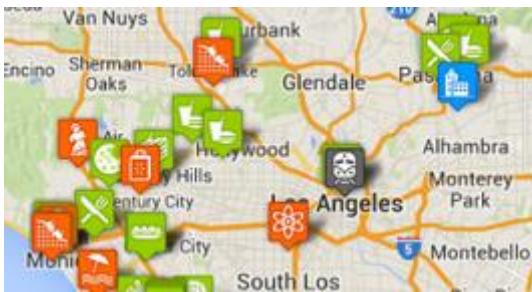
Text Relevance Learning



POI recommendation

- Given a set of **POIs**, and a set of **users** each associated with a set of **visited POIs**, POI recommendation is to recommend for each user **new POIs** that are likely to be visited.

A large number of POIs



Users with different interests

POI
recommendation



Geospatial database

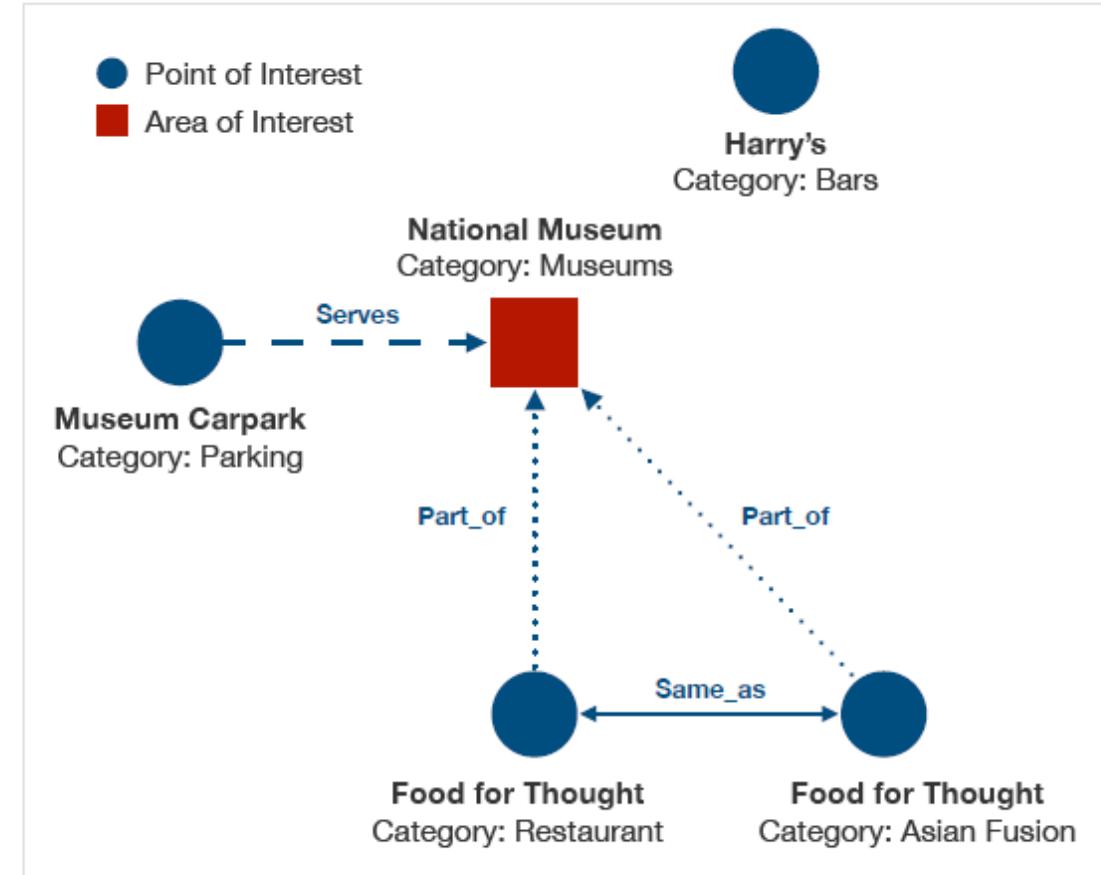
Name	Lat	Long	Address	Category
National Museum	1.29682	103.84877	93 Stamford Rd, 178897	Museums
Food for Thought	1.2963	103.84876	93 Stamford Road #01-04, National Museum, 178897	Asian Fusion
Museum Carpark	1.296509	103.84794		Parking
Harry's	1.2976	103.84905	90 Stamford Rd, 178903	Bars
Food for Thought	1.29675	103.8486		Restaurant

Geospatial DB

Although convenient, this representation hinders the exploration of **geospatial relationships** between the entities

Geospatial KG

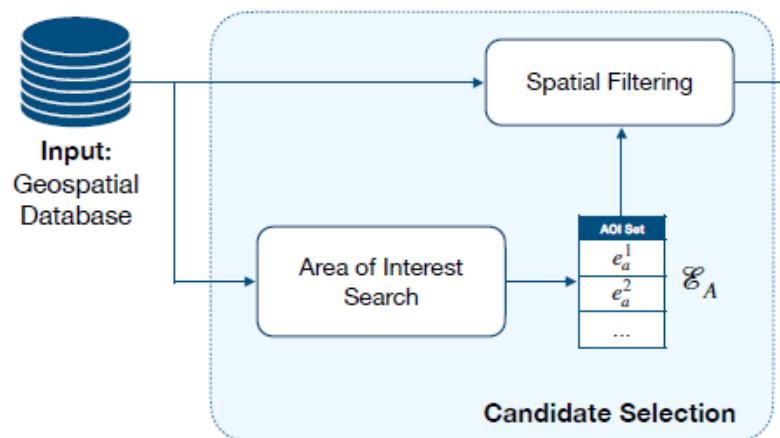
- Relationships between the entities exist and can be captured in a KG representation
- **Knowledge Graphs** are ubiquitous today and offer several advantages:
 - Machine-readable format
 - Can represent both entities and their relations
 - Widely adopted in AI applications
- Existing geoKGs represent only *coarse-grained* relationships



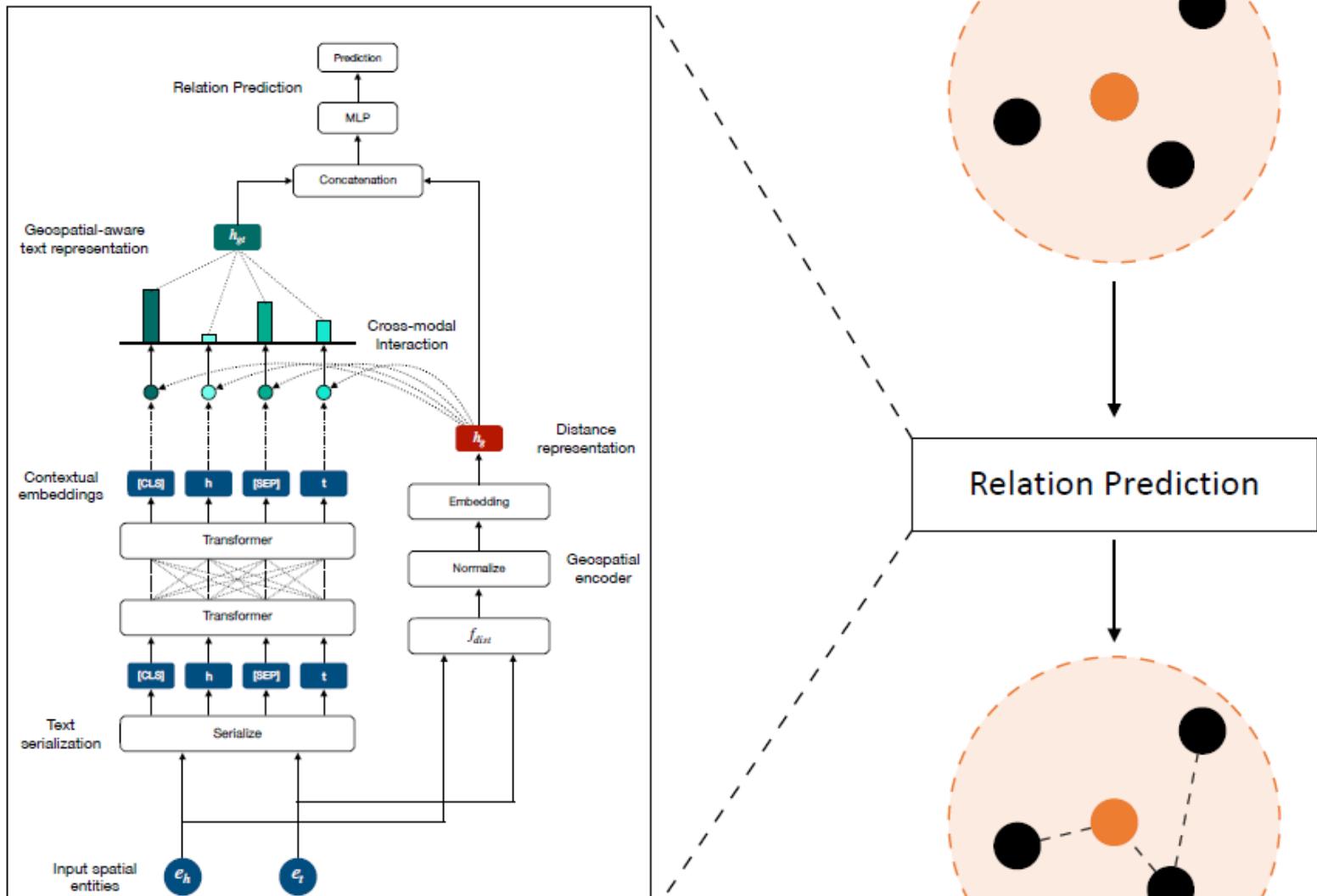
YAGO2Geo

DBpedia

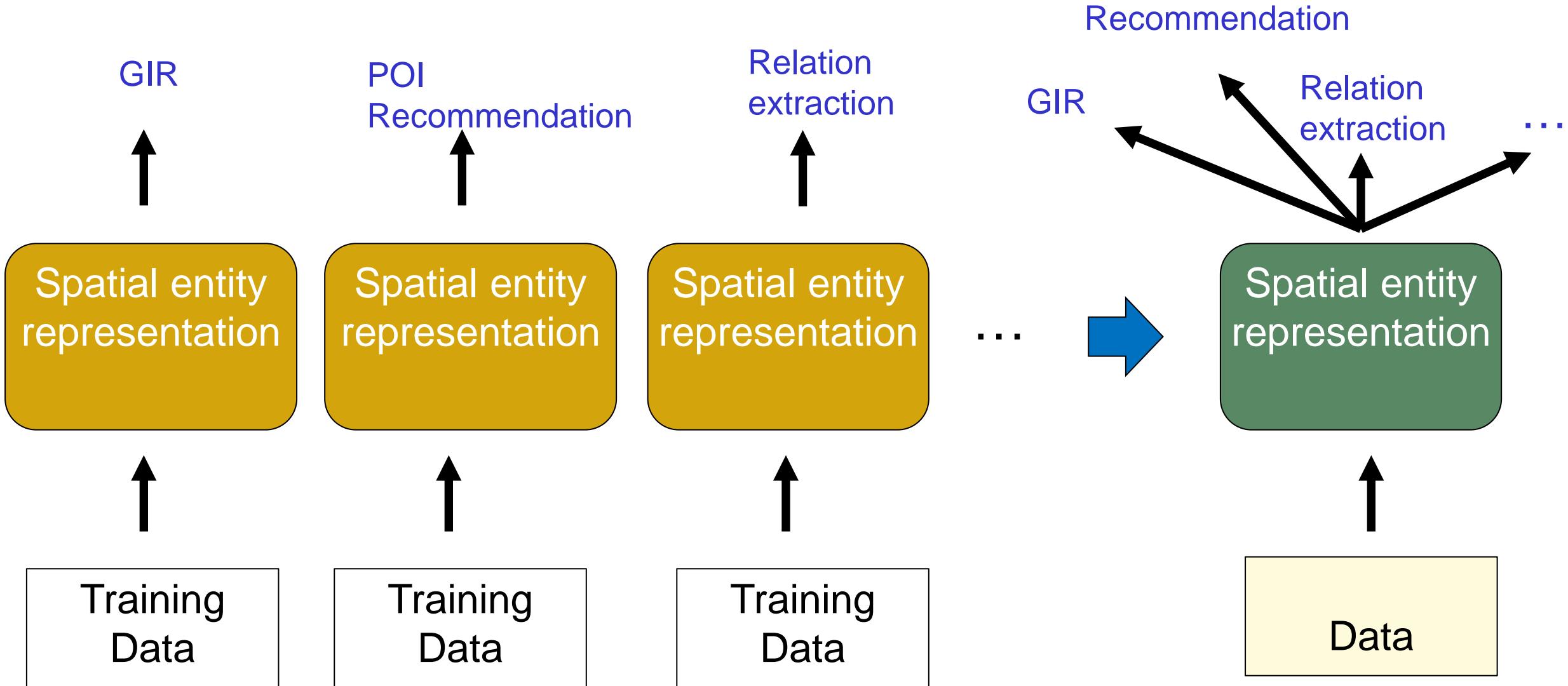
Proposed solution



- **Candidate Selection Step:** Aim at relationships
- **Relation Prediction:** Aim at identifying relations
- **The KG refinement:** Aim to ensure correctness



A Summary

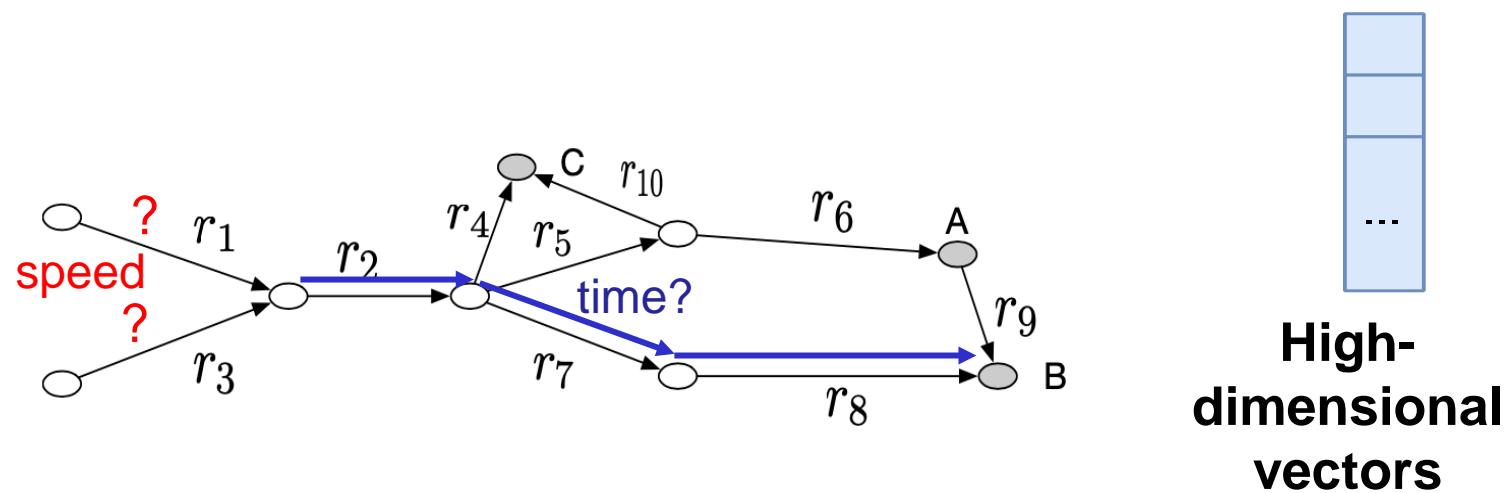


Outline

- Target on a specific problem on **point spatial entity**
 - Geospatial IR or Spatial Keyword Search (VLDB'09--SIGMOD'23)
 - POI recommendations (SIGIR'13 --)
 - Spatial relationship extraction (SIGMOD'23)
- Self-supervised learning for geospatial entity representation
 - Road Network Representation for Road Network Applications (CIKM'21)
 - Region Representation for Region-Level Applications (KDD'23)
 - Application of Foundation Models for Geospatial Applications
 - Efforts toward City Foundation Models.

