The KDD Tutorial on FM4ST

Foundation Models for Spatio-Temporal Data Science

Yuxuan Liang
Assistant Professor
香港科技大学(广州)
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Qingsong Wen
Head of AI
Squirrel Ai Learning

August 4th, 2025, Toronto, Canada

Held in conjunction with the 31st ACM SIGKDD 2025

Organizers

Yuxuan Liang
HKUST (Guangzhou)

Haomin Wen
Carnegie Mellon University

Yutong Xia
National University of Singapore

Ming Jin
Griffith University

Bin Yang
East China Normal University

Flora Salim
UNSW

Qingsong Wen
Squirrel AI, USA

Shirui Pan
Griffith University

Gao Cong
Nanyang Technological University

香港科技大学(广州) THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU) · Griffith UNIVERSITY · Squirrel Ai Learning · UNSW SYDNEY · Carnegie Mellon University · NUS National University of Singapore · NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Schedule

Time	Talk	Speaker
8:00-9:30	**Background of FM/LLM for TS/ST Data** **FM/LLM for Time Series Data**	Qingsong Wen Squirrel Ai Learning
9:30-10:00	Coffee Break	-
10:00-11:00	**When Foundation Models Meets Spatio-Temporal Data**	Yuxuan Liang HKUST (Guangzhou)

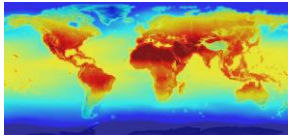
Schedule

Time	Talk	Speaker
8:00-9:30	**Background of FM/LLM for TS/ST Data** **FM/LLM for Time Series Data**	Qingsong Wen Squirrel Ai Learning
9:30-10:00	Coffee Break	-
10:00-11:00	When Foundation Models Meets Spatio-Temporal Data	Yuxuan Liang HKUST (Guangzhou)

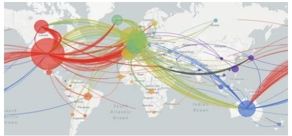
What is Time Series (TS) Data?

- With recent advances in sensing technologies, a myriad of Time Series (TS) Data has been collected and contributed to various disciplines

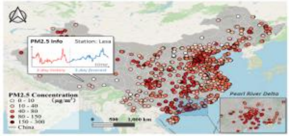
- Time series is a sequence of data points collected or recorded at specific time intervals, showing how a variable changes over time



Climate



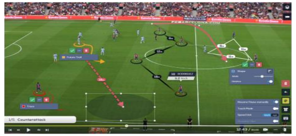
Epidemiology



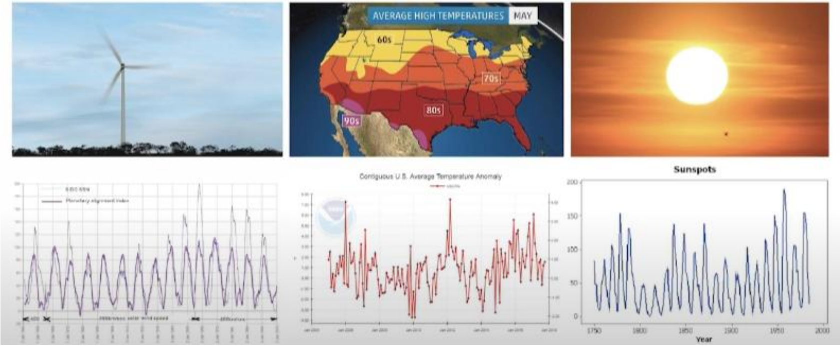
Environment

Social Science

Transportation

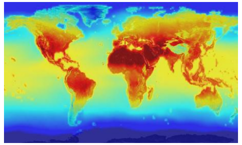


Sports Analysis



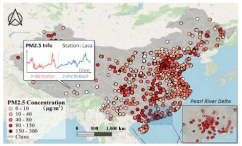
What is Spatio-Temporal (ST) Data?

- With recent advances in sensing technologies, a myriad of Spatio-Temporal Data has been collected and contributed to various disciplines



Climate

Epidemiology

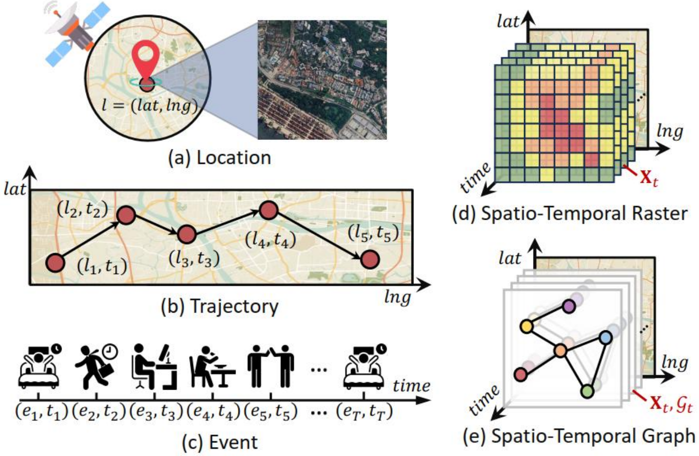


Environment

Social Science

Transportation

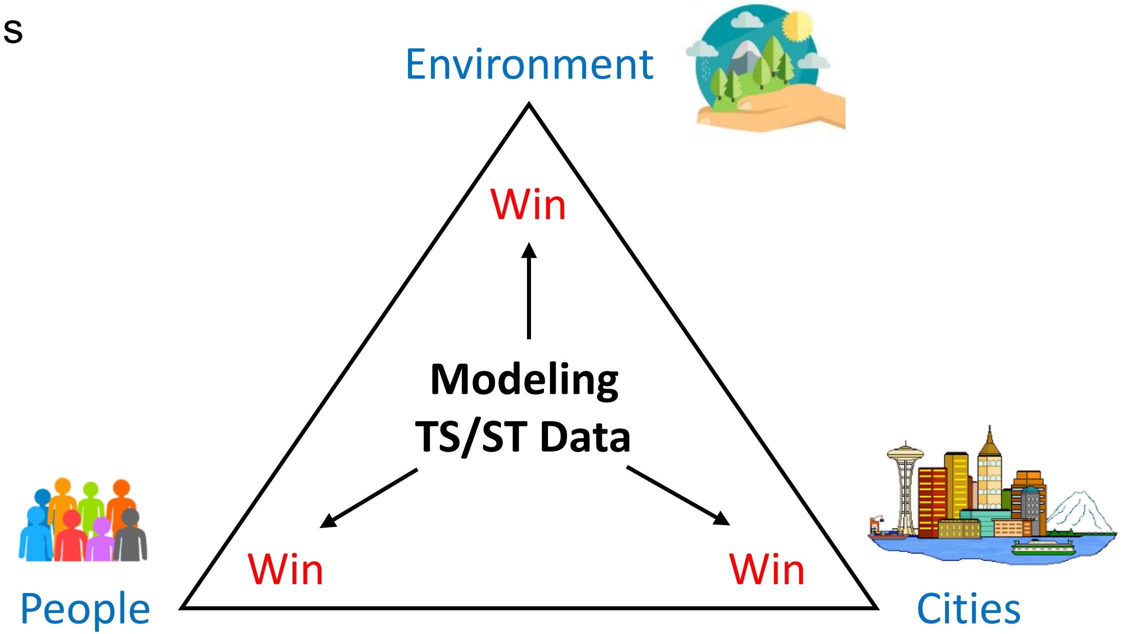
Sports Analysis



Time, Location, Event

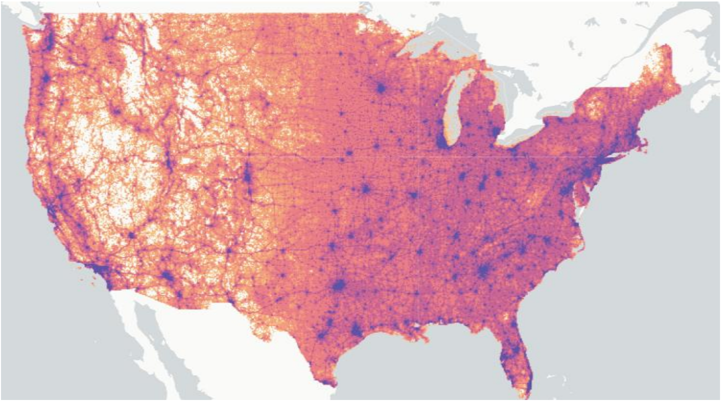
Motivation

- Modeling ST/ST data is the foundation of many real-world applications with high social impacts.
- Creating win-win-win solutions that improve the environment, human life quality, and city operation systems



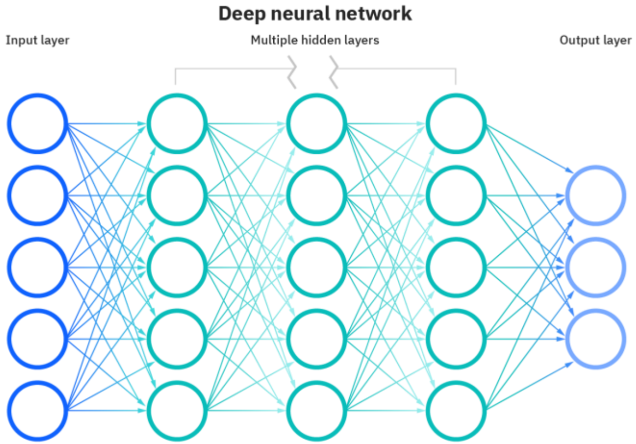
Time Series & Spatio-Temporal Data Mining (TS/ST DM)

- TS/ST DM aims to harness the power of ST data with DM/AI techniques by
  - transforming them into **useful knowledge**, i.e., representations
  - leveraging the extracted knowledge to support a diversity of real-world, such as AIOps, energy system, intelligent transportation, smart environment, etc.



Big TS/ST data

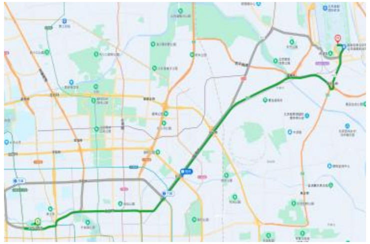
+



DM/AI

Our ST/ST Methodologies & Applications

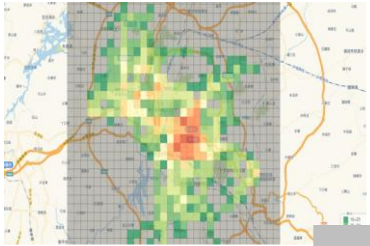
Modeling ST Trajectory



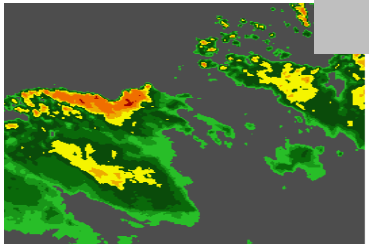
ControlTraj [KDD’24]

TrajODE [IJCAI’21]

Modeling ST Grid Data

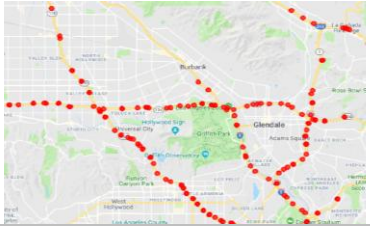


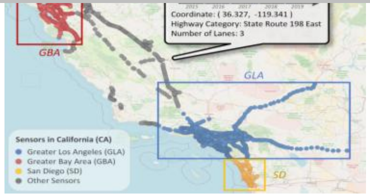
PhysicNet [TKDE’23



NuwaDynamic [ICLR’24]

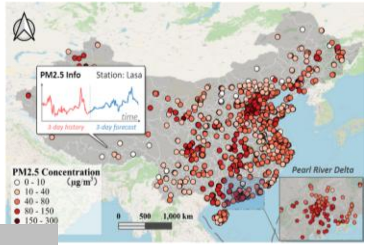
Modeling ST Graphs



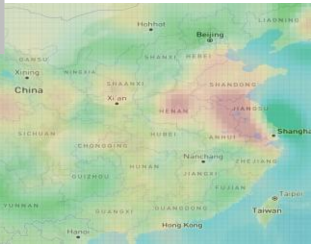


LargeST [NeurIPS’23]

Modeling ST Series



AirFormer [AAAI’23]

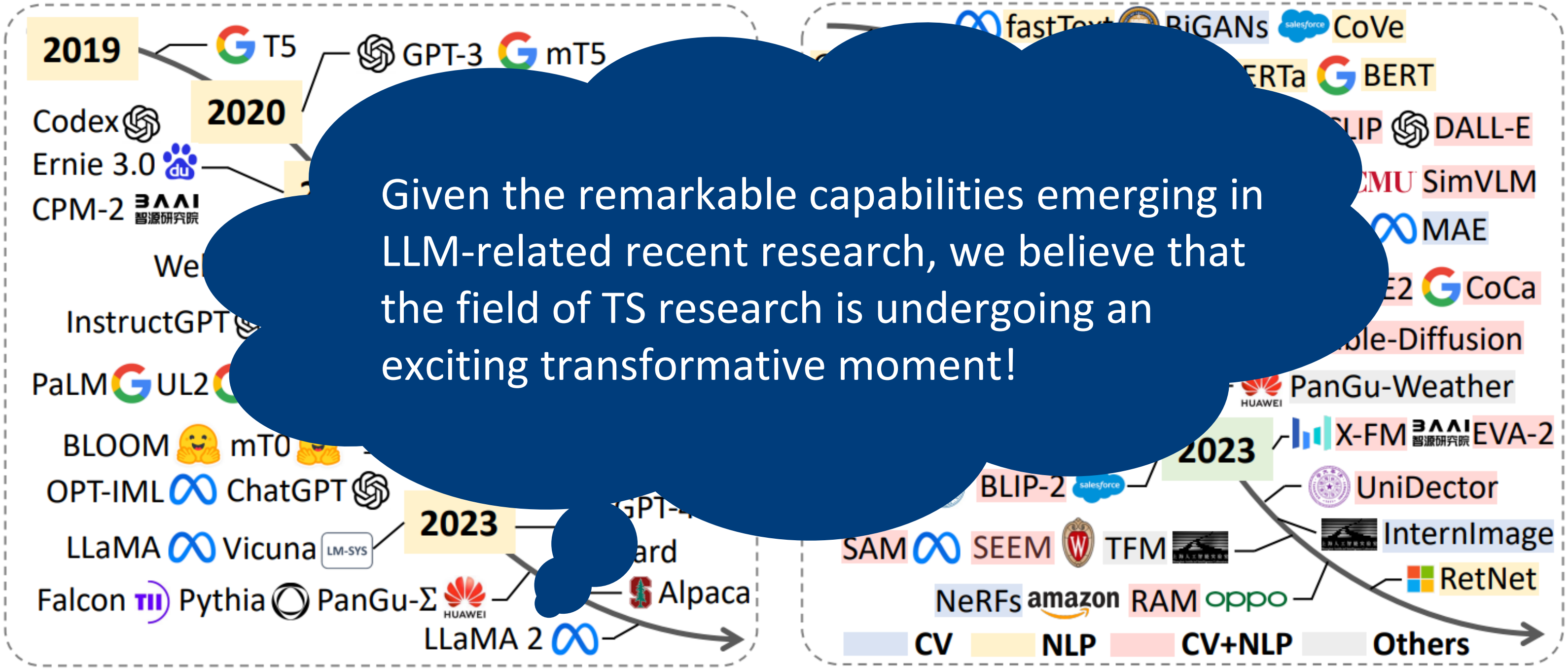


STGNP [KDD’23]

Multivariate Time Series

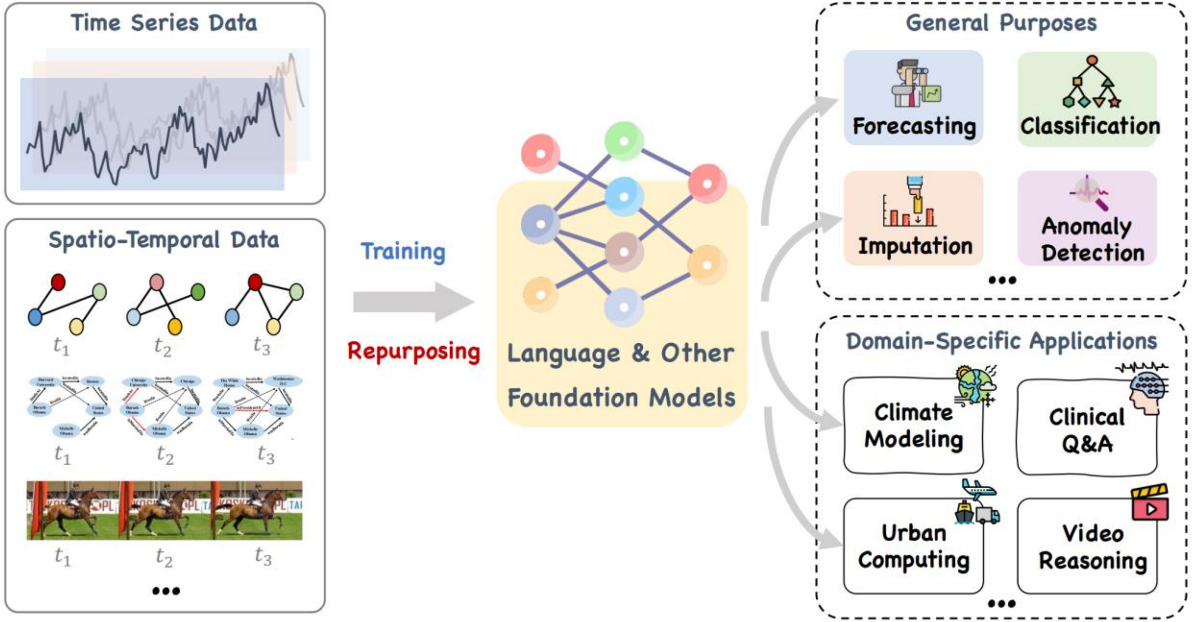
Roadmap of LLMs

- LLMs and Foundation Models



Towards General Intelligence for TS/ST

- LLMs can be either trained or adeptly repurposed to handle TS data for a range of general-purpose tasks and specialized domain applications.



M. Jin et al., Large Models for Time Series and Spatio-Temporal Data: A Survey and Outlook. arXiv 2023.

Frontier Approaches: FM and LLM for TS Analysis

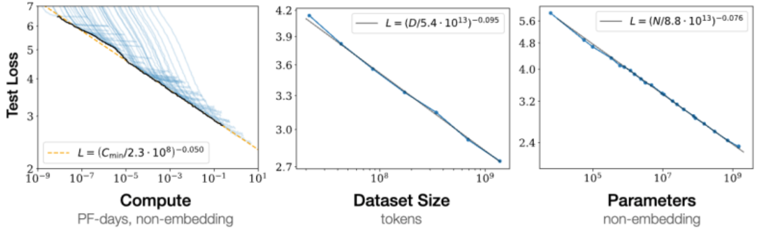
- **FM for Time Series Forecasting**
  - **Train from scratch** based on time series data
  - E.g.: **Time-MoE**, Moirai, TimesFM, Chronos, Moment,...
- **LLM for Time Series Forecasting**
  - Leverage/repurpose **LLM** for time series forecasting
  - E.g.: **Time-LLM**, LLMTime, AutoTimes, OFA, UniST, ...
- **LLM for Multi-Task Time Series Analysis**
  - Leverage **LLM** for general multi-task time series analysis
  - E.g.: **Time-MQA**

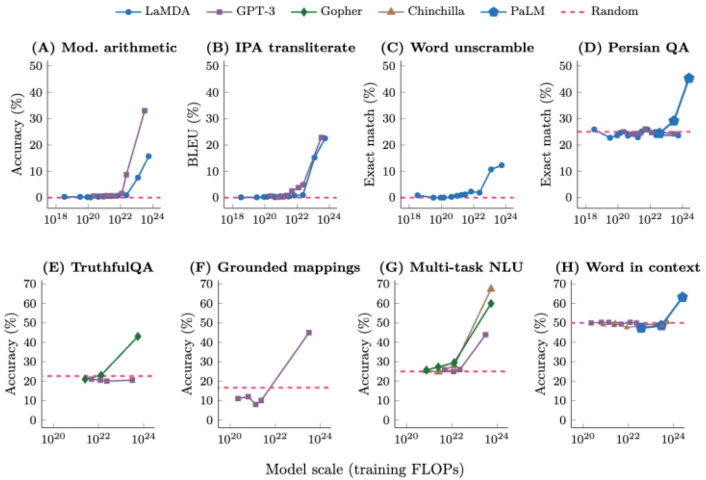
[1] Yuxuan Liang, Haomin Wen, Yuqi Nie, Yushan Jiang, Ming Jin, Dongjin Song, Shirui Pan, Qingsong Wen*, **"Foundation Models for Time Series Analysis: A Tutorial and Survey", KDD 2024**.

[2] Ming Jin, Qingsong Wen*, Yuxuan Liang, Chaoli Zhang, Siqiao Xue, Xue Wang, James Zhang, Yi Wang, Haifeng Chen, Xiaoli Li, Shirui Pan, Vincent S. Tseng, Yu Zheng, Lei Chen, Hui Xiong, **"Large Models for Time Series and Spatio-Temporal Data: A Survey and Outlook", arXiv 2023**.

FM for Time Series: Motivation

- Scaling Laws & Capabilities in LLM





How about Time Series ?

