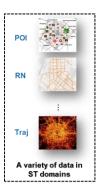


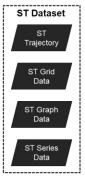
## Summary

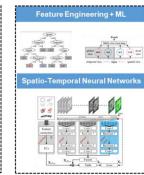


#### Future directions

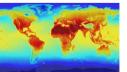
- Novel applications: big ST data + rich scenarios
- Efficient AI for large-scale ST Applications
- Explainable AI for ST Data Science
- Exploring LLMs for ST data
- Uncertainty-Aware Generative AI for ST data





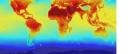






Climate

Social Science







Environment

**Epidemiology** 

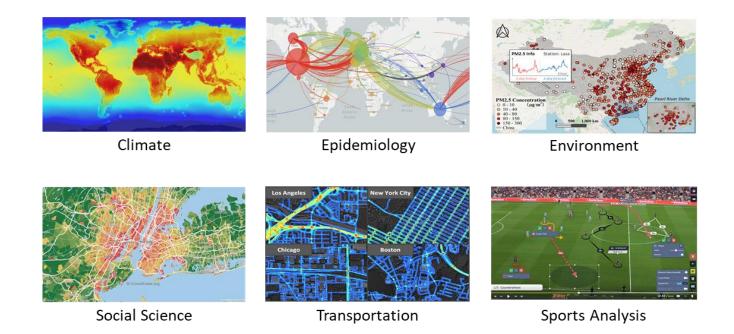


Transportation

Sports Analysis



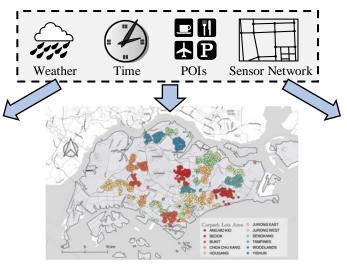
 Novel interdisciplinary research: big ST data + rich scenarios + domain knowledge





- Efficient AI for large-scale ST Applications
  - Produce new benchmarks for the community: large-scale ST data + external factors
  - Designing scalable AI methods

#### **Cross-domain data**



Carpark availability data in							
Singapore (	~2,000 carparks)						

Datasets	#Sensors	#Edges	
PeMSD7(M)	228	1132	
PeMSD7(L)	1026	10150	
PeMS03	358	547	
PeMS04	307	340	
PeMS07	883	866	
PeMS08	170	295	

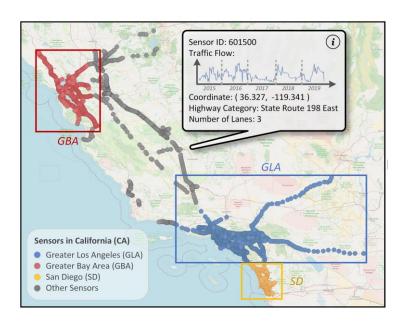


**Traffic speed/flow** in California (4,000+ sensors)

## LargeST: a Large-Scale Traffic Dataset



- We crawled, processed, and released a large-scale traffic dataset
  - Scalability, Generalizability, Robustness, etc.



Dataset	Nodes	Edges	Degree	Meta	Time Range	Frames	Data Points
PeMSD7(M)	228	1,664	7.3	6	05/01/2012 - 06/30/2012	12,672	2.89M
PeMSD7(L)	1,026	14,534	14.2	0	05/01/2012 - 06/30/2012	12,672	13.00M
METR-LA	207	1,515	7.3	3	03/01/2012 - 06/27/2012	34,272	7.09M
PEMS-BAY	325	2,369	7.3	3	01/01/2017 - 06/30/2017	52,116	16.94 <mark>M</mark>
PEMS03	358	546	1.5	1	09/01/2018 - 11/30/2018	26,208	9.38M
PEMS04	307	338	1.1	0	01/01/2018 - 02/28/2018	16,992	5.22M
PEMS07	883	865	1.0	0	05/01/2017 - 08/06/2017	28,224	24.92M
PEMS08	170	276	1.6	0	07/01/2016 - 08/31/2016	17,856	3.04M
CA	8,600	201,363	23.4	9	01/01/2017 - 12/31/2021	525,888	4.52B
GLA	3,834	98,703	25.7	9	01/01/2017 - 12/31/2021	525,888	2.02B
GBA	2,352	61,246	26.0	9	01/01/2017 - 12/31/2021	525,888	1.24B
SD	716	17,319	24.2	9	01/01/2017 - 12/31/2021	525,888	0.38 <mark>B</mark>