Representation Learning for Road Networks

- ▷ **Motivation:** numerous applications are built upon road networks, such as travel time estimation, traffic inference, etc.

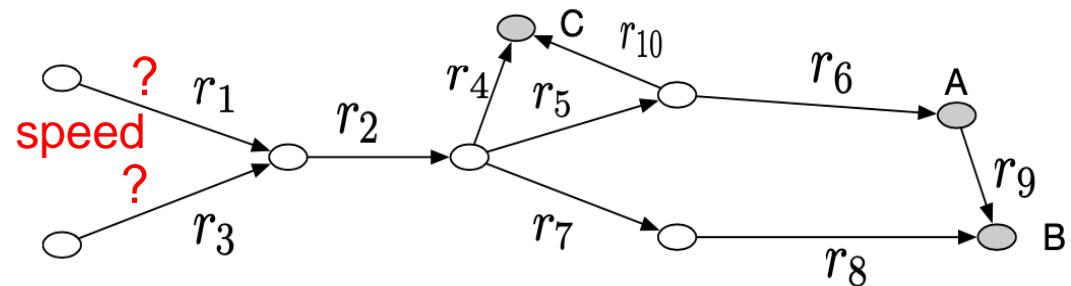


- ▷ **Objective:** derive effective representations that are robust and generic for downstream applications.
 - Road segment-based & trajectory-based applications

Experiments

▷ Road segment-based application result:

Task	Road Label Classification				Traffic Inference			
	Chengdu		Xi'an		Chengdu		Xi'an	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	MAE	RMSE	MAE	RMSE
DW	0.522	0.493	0.552	0.524	7.32	9.14	6.78	8.57
node2vec	0.524	0.495	0.586	0.559	7.12	9.00	6.41	8.22
GAE	0.432	0.328	0.447	0.339	6.91	8.72	6.41	8.39
GraphSAGE	0.452	0.324	0.466	0.347	6.48	8.52	6.12	7.98
RFN	0.516	0.484	0.577	0.570	6.89	8.77	6.57	8.43
IRN2Vec	0.497	0.458	0.531	0.506	6.52	8.52	6.60	8.59
HRNR	0.541	0.527	0.631	0.609	7.03	8.82	6.52	8.45
Toast	0.602	0.599	0.692	0.659	5.95	7.70	5.71	7.44



Experiments

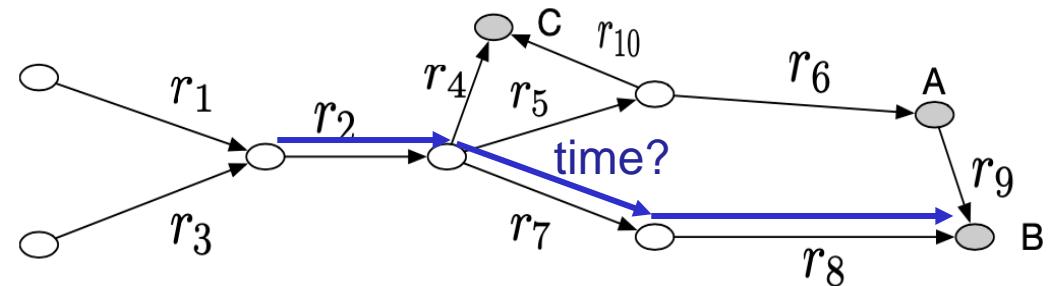
▷ Trajectory-based application result

Trajectory similarity search

	Chengdu		Xi'an	
	MR	HR@10	MR	HR@10
para2vec	216.92	0.251	279.38	0.205
t2vec	46.17	0.781	38.67	0.806
LCSS	67.72	0.487	83.94	0.469
EDR	458.20	0.174	529.74	0.119
Fréchet	21.17	0.847	22.79	0.894
Toast	10.10	0.885	13.71	0.905

Travel time estimation

	Chengdu		Xi'an	
	MAE	RMSE	MAE	RMSE
para2vec	220.45	302.72	244.73	345.49
t2vec	165.18	240.72	207.56	311.04
Road-Pool	151.80	223.02	185.47	293.82
Toast	127.80	190.86	175.68	265.09

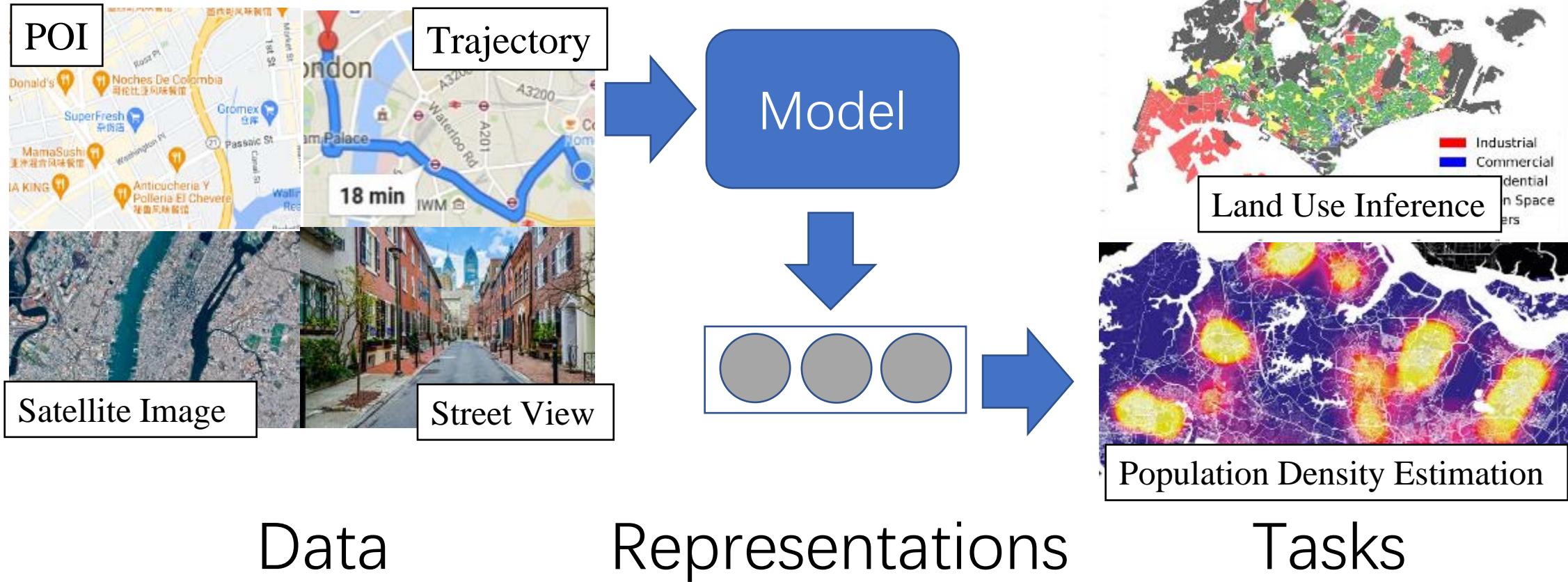


Outline

- Target on a specific problem on **point spatial entity**
 - Geospatial IR or Spatial Keyword Search (VLDB'09--SIGMOD'23)
 - POI recommendations (SIGIR'13 --)
 - Spatial relationship extraction (SIGMOD'23)
- Self-supervised learning for geospatial entity representation
 - Road Network Representation for Road Network Applications (CIKM'21)
 - Region Representation for Region-Level Applications (KDD'23)
 - Application of Foundation Models for Geospatial Applications
 - Efforts toward City Foundation Models.

Problem of Urban Region Representation Learning

- Urban Region Representation Learning aims at learning effective feature vectors for urban regions to serve various downstream tasks.



Our motivations

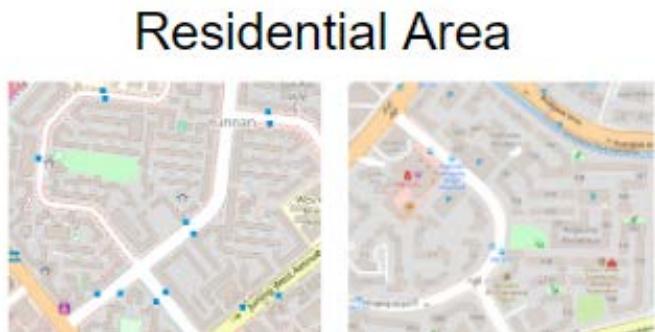
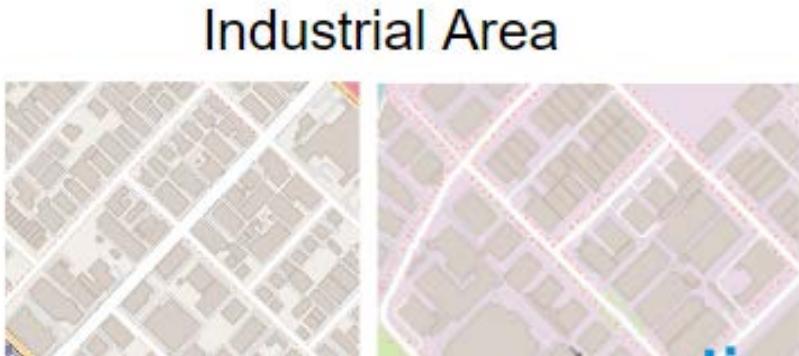


An Example Building Group
(Singapore Public House)

We focus on **OSM buildings**.

- **Buildings**, (or formally, **building footprints**), refer to the 2-D building polygon on the map
 - size, height, type, name...
- **Building groups** refers to the **collection of buildings** in a defined spatial area.
 - We use OSM road networks to partition buildings into building groups.

Introduction



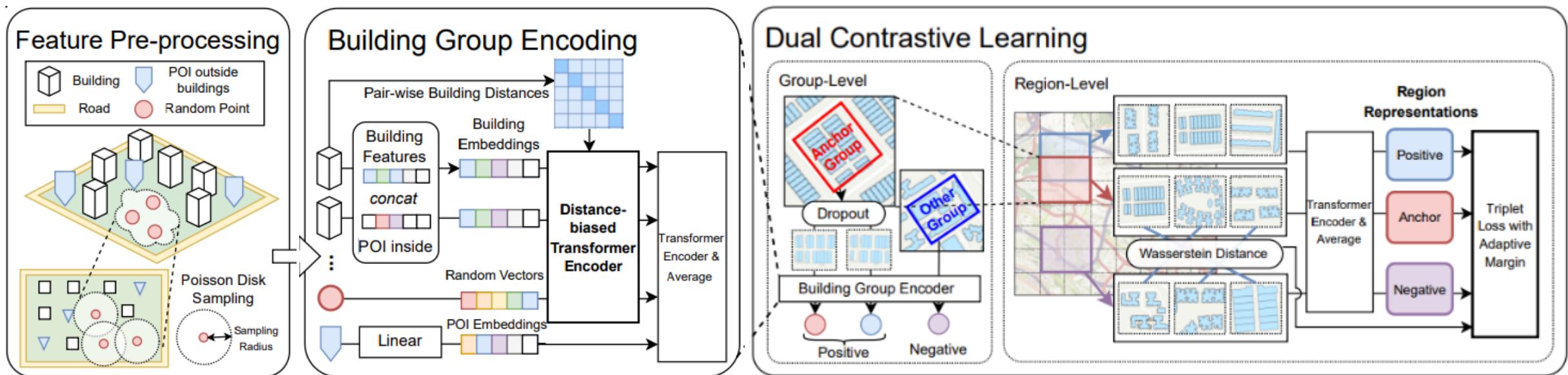
Example Building Groups with
Specific Urban Functions

Comparing to other data types, building data has **advantages**:

- **Effectiveness**
 - Buildings directly carrying urban functions.
- **Availability**
 - Buildings are readily available in OSM

Method

1. **Partition** the city into building groups with road network.
2. **Encode** building groups with POIs and regions with Transformer-based encoders.
3. **Train** the encoder with Group-level and Region-level contrastive learning



Experiments: Land Use Inference

- Infer 5 types of land use (Residential, Industrial, Commercial, Open Space, Other)

Table 2: Land Use Inference in Singapore and New York City

Models	Singapore			New York City		
	L1↓	KL↓	Cosine↑	L1↓	KL↓	Cosine↑
Urban2Vec	0.657±0.033	0.467±0.043	0.804±0.017	0.473±0.018	0.295±0.015	0.890±0.007
Place2Vec	0.645±0.039	0.451±0.047	0.812±0.018	0.518±0.016	0.308±0.012	0.878±0.005
Doc2Vec	0.679±0.050	0.469±0.058	0.789±0.027	0.506±0.015	0.299±0.016	0.885±0.008
GAE	0.759±0.040	0.547±0.051	0.765±0.022	0.589±0.011	0.365±0.011	0.855±0.007
DGI	0.598±0.029	0.372±0.032	0.846±0.012	0.433±0.009	0.237±0.012	0.907±0.005
Transformer	0.556±0.046	0.357±0.070	0.850±0.026	0.436±0.020	0.251±0.018	0.903±0.008
RegionDCL-no random	0.535±0.054	0.321±0.066	0.863±0.030	0.422±0.011	0.234±0.010	0.910±0.005
RegionDCL-fixed margin	0.515±0.042	0.303±0.040	0.872±0.020	0.426±0.011	0.248±0.018	0.905±0.008
RegionDCL	0.498±0.038	0.294±0.047	0.879±0.021	0.418±0.010	0.229±0.008	0.912±0.004