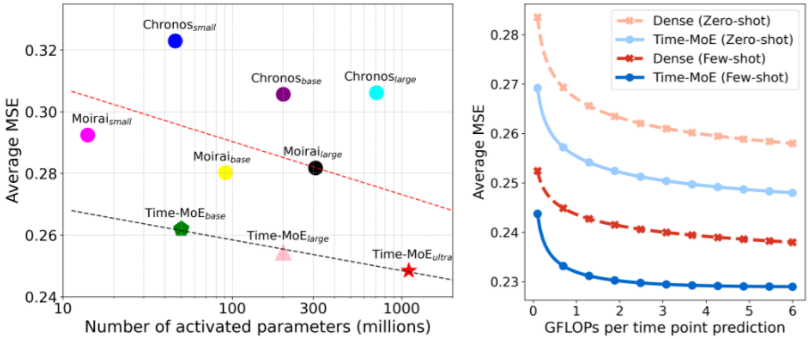


Figure 1: Performance overview. (**Left**) Comparison between TIME-MOE models and state-of-the-art time series foundation models, reporting the average zero-shot performance across six benchmark datasets. (**Right**) Comparison of few- and zero-shot performance between TIME-MOE and dense variants, with similar effective FLOPs per time series token, across the same six benchmarks.

➢ time series forecasting benefits from the scaling laws

Jared Kaplan, et al. "Scaling laws for neural language models."arXiv:2001.08361, 2020.
Jason Wei, et al., Emergent Abilities of Large Language Models. *TMLR*, 2022.

Time-300B, Time-MoE

Key statistics of the pre-training dataset Time-300B from various domains.

	Energy	Finance	Healthcare	Nature	Sales	Synthetic	Transport	Web	Other	Total
# Seqs.	2,875,335	1,715	1,752	31,621,183	110,210	11,968,625	622,414	972,158	40,265	48,220,929
# Obs.	15.981 B	413.696 K	471.040 K	279.724 B	26.382 M	9.222 B	2.130 B	1.804 B	20.32 M	309.09 B
%	5.17 %	0.0001%	0.0001%	90.50 %	0.008 %	2.98%	0.69 %	0.58 %	0.006 %	100%

A high-level summary of TIME-MOE model configurations.

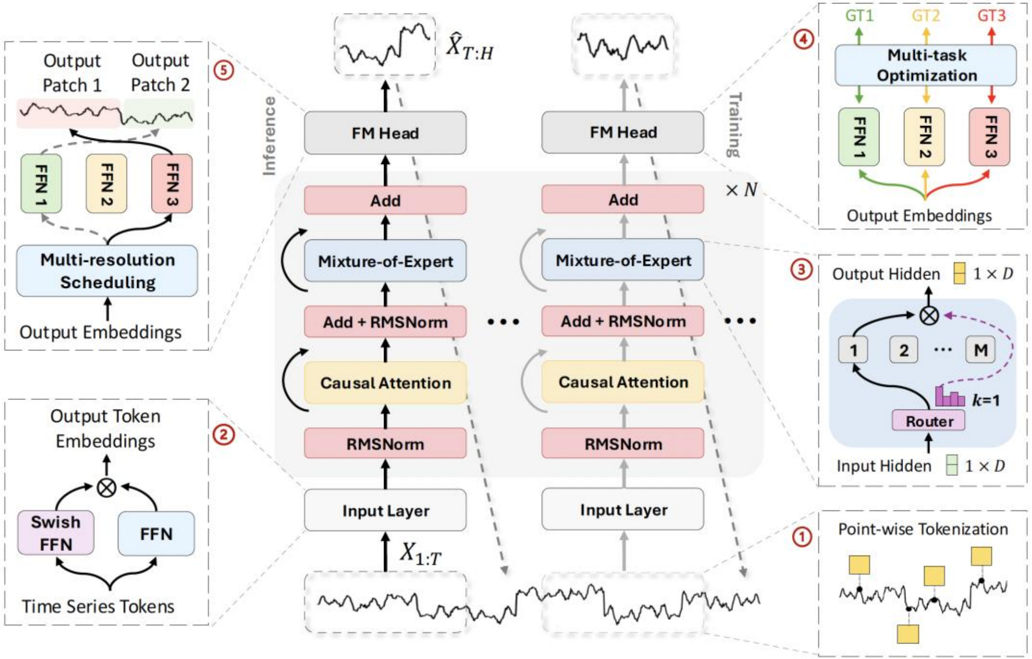
	Layers	Heads	Experts	K	d_{model}	d_{ff}	d_{expert}	Activated Params	Total Params
TIME-MOE$_{base}$	12	12	8	2	384	1536	192	50 M	113 M
TIME-MOE$_{large}$	12	12	8	2	768	3072	384	200 M	453 M
TIME-MOE$_{ultra}$	36	16	8	2	1024	4096	512	1.1 B	2.4 B

- Time-300B: the largest open-access time series data collection
- Time-MoE: the first work to scale time series foundation models up to 2.4 billion parameters

Shi, Xiaoming, Shiyu Wang, Yuqi Nie, Dianqi Li, Zhou Ye, Qingsong Wen*, and Ming Jin*. "Time-MoE: Billion-Scale Time Series Foundation Models with Mixture of Experts." *ICLR 2025*.

Time-MoE: Architecture

$$
\begin{aligned}
\mathbf{h}_t^0 &= \mathrm{SwiGLU}(x_t) = \mathrm{Swish}\left(W x_t\right) \otimes \left(V x_t\right) \\
\mathbf{u}_t^l &= \mathrm{SA}\left(\mathrm{RMSNorm}\left(\mathbf{h}_t^{l-1}\right)\right) + \mathbf{h}_t^{l-1}, \\
\bar{\mathbf{u}}_t^l &= \mathrm{RMSNorm}\left(\mathbf{u}_t^l\right), \\
\mathbf{h}_t^l &= \mathrm{Mixture}\left(\bar{\mathbf{u}}_t^l\right) + \mathbf{u}_t^l.
\end{aligned}
$$



The architecture of TIME-MOE, which is a decoder-only model. Given an input time series of arbitrary length, ① we first tokenize it into a sequence of data points, ② which are then encoded. These tokens are processed through N-stacked backbone layers, primarily consisting of causal multi-head self-attention and ③ sparse temporal mixture-of-expert layers. During training, ④ we optimize forecasting heads at multiple resolutions. For model inference, TIME-MOE provides forecasts of flexible length by ⑤ dynamically scheduling these heads.

MoE Structure (-> efficiency and capacity)

$$
\mathrm{Mixture}\left(\bar{\mathbf{u}}_t^l\right) = g_{N+1,t}\,\mathrm{FFN}_{N+1}\left(\bar{\mathbf{u}}_t^l\right) + \sum_{i=1}^{N}\left(g_{i,t}\,\mathrm{FFN}_i\left(\bar{\mathbf{u}}_t^l\right)\right)
$$

$$
g_{i,t} = \begin{cases} s_{i,t}, & s_{i,t} \in \mathrm{Topk}(\{s_{j,t} \mid 1 \le j \le N\}, K), \\ 0, & \text{otherwise}, \end{cases}
$$

$$
g_{N+1,t} = \mathrm{Sigmoid}\left(\mathbf{W}_{N+1}^l \bar{\mathbf{u}}_t^l\right),
$$

$$
s_{i,t} = \mathrm{Softmax}_i\left(\mathbf{W}_i^l \mathbf{u}_t^l\right),
$$

Huber loss (-> outlier)
Auxiliary loss (-> routing collapse)
Multi-resolution forecasting (-> various horizons)

$$
\mathcal{L} = \frac{1}{P}\sum_{j=1}^{P} \mathcal{L}_{\mathrm{ar}}\left(\mathbf{X}_{t+1:t+p_j}, \hat{\mathbf{X}}_{t+1:t+p_j}\right) + \alpha \mathcal{L}_{\mathrm{aux}},
$$

$$
\mathcal{L}_{\mathrm{ar}}\left(x_t, \hat{x}_t\right) = \begin{cases} \frac{1}{2}\left(x_t - \hat{x}_t\right)^2, & \text{if } |x_t - \hat{x}_t| \le \delta, \\ \delta \times \left(|x_t - \hat{x}_t| - \frac{1}{2} \times \delta\right), & \text{otherwise}, \end{cases} \qquad \mathcal{L}_{\mathrm{aux}} = N \sum_{i=1}^{N} f_i r_i,
$$

$$
f_i = \frac{1}{KT}\sum_{t=1}^{T} \mathbb{I}\left(\text{Time point } t \text{ selects Expert } i\right), \quad r_i = \frac{1}{T}\sum_{t=1}^{T} s_{i,t},
$$

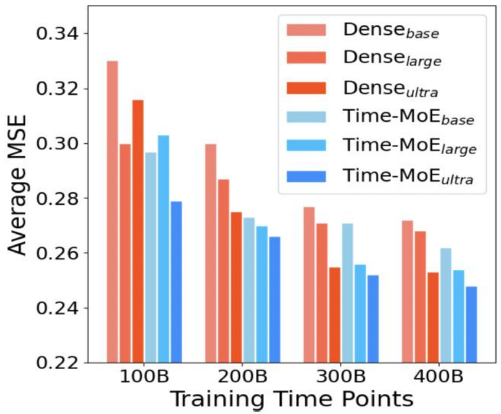
Time-MoE: Zero-shot Performance

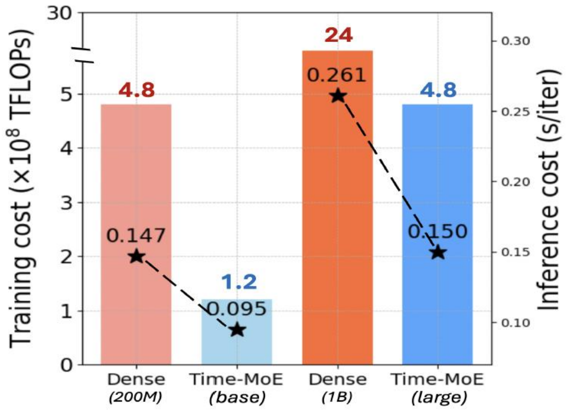
Table 3: Full results of zero-shot forecasting experiments. A lower MSE or MAE indicates a better prediction. TimesFM, due to its use of Weather datasets in pretraining, is not evaluated on these two datasets and is denoted by a dash (–). **Red**: the best, Blue: the 2nd best.

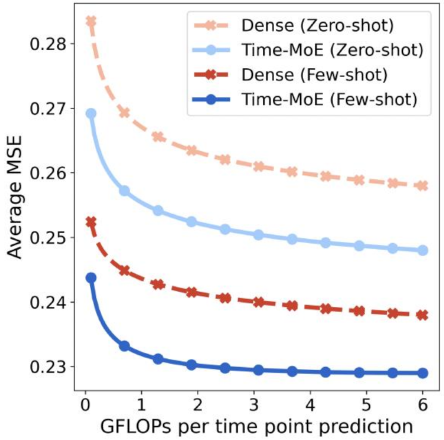
Models		Time-MoE (Ours)						Zero-shot Time Series Models															
		Time-MoE$_{base}$		Time-MoE$_{large}$		Time-MoE$_{ultra}$		Moirai$_{small}$		Moirai$_{base}$		Moirai$_{large}$		TimesFM		Moment		Chronos$_{small}$		Chronos$_{base}$		Chronos$_{large}$	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.357	0.381	0.350	0.382	**0.349**	**0.379**	0.401	0.402	0.376	0.392	0.381	0.388	0.414	0.404	0.688	0.557	0.466	0.409	0.440	0.393	0.441	0.390
	192	**0.384**	**0.404**	0.388	0.412	0.395	0.413	0.435	0.421	0.412	0.413	0.434	0.415	0.465	0.434	0.688	0.560	0.530	0.450	0.492	0.426	0.502	0.424
	336	**0.411**	0.434	**0.411**	0.430	0.447	0.453	0.438	0.434	0.433	**0.428**	0.495	0.445	0.503	0.456	0.675	0.563	0.570	0.486	0.550	0.462	0.576	0.467
	720	0.449	0.477	**0.427**	0.455	0.457	0.462	0.439	0.454	0.447	**0.444**	0.611	0.510	0.511	0.481	0.683	0.585	0.615	0.543	0.882	0.591	0.835	0.583
	AVG	0.400	0.424	**0.394**	**0.419**	0.412	0.426	0.428	0.427	0.417	**0.419**	0.480	0.439	0.473	0.443	0.683	0.566	0.545	0.472	0.591	0.468	0.588	0.466
ETTh2	96	0.305	0.359	0.302	0.354	**0.292**	0.352	0.297	0.336	0.294	**0.330**	0.296	**0.330**	0.315	0.349	0.342	0.396	0.307	0.356	0.308	0.343	0.320	0.345
	192	0.351	0.386	0.364	0.385	**0.347**	0.379	0.368	0.381	0.365	0.375	0.361	**0.371**	0.388	0.395	0.354	0.402	0.376	0.401	0.384	0.392	0.406	0.399
	336	0.391	0.418	0.417	0.425	0.406	0.419	0.370	0.393	0.376	**0.390**	0.390	**0.390**	0.422	0.427	**0.356**	0.407	0.408	0.431	0.429	0.430	0.492	0.453
	720	0.419	0.454	0.537	0.496	0.439	0.447	0.411	0.426	0.416	0.433	0.423	**0.418**	0.443	0.454	**0.395**	0.434	0.604	0.533	0.501	0.477	0.603	0.511
	AVG	0.366	0.404	0.405	0.415	0.371	0.399	**0.361**	0.384	0.362	0.382	0.367	**0.377**	0.392	0.406	**0.361**	0.409	0.424	0.430	0.405	0.410	0.455	0.427
ETTm1	96	0.338	0.368	0.309	0.357	**0.281**	**0.341**	0.418	0.392	0.363	0.356	0.380	0.361	0.361	0.370	0.654	0.527	0.511	0.423	0.454	0.408	0.457	0.403
	192	0.353	0.388	0.346	0.381	**0.305**	**0.358**	0.431	0.405	0.388	0.375	0.412	0.383	0.414	0.405	0.662	0.532	0.618	0.485	0.567	0.477	0.530	0.450
	336	0.381	0.413	0.373	0.408	**0.369**	0.395	0.433	0.412	0.416	**0.392**	0.436	0.400	0.445	0.429	0.672	0.537	0.683	0.524	0.662	0.525	0.577	0.481
	720	0.504	0.493	0.475	0.477	0.469	0.472	0.462	0.432	**0.460**	**0.418**	0.462	0.420	0.512	0.471	0.692	0.551	0.748	0.566	0.900	0.591	0.660	0.526
	AVG	0.394	0.415	0.376	0.405	**0.356**	0.391	0.436	0.410	0.406	**0.385**	0.422	0.391	0.433	0.418	0.670	0.536	0.640	0.499	0.645	0.500	0.555	0.465
ETTm2	96	0.201	0.291	**0.197**	0.286	0.198	0.288	0.214	0.288	0.205	0.273	0.211	0.274	0.202	**0.270**	0.260	0.335	0.209	0.291	0.199	0.274	**0.197**	0.271
	192	0.258	0.334	0.250	0.322	**0.235**	**0.312**	0.284	0.332	0.275	0.316	0.281	0.318	0.289	0.321	0.289	0.350	0.280	0.341	0.261	0.322	0.254	0.314
	336	0.324	0.373	0.337	0.375	**0.293**	**0.348**	0.331	0.362	0.329	0.350	0.341	0.355	0.360	0.366	0.324	0.369	0.354	0.390	0.326	0.366	0.313	0.353
	720	0.488	0.464	0.480	0.461	0.427	0.428	0.402	**0.408**	0.437	0.411	0.485	0.428	0.462	0.430	**0.394**	0.409	0.553	0.499	0.455	0.439	0.416	0.415
	AVG	0.317	0.365	0.316	0.361	**0.288**	0.344	0.307	0.347	0.311	**0.337**	0.329	0.343	0.328	0.346	0.316	0.365	0.349	0.380	0.310	0.350	0.295	0.338
Weather	96	0.160	0.214	0.159	0.213	**0.157**	**0.211**	0.198	0.222	0.220	0.217	0.199	**0.211**	-	-	0.243	0.255	0.211	0.243	0.203	0.238	0.194	0.235
	192	0.210	0.260	0.215	0.266	**0.208**	0.256	0.247	0.265	0.271	0.259	0.246	**0.251**	-	-	0.278	0.329	0.263	0.294	0.256	0.290	0.249	0.285
	336	0.274	0.309	0.291	0.322	**0.255**	**0.290**	0.283	0.303	0.286	0.297	0.274	0.291	-	-	0.306	0.346	0.321	0.339	0.314	0.336	0.302	0.327
	720	0.418	0.405	0.415	0.400	0.405	0.397	0.373	0.354	0.373	0.354	**0.337**	**0.340**	-	-	0.350	0.374	0.404	0.397	0.397	0.396	0.372	0.378
	AVG	0.265	0.297	0.270	0.300	**0.256**	0.288	0.275	0.286	0.287	0.281	0.264	**0.273**	-	-	0.294	0.326	0.300	0.318	0.292	0.315	0.279	0.306
Global Temp	96	0.211	0.343	**0.210**	**0.342**	0.214	0.345	0.227	0.354	0.224	0.351	0.224	0.351	0.255	0.375	0.363	0.472	0.234	0.361	0.230	0.355	0.228	0.354
	192	0.257	0.386	0.254	0.385	**0.246**	**0.379**	0.269	0.396	0.266	0.394	0.267	0.395	0.313	0.423	0.387	0.489	0.276	0.400	0.273	0.395	0.276	0.398
	336	0.281	0.405	0.267	**0.395**	**0.266**	0.398	0.292	0.419	0.296	0.420	0.291	0.417	0.362	0.460	0.430	0.517	0.314	0.431	0.324	0.434	0.327	0.437
	720	0.354	0.465	0.289	**0.420**	**0.288**	0.421	0.351	0.437	0.403	0.498	0.387	0.488	0.486	0.545	0.582	0.617	0.418	0.504	0.505	0.542	0.472	0.535
	AVG	0.275	0.400	0.255	**0.385**	**0.253**	**0.385**	0.285	0.409	0.297	0.416	0.292	0.413	0.354	0.451	0.440	0.524	0.311	0.424	0.333	0.431	0.326	0.431
Average		0.336	0.384	0.336	0.380	**0.322**	0.372	0.349	0.377	0.347	**0.370**	0.359	0.373	0.396	0.413	0.461	0.454	0.428	0.420	0.429	0.412	0.416	0.405
1st Count		3		10		**28**		2		11		10		1		4		0		0		1	

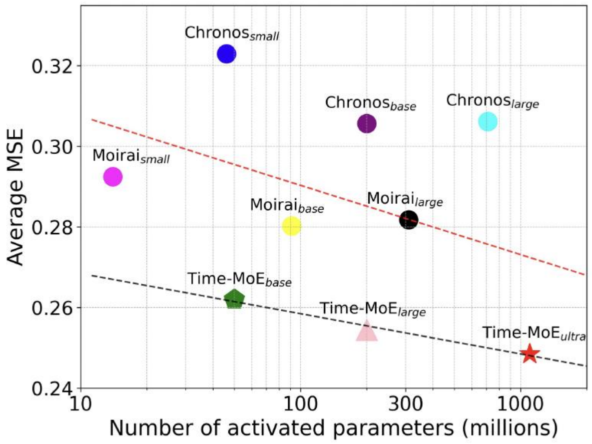
All pretrained time series models were evaluated without further training for different forecasting horizons

Time-MoE: Scalability Analysis





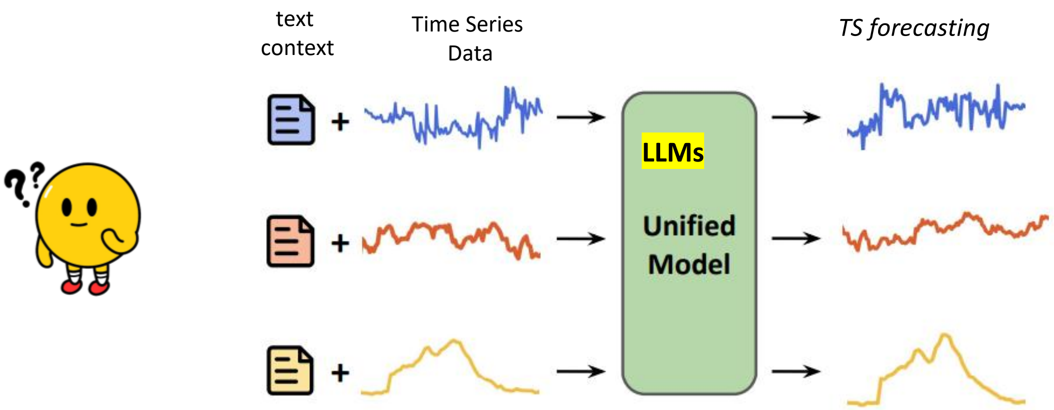




LLM for Time Series: Motivation

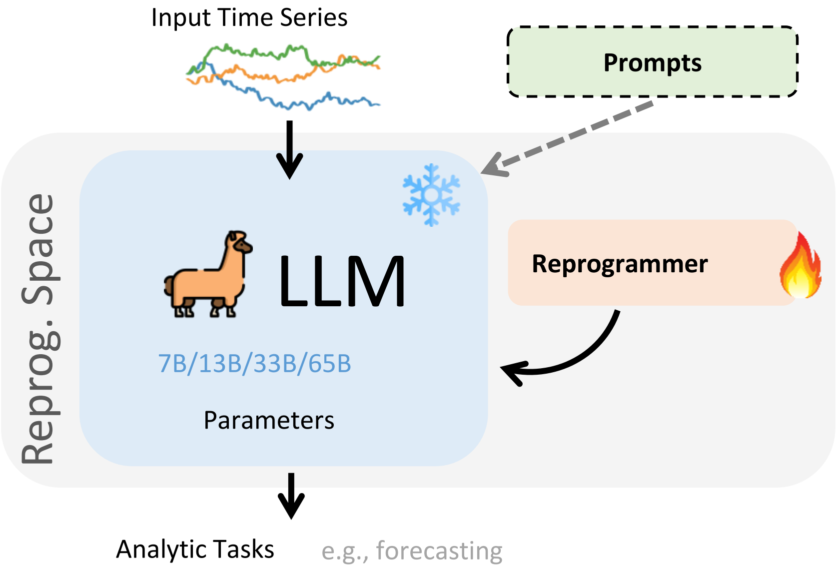
- Multimodal large language models

How time series forecasting benefits from the recent advances of LLMs?