#### LaDe Dataset

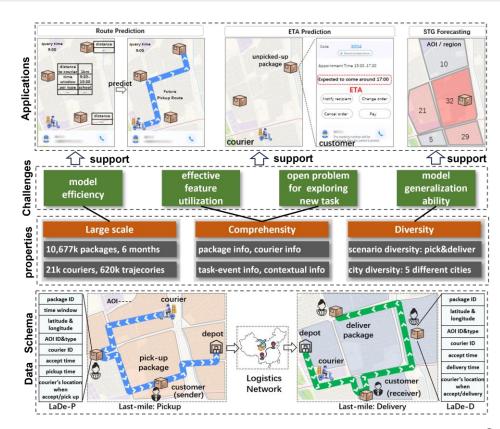


- We released The First Last-Mile Delivery Dataset from Industry
  - Large scale
  - Comprehensity
  - Diversity



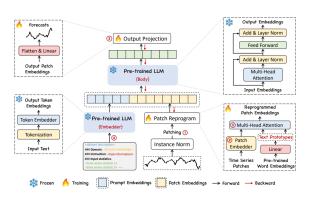


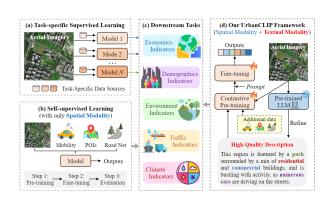


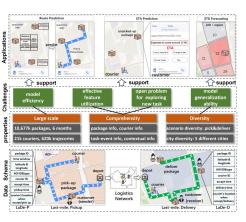




 Investigating Large Language Models (LLMs) for learning human mobility data



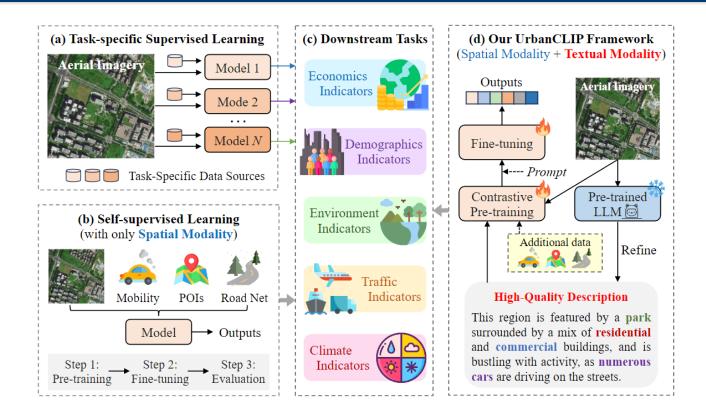




Time-LLM Urban-LLM Logistic-LLM

### Urban LLM

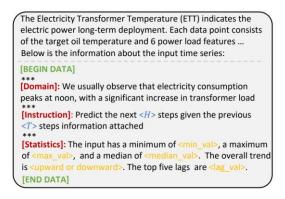


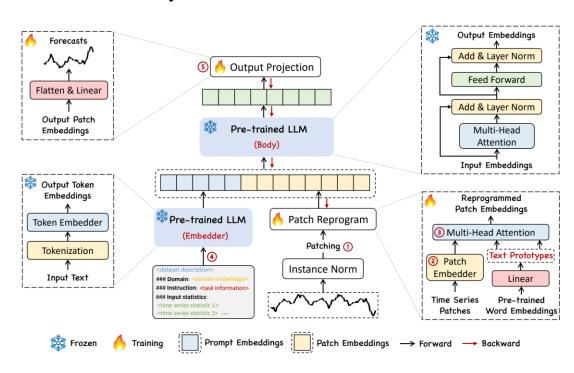


### Time-LLM: LLM for Time Series



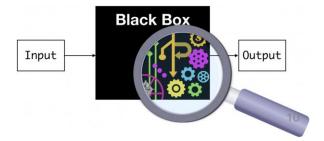
Time-LLM supports general time series analysis