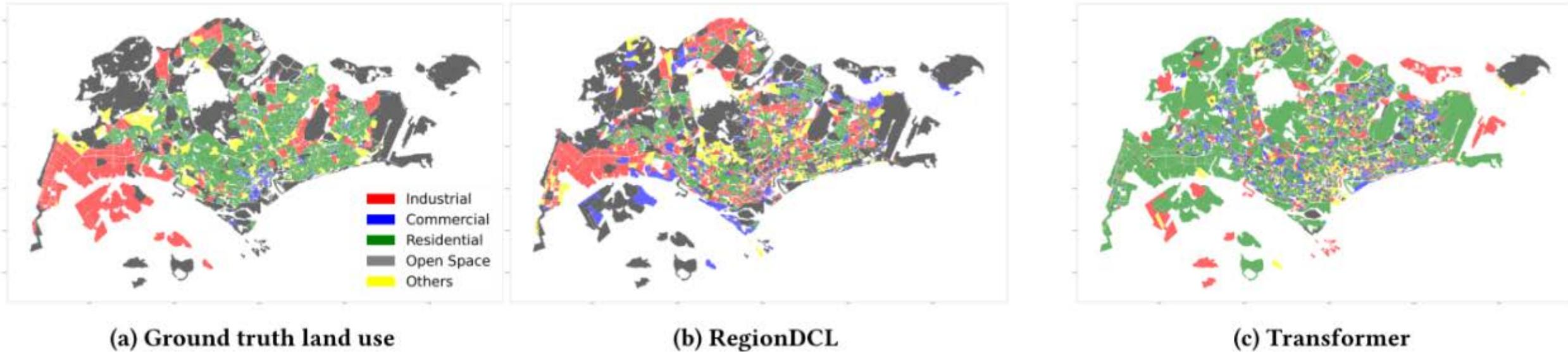
Experiments: Population Density Inference

- Similar results in inferring the population density within regions

Table 3: Population Density Inference in Singapore and New York City

Models	Singapore			New York City		
	MAE↓	RMSE↓	R ² ↑	MAE↓	RMSE↓	R ² ↑
Urban2Vec	6667.84±623.27	8737.27±902.41	0.303±0.119	5328.38±200.58	7410.42±261.89	0.522±0.028
Place2Vec	6952.34±713.30	9696.31±1239.65	0.171±0.121	8109.79±175.18	10228.61±261.43	0.096±0.043
Doc2Vec	6982.85±650.76	9506.81±1052.25	0.206±0.062	7734.56±247.99	9827.56±354.51	0.166±0.031
GAE	7183.24±579.82	9374.20±913.56	0.163±0.112	8010.73±290.33	10341.09±362.28	0.071±0.027
DGI	6423.44±671.25	8495.16±972.87	0.305±0.151	5330.11±261.77	7381.92±358.09	0.526±0.032
Transformer	6837.67±716.28	9042.02±1032.99	0.269±0.081	5345.17±216.30	7379.47±308.36	0.522±0.039
RegionDCL-no random	6400.50±630.35	8437.89±993.41	0.364±0.075	5228.27±210.46	7278.70±322.85	0.535±0.040
RegionDCL-fixed margin	6237.61±647.54	8387.56±948.78	0.365±0.107	5125.66±184.27	7159.65±250.12	0.551±0.033
RegionDCL	5807.54±522.74	7942.74±779.44	0.427±0.108	5020.20±216.63	6960.51±282.35	0.575±0.039

Visualization



(a) Ground truth land use

(b) RegionDCL

(c) Transformer

- Cluster the building group embeddings via K-Means
- Ours are visually close to the Singapore land use ground truth
- Baseline fails.

Outline

- Target on a specific problem on **point spatial entity**
 - Geospatial IR or Spatial Keyword Search (VLDB'09--SIGMOD'23)
 - POI recommendations (SIGIR'13 --)
 - Spatial relationship extraction (SIGMOD'23)
- Self-supervised learning for geospatial entity representation
 - Road Network Representation for Road Network Applications (CIKM'21)
 - Region Representation for Region-Level Applications (KDD'23)
 - Application of Foundation Models for Geospatial Applications
 - Efforts toward City Foundation Models.

Foundation Models

Task-Specific Models

Training **specific** models for **specific** tasks

Question Answering Models

Machine Translation Models

Common Sensing Reasoning Models

Reading Comprehension Models

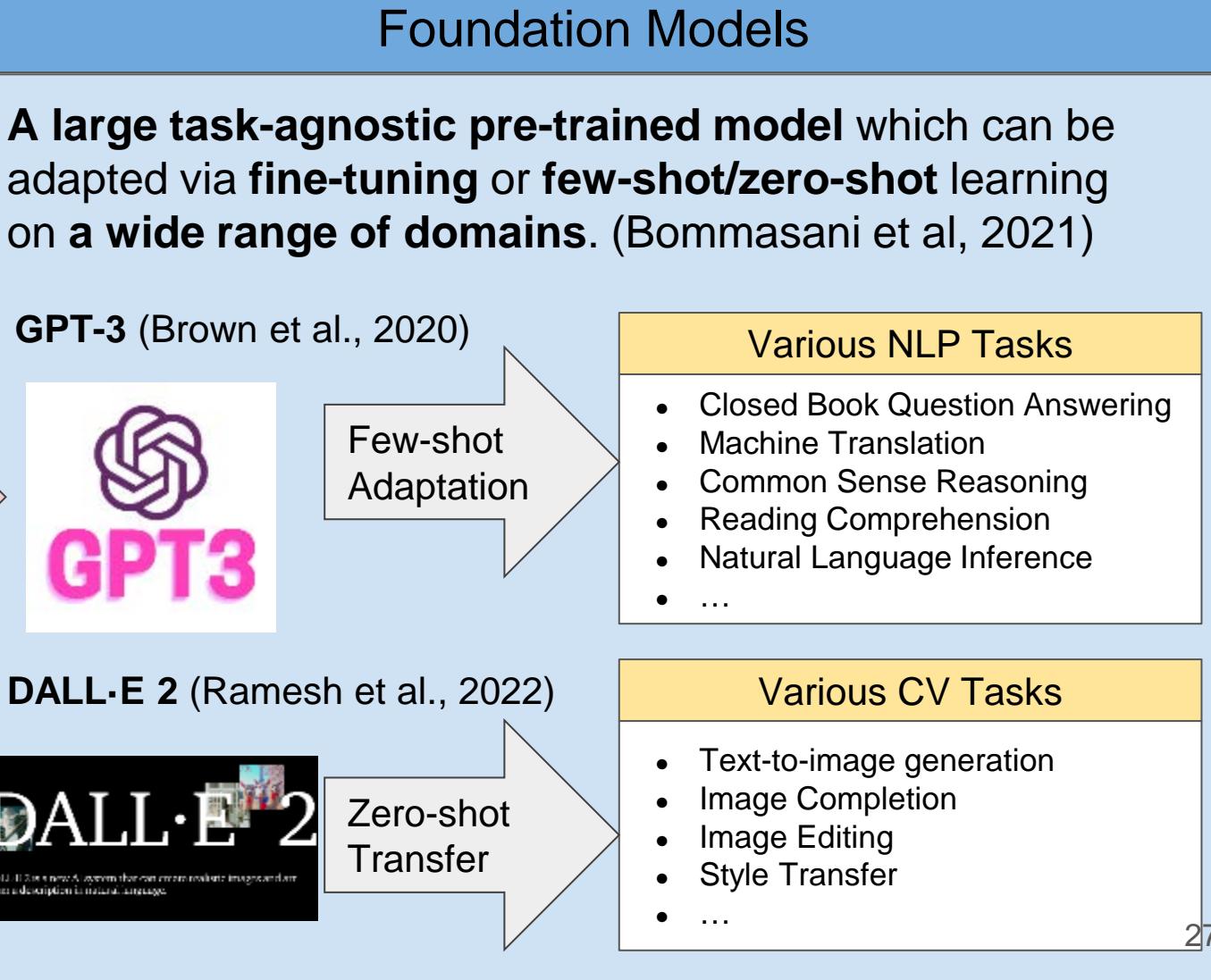
Natural Language Inference Models

Image Classification Models

Text-to-image Generation Models

Image Editing Models

Paradigm shift

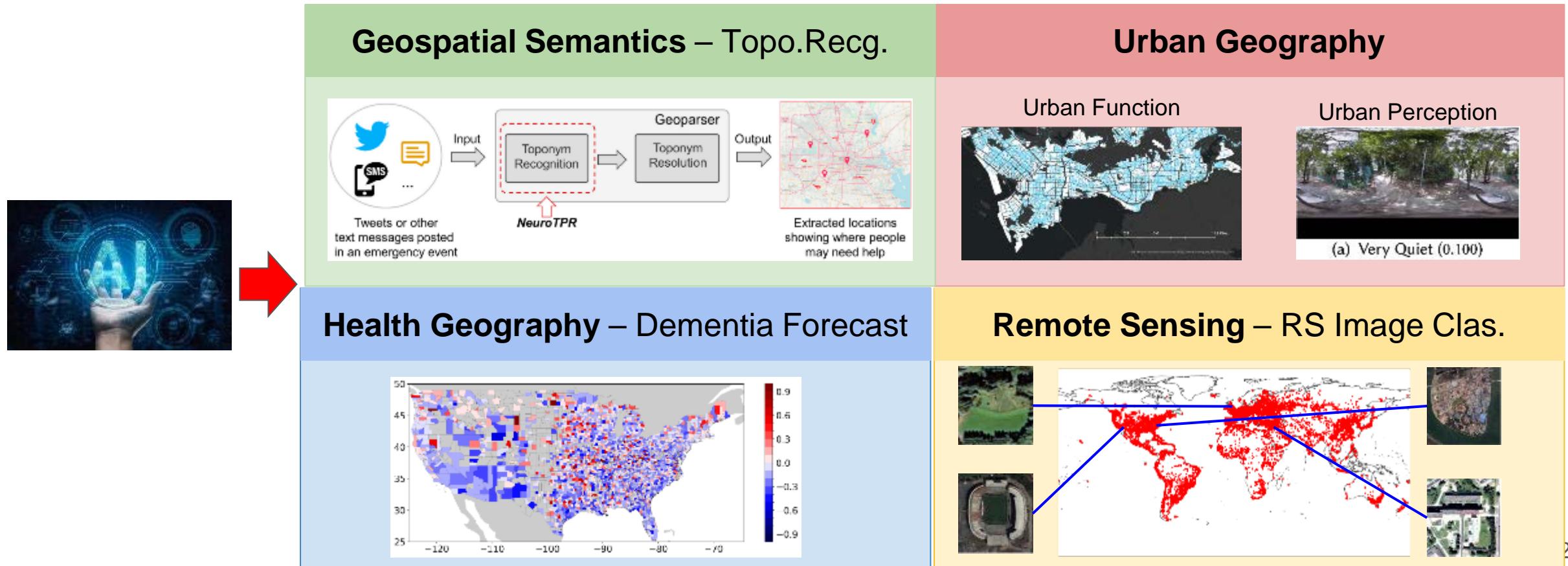


The **Big**
Question



AGI on Geospatial Problems

How do the existing cutting-edge foundation models perform when compared with the state-of-the-art fully supervised task-specific models on various geospatial tasks?



Geospatial Semantics

- Investigate the performance of **GPT-3** on some well established **geospatial semantic tasks**:

Typonym Recognition

[Instruction] ...

Paragraph: Alabama State Troopers say a Greenville man has died of his injuries
→ after being hit by a pickup truck on Interstate 65 in Lowndes County.

Q: Which words in this paragraph represent named places?

A: Alabama; Greenville; Lowndes

...

--

Paragraph: The Town of Washington is to what Williamsburg is to Virginia.

Q: Which words in this paragraph represent named places?

A: Washington; Williamsburg; Virginia

Location Description Recognition

[Instruction] ...

Paragraph: Papa stranded in home. Water rising above waist. HELP 8111 Woodlyn Rd
→ , 77028 #houstonflood

Q: Which words in this paragraph represent location descriptions?

A: 8111 Woodlyn Rd, 77028

...

--

Paragraph: HurricaneHarvey Help Need AT 7506 Jackrabbit Rd, Houston, TX 77095.

Q: Which words in this paragraph represent location descriptions?

A: 7506 Jackrabbit Rd, Houston, TX 77095

*toponyms: proper names of places, also known as place names and geographic names.

GPT-3 Fewshot Learning for Geospatial Semantic Tasks

Task 1 & 2: Toponym Recognition & Location Description Recognition

- **Toponym recognition:** FMs (e.g., GPT-2/3) consistently outperform the **fully-supervised** baselines with only **8 few-shot** examples
- **Location Description Recognition:** GPT-3 achieves the best Recall score across all methods

		Toponym Recognition		Location Description Recognition			
Model		#Param	Hu2014	Ju2016	HaveyTweet2017		
			Accuracy ↓	Accuracy ↓	Precision ↓	Recall ↓	F-Score ↓
(A)	Stanford NER (nar. loc.) [30]	-	0.787	0.010	0.828	0.399	0.539
	Stanford NER (bro. loc.) [30]	-	-	0.012	0.729	0.44	0.548
	Retrained Stanford NER [30]	-	-	0.078	0.604	0.410	0.489
	Caseless Stanford NER (nar. loc.) [30]	-	-	0.460	0.803	0.320	0.458
	Caseless Stanford NER (bro. loc.) [30]	-	-	0.514	0.721	0.336	0.460
	spaCy NER (nar. loc.) [44]	-	0.681	0.000	0.575	0.024	0.046
	spaCy NER (bro. loc.) [44]	-	-	0.006	0.461	0.304	0.366
	DBpedia Spotlight[99]	-	0.688	0.447	-	-	-
(B)	Edinburgh [7]	-	0.656	0.000	-	-	-
	CLAVIN [134]	-	0.650	0.000	-	-	-
	TopoCluster [23]	-	0.794	0.158	-	-	-
(C)	CamCoder [33]	-	0.637	0.004	-	-	-
	Basic BiLSTM+CRF [77]	-	-	0.595	0.703	0.600	0.649
	DM NLP (top. rec.) [139]	-	-	0.723	0.729	0.680	0.703
	NeuroTPR [135]	-	0.675 [†]	0.821	0.787	0.678	0.728
(D)	GPT2 [115]	117M	0.556	0.650	0.540	0.413	0.468
	GPT2-Medium [115]	345M	0.806	0.802	0.529	0.503	0.515
	GPT2-Large [115]	774M	0.813	0.779	0.598	0.458	0.518
	GPT2-XL [115]	1558M	0.869	0.846	0.492	0.470	0.481
	GPT-3 [15]	175B	0.881	0.811*	0.603	0.724	0.658
	InstructGPT [106]	175B	0.863	0.817*	0.567	0.688	0.622
	ChatGPT (Raw.) [104]	176B	0.800	0.696*	0.516	0.654	0.577
	ChatGPT (Con.) [104]	176B	0.806	0.656*	0.548	0.665	0.601

Health Geography

Task 4: US County-Level Dementia Time Series Forecasting

[Instruction] This is a set of time series forecasting problems.