Motivation

- Reprogramming makes LLMs **instantly ready, more powerful** for time series tasks



We keep pretrained LLMs intact and **only fine-tune reprogrammer** to achieve certain alignments

Reprogramming ≈ Adaptation + Alignment

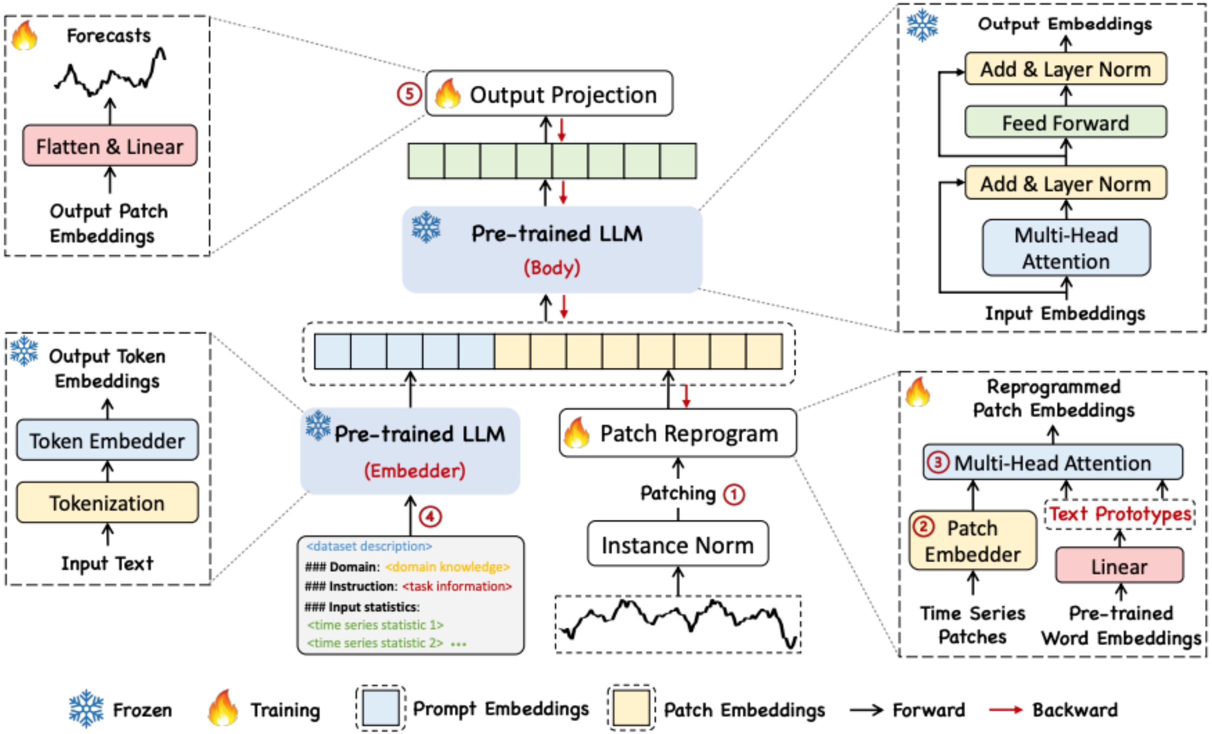
Adaptation makes LLMs to understand how to process the time series data → Breaking domain isolation and enabling knowledge sharing

Alignment further eliminates domain boundary to facilitate knowledge acquiring

Chen, Pin-Yu. "Model reprogramming: Resource-efficient cross-domain machine learning." AAAI Conference on Artificial Intelligence, 2024.

Time-LLM: Architecture

TL;DR Domain expert knowledge & Task instructions + Reprogrammed input time series = Significantly better forecasts



Unlocking the LLM's ability for time series

Cross-modal Adaptation: ① ② ⑤

Cross-modal Alignment: ③ ④

- **Patch Reprogramming:** we reprogram TS patch embeddings into the source data representation space to align the modalities of time series and natural language to activate the backbone's time series understanding and reasoning capabilities.
- **Prompt-as-Prefix:** natural language-based prompts (domain knowledge & task instructions & input statistics) can act as prefixes to enrich the input context and guide the transformation of reprogrammed TS patches

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan*, Qingsong Wen*, "Time-LLM: Time Series Forecasting by Reprogramming Large Language Models", **ICLR 2024**.

Time-LLM: Few-shot & Zero-shot

Table 3: Few-shot learning on 10% training data. We use the same protocol in Tab. 1. All results are averaged from four different forecasting horizons: $H \in \{96, 192, 336, 720\}$. Our full results are in Appendix E.

Methods	TIME-LLM (**Ours**)		GPT4TS (2023a)		DLinear (2023)		PatchTST (2023)		TimesNet (2023)		FEDformer (2022)		Autoformer (2021)		Stationary (2022)		ETSformer (2022)		LightTS (2022a)		Informer (2021)		Reformer (2020)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$ETTh1$	**0.556**	**0.522**	0.590	0.525	0.691	0.600	0.633	0.542	0.869	0.628	0.639	0.561	0.702	0.596	0.915	0.639	1.180	0.834	1.375	0.877	1.199	0.809	1.249	0.833
$ETTh2$	**0.370**	**0.394**	0.397	0.421	0.605	0.538	0.415	0.431	0.479	0.465	0.466	0.475	0.488	0.499	0.462	0.455	0.894	0.713	2.655	1.160	3.872	1.513	3.485	1.486
$ETTm1$	**0.404**	**0.427**	0.464	0.441	0.411	0.429	0.501	0.466	0.677	0.537	0.722	0.605	0.802	0.628	0.797	0.578	0.980	0.714	0.971	0.705	1.192	0.821	1.426	0.856
$ETTm2$	**0.277**	**0.323**	0.293	0.335	0.316	0.368	0.296	0.343	0.320	0.353	0.463	0.488	1.342	0.930	0.332	0.366	0.447	0.487	0.987	0.756	3.370	1.440	3.978	1.587
$Weather$	**0.234**	**0.273**	0.238	0.275	0.241	0.283	0.242	0.279	0.279	0.301	0.284	0.324	0.300	0.342	0.318	0.323	0.318	0.360	0.289	0.322	0.597	0.495	0.546	0.469
ECL	**0.175**	0.270	0.176	**0.269**	0.180	0.280	0.180	0.273	0.323	0.392	0.346	0.427	0.431	0.478	0.444	0.480	0.660	0.617	0.441	0.489	1.195	0.891	0.965	0.768
$Traffic$	**0.429**	0.306	0.440	0.310	0.447	0.313	0.430	**0.305**	0.951	0.535	0.663	0.425	0.749	0.446	1.453	0.815	1.914	0.936	1.248	0.684	1.534	0.811	1.551	0.821
1st Count	**8**		1		0		1		0		0		0		0		0		0		0		0	

Table 5: Zero-shot learning results. **Red**: the best, Blue: the second best. Appendix E shows our detailed results.

Methods	TIME-LLM (**Ours**)		GPT4TS (2023a)		LLMTime (2023)		DLinear (2023)		PatchTST (2023)		TimesNet (2023)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$ETTh1 \rightarrow ETTh2$	**0.353**	**0.387**	0.406	0.422	0.992	0.708	0.493	0.488	0.380	0.405	0.421	0.431
$ETTh1 \rightarrow ETTm2$	**0.273**	**0.340**	0.325	0.363	1.867	0.869	0.415	0.452	0.314	0.360	0.327	0.361
$ETTh2 \rightarrow ETTh1$	**0.479**	**0.474**	0.757	0.578	1.961	0.981	0.703	0.574	0.565	0.513	0.865	0.621
$ETTh2 \rightarrow ETTm2$	**0.272**	**0.341**	0.335	0.370	1.867	0.869	0.328	0.386	0.325	0.365	0.342	0.376

Time-LLM: Ablation & Efficiency

Table 6: Ablations on ETTh1 and ETTm1 in predicting 96 and 192 steps ahead (MSE reported). **Red**: the best.

Variant	Long-term Forecasting				Few-shot Forecasting			
	ETTh1-96	ETTh1-192	ETTm1-96	ETThm1-192	ETTh1-96	ETTh1-192	ETTm1-96	ETThm1-192
A.1 Llama (**Default**; 32)	**0.362**	**0.398**	**0.272**	**0.310**	**0.448**	**0.484**	**0.346**	**0.373**
A.2 Llama (8)	0.389	0.412	0.297	0.329	0.567	0.632	0.451	0.490
A.3 GPT-2 (12)	0.385	0.419	0.306	0.332	0.548	0.617	0.447	0.509
A.4 GPT-2 (6)	0.394	0.427	0.311	0.342	0.571	0.640	0.468	0.512
B.1 w/o Patch Reprogramming	0.410	0.412	0.310	0.342	0.498	0.570	0.445	0.487
B.2 w/o Prompt-as-Prefix	0.398	0.423	0.298	0.339	0.521	0.617	0.432	0.481
C.1 w/o Dataset Context	0.402	0.417	0.298	0.331	0.491	0.538	0.392	0.447
C.2 w/o Task Instruction	0.388	0.420	0.285	0.327	0.476	0.529	0.387	0.439
C.3 w/o Statistical Context	0.391	0.419	0.279	0.347	0.483	0.547	0.421	0.461

Table 7: Efficiency analysis of TIME-LLM

Length	ETTh1-96			ETTh1-192		
Metric	Param. (M)	Mem. (MiB)	Speed(s/iter)	Param. (M)	Mem. (MiB)	Speed(s/iter)
D.1 LLama (32)	3404.53	32136	0.517	3404.57	33762	0.582
D.2 LLama (8)	975.83	11370	0.184	975.87	12392	0.192
D.3 w/o LLM	6.39	3678	0.046	6.42	3812	0.087

Table 17: Efficiency comparison between model reprogramming and parameter-efficient fine-tuning (PEFT) with QLoRA (Dettmers et al., 2023) on ETTh1 dataset in forecasting two different steps ahead.

Length		ETTh1-96			ETTh1-336		
Metric		Trainable Param. (M)	Mem. (MiB)	Speed(s/iter)	Trainable Param. (M)	Mem. (MiB)	Speed(s/iter)
Llama (8)	QLoRA	12.60	14767	0.237	12.69	15982	0.335
	Reprogram	5.62	11370	0.184	5.71	13188	0.203
Llama (32)	QLoRA	50.29	45226	0.697	50.37	49374	0.732
	Reprogram	6.39	32136	0.517	6.48	37988	0.632

- **Language model variants:** The scaling law retains with the LLM reprogramming.
- **Cross-modality alignment:** (1) we find that the alignment is crucial (see B.1 and B.2); (2) domain knowledge and task instructions are both valuable (C.1-C.3) and can be integrated via Prompt-as-Prefix (PaP)
- **Reprogramming efficiency:** (1) our reprogramming network is lightweight in activating the LLM's ability for time series forecasting (see D.3 -- i.e., fewer than 6.6M trainable parameters; only around 0.2% of the parameters in Llama-7B); (2) this is favorable even compared to parameter-efficient fine-tuning (PEFT; Tab. 17)

LLM for Multi-Task Time Series Analysis: Motivation

Motivation

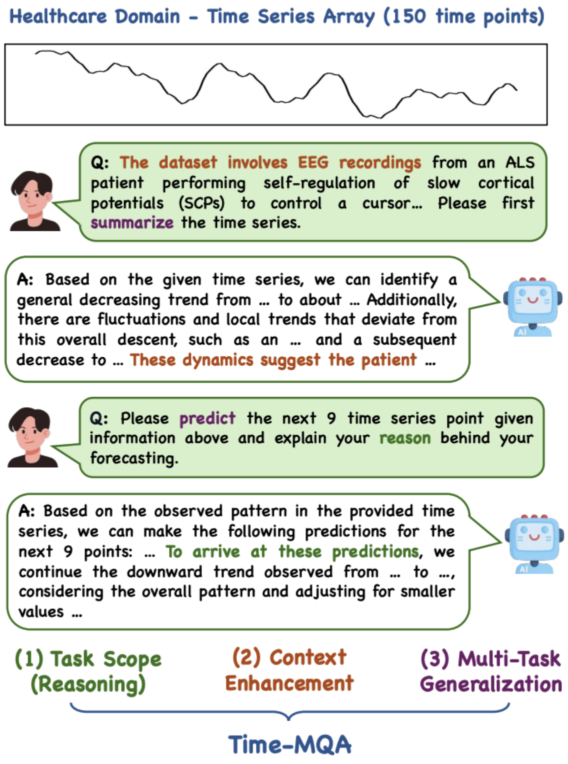
- Current time series analysis focus on **narrow and specific tasks** such as forecasting or anomaly detection.
- Need **frameworks that address a broader array** of analytical and reasoning tasks on time series data.

Designed Framework: Time-MQA

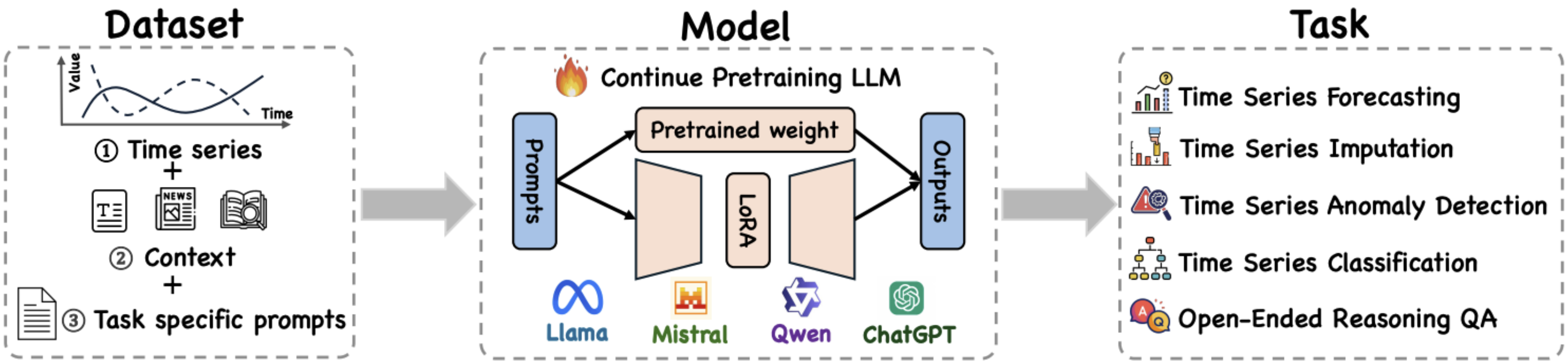
- Unified framework for **multi-task time series question answering (QA)**, enabling natural language queries across common time series tasks, including both numerical analysis and open-ended reasoning.

Data Contribution: TSQA

- Around 200k question-answer pairs across diverse domains and a variety of time series tasks.



Time-MQA: Overall Framework



Unified QA Function

Given a time series **X**, contextual information **C**, and a natural language question **Q**, the Time-MQA model generates an answer **A**.

Tasks can involve data forecasting, classification, anomaly detection, or reasoning responses.

Key Innovations

Time-MQA distinguishes itself by supporting **multitask QA**, leveraging contextual information to enhance robustness and adaptivity, and dynamically generalizing across various question types **within a single, flexible architecture**.

Yaxuan Kong, Yiyuan Yang, Yoontae Hwang, Wenjie Du, Stefan Zohren, Zhangyang Wang, Ming Jin*, Qingsong Wen*,
"Time-MQA: Time Series Multi-Task Question Answering with Context Enhancement," ACL 2025 (Main).

TSQA Dataset: ~200k QA pairs

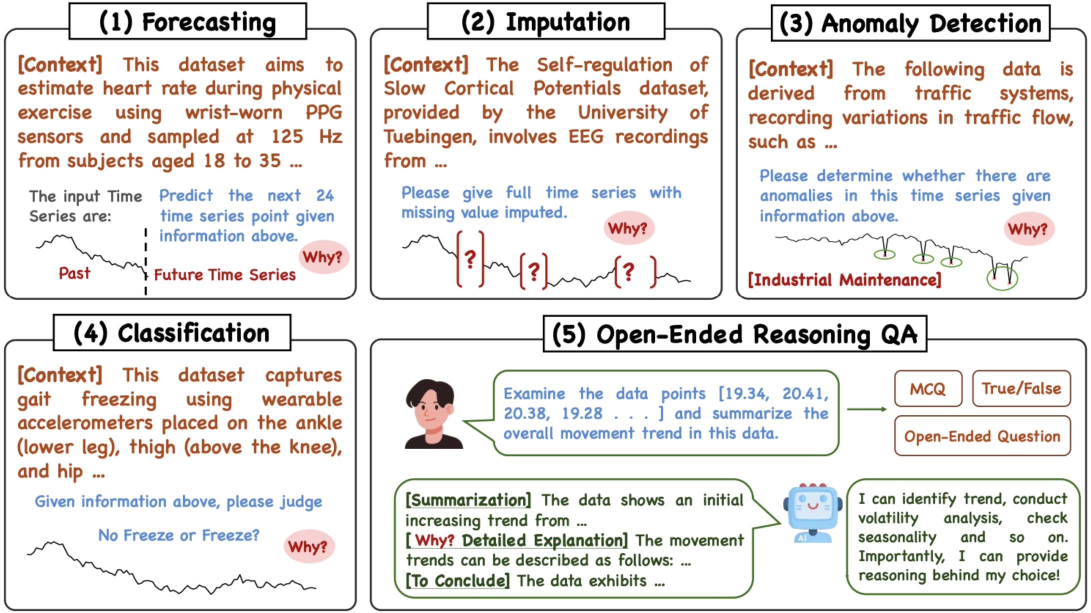


Figure 3: The demonstration of the Time-MQA and example of TSQA dataset with context enhancement.

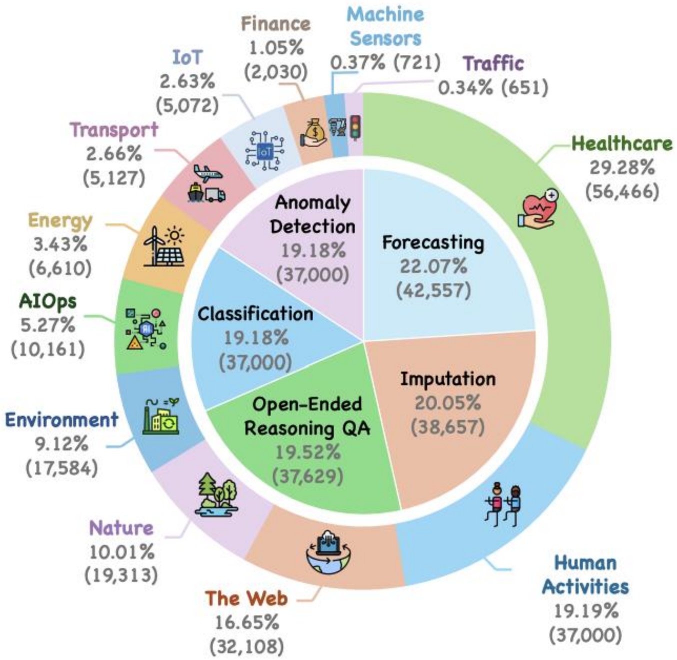


Figure 4: The distribution of data statistics in the TSQA dataset. The inner ring shows task types, and the outer ring shows domains.

Note: the TSQA dataset can be found at https://huggingface.co/Time-MQA.

Open-Ended Reasoning QA

MCQ (multiple-choice question)

Question: Determine if there's a structural break in the data points *[Time Series Data Points]*. A) Yes, between 0.71 and 0.72 B) Yes, between 0.65 and 0.63 C) No structural break.

Answer: C) No structural break. The data follows a gradual upward trend and subsequent flattening with slight variability, but no abrupt change in the level or direction indicating a structural break.