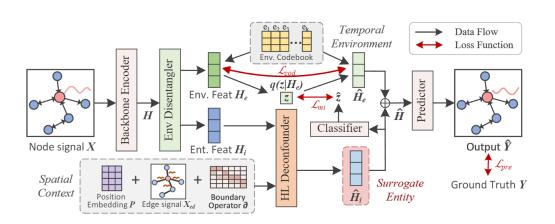
- Novel Interdisciplinary Research: big ST data + rich scenarios + domain knowledge
- Efficient AI for large-scale ST Applications
  - Produce new benchmarks for the community: large-scale ST data + external factors
  - Designing scalable AI methods
- Leveraging the power of Large Language Models for ST data
- Explainable AI for ST Data Science, e.g., causal inference

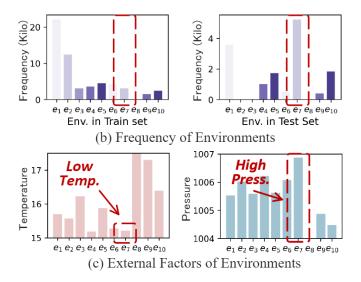


### **Future Trends**



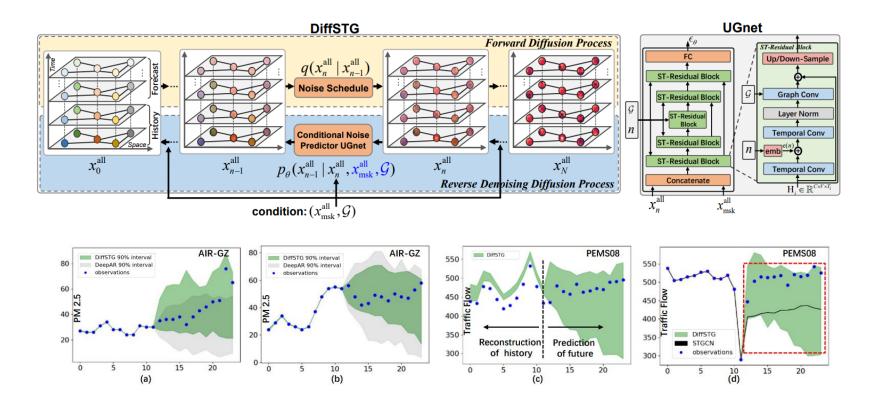
- Explainable AI for ST Data Science
  - Causal inference
  - Gradient-based approaches





# Generative Modeling (DDPM) for ST Graphs





12

# Generative Modeling (DDPM) for Trajectories



- Using real-world human trajectories usually has privacy concerns
  - Protect users' privacy
  - DDPM outperforms GAN and VAE





 $\begin{array}{c}
q(x_t|x_{t-1}, x_0^{co}) \\
\hline
p_{\theta}(x_{t-1}^{\overline{s}}|x_t^{\overline{s}}, x_0^{\overline{co}})
\end{array}$ 





Same origin and destination

Different origin and destination

# Project Presentation



- Each presentation should consist of an 15-minute presentation followed by a 5-minute QA session, totally of 20 minutes per project
- Each questioner may be awarded additional scores
- Orders
  - Team 1: Jiaxi Hu, Yongzi Yu
  - Team 2: Gangyong Zhu, Jiahui Llang
  - Team 3: Yixuan Wang, Ruiguo Zhong
  - Break time
  - Team 4: Pei Liu, Jingtao He
  - Team 5: Zhixiong Wang, Tianyu Wei
  - Team 6: Yongkai Gao
  - Team 7: Qiongyan Wang, Weilin Ruan



