The "Paragraph" is a time series of the numbers of deaths from

→ alzheimer's disease for one of US counties from 1999 to 2019.

The goal is to predict the number of deaths from alzheimer's disease at

→ this county in 2020. Please give a single number as the

→ prediction.

--

--

Paragraph: At Santa Barbara County, CA, from 1999 to 2019, the numbers

→ of deaths from alzheimer's disease are

→ 126 in 1999, 114 in 2000, 124 in 2001, 127 in 2002, 156 in 2003,

→ 154 in 2004, 175 in 2005, 172 in 2006, 171 in 2007, 248 in 2008, 204

→ in 2009, 241 in 2010, 260 in 2011, 297 in 2012, 283 in 2013, 308 in

→ 2014, 358 in 2015, 365 in 2016, 334 in 2017, 363 in 2018,

→ and 328 in 2019.

Q: Please forecast the number in 2020 at Santa Barbara County, CA?

A: 345

Listing 4. US county-level Alzheimer time series forecasting with LLMs by zero-shot learning. Yellow block: the historical time series data of one US county. Orange box: the outputs of InstructGPT. Here, we use Santa Barbara County, CA as an example and the correct answer is 373.

Table 3. Evaluation results of various GPT models and baselines on the US county-level dementia time series forecasting task. We use same model set and evaluation metrics as Table 2.

	Model	#Param	MSE ↓	MAE ↓	MAPE ↓	R ² ↑
(A) Simple	Persistence [103, 107]	-	1,648	16.9	0.189	0.979
(B) Supervised ML	ARIMA [58]	-	1,133	15.1	0.193	0.986
	GPT2 [115]	117M	77,529	92.0	0.587	-0.018
	GPT2-Medium [115]	345M	226,259	108.1	0.611	-2.824
	GPT2-Large [115]	774M	211,881	94.3	0.581	-1.706
	GPT2-XL [115]	1558M	162,778	99.8	0.627	-1.082
(C) Zero shot LLMs	GPT-3 [15]	175B	1,105	14.5	0.180	0.986
	InstructGPT [106]	175B	831	13.3	0.179	0.989
	ChatGPT (Raw.) [104]	176B	4,115	23.2	0.217	0.955
	ChatGPT (Con.) [104]	176B	3,402	20.7	0.231	0.944

GPT-3 Fewshot Learning for Geospatial Semantic Tasks

- **Shortcoming of text FMs:** by design they are unable to handle other data modality, e.g., geo-coordinates, toponym resolution/geoparsing

- The predicted coordinates are not accurate

Geoparsing

[Instruction] ...

Paragraph: San Jose was founded in 1803 when allotments of land were made ...

Q: Which words in this paragraph represent named places?

A: San Jose; New Mexico

Q: What is the location of San Jose?

A: 35.39728, -105.47501

...
--

Paragraph: the city of fairview had a population of 260 as of july 1, 2015. ...

Q: Which words in this paragraph represent named places?

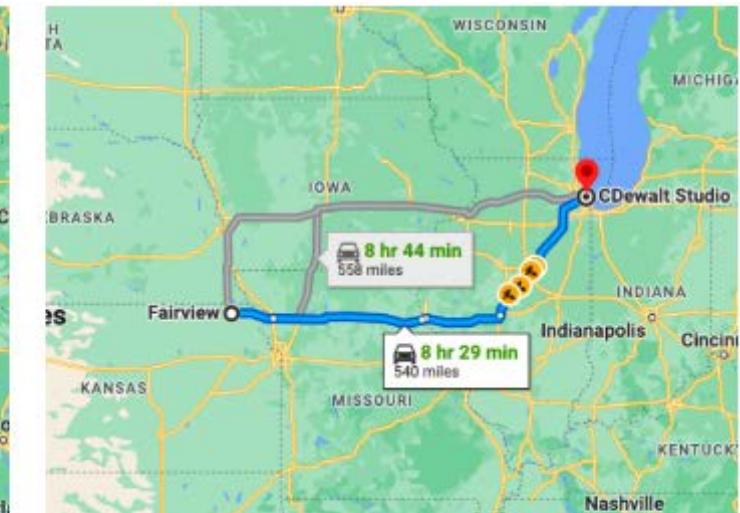
A: Fairview

Q: What is the location of Fairview?

A: 41.85003, -87.65005



(a) [TEXT]: Franklin is a city in and the county seat of simpson county, ...



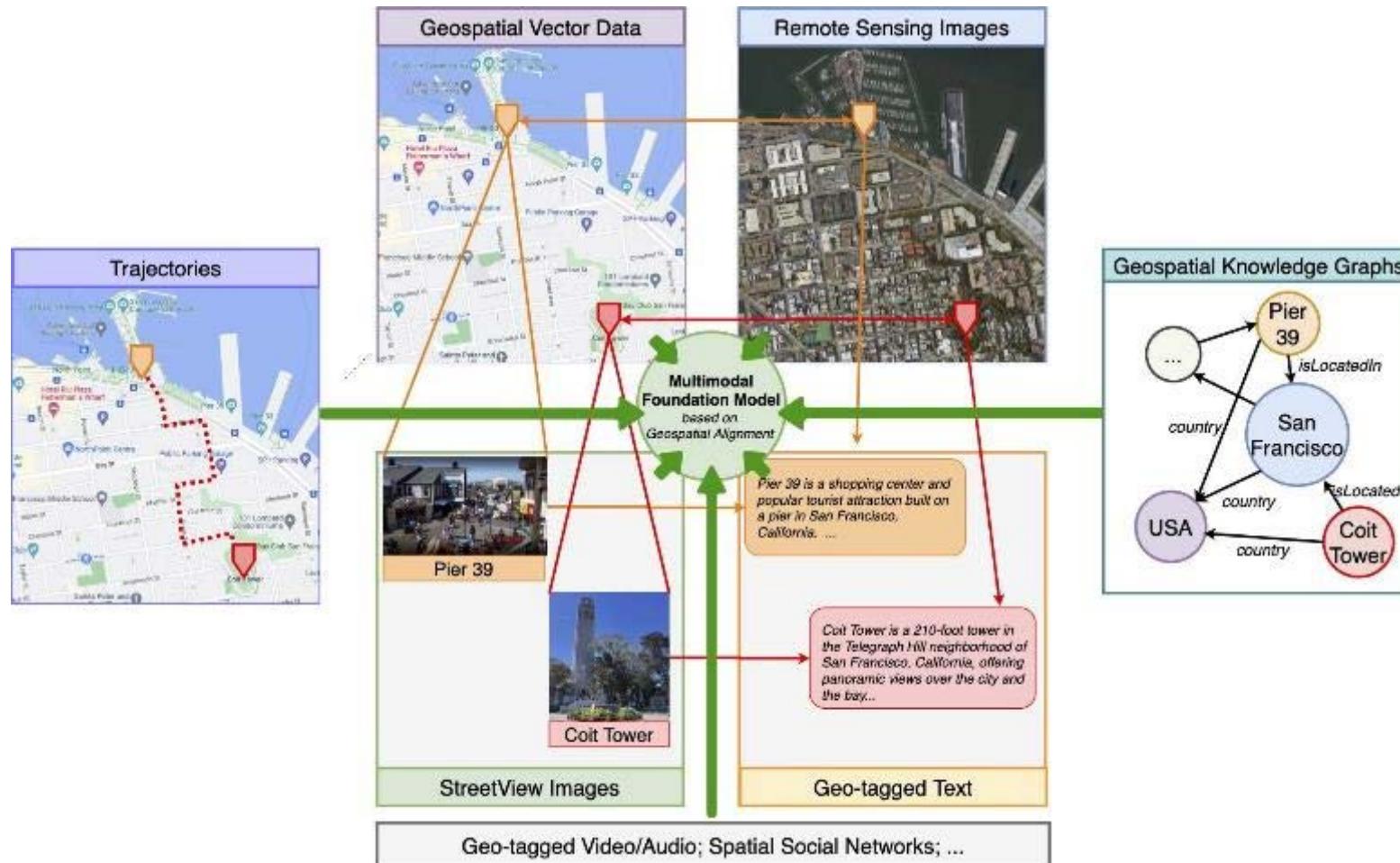
(b) [TEXT]: the city of Fairview had a population of 260 as of july 1, 2015. ...

Outline

- Target on a specific problem on **point spatial entity**
 - Geospatial IR or Spatial Keyword Search (VLDB'09--SIGMOD'23)
 - POI recommendations (SIGIR'13 --)
 - Spatial relationship extraction (SIGMOD'23)
- Self-supervised learning for geospatial entity representation
 - Road Network Representation for Road Network Applications (CIKM'21)
 - Region Representation for Region-Level Applications (KDD'23)
 - Application of Foundation Models for Geospatial Applications
 - **Efforts toward City Foundation Models.**

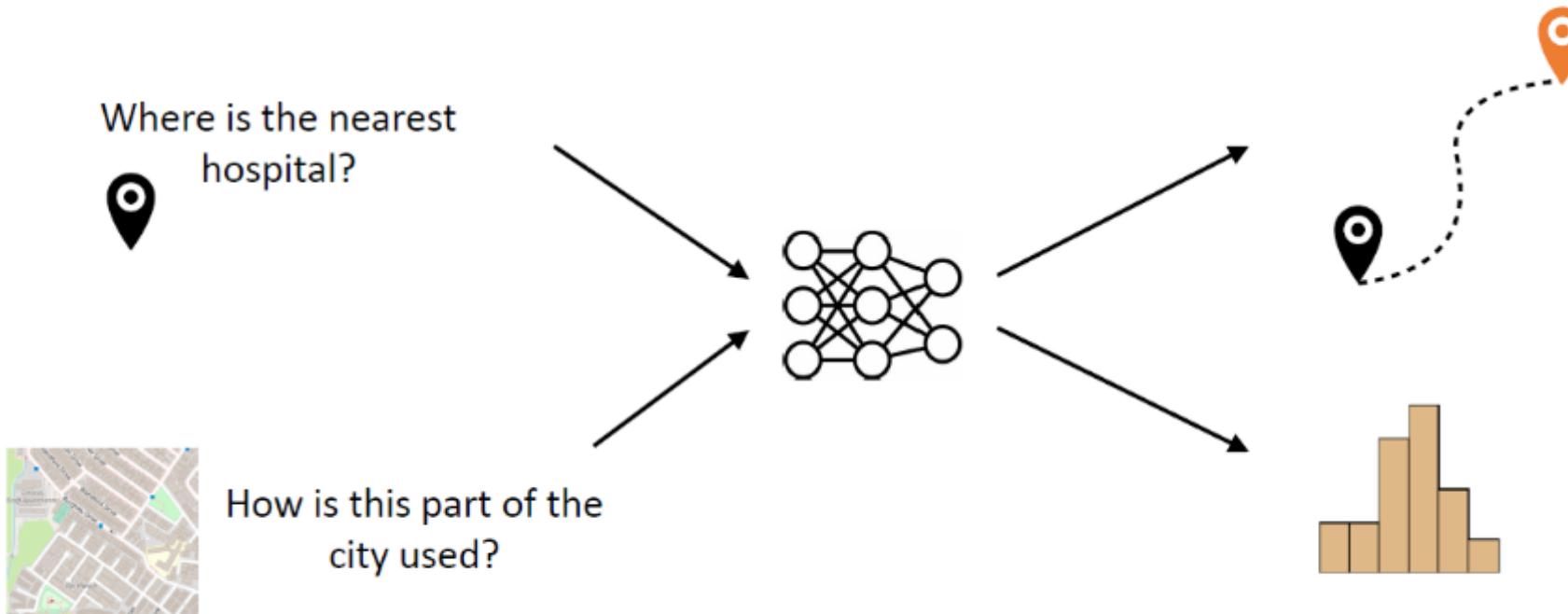
A Multimodal City FM for GeoAI

Vision: a multimodal City FM for GeoAI that use their **geospatial relationships as alignments** among different data modalities.



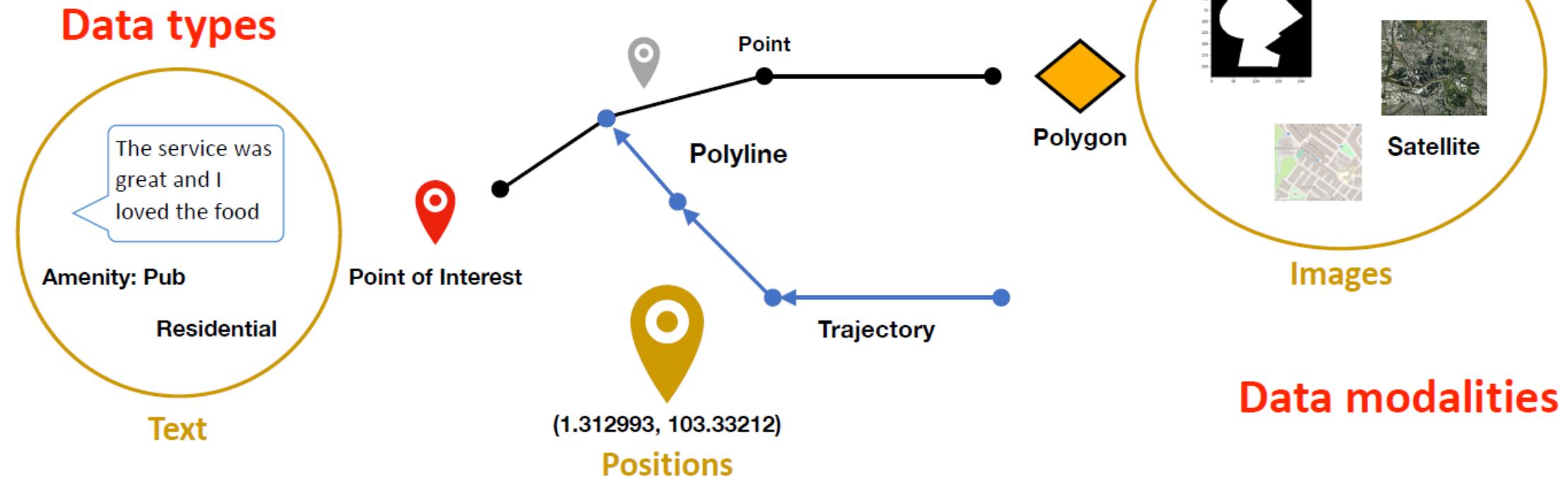
Motivations of City Foundation Models

FMs have the potential to revolutionise the way we use **geospatial data**



Challenges

A slower adoption of FMs in the **geospatial** domain... why?