Judgment (true or false)

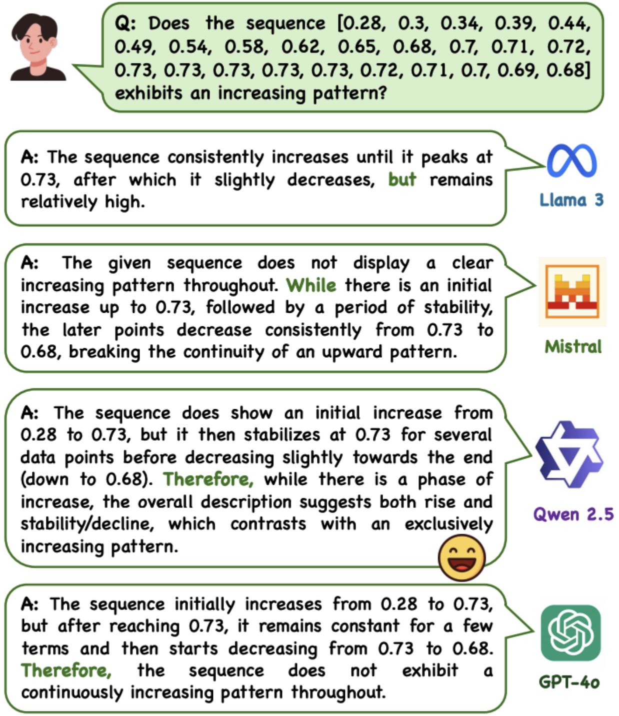
Question: The data points *[Time Series Data Points]* exhibit a clear seasonal pattern. True or False?

Answer: False. While there are fluctuations and some repeated values, no clear cyclical pattern emerges that would indicate seasonality. The variations appear to be more random than driven by any seasonal influences.

Open-Ended QA

Question: Summarize the overall behavior observed in the data points *[Time Series Data Points]*.

Answer: The data initially shows a slow increase followed by a decreasing trend. The middle part of the data points represents a peak with few higher fluctuations and later shows a slightly decline indicating a volatile pattern.



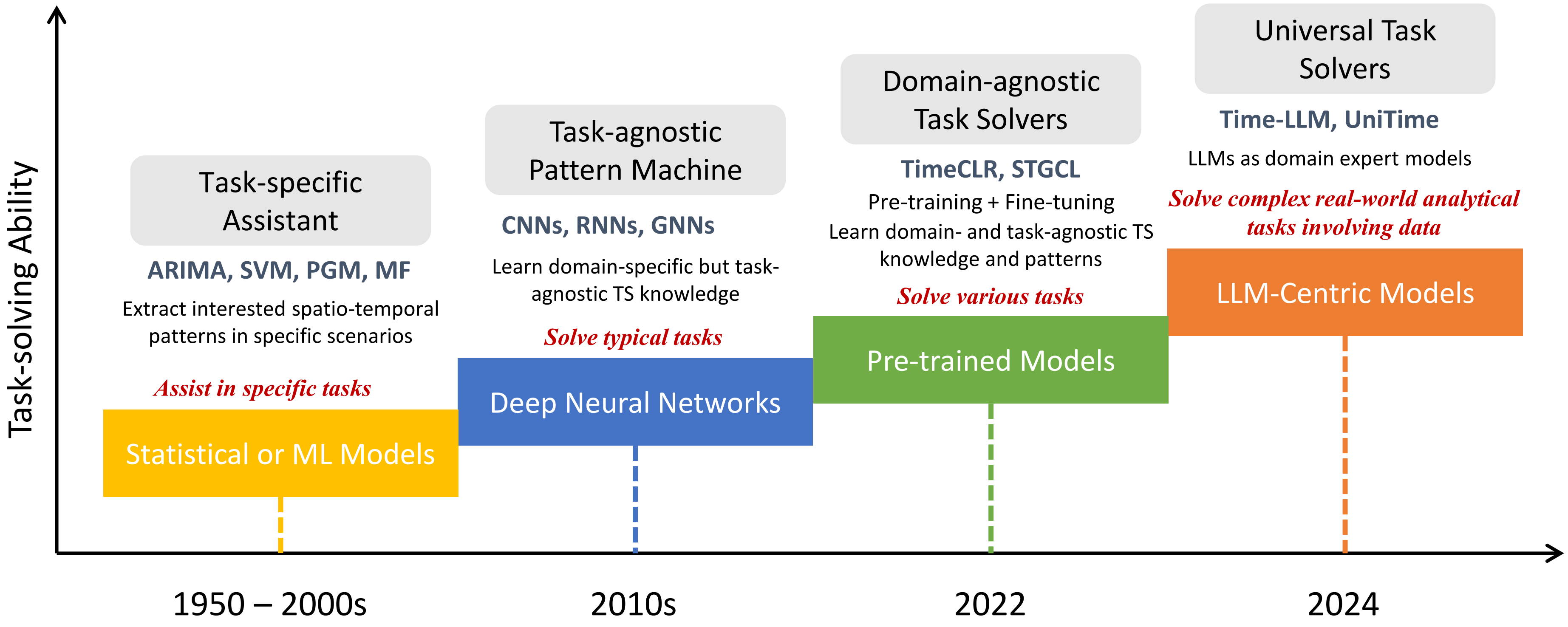
Left: examples of three main types of open-ended reasoning QA. Right: an example of an open-ended reasoning question answering and generated model responses comparison between our fine-tuned models with those from GPT-4o.

Time-MQA: Result Comparison

Table 3: Comparison of our three fine-tuned models, GPT-4o, and Doubao across diverse tasks. Forecasting and imputation tasks were evaluated using average MSE, while anomaly detection, classification, and open-ended reasoning tasks (including multiple-choice questions (MCQs) and true-false questions (Judgment)) were measured by accuracy. A lower value of MSE ↓ and a higher value of accuracy ↑ indicate better performance. * Doubao uses simple mean forecasting, which outputs the same value for all forecasting.

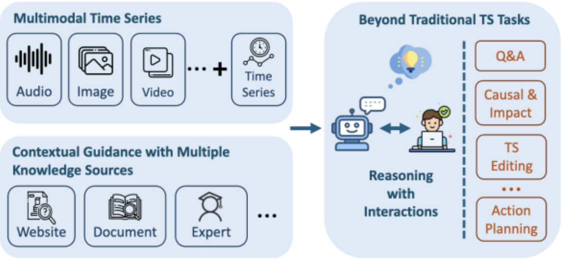
Backbone	**Classical Numerical Task**				**Open-Ended Reasoning QA**	
	Forecasting ↓	**Imputation ↓**	**Anomaly Detection ↑**	**Classification ↑**	**Judgment ↑**	**MCQ ↑**
Doubao	—*	0.018	0.52	0.44	0.78	0.56
GPT-4o	1.79	0.018	0.64	0.32	0.72	0.58
Llama-3 8B	2.01	0.020	0.54	0.24	0.74	0.48
Qwen-2.5 7B	1.82	0.016	**0.68**	**0.52**	**0.82**	0.54
Mistral 7B	**1.35**	**0.014**	0.58	0.44	0.80	**0.64**

Roadmap



Ming Jin, Yifan Zhang, Wei Chen, Kexin Zhang, Yuxuan Liang*, Bin Yang, Jindong Wang, Shirui Pan*, Qingsong Wen*, "Position: What Can Large Language Models Tell Us about Time Series Analysis", **ICML 2024**

➢ Future: Empowering Time Series Reasoning with Multimodal LLMs



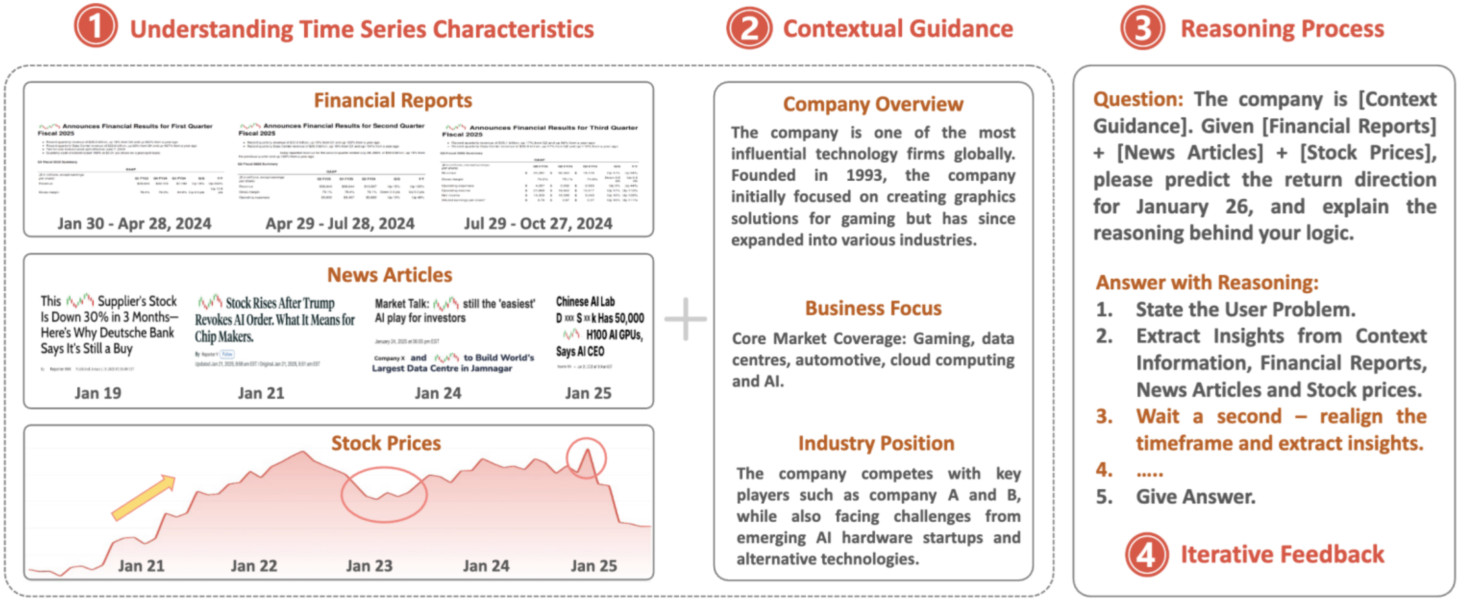


Figure 2. Key components for achieving time series reasoning (illustrated with financial time series example).

Yaxuan Kong, Yiyuan Yang, Shiyu Wang, Chenghao Liu, Yuxuan Liang, Ming Jin, Stefan Zohren, Dan Pei, Yan Liu, Qingsong Wen*, **"Position: Empowering Time Series Reasoning with Multimodal LLMs", arXiv 2025**.

Coffee Break

Schedule

Time	Talk	Speaker
8:00-9:30	Background of FM/LLM for TS/ST Data FM/LLM for Time Series Data	Qingsong Wen Squirrel Ai Learning
9:30-10:00	Coffee Break	-
10:00-11:00	**When Foundation Models Meets Spatio-Temporal Data**	Yuxuan Liang HKUST (Guangzhou)

Our tutorial on FM for Time Series (FM4TS) at KDD'24

When Foundation Models Meets Spatio-Temporal Data

Yuxuan Liang

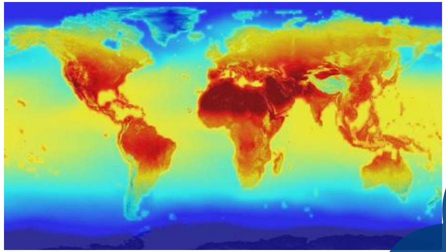
Assistant Professor, INTR & DSA Thrust

yuxuanliang@hkust-gz.edu.cn

香港科技大学(广州)
THE HONG KONG
UNIVERSITY OF SCIENCE AND
TECHNOLOGY (GUANGZHOU)

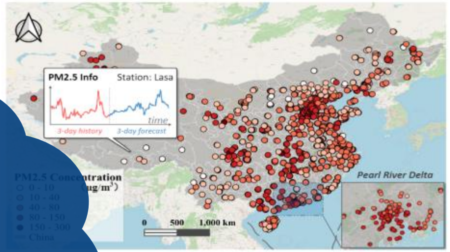
What is Spatio-Temporal (ST) Data?

- With recent advances in sensing technologies, a myriad of Spatio-Temporal Data has been collected and contributed to various disciplines



Climate

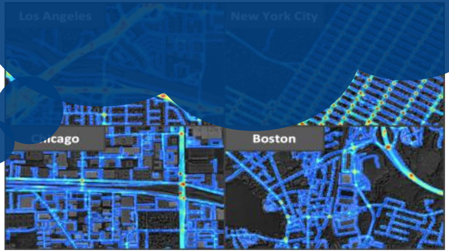
Epidemiology



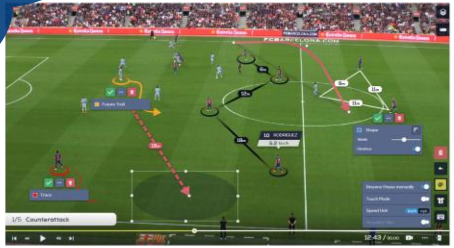
Environment

Time, Location, Event

Social Science



Transportation



Sports Analysis

ST Modeling Pipeline

Credit to Junbo Zhang

Raw ST Data → Data Transformation → Standard ST Dataset → Model Selection → Spatio-Temporal AI Model → Training & deploy → Real-world Application

POI

RN

⋮

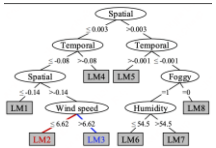
Traj

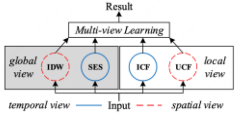
A variety of data in ST domains

ST Dataset

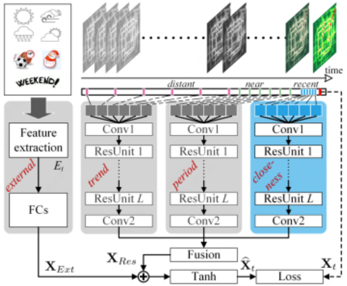
- ST Trajectory
- ST Grid Data
- ST Graph Data
- ST Series Data

Feature Engineering + ML





Spatio-Temporal Neural Networks

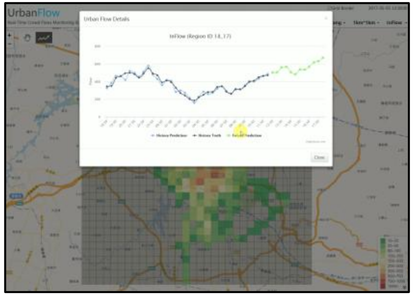


Tasks

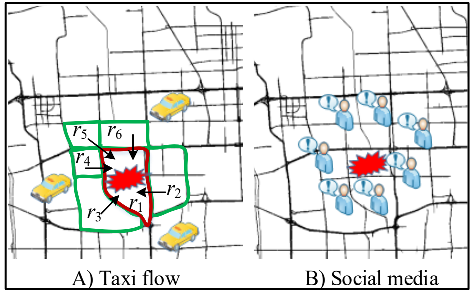
- Prediction
- Anomaly
- Interpolation
- Classification
- Recommendation
- Clustering
- Scheduling
- ⋮

Popular Downstream Tasks

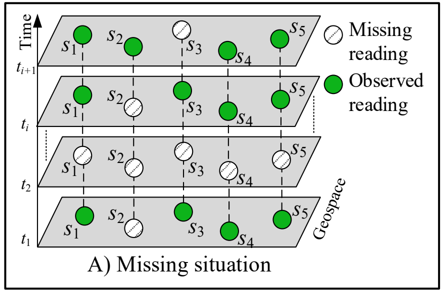
ST Prediction



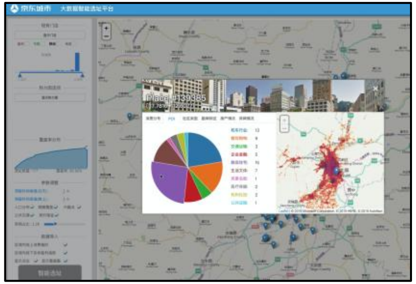
Anomaly Detection



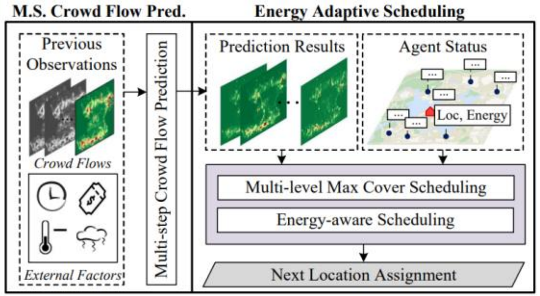
ST Interpolation



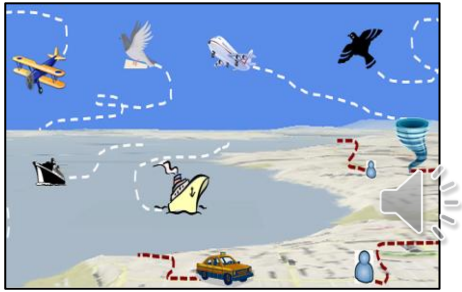
ST Recommendation



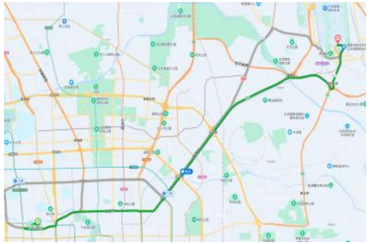
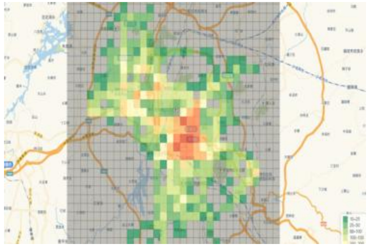
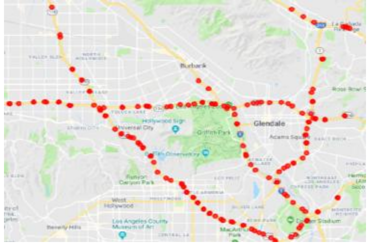
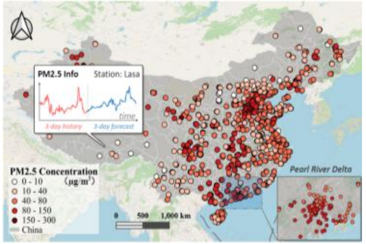
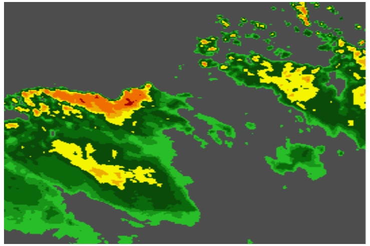
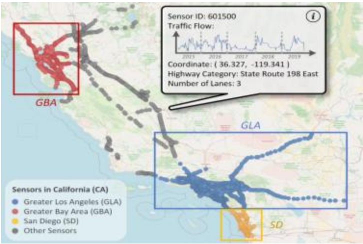
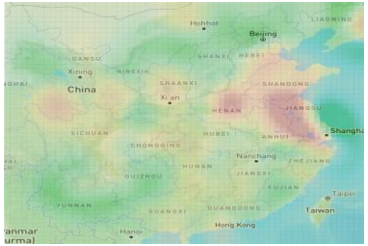
Scheduling



Classification

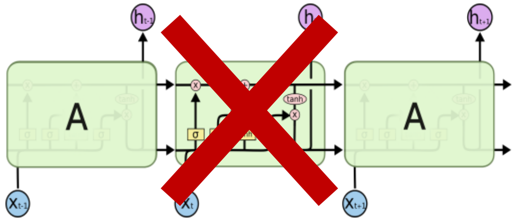


A Revisit on Existing Methods

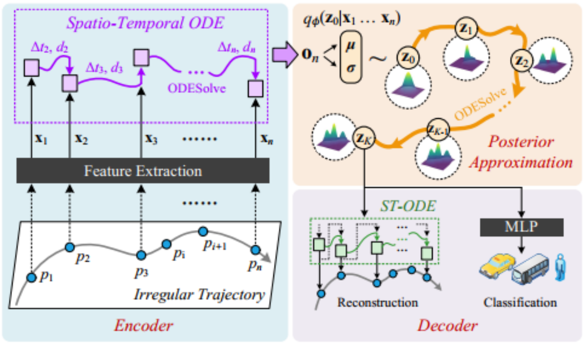
Modeling ST Trajectory	Modeling ST Grid Data	Modeling ST Graphs	Modeling ST Series
			
ControlTraj [KDD'24]	PhysicNet [TKDE'24]	CaST [NeurIPS'23]	AirFormer [AAAI'23]
			
OmniTraj [KDD'25]	NuwaDynamic [ICLR'24]	EAC [ICLR'25]	STGNP [KDD'23]

Methodologies for Learning ST Trajectories

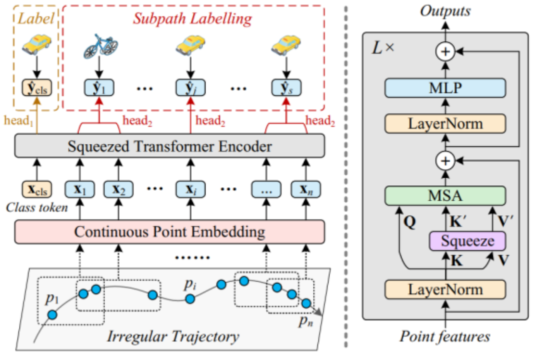
- Existing AI approaches for CV/NLP are NOT always good choices for modeling trajectories
- Capturing the irregularity of trajectories is of great importance to trajectory modeling
- We demonstrate how to encode the domain knowledge (i.e., irregularity) into existing AI methods, including RNNs and Transformers



Classic RNNs



Continuous trajectory modeling

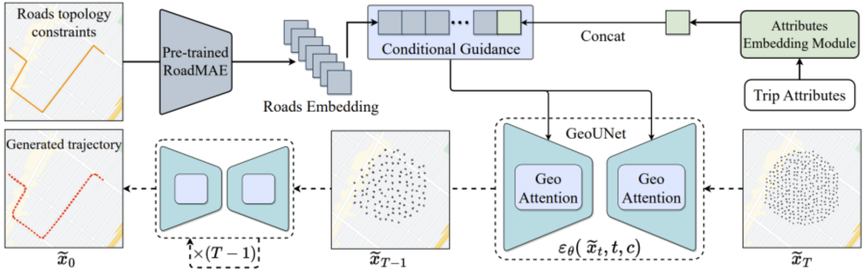


Efficient trajectory modeling

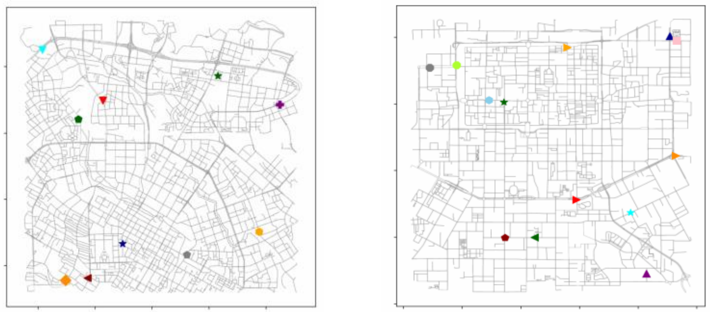
Y. Liang et al., Modelling Trajectories with Neural Ordinary Differential Equations. **IJCAI** 2021.
Y. Liang et al., TrajFormer: Efficient Trajectory Classification with Transformers. **CIKM** 2022.

Methodologies for Generating ST Trajectories

- Using real-world human trajectories usually has privacy concerns
  - Generation helps protect users' privacy
  - **DDPM outperforms GAN and VAE**



Same origin and destination



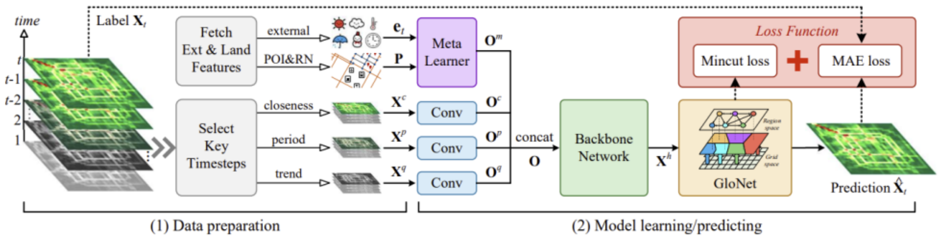
Different origin and destination

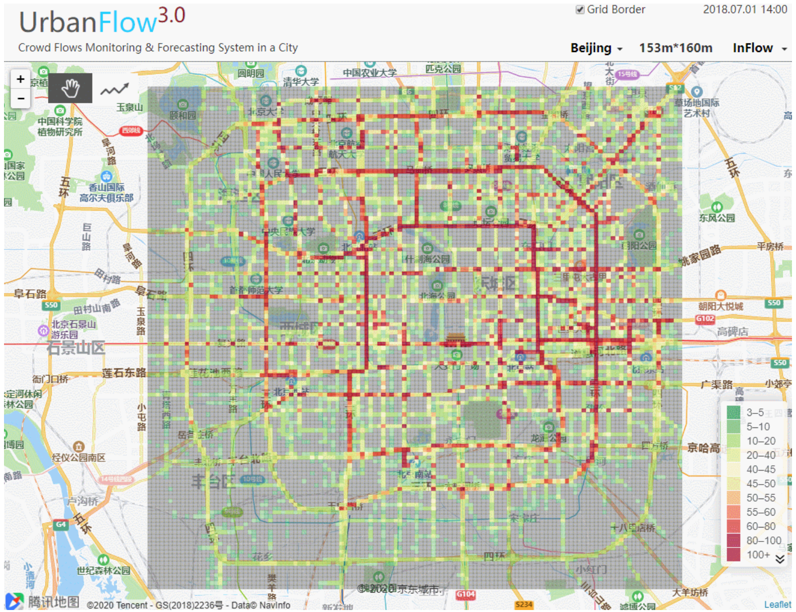
Y. Zhu et al., ControlTraj: Controllable Trajectory Generation with Topology-Constrained Diffusion Model. **KDD** 2024.

Methodologies for Learning ST Rasters

Citywide crowd flow prediction

- Predicting the inflow/outflow of every region in several hours
- Challenges
  - Complex ST dependencies
  - Long-range spatial dependencies

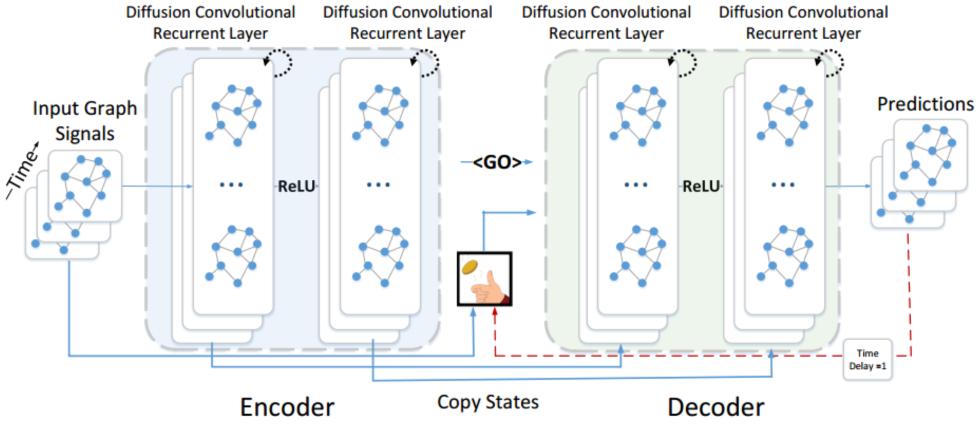


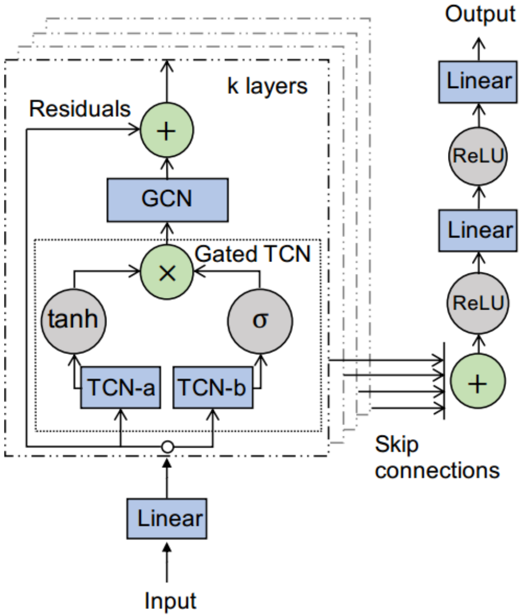


Y. Liang et al., Fine-grained Urban Flow Prediction. WWW 2021.

Methodologies for Learning ST Graphs

- Spatio-Temporal Graph Neural Network (STGNN)
  - Capturing spatial correlations by Graph Neural Networks (GNN)
  - Learning temporal dependencies by RNN or CNN





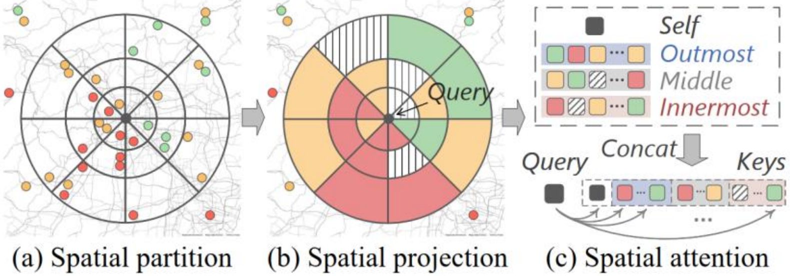
G. Jin et al., Spatio-Temporal Graph Neural Networks for Predictive Learning in Urban Computing: A Survey. **TKDE** 2023.

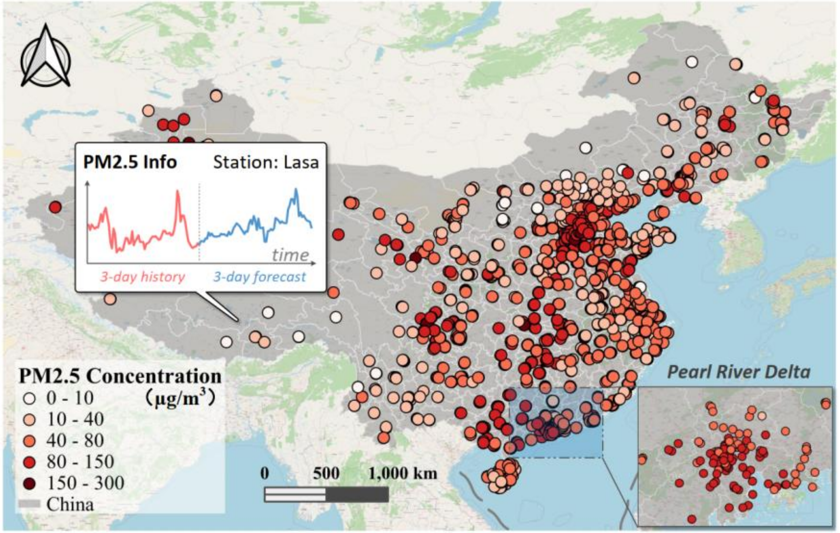
Methodologies for Learning ST Series

Nationwide Air Quality Prediction in China

- We present the first attempt to *collectively* predict air quality in the Chinese mainland with an **unprecedented fine spatial granularity**, covering 1,000+ stations.

- To capture dynamic spatial correlations
  - Using **self-attention mechanism**
  - Challenge: quadratic complexity w.r.t #locations

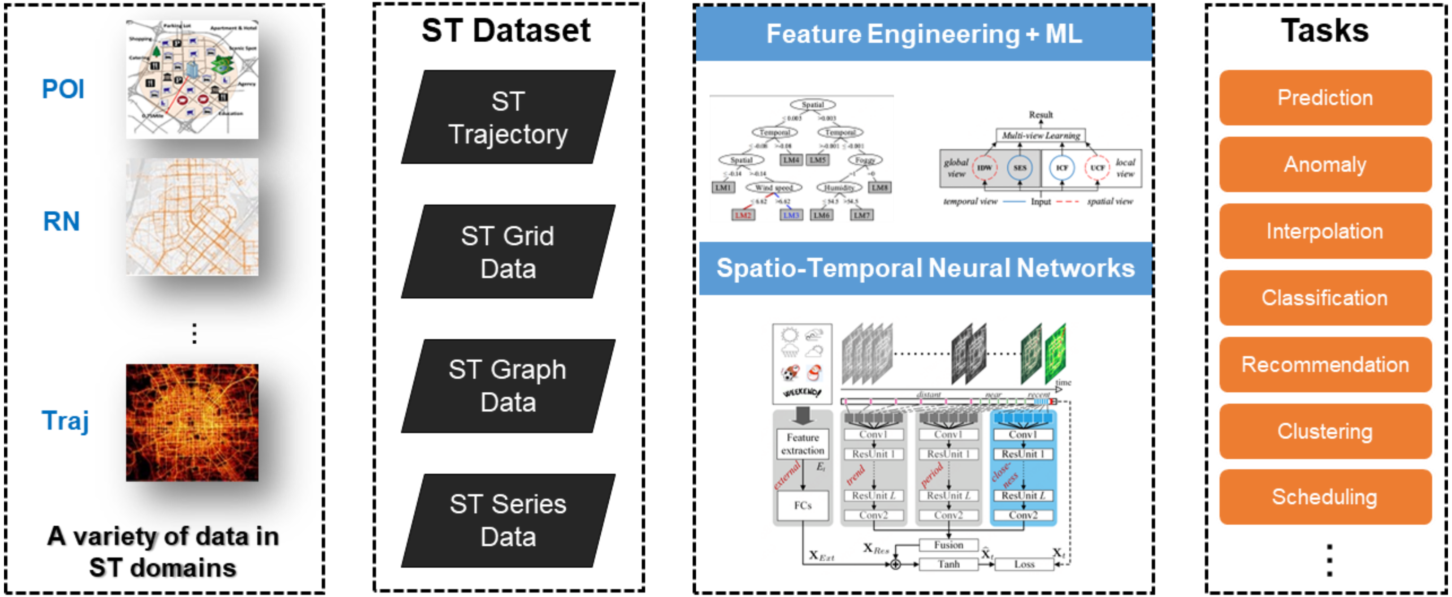




Y. Liang et al., AirFormer: Predicting Nationwide Air Quality in China with Transformers. **AAAI** 2023.

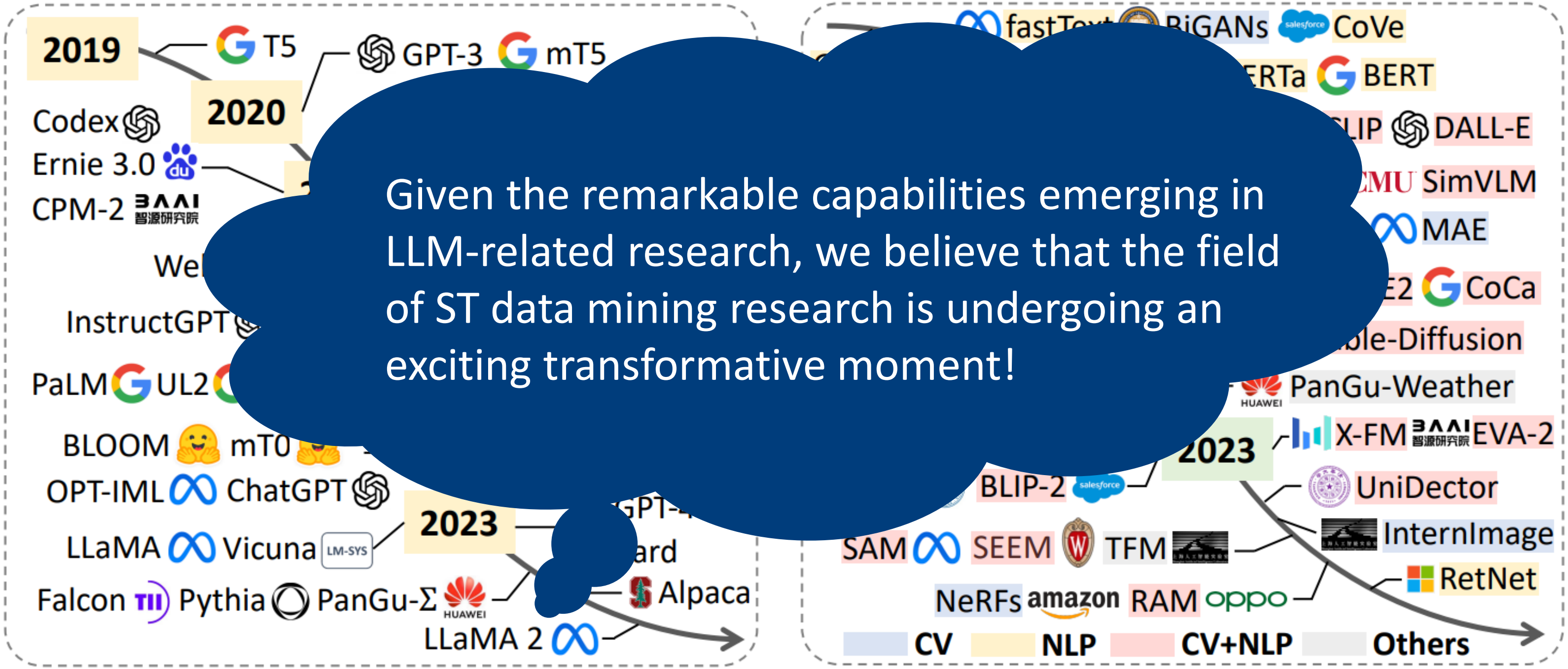
Key Limitation

- **There is no free lunch**
- Prior literature mostly concentrated on solving specific tasks



Roadmap of LLMs

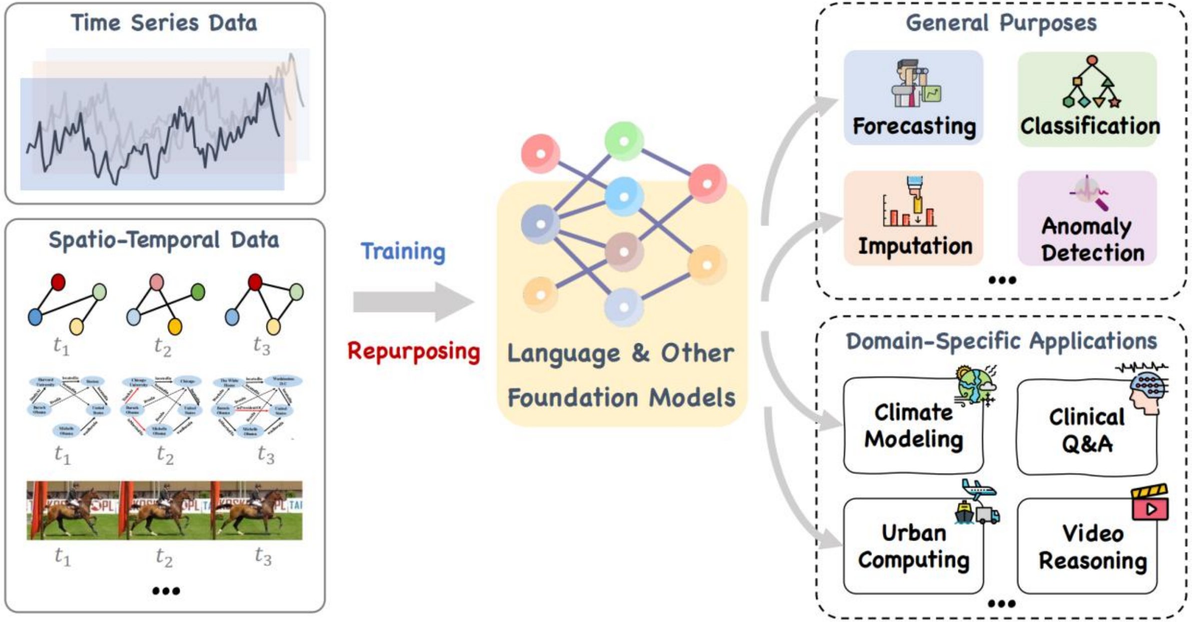
- LLMs and Foundation Models



Given the remarkable capabilities emerging in LLM-related research, we believe that the field of ST data mining research is undergoing an exciting transformative moment!

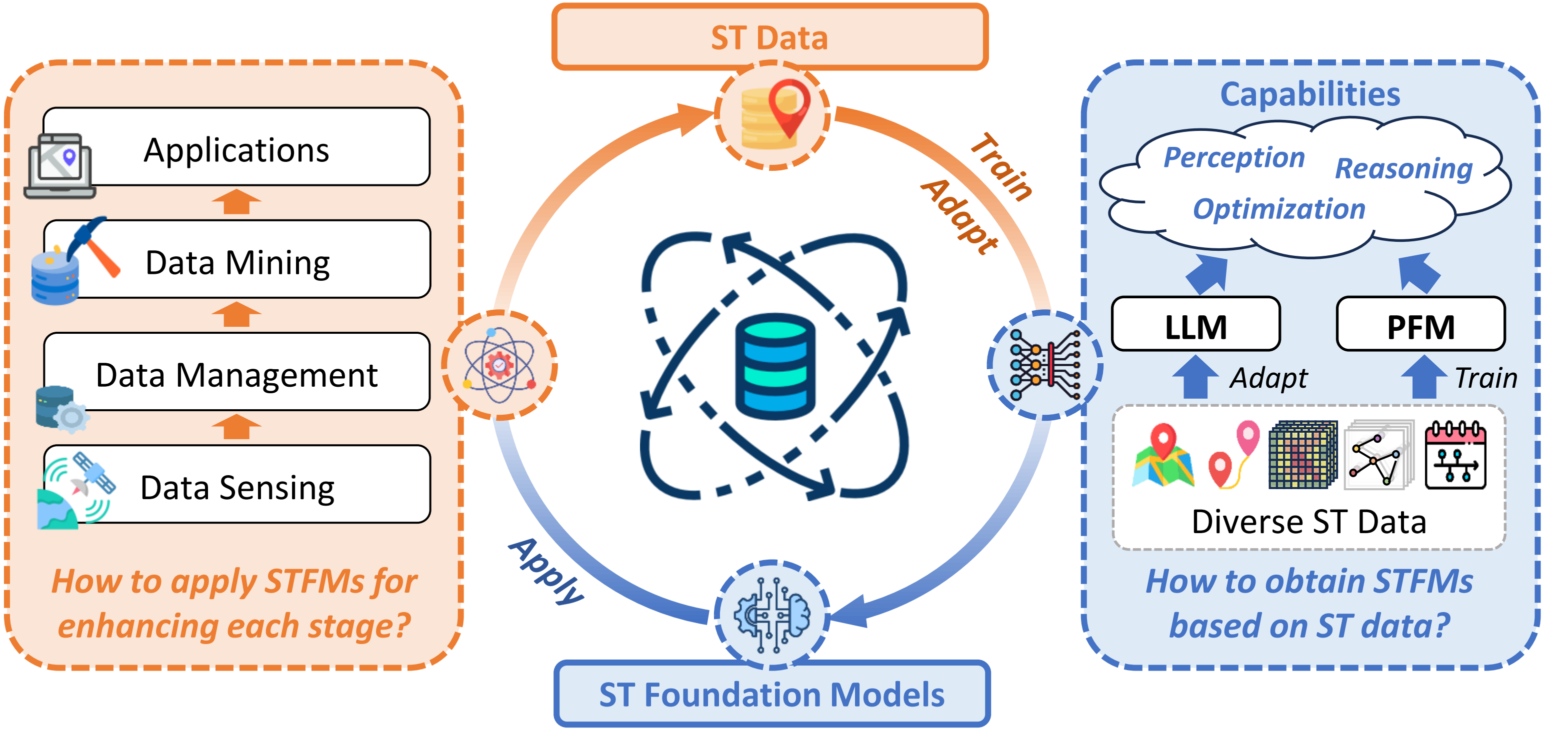
Towards ST Foundation Models (STFM)

- **Foundation models** can be either trained or adeptly repurposed to handle ST data for a range of general-purpose tasks and specialized domain applications.