Challenges

A slower adoption of FMs in the **geospatial** domain... why?

Data sources also present a challenge, different data comes from different providers, and is available in different places!

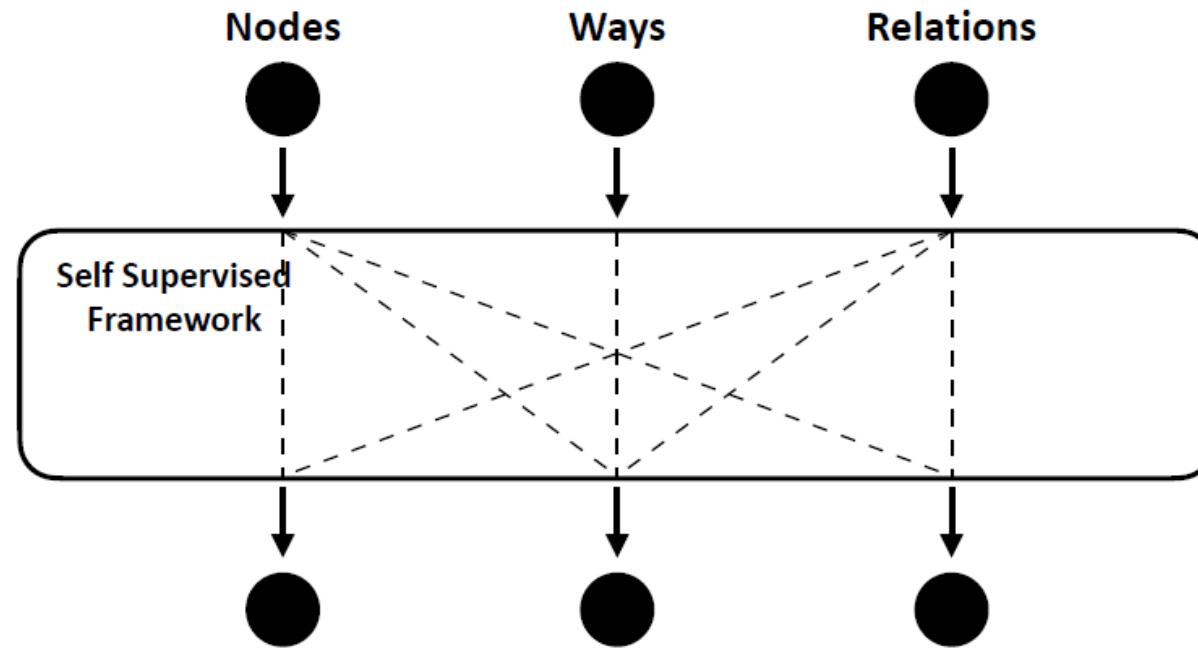


Our Attempts

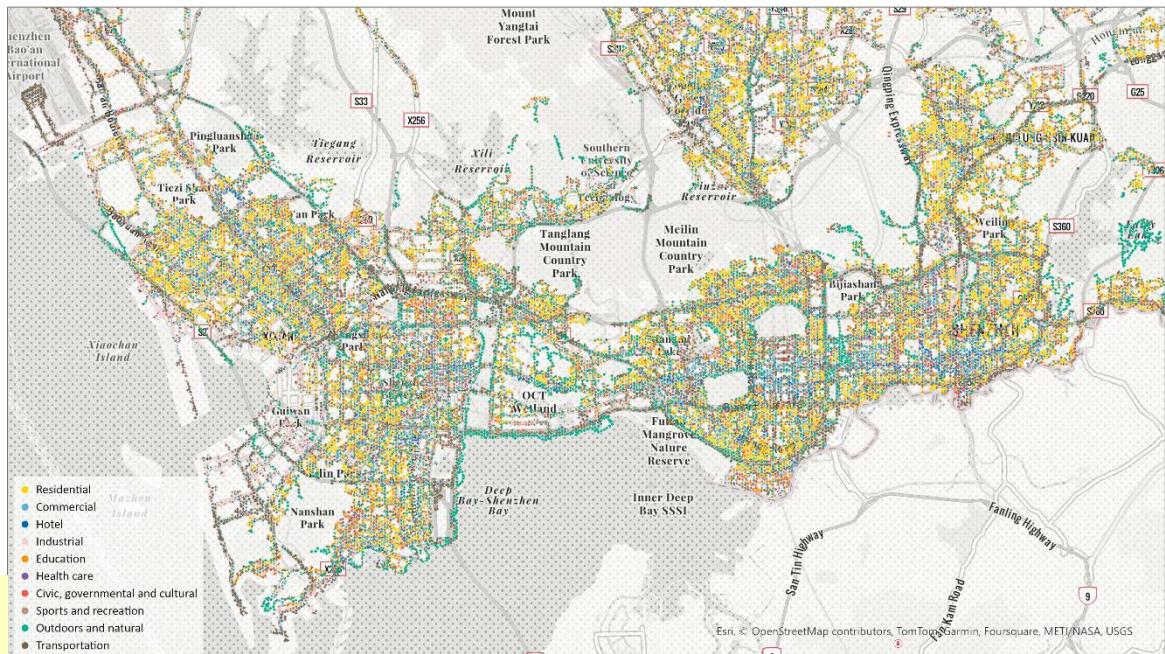
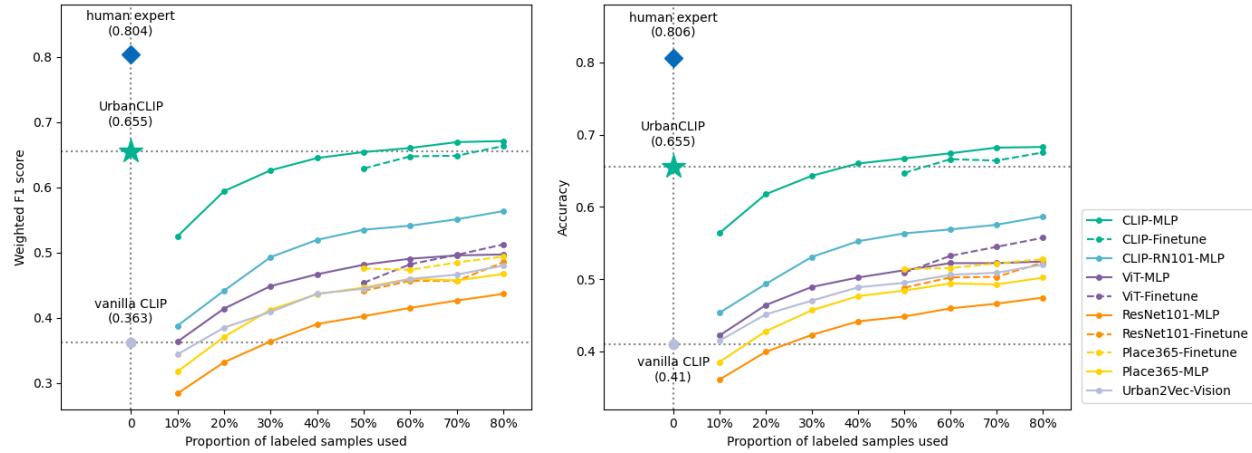
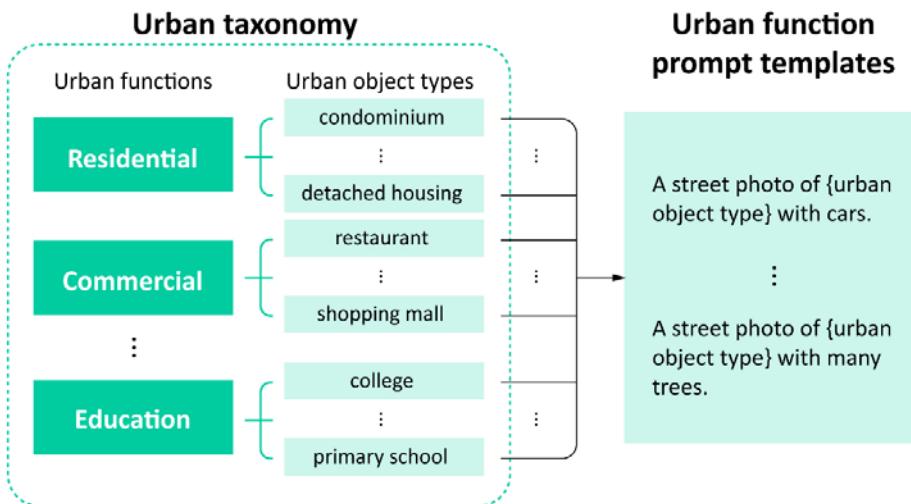
- Attempt 1: Build a “Foundation Model” from the scratch (**decoding part only, 2022**)
- Attempt 2: Integrate Domain Knowledge with existing Foundation Models (**prompting**)
- Attempt 3: Finetune existing Foundation Models with domain data/knowledge (**finetuning**)
 - Data preparation
 - Tokenization
 - Mixture-of-Experts
- Attempt 4: **AI Agents**

Attempt1: Use OpenStreetMap to Build a City Foundation Model

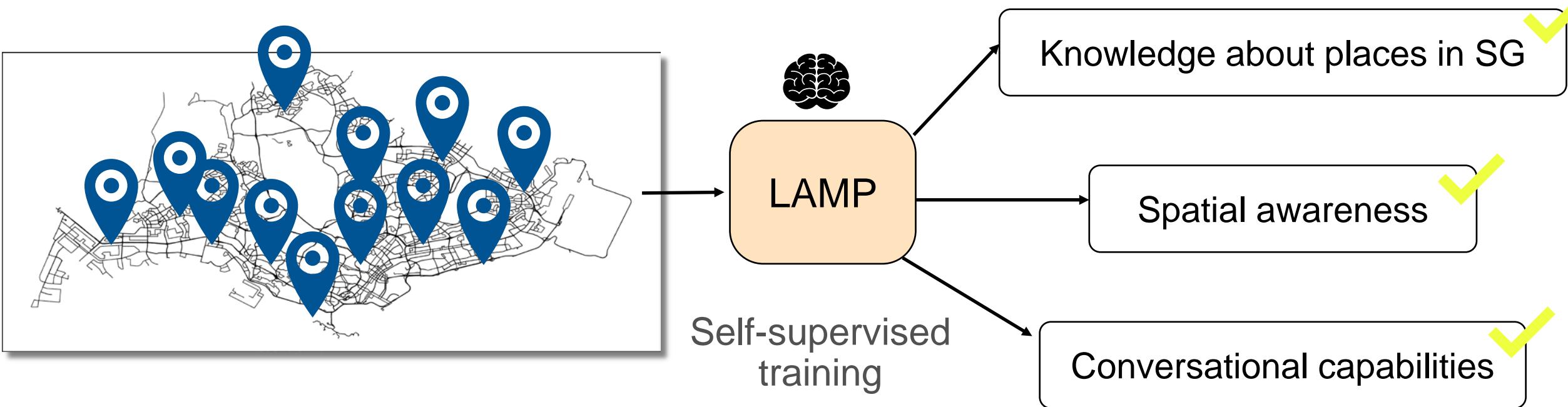
How to leverage the different data types and modalities in OSM, to
pre-train a geospatial FM?



Attemp2: UrbanCLIP – a prompting framework for zero-shot urban land use inference



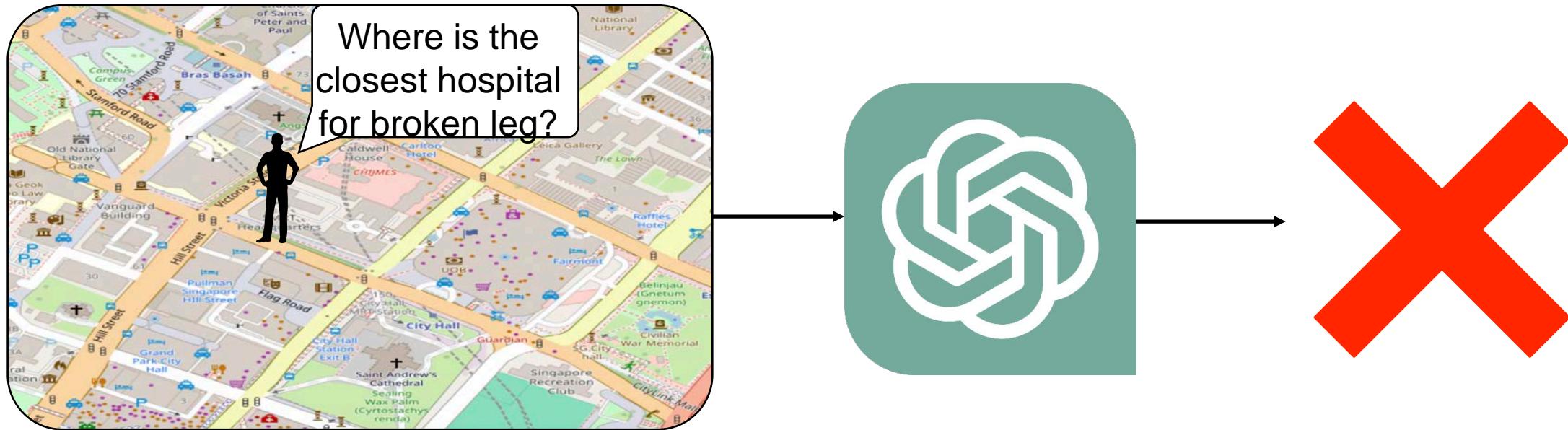
Attemp3: LAMP: A Language Model on the Map



We trained LAMP on Singapore places data, including POIs, addresses, opening hours, positions...

In this way, LAMP becomes aware of fine-grained urban places in Singapore, with a spatial sense (proximity)

LLM for conversational city search



...such general models lack **city-specific knowledge!**

LAMP: A Language Model on the Map

Current position: 81, Victoria Street, 188065, Singapore

Query: I'd like to have Japanese food. Could you find a place nearby?

LLaMa-2-70B-Chat: Sure, I'd be happy to help! There are several Japanese restaurants near Victoria Street, Museum, Singapore. Here are a few options:

1. Sushi Tei - Located at 100 Victoria Street, #01-01, Singapore 188064.
2. Tsukemen Ginza - Located at 111 Middle Road, #01-01, Singapore 188969.
3. Ramen Nagi - Located at 100 Victoria Street, #01-02, Singapore 188064.

ChatGPT 3.5: Sure, one nearby Japanese restaurant to 81 Victoria Street, Singapore 188065 is "Shin Minori Japanese Restaurant" located at 81 Victoria Street, #01-01, Singapore 188065 .