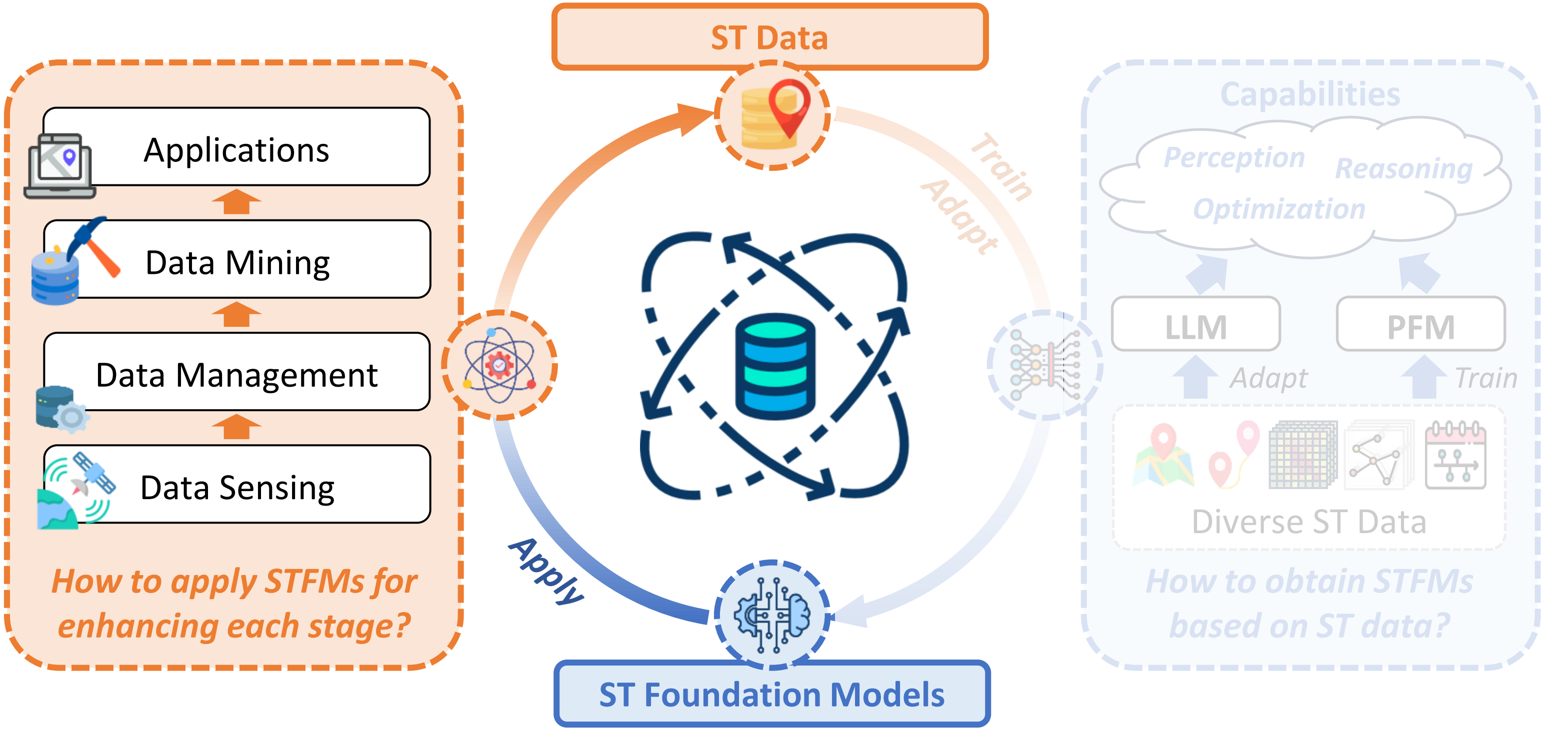
M. Jin et al., Large Models for Time Series and Spatio-Temporal Data: A Survey and Outlook. arXiv 2023.

Overview – Two Key Questions



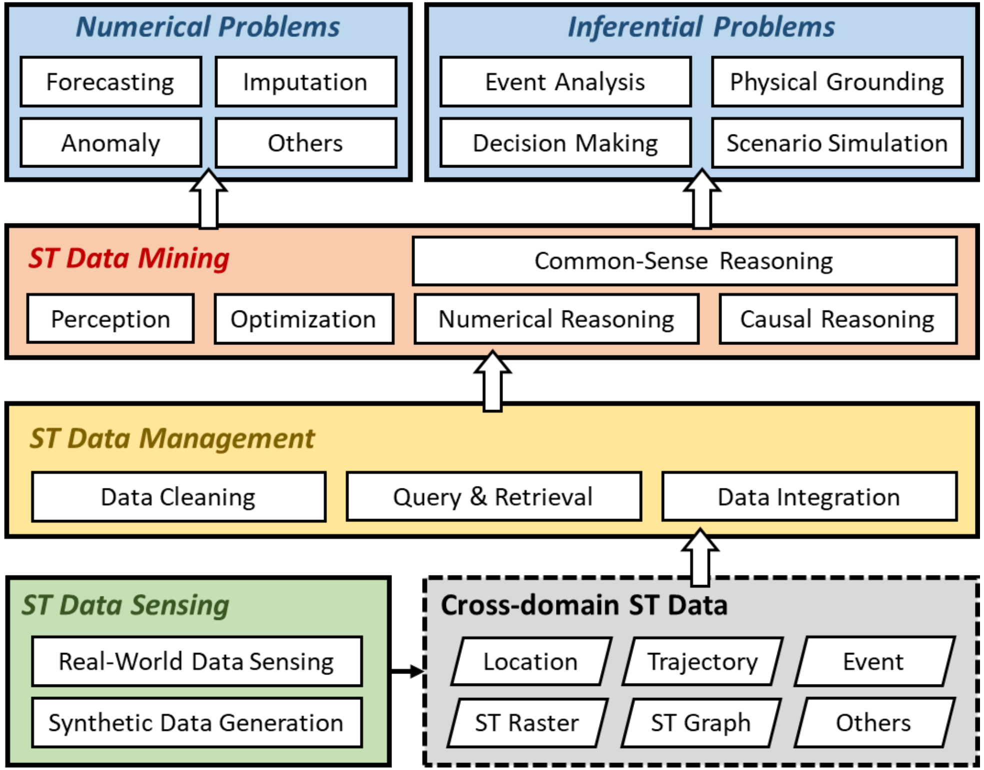
Y. Liang et al., Foundation Models for Spatio-Temporal Data Science: A Tutorial and Survey. **KDD** 2025.

How to Apply STFMs for ST Data Science?



Leveraging FMs for ST Data Science

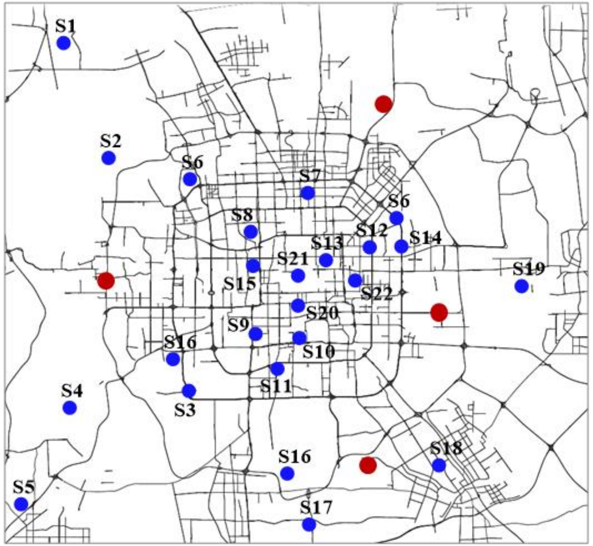
- Overview
  - ST data sensing
  - ST data management
  - ST data mining
  - Applications

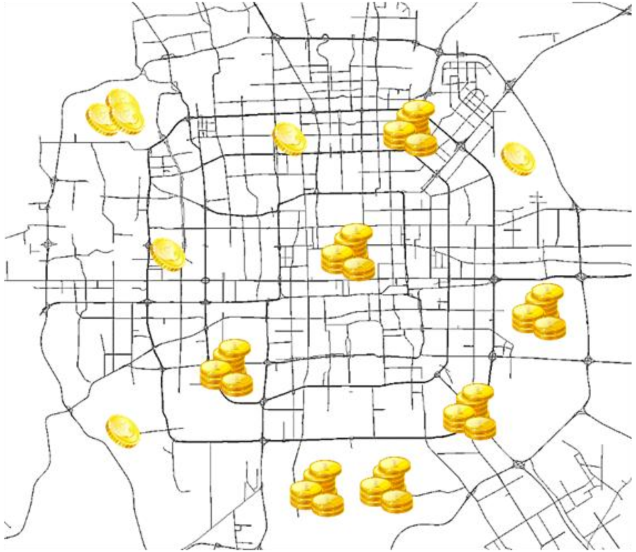


Y. Liang et al., Foundation Models for Spatio-Temporal Data Science: A Tutorial and Survey. **KDD** 2025.

ST Data Sensing: Current Challenges

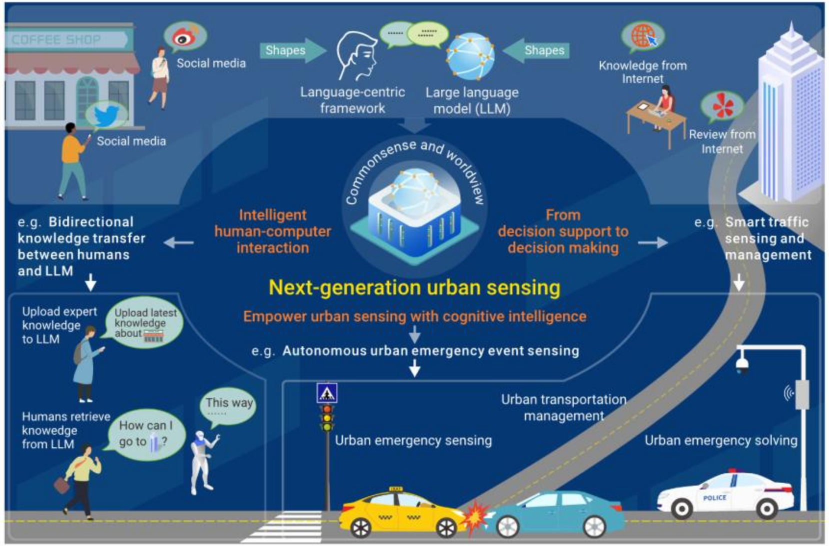
- A limited resource (budget, labors, land…)
  - Static sensing: Where to deploy sensor to maximize the gain?
  - Crowdsensing: How to arrange the incentives dynamically?



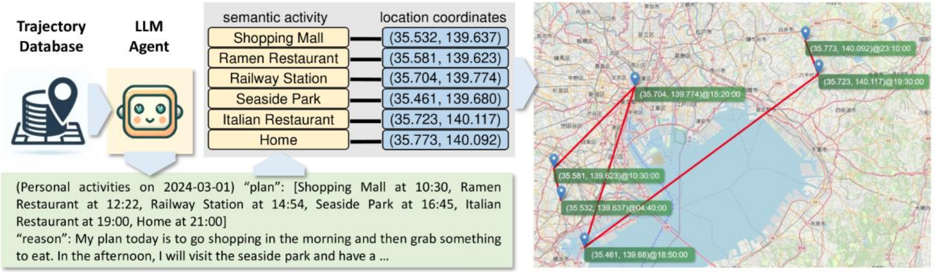


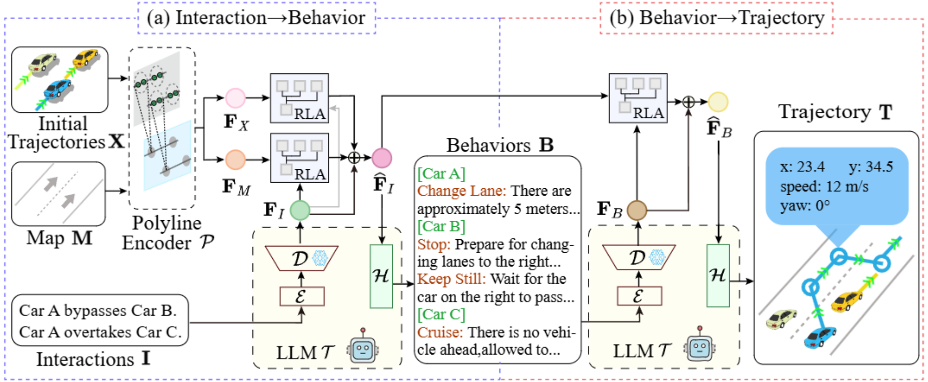
FMs for ST Data Sensing

- **Real-world data sensing** [1]



- **Synthetic data generation** [2, 3]

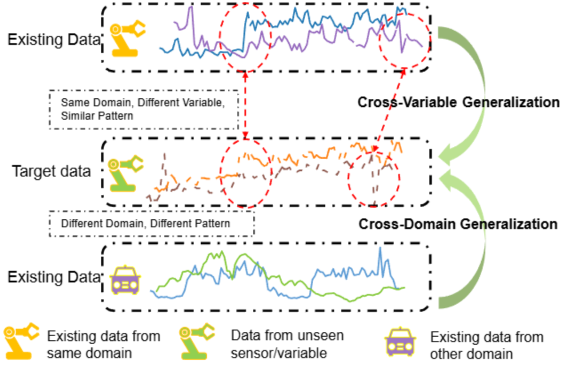




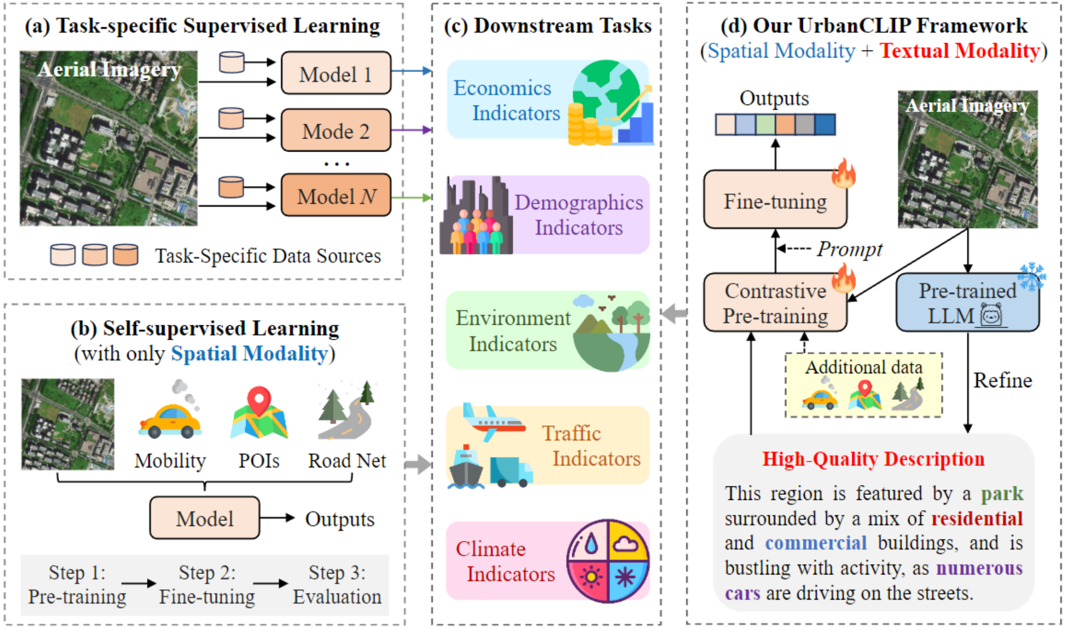
[1] C. Hou et al., Urban sensing in the era of large language models. **The Innovation** 2025.
[2] W. Wang et al., Large Language Models as Urban Residents: An LLM Agent Framework for Personal Mobility Generation. **NeurIPS** 2024.
[3] K. Yang et al., Trajectory-LLM: A Language-based Data Generator for Trajectory Prediction in Autonomous Driving. **ICLR** 2025.

FMs for ST Data Management

- **Data Cleaning**
  - Filling missing values
  - Filling missing views



(a) Cross-Variable Generalization and Cross-Domain Generalization

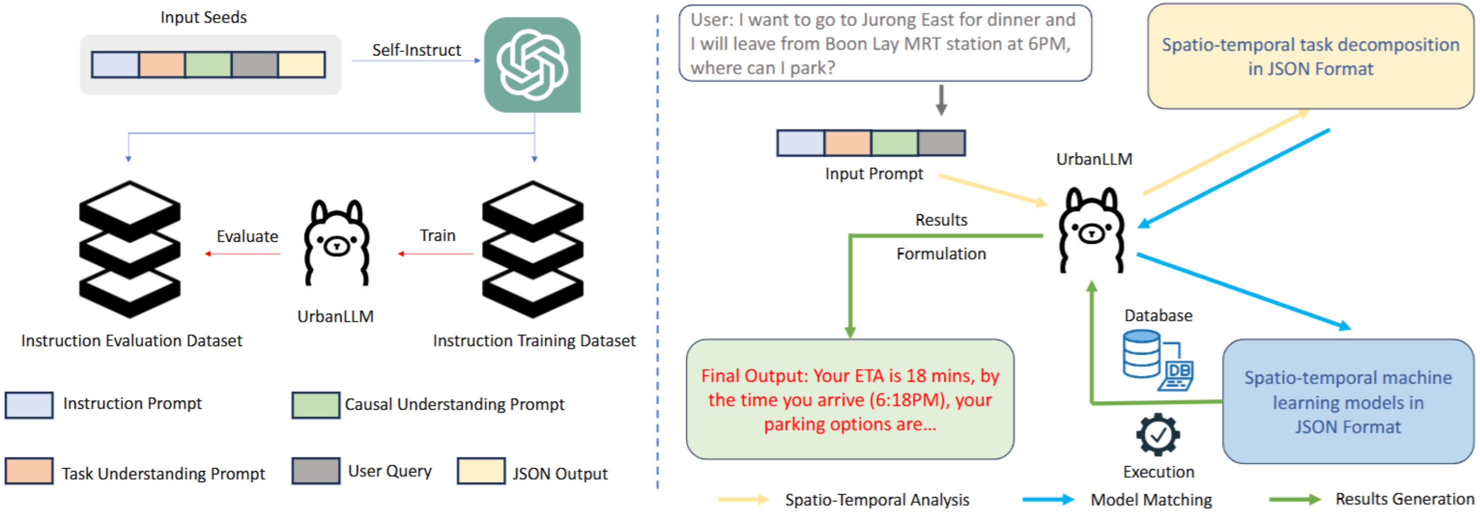


J. Cheng et al., NuwaTS: a Foundation Model Mending Every Incomplete Time Series. arXiv 2024

Y. Yan et al., UrbanCLIP: Learning Text-enhanced Urban Region Profiling with Contrastive Language-Image Pretraining from the Web. **WWW** 2024

FMs for ST Data Management

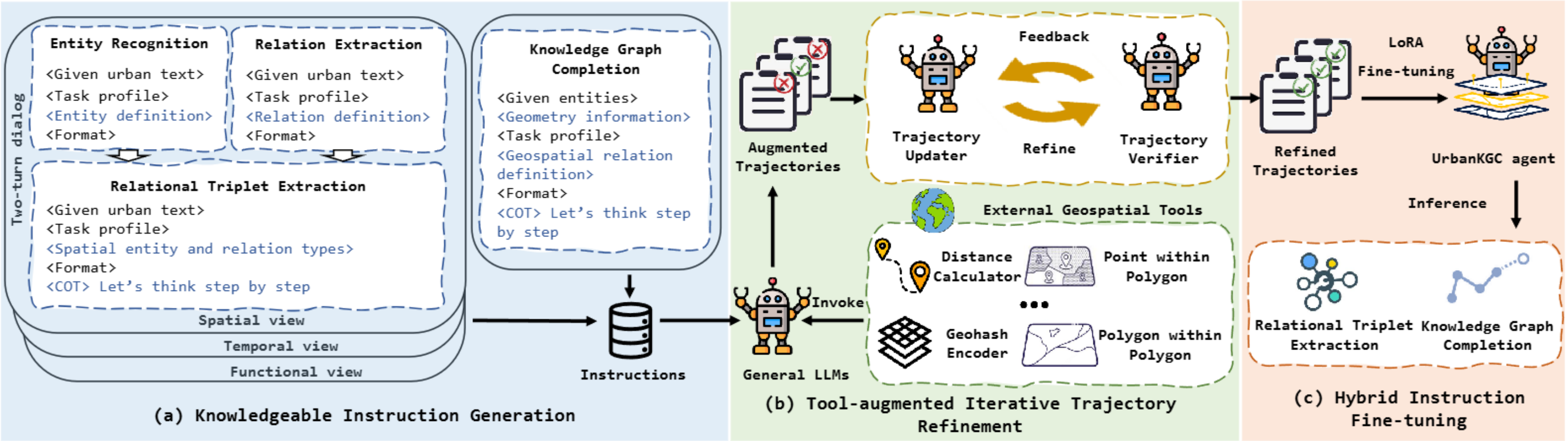
- **Querying & Retrieval**
  - E.g., UrbanLLM functions as a problem solver by decomposing urban-related queries into manageable sub-tasks, identifying suitable AI models for each sub-task, and generating comprehensive responses to the given queries.



Jiang et al., UrbanLLM: Autonomous Urban Activity Planning and Management withLarge Language Models. **EMNLP** 2024.

FMs for ST Data Management

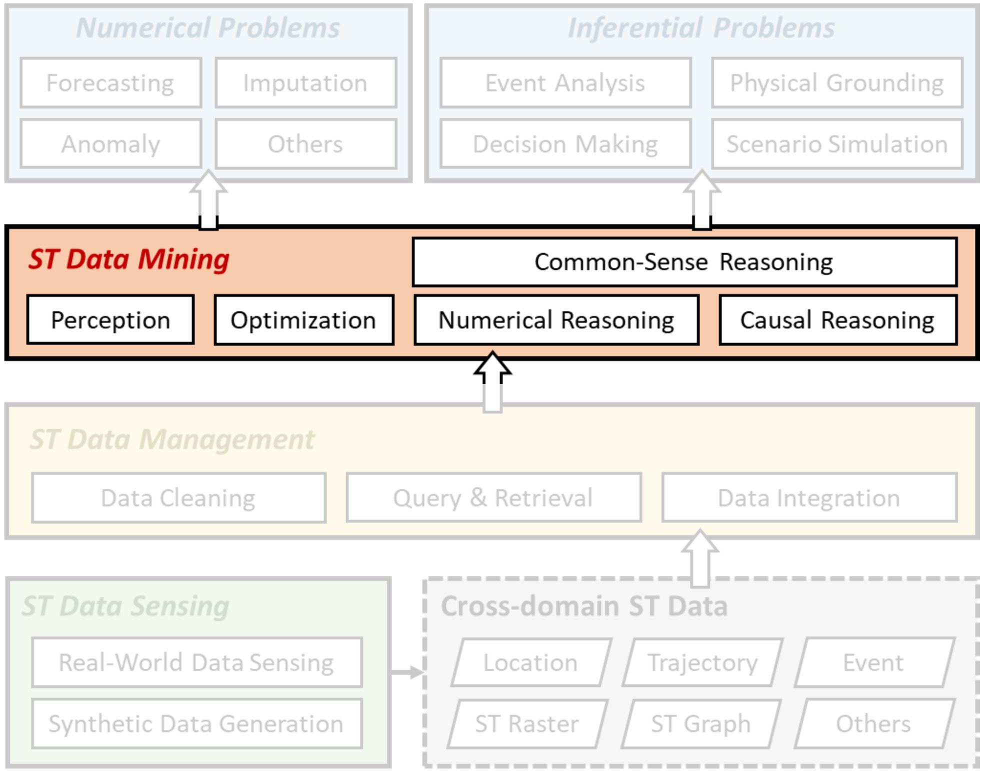
- **Data Integration**
  - It aims to combine information from disparate sources, necessitating the understanding and mapping of relationships between entities in heterogeneous datasets
  - Example: Building knowledge graphs for urban data



Y. Ning et al., UrbanKGent: A Unified Large Language Model Agent Framework for Urban Knowledge Graph Construction. **NeurIPS** 2024

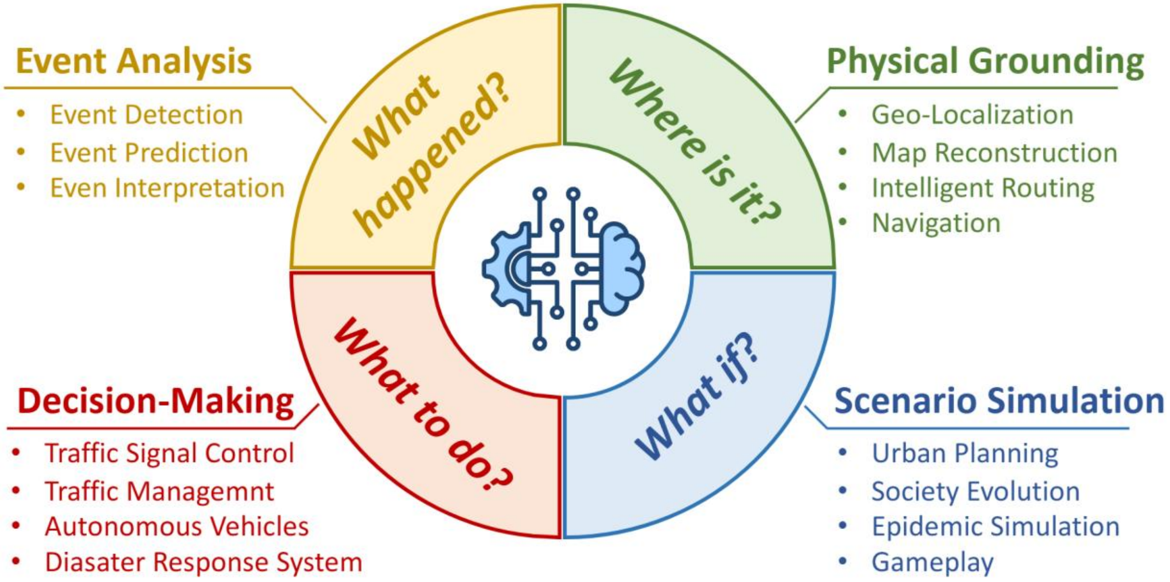
FMs for Spatio-Temporal Data Mining

- **Key capabilities**
  - Perception
  - Optimization
  - Reasoning
    - Common-sense reasoning
    - Numerical reasoning
    - Causal reasoning

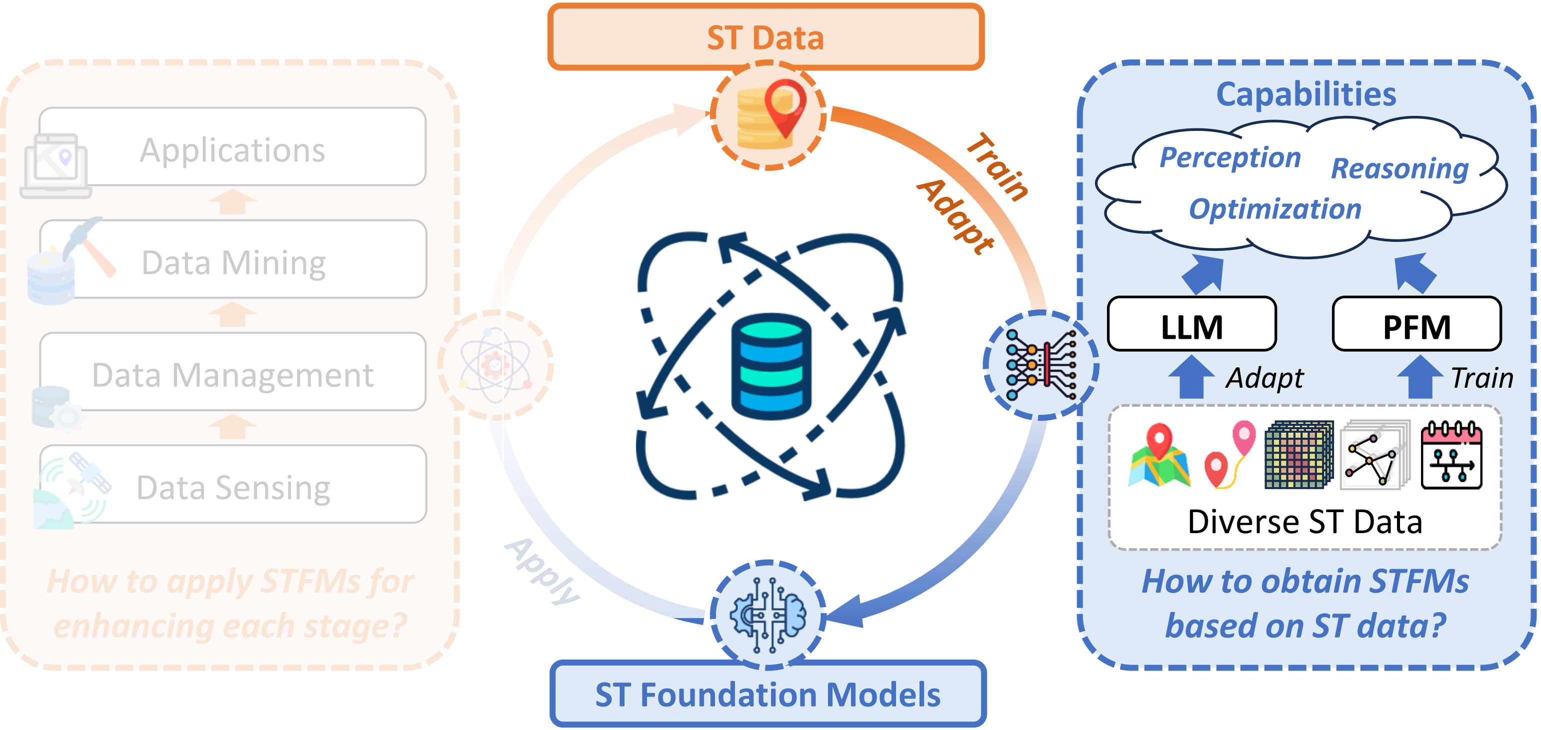


Downstream Applications

- **Numerical problems**
  - E.g., forecasting, imputation, anomaly detection
- **Inferential problems**

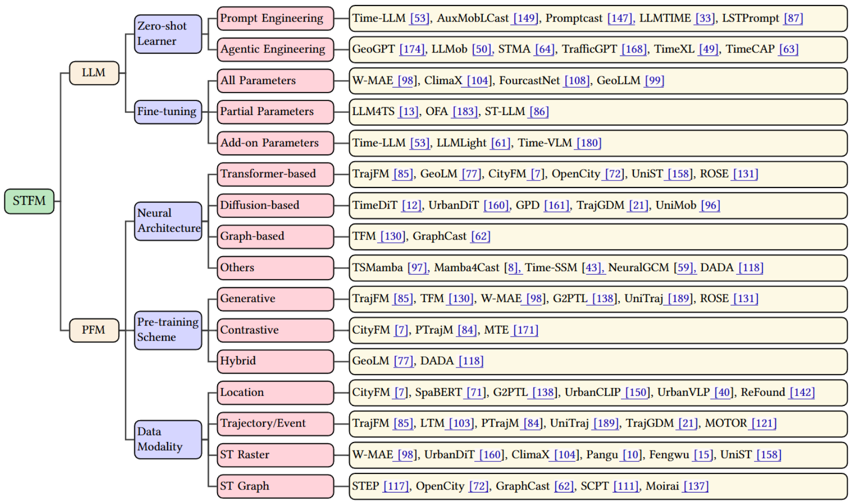


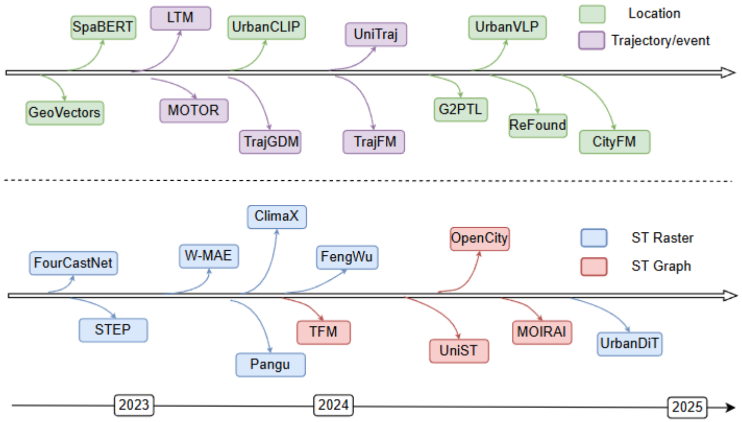
How to obtain STFMs based on ST Data?



Two Ways of STFMs

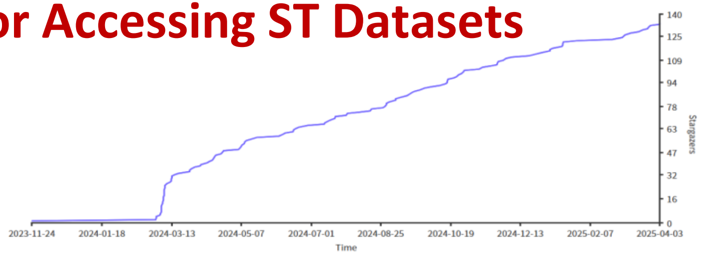
- By harnessing Large Language Models (**LLMs**), it becomes possible to develop more generalized, adaptable solutions that can be fine-tuned for specific tasks with minimal data.
- Another prominent approach involves Pretrained Foundation Models (**PFMs**) on cross-domain ST data and adapting them for particular domains.

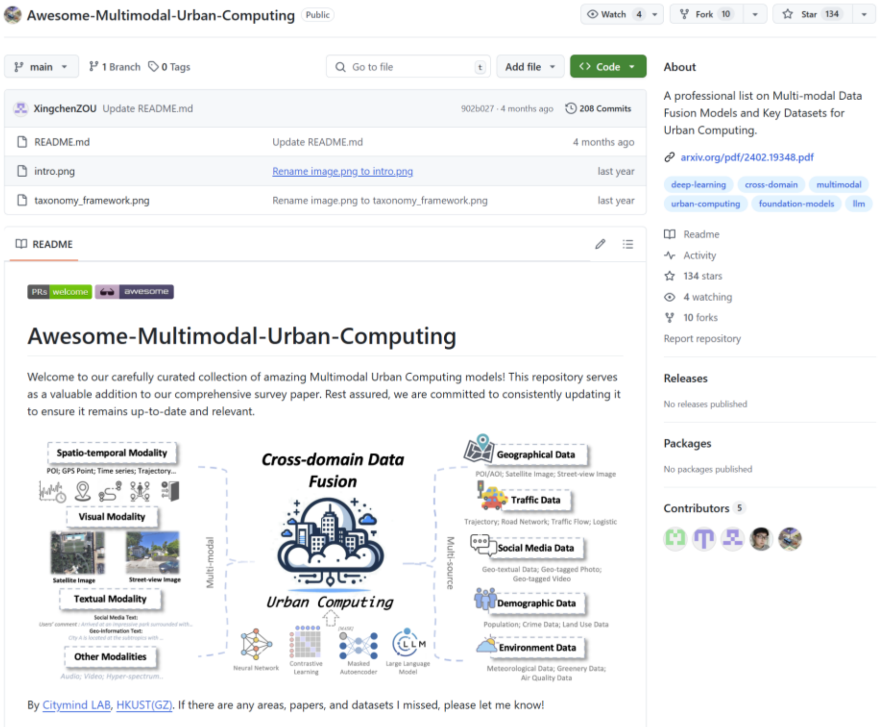




Y. Liang et al., Foundation Models for Spatio-Temporal Data Science: A Tutorial and Survey. **KDD** 2025

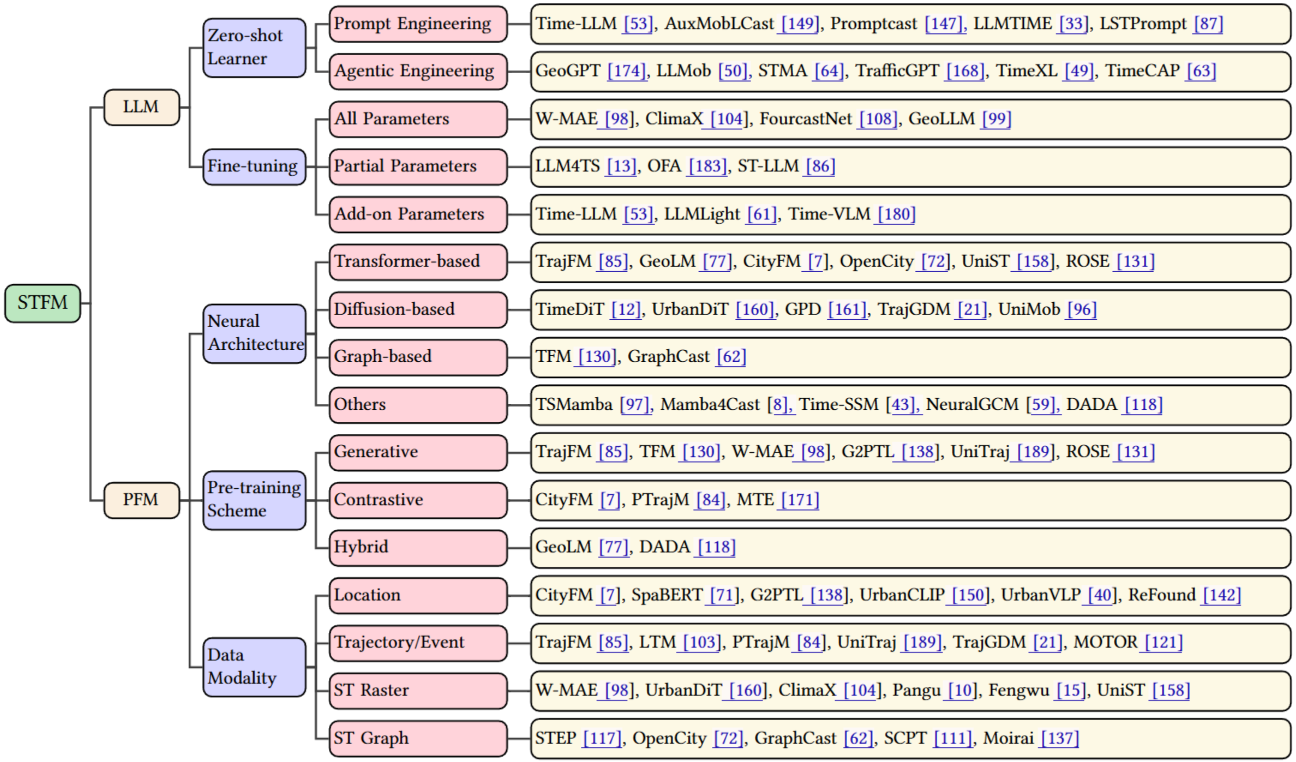
Scan it for Accessing ST Datasets





Category	Content	Format	Dataset	Link	Reference
Geographical Data	Satellite Image	Image	ArcGIS	https://developers.arcgis.com	[186]
			PlanetScope	https://developers.planet.com/docs/data/planetscope/	[154]
			Google Earth	https://developers.google.com/maps/documentation/	[116]
			OpenStreetMap	https://www.openstreetmap.org/	[337]
			Baidu Maps	https://lbsyun.baidu.com	[324, 313]
	Street-View Image	Image	Baidu Map	https://lbsyun.baidu.com	[186, 124]
			Google Street	https://developers.google.com/maps/	[186, 4]
			Tencent Map	https://lbs.qq.com/tool/streetview/index.html	[112]
	POIs	Point Vector	Tencent Map Service	https://lbs.qq.com/getPoint/	[309, 235]
			WeChat POIs	https://open.weixin.qq.com	[277]
			Baidu Map POIs	https://lbsyun.baidu.com	[154, 172, 175, 110, 313]
			NYC Open POIs	https://opendata.cityofnewyork.us/	[170, 272, 20, 366, 288]
			Foursquare	https://developer.foursquare.com/docs/checkins/checkins	[20, 381, 13, 42, 107, 116]
			Wikipedia POIs	https://www.wikipedia.org	[386]
			AMap Service	https://lbs.amap.com	[10]
			Yelp POIs	https://www.yelp.com/developers	[13, 380, 383]
			Dianping POIs	https://api.dianping.com/	[33, 63]
			Weibo POIs	https://open.weibo.com/wiki/API	[33, 134, 77]
			Flickr POIs	https://www.flickr.com/services/developer/api/	[99]
			Bing Map POIs	https://www.bingmapsportal.com	[37]
Traffic Data	Traffic Trajectory	Spatio-temporal Trajectory	Shenzhou UCar	https://bit.ly/2MG47xz	[93]
			Chicago Transportation	https://data.cityofchicago.org/	[272, 288, 116]
			VED	https://github.com/gsoh/VED	[209, 372]
			Taxi Shenzhen	https://github.com/cbdog94/STL	[113, 302]
			NYC Open Taxi Data	https://opendata.cityofnewyork.us/how-to/	[368, 366]
			GeoLife	http://urban-computing.com/index-893.htm	[96, 398, 400, 394, 347]
			T-Drive Taxi	http://urban-computing.com/index-58.htm	[350, 351, 217, 191]
			DiDi Traffic	https://outreach.didichuxing.com/research/opendata/	[349, 188, 228, 328, 261]
			Xiamen Taxi	https://data.mendeley.com/datasets/6xg39x9vgd/1	[342, 40, 124, 39]
			Grab-Posisi	https://goo.su/W3yD5m	[337, 339]
	Taffic Flow	Spatio-temporal Graph	California-PEMS	http://pems.dot.ca.gov	[9, 254]
			METR-LA	https://www.metro.net	[143, 171]
			Large-ST	https://github.com/liuxu77/LargeST	[182]
			MobileBJ	https://github.com/FIBLAB/DeepSTN/issues/4	[170, 134, 33]
			TaxiBJ	https://goo.su/aQyjTAz	[164, 11, 226, 120, 368, 74]
			BikeNYC	https://citibikenyc.com/	[170, 11, 226, 120]
	Road Network	Spatial Graph	OpenStreetMap	https://www.openstreetmap.org	[339, 13, 188, 349, 84]
			US Census Bureau	https://www.census.gov/data.html	[366]
	Logistics	Spatio-temporal Trajectory	LaDe	https://cainiaotechai.github.io/LaDe-website/	[305]
			JD Logistics	https://corporate.jd.com/ourBusiness#jdLogistics	[235]
Social Media Data	Text	Text	Twitter	https://developer.twitter.com/en/docs	[20, 381, 383, 352, 270, 301, 240] [289, 283, 285, 284, 200, 184]
			Common Crawl	https://registry.opendata.aws/commoncrawl/	[380]
			Yelp Reviews	https://www.yelp.com/dataset	[380, 383]
			Weibo Traffic Police	http://open.weibo.com/developers/	[342]
	Geo-tagged Image & Video	Image&Video	YFCC100M	https://goo.su/jzaDU	[386, 340, 99]
			NUS-WIDE	https://goo.su/dWPQZcD	[340, 338]
			GeoUGV	https://qualinet.github.io/databases/video/	[187]
	Users' Info	Time Series	Jiepang User Check-in	https://jiepang.app/	[74]
			Gowalla User Location	http://konect.cc/networks/loc-gowalla_edges/	[42, 352]
			WeChat Mobility	https://open.weixin.qq.com/	[277]
Demographic Data	Crime	Time Series	NYC Crime	https://opendata.cityofnewyork.us/	[368]
	Land Use	Time Series	Land Use SG	https://www.ura.gov.sg/Corporate/Planning/Master-Plan	[156]
			Land Use NYC	https://goo.su/puTuG	[156]
	Population	Time Series	WorldPop	https://www.worldpop.org/	[309, 154, 10]
Environment Data	Meteorology	Time Series	TipDM China Weather	https://www.tipdm.org/	[178]
			DarkSky Weather	https://support.apple.com/en-us/102594	[349]
			WeatherNY	https://opendata.cityofnewyork.us/	[272]
			WeatherChicago	https://data.cityofchicago.org/	[272]
			Weather Underground	https://www.wunderground.com/	[342]
			DidiSY	https://www.didiglobal.com/	[12]
			WD_BJ weather	https://goo.su/DmHFHd	[192]
			WD_USA weather	https://goo.su/RVhBA	[192]
	Greenery	Time Series	Google Earth	https://earth.google.com/	[342]
	Air Quality	Time Series	UrbanAir	https://goo.su/hfzNB53	[399, 396, 392]
			KnowAir	https://github.com/shuowang-ai/PM2.5-GNN	[286, 346, 370, 318]