LAMP: I found two Japanese restaurants near your location:

- Shirokiya, located at 30 Victoria Street Chijmes, #01-05/06, 187996 ,
- Gyu Kaku Japanese BBQ, located at 30 Victoria Street, CHIJMES #01-01/03, 187996. It offers takeaway services and it is open until 11pm.

Let me know if you need any more recommendations!

When queried to find a specific POI, ChatGPT and LLaMa-2 are prone to hallucinating, and create places that do not exist, to satisfy the user's query, often replicating the address of the user or their zip code.

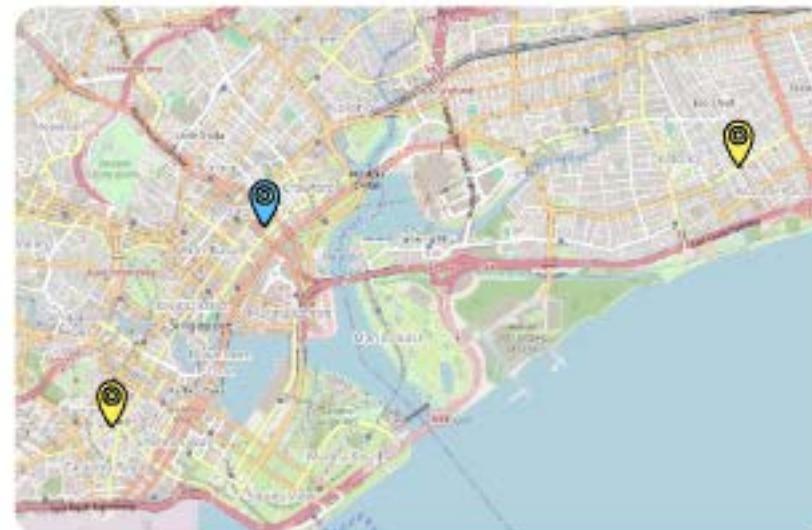
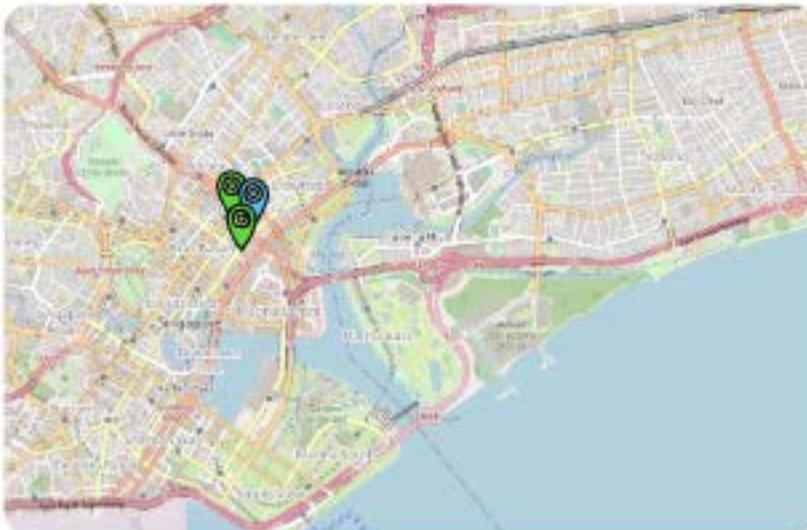
hallucinated information shaded in Orange, and correct information in Green .

LAMP: A Language Model on the Map

Current position: 100 High St, Singapore 179434

Query: Hi LAMP, we are planning a day out. We would like to find a nearby gym to exercise then have something at a bar for recommendations.

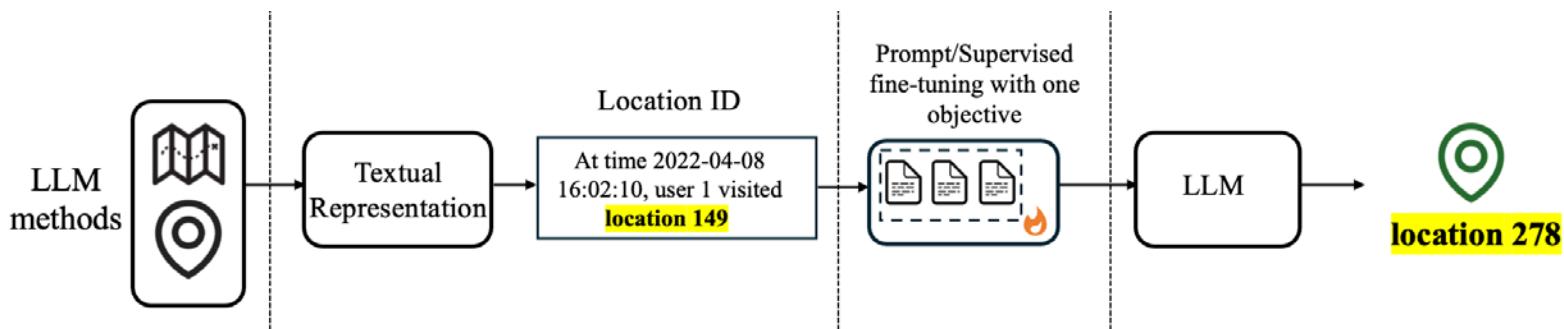
LAMP: Sure
1. First, visit Boulevard, #
2. Then, go to the floor of Marina Bay Sands at 039594. It is open 24 hours.
3. Finally, Fraser Street offers wheelchair accessibility and a cosy/c
Please let me know if you want more places.



The left places are recommended by LAMP, while the right ones are from GPT-4.

Attempt 3: QTMob: tokenization

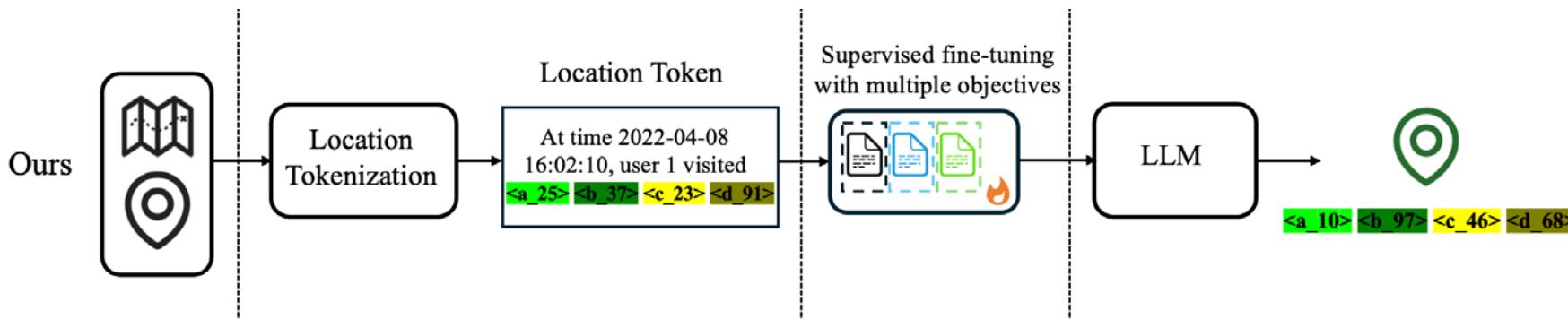
- LLMs naturally possess a comprehensive understanding of general geospatial knowledge. This has led to increasing interest in adapting LLMs for mobility data for downstream analytical tasks.
- Previous LLM-based paradigm:



- AgentMove (NAACL 2025), LLM-Move (arxiv 2024), LLM4POI (SIGIR 2024), etc.

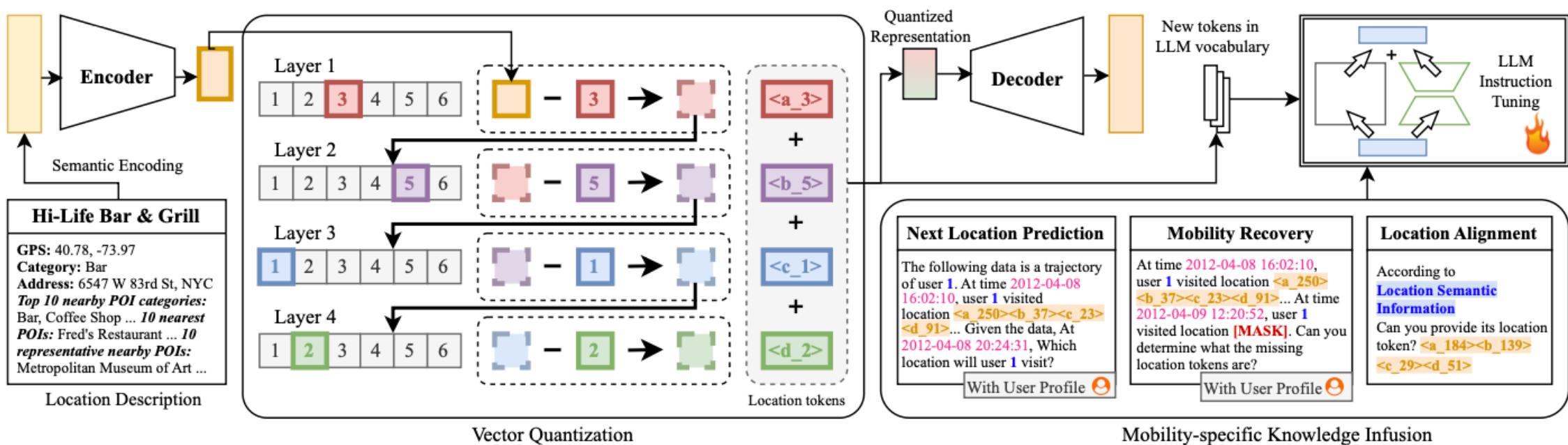
Introduction – QTMob

- Limitations
 - Locations as discrete IDs (e.g., “location 124”)
 - Fails to encode the semantic richness and contextual information of locations
 - LLMs are not fully adapted to mobility knowledge
 - Utilize prompt engineering without additional knowledge, or fine-tuning LLMs on limited mobility instruction dataset.
- Our New Paradigm:



Framework

- QT-Mob
 - Semantic Location Tokenization + Diverse instruction-tuning objectives
 - Discrete inputs with semantics
 - LLM alignment with fine-tuning



Experiments

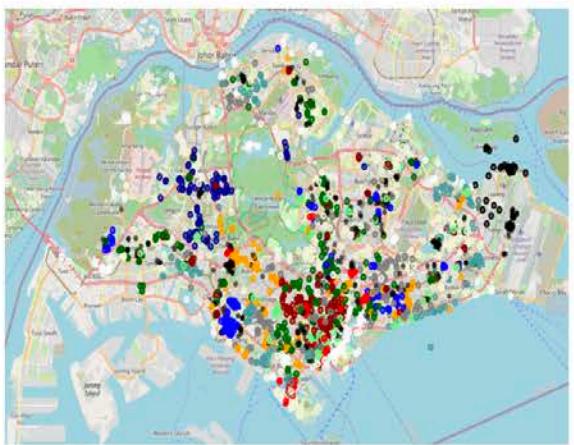
- Performance comparison

Next location prediction

Dataset	NYC						SG						CE					
	Model	Hit@1	Hit@5	Hit@10	N@5	N@10	Hit@1	Hit@5	Hit@10	N@5	N@10	Hit@1	Hit@5	Hit@10	N@5	N@10		
FPMC	0.0852	0.2216	0.2661	0.1577	0.1721	0.0489	0.0966	0.1287	0.0727	0.0830	0.0317	0.0856	0.1219	0.0590	0.0706			
DeepMove	0.1108	0.2105	0.2420	0.1635	0.1738	0.0520	0.1131	0.1484	0.0843	0.0956	0.0551	0.1424	0.1979	0.1000	0.1178			
CTLE	0.1367	0.3031	0.3812	0.2255	0.2473	0.0628	0.1560	0.2147	0.1350	0.1538	0.0728	0.1628	0.2205	0.1232	0.1458			
TrajFormer	0.0693	0.1544	0.1782	0.1283	0.1435	0.0321	0.0638	0.0980	0.0517	0.0634	0.0284	0.0547	0.0942	0.0468	0.0598			
GETNext	0.1542	0.3347	0.4066	0.2433	0.2668	0.0921	0.1960	0.2509	0.1447	0.1626	0.1029	0.2426	0.3098	0.1763	0.1980			
MCLP	0.1804	0.3662	0.4356	0.2791	0.3017	0.0839	0.1973	0.2590	0.1422	0.1622	0.0964	0.1929	0.2399	0.1468	0.1620			
PLSPL	0.1935	0.3751	0.4507	0.2894	0.3140	0.1055	0.2061	0.2541	0.1571	0.1726	0.0976	0.2261	0.2967	0.1645	0.1872			
STHGCN	0.1982	0.3821	0.4469	0.2970	0.3180	0.1246	0.2428	0.2957	0.1869	0.2041	0.0844	0.2045	0.2669	0.1462	0.1664			
LLM-Move	0.1717	0.3559	0.4397	0.2803	0.3071	0.1010	0.1961	0.2457	0.1513	0.1674	0.0823	0.2078	0.2657	0.1432	0.1629			
AgentMove	0.1753	0.3536	0.4377	0.2804	0.3077	0.1118	0.2046	0.2526	0.1558	0.1720	0.0836	0.2093	0.2695	0.1455	0.1650			
GenUP	0.2216	0.3550	0.3784	0.2946	0.3023	0.1227	0.2152	0.2375	0.1726	0.1798	0.0935	0.2008	0.2324	0.1502	0.1605			
MobilityLLM	0.1840	0.3944	0.4781	0.2949	0.3221	0.1077	0.2367	0.3045	0.1749	0.1967	0.1084	0.2623	0.3343	0.1883	0.2116			
QT-Mob	0.2550	0.4073	0.4562	0.3366	0.3525	0.1514	0.2629	0.3171	0.2105	0.2281	0.1243	0.2728	0.3418	0.2021	0.2245			

Experiments

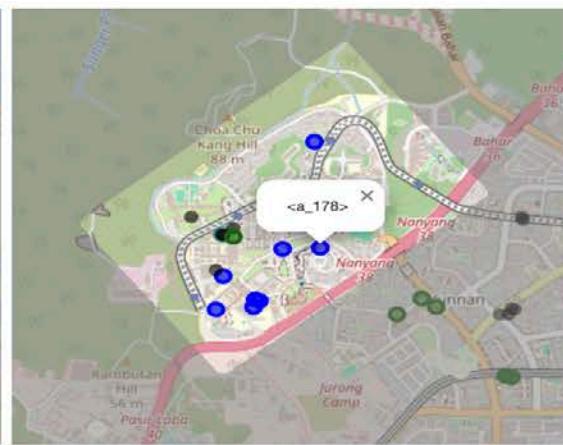
- Case studies reveal fascinating properties of the location tokens.



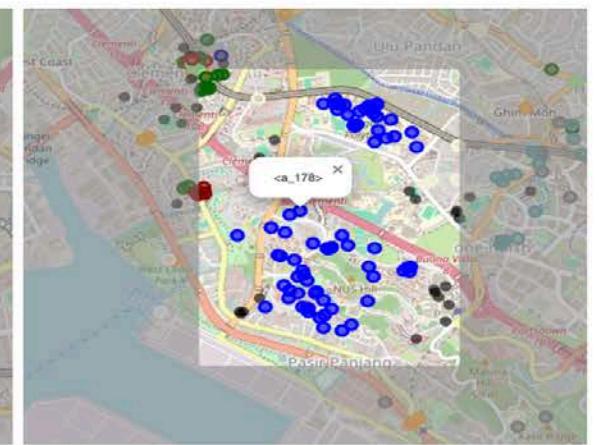
(a) Location tokens in SG



(b) Changi Airport

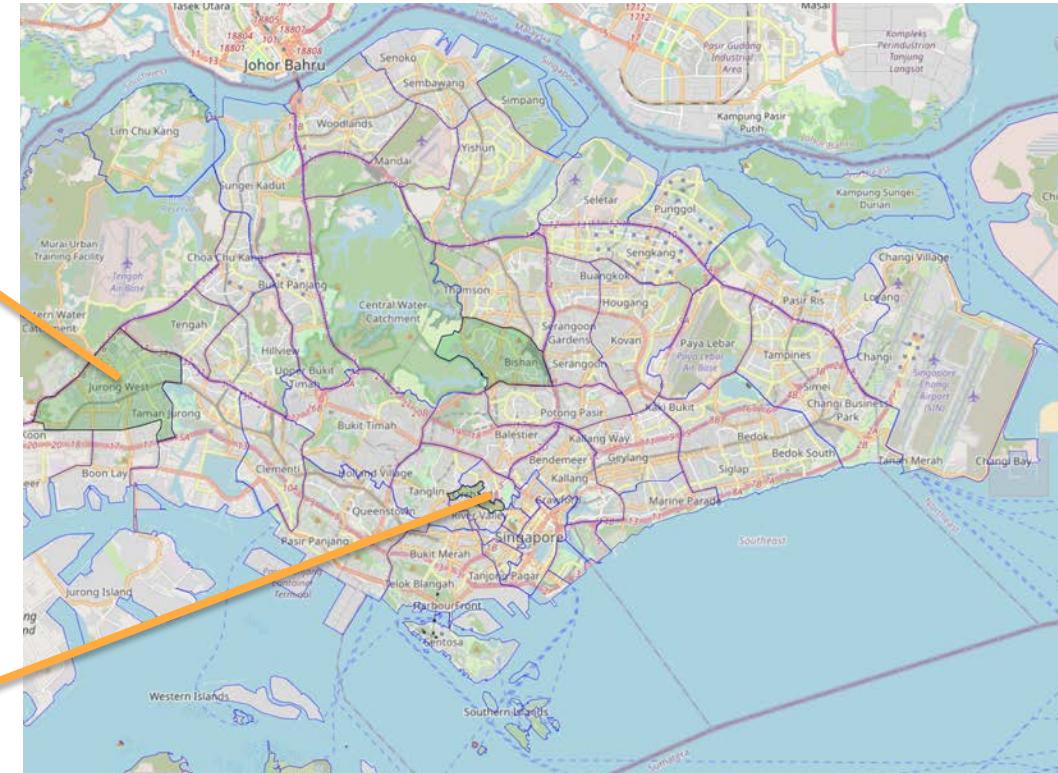
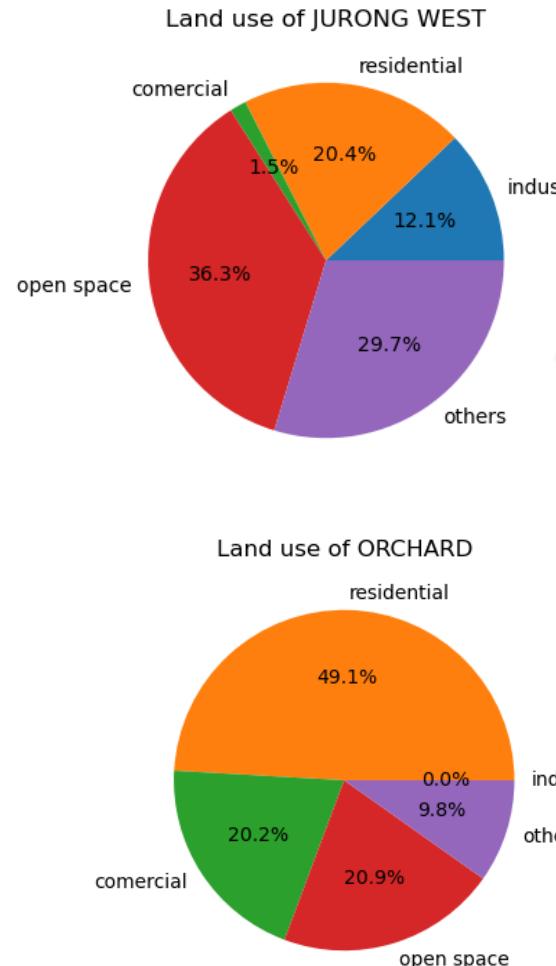


(c) University I



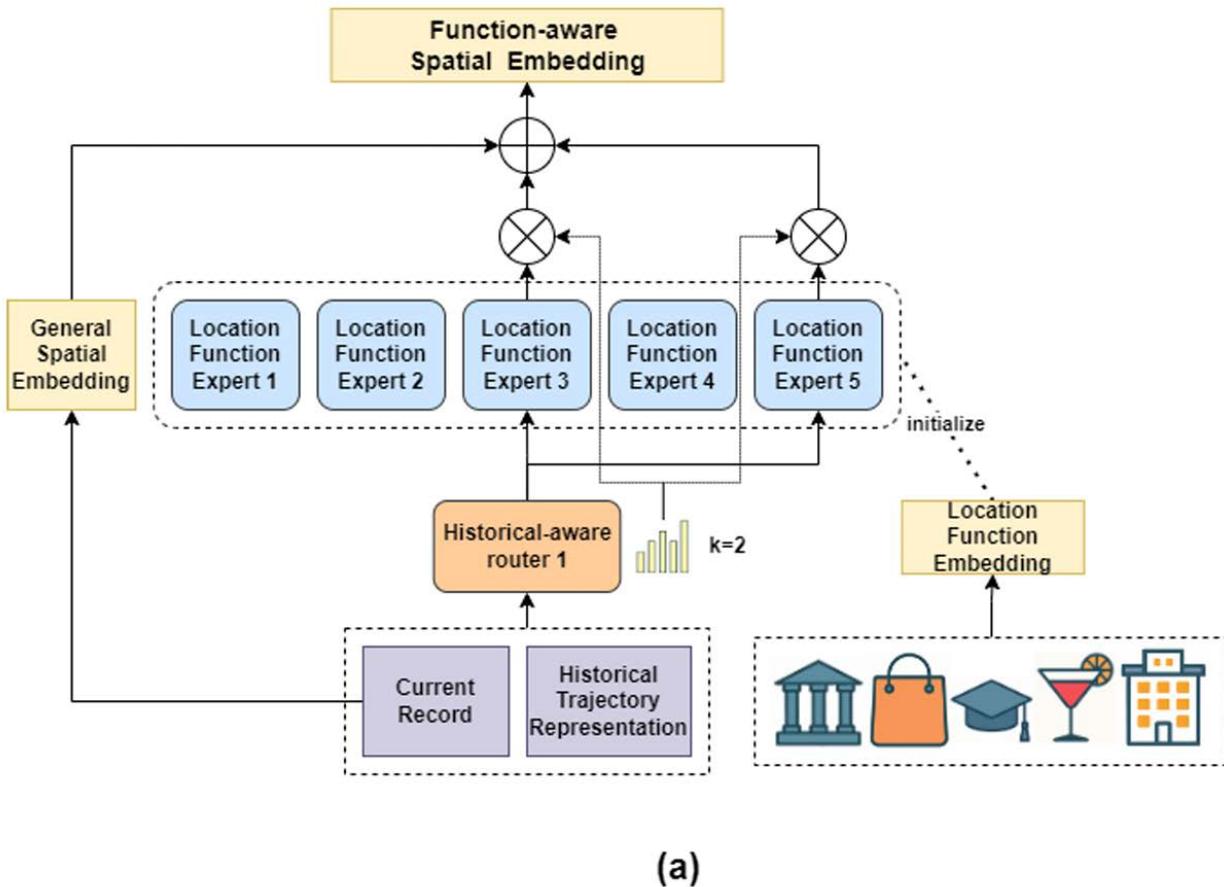
(d) University II

Attempt 3: NextLocMoE: semantic Mixture-of-Experts for Mobility Prediction



- One location may serve multiple semantic roles
- A single location embedding cannot capture such multi-functional semantics

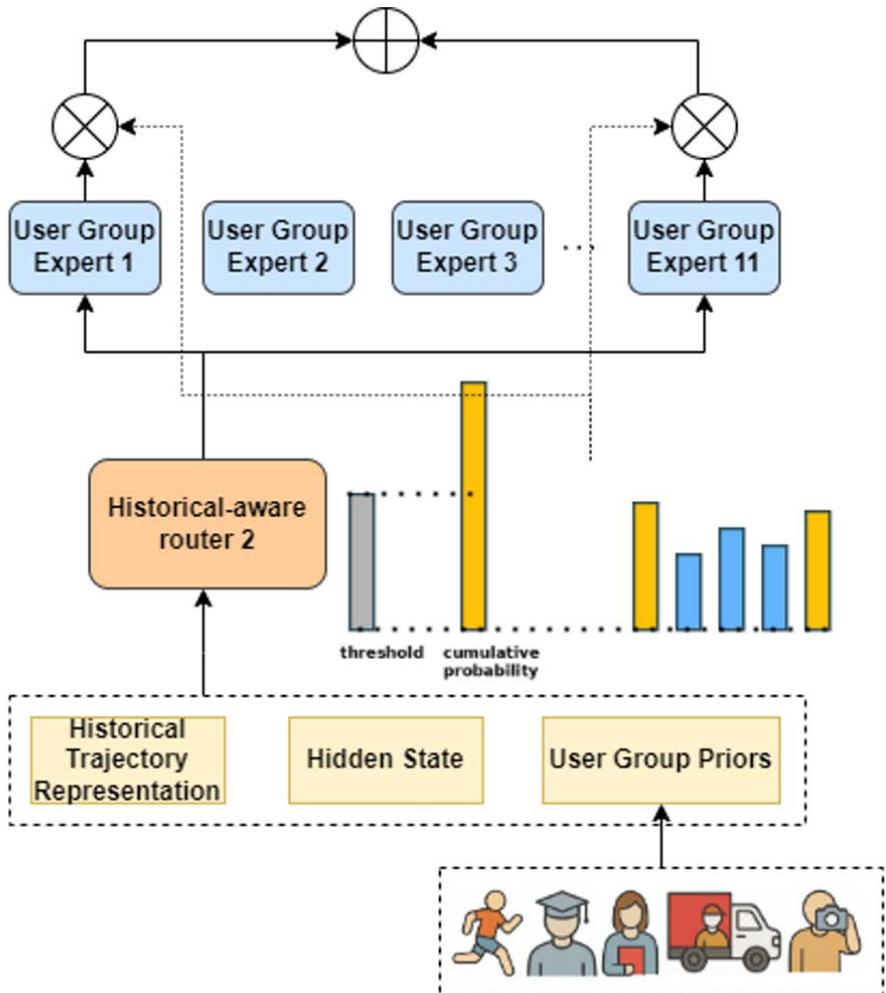
NextLocMoE: semantic Mixture-of-Experts for Mobility Prediction



Location Semantics MoE

- Enrich general spatial embedding with location function experts
- Each expert initialized by LLM-encoded function descriptions
- Historical-aware router selects top-k relevant experts

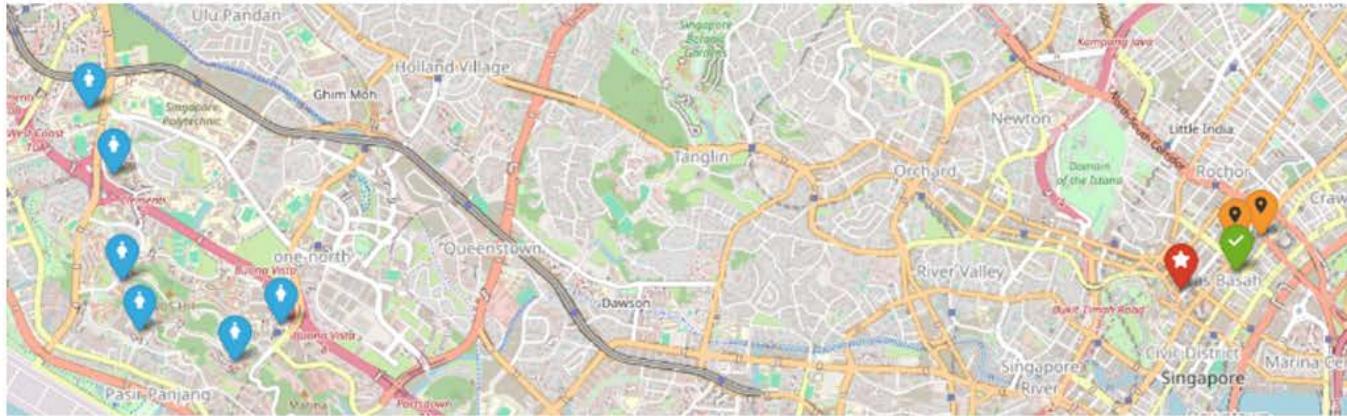
NextLocMoE: semantic Mixture-of-Experts for Mobility Prediction



Personalized MoE

- Model user heterogeneity with user-group experts
- LLM-encoded user group descriptions provide semantic priors
- Historical-aware router → threshold-based dynamic expert selection