Taxonomy on Methodology

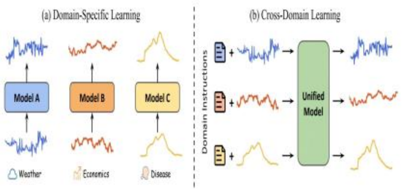


Y. Liang et al., Foundation Models for Spatio-Temporal Data Science: A Tutorial and Survey. **KDD** 2025.

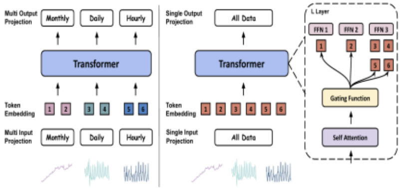
Our Methodologies

PFMs

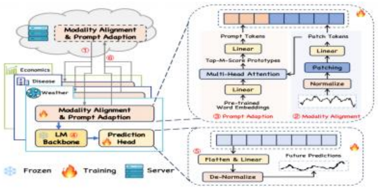
LLMs



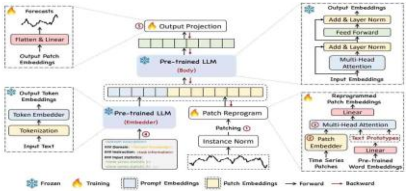
UniTime [WWW’24]



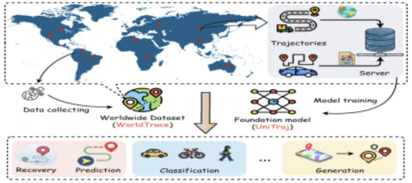
Moirai-MoE [ICML’25]



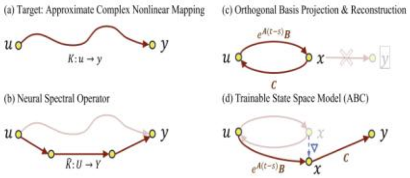
Time-FFM [NeurIPS’24]



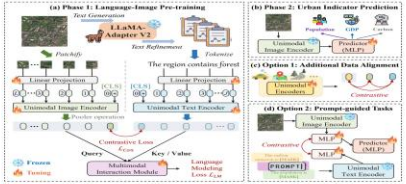
Time-LLM [ICLR’24]



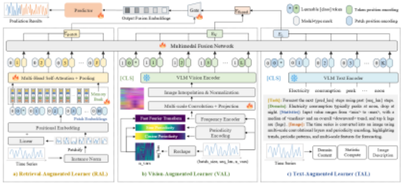
UniTraj [arXiv’24]



Time-SSM [arXiv’24]



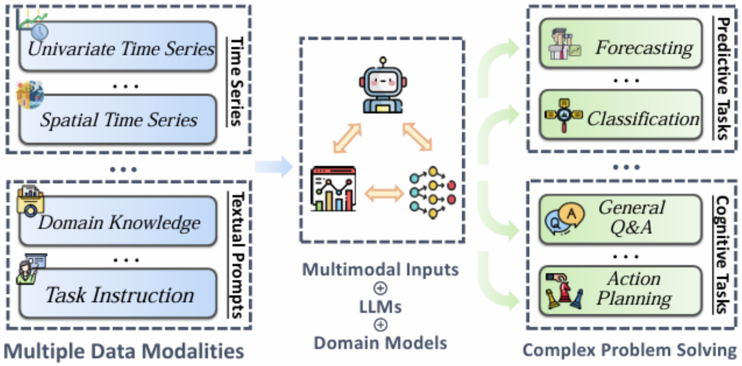
UrbanCLIP [WWW’24]



Time-VLM [ICML’25]

LLMs for ST Data Science

Survey on LLMs



[ICML'24]

M. Jin, Y. Zhang, W. Chen et al., Position: What Can Large Language Models Tell Us about Time Series Analysis. **ICML** 2024.

LLMs for ST Data Science

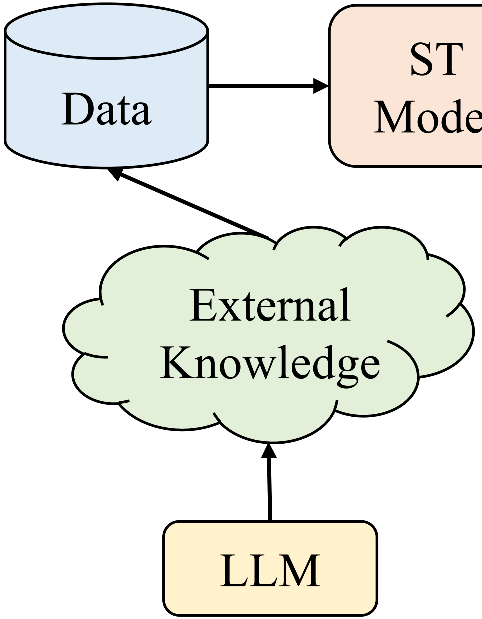
- Our standpoint is that LLMs can serve as the central hub for understanding and advancing ST Data Science in three principal ways
  - LLM-as-Enhancers: augmenting ST data and existing approaches with enhanced external knowledge and analytical prowess
  - LLM-as-Predictors: utilizing their extensive internal knowledge and emerging reasoning abilities to benefit a range of downstream tasks, e.g., forecasting
  - LLM-as-Agents: transcending conventional roles to actively engage in and transform spatio-temporal data mining

LLM-as-Enhancers

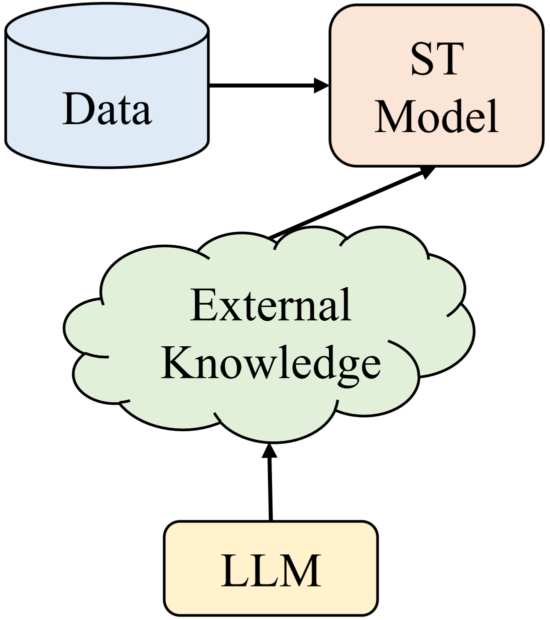
- Our standpoint is that LLMs can serve as the central hub for understanding and advancing ST Data Science in three principal ways
  - LLM-as-Enhancers: augmenting ST data and existing approaches with enhanced external knowledge and analytical prowess
  - LLM-as-Predictors: utilizing their extensive internal knowledge and emerging reasoning abilities to benefit a range of downstream tasks, e.g., forecasting
  - LLM-as-Agents: transcending conventional roles to actively engage in and transform spatio-temporal data mining

Categories of LLM-as-Enhancers

- Data-centric enhancer



- Model-centric enhancer



Data-Centric Enhancers: An Example in Finance

- This study focuses on NASDAQ-100 stock price prediction, supplemented by metadata about the stock company and relevant financial news data

Prompt: Generate a short description for stock AAPL' s company. Also list general positive and negative factors that might impact the stock price; be brief and use keywords. Consider diverse general factors, such as macro economic situation (e.g. inflation, CPI growth), business factors (e.g. sales, investment, products), technology factors (e.g. innovation), and others. Use format Description: ..., Positive Factors: ..., Negative factors: ...

GPT-4 Response:
Description:
AAPL (Apple Inc.) is a world-renowned technology company based in Cupertino, California. Known for its sleek designs and innovative technologies, the company produces a wide range of products including the iPhone, iPad, Mac, Apple Watch, and services like iCloud, App Store, and more.

Positive Factors:
Macroeconomic: Economic growth, increased consumer spending
Business: Strong sales, expansion into new markets, positive earnings reports, strategic acquisitions
Technology: Innovation, new product launches, advancements in artificial intelligence and augmented reality
Brand: Strong global brand, loyal customer base
Legal/Regulatory: Favorable policy changes, successful patent applications

Negative Factors:
Macroeconomic: Inflation, economic downturn, trade restrictions
Business: Decreased sales, poor earnings reports, lack of new investments
Technology: Slow pace of innovation, delays in product launches
Brand: Damage to brand reputation, loss of customer loyalty
Legal/Regulatory: Antitrust investigations, unfavorable policy changes, patent disputes
Environmental: Failure to meet sustainability goals, negative environmental impact

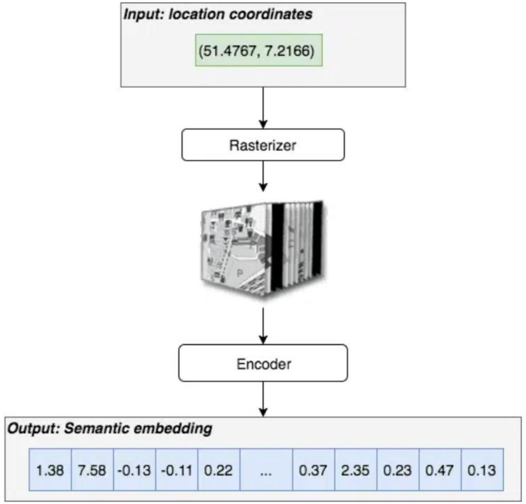
An example of a stock's company profile consisting of the company description

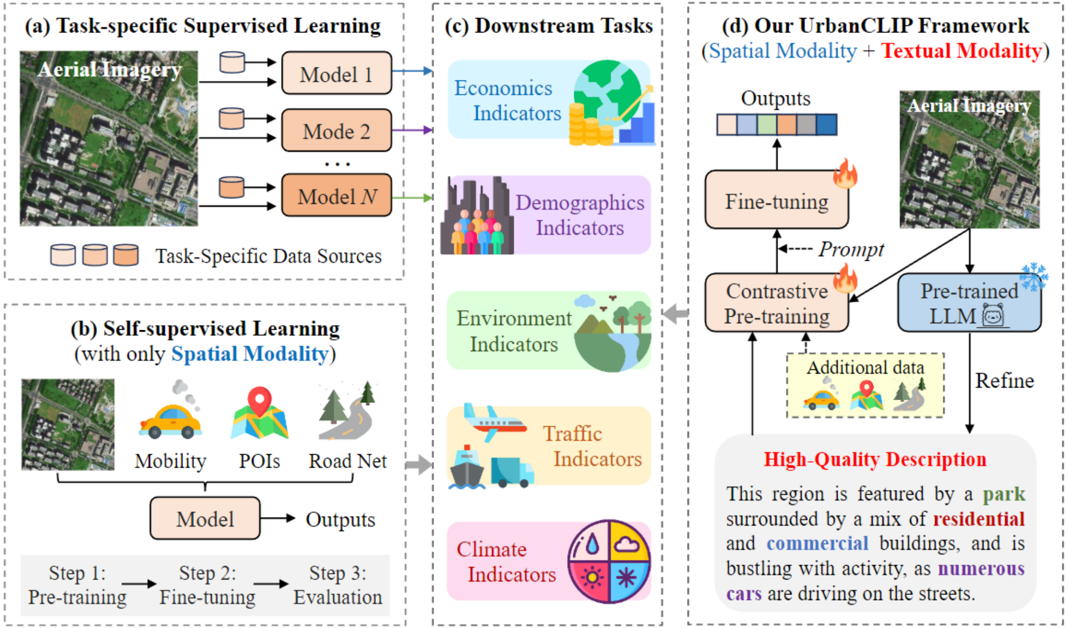
X. Xu et al., Temporal Data Meets LLM -Explainable Financial Time Series Forecasting. arXiv 2023.

Model-Centric Enhancers: UrbanCLIP as Example

- **Urban Contrastive Language-Image Pre-training** (UrbanCLIP) is the first framework that integrates the knowledge of text modality into **urban region profiling**

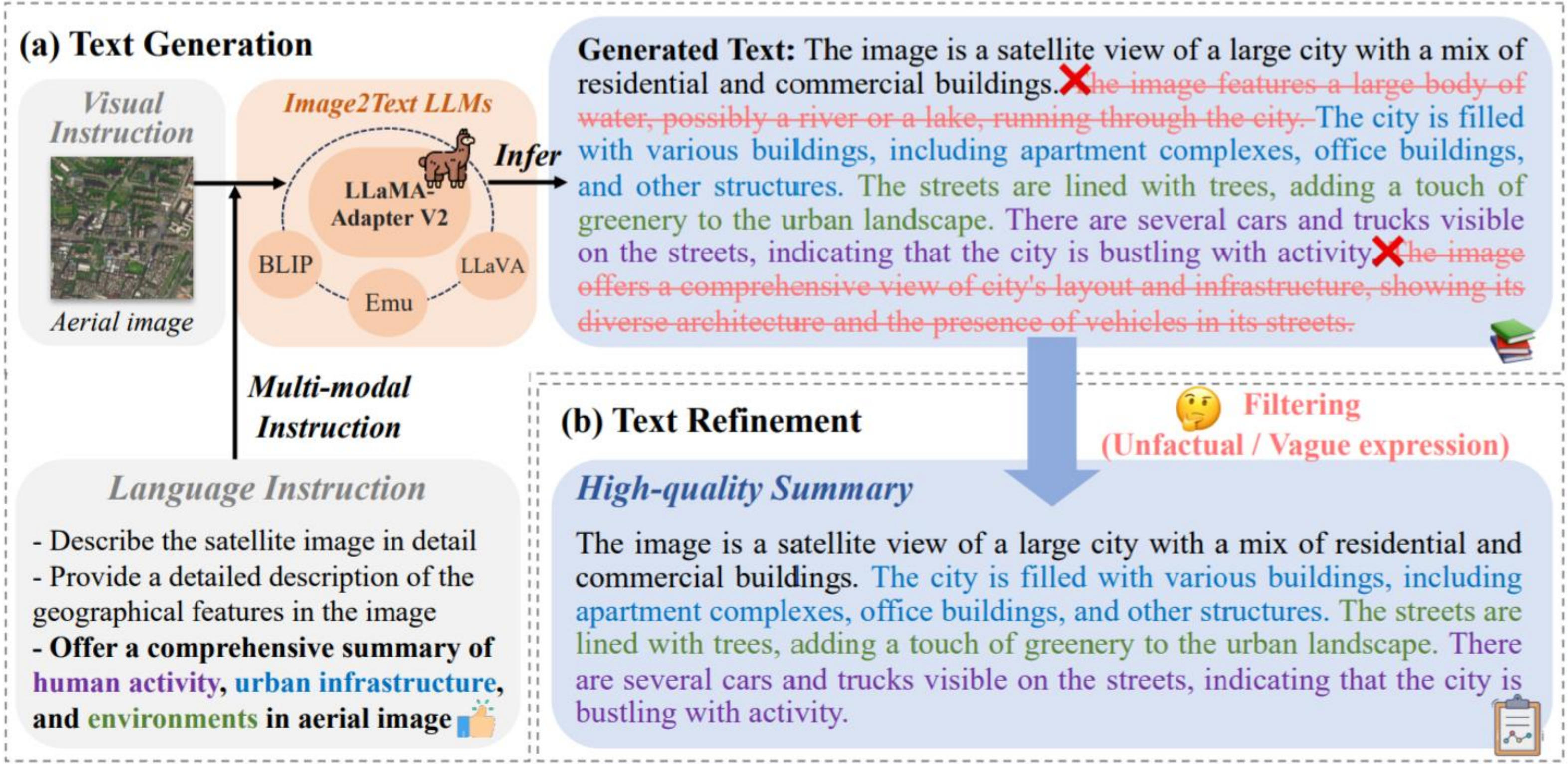
Definition





Y. Yan et al., UrbanCLIP: Learning Text-enhanced Urban Region Profiling with Contrastive Language-Image Pretraining from the Web. **WWW** 2024.

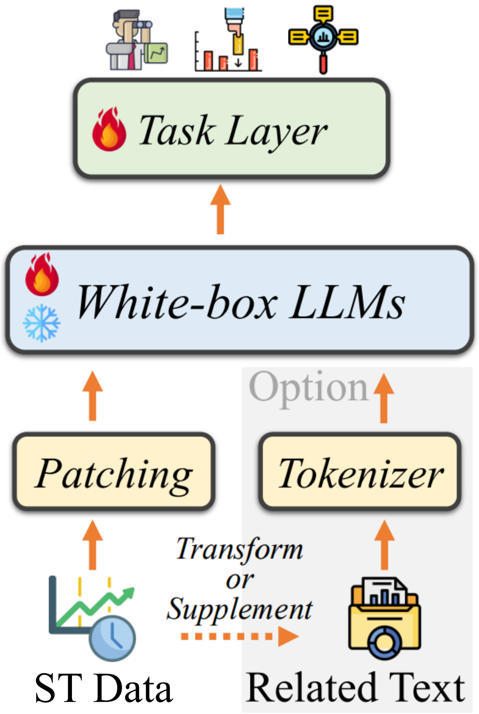
Text Generation & Refinement



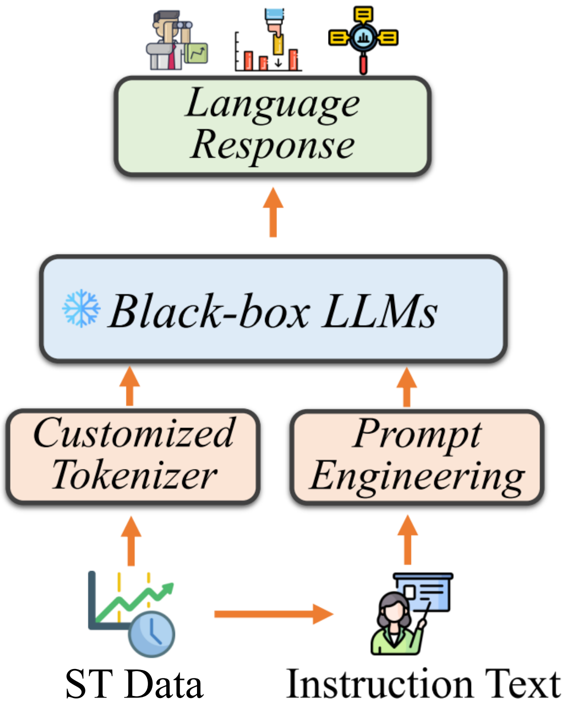
LLM-as-Predictors

- Our standpoint is that LLMs can serve as the central hub for understanding and advancing ST Data Science in three principal ways
  - LLM-as-Enhancers: augmenting ST data and existing approaches with enhanced external knowledge and analytical