NextLocMoE: semantic Mixture-of-Experts for Mobility Prediction



History



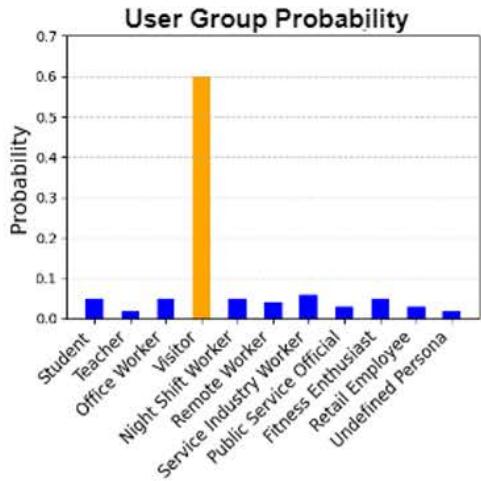
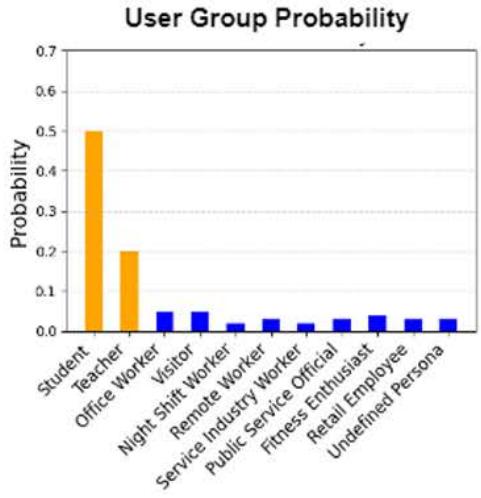
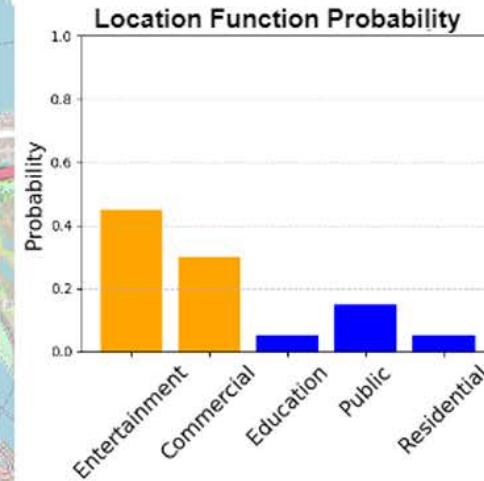
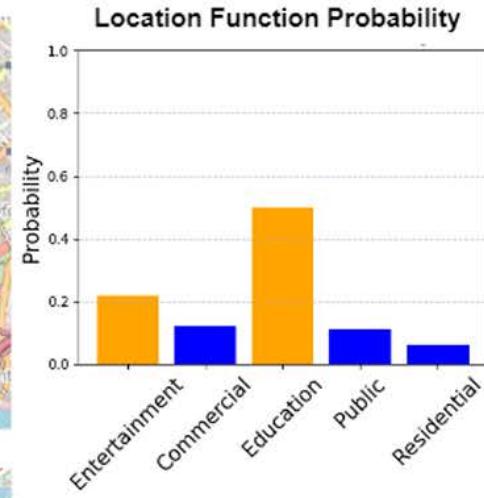
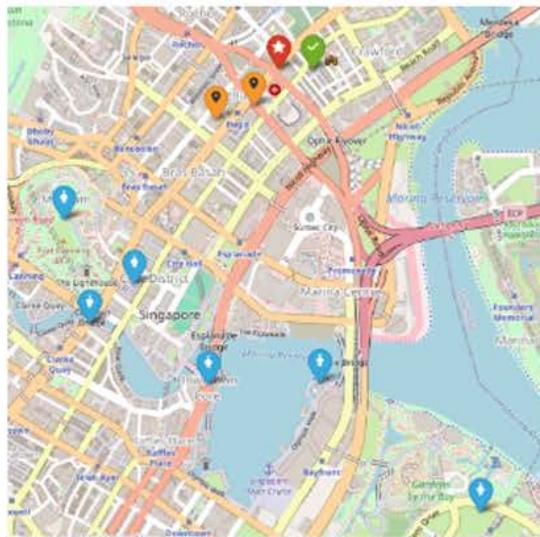
Current



GroundTruth

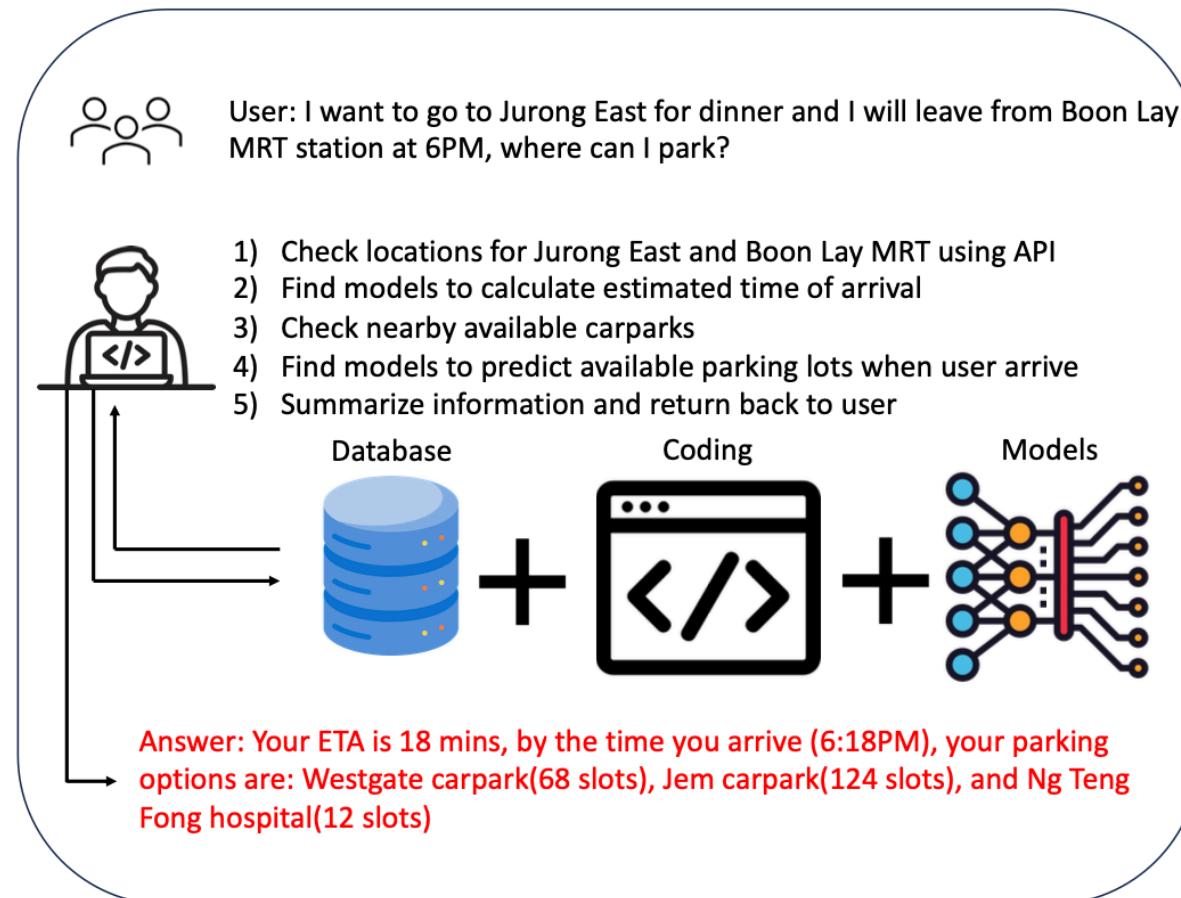


Prediction



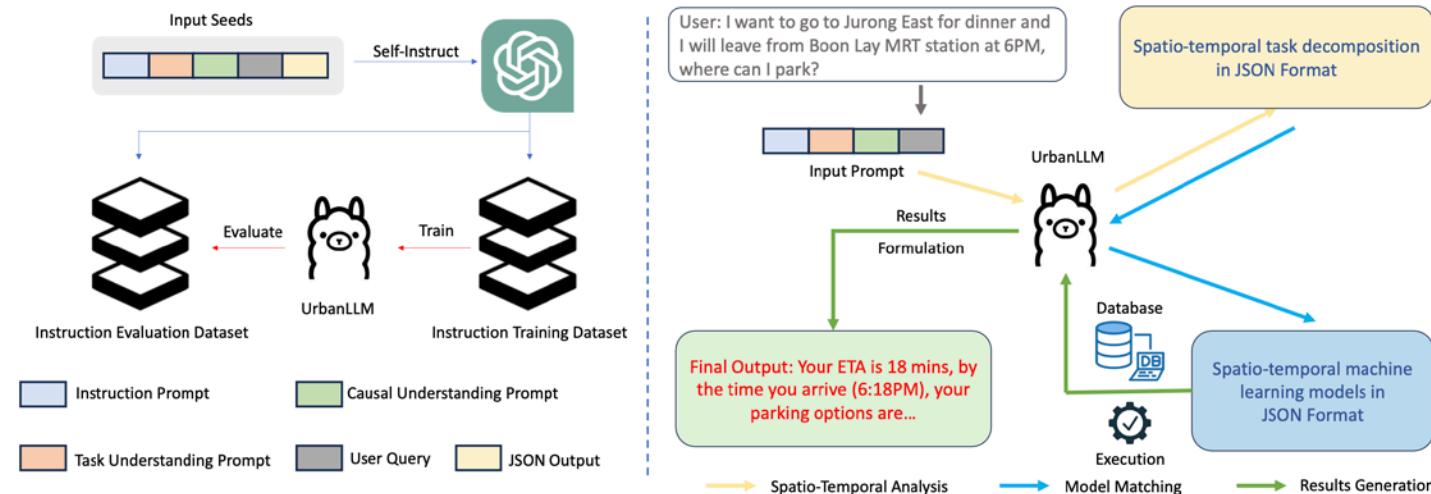
Attempt 4: Using AI agents

- Real-world Urban Activity Planning and Management problem are often complex
 - Involved multiple models
 - Need experienced software engineer and spatio-temporal domain experts



Framework

- UrbanLLM is a fine-tuned LLaMA-Series-based LLM designed to autonomously plan, manage, and reason about urban activities. Its core ideas are:
 - Instruction-Tuned Urban LLM
 - Fine-tuned on a large self-instruct dataset ($\approx 17K$ examples) curated from real Singapore urban scenarios.
 - Three-Stage Inference Pipeline
 - Spatio-Temporal Analysis: Decomposes a natural language query into structured JSON sub-tasks
 - Model Matching: Leverages a model zoo with 50+ specialized spatio-temporal AI models and selects the most suitable model for each sub-task
 - Results Generation: Executes the selected models, retrieves intermediate outputs, and composes a final, natural language response.



Experiments

📌 We have prepared 13 spatial/temporal sub-tasks and 34 combination problems of 1700 testing scenarios.

Table 1: Evaluation for Spatio-Temporal Task Analysis

	Accuracy	Precision	Recall	F1
Llama2-7b	0.18%	10.52%	8.75%	9.18%
Vicuna-7b-v1.5	8.44%	14.08%	13.89%	13.95%
Llama3-8b	5.31%	12.96%	15.50%	13.08%
GPT-3.5	17.95%	23.25%	22.35%	22.54%
GPT-4o	49.99%	55.31%	54.42%	54.63%
UrbanLLM	68.30%	80.05%	79.26%	79.49%
% Improve	36.63%	44.73%	45.64%	45.50%

Table 2: Evaluation for Spatio-Temporal Single-Task Analysis

	Accuracy	Precision	Recall	F1
Llama2-7b	0.47%	0.57%	0.57%	0.57%
Vicuna-7b-v1.5	33.26%	33.26%	33.26%	33.26%
Llama3-8b	15.46%	17.51%	21.19%	17.97%
GPT-3.5	13.58%	13.70%	13.74%	13.71%
GPT-4o	67.44%	68.56%	68.60%	68.57%
UrbanLLM	95.78%	96.78%	96.84%	96.80%
% Improve	42.02%	41.16%	41.17%	41.17%

Table 3: Evaluation for Spatio-Temporal Multi-Task Analysis

	Accuracy	Precision	Recall	F1
Llama2-7b	0.00%	13.80%	11.44%	12.01%
Vicuna-7b-v1.5	0.08%	7.62%	7.36%	7.45%
Llama3-8b	1.81%	11.36%	13.52%	11.37%
GPT-3.5	19.35%	26.40%	25.20%	25.45%
GPT-4o	40.13%	50.89%	49.68%	49.97%
UrbanLLM	59.08%	74.47%	73.40%	73.71%
% Improve	47.22%	46.33%	47.75%	47.51%

Open problems

- What are the new applications that LLMs can bring to Spatial/GIS?
 - Beyond the current applications
 - Beyond a chatbot
 - Benchmark datasets
- How to make LLMs to have spatial abilities?
 - Spatial data/computation?
 - Spatial inference ability
 - Data preparation
 - Finetune/retrain
- Agents/Agentic Spatial RAG
 - Interact with spatial databases

References and Acknowledgement

- **Shang Liu**, Gao Cong, **Kaiyu Feng**, Wanli Gu, Fuzheng Zhang. **Effectiveness Perspectives and a Deep Relevance Model for Spatial Keyword Queries**. SIGMOD 2023
- **Pasquale Balsebre**, Dezhong Yao, **Weiming Huang**, Gao Cong, Zhen Hai. **Mining Geospatial Relationships from Text**. SIGMOD 2023
- **Yile Chen**, **Xiucheng Li**, Gao Cong, Zhifeng Bao, et al.: **Robust Road Network Representation Learning: When Traffic Patterns Meet Traveling Semantics**. CIKM 2021
- **Yi Li**, **Weiming Huang**, Gao Cong, **Hao Wang**, and **Zheng Wang**. **Urban Region Representation Learning with OpenStreetMap Building Footprints**. SIGKDD 2023
- Gengchen Mai, et al. . **On the Opportunities and Challenges of Foundation Models for Geospatial Artificial Intelligence**. ACM TSAS 2024
- **Weiming Huang**, Jing Wang, Gao Cong. **Zero-shot urban function inference with street view images through prompting a pre-trained vision-language model**. IJGIS 2024.
- **Pasquale Balsebre**, **Weiming Huang**, Gao Cong. **LAMP: A Language Model on the Map**. On arXiv.
- **Pasquale Balsebre**, **Weiming Huang**, Gao Cong, **Yi Li**. **Towards City Foundation Models**. CIKM'24
- **Yue Jiang**, Qin Chao, **Yile Chen**, **Xiucheng Li**, **Shuai Liu**, Gao Cong, UrbanLLM: **Autonomous Urban Activity Planning and Management with Large Language Models**. EMNLP findings 2024
- **Yile Chen**, **Yicheng Tao**, **Yue Jiang**, **Shuai Liu**, Han Yu, Gao Cong. “**Enhancing Large Language Models for Mobility Analytics with Semantic Location Tokenization**, KDD'25
- **Shuai Liu**, Ning Cao, **Yile Chen**, **Yue Jiang**, Gao Cong. **Mixture-of-Experts for Personalized and Semantic-Aware Next Location Prediction**. On arXiv

Thanks my
students,
postdoc and
collaborators

Our group website:
<https://stdataminer.github.io/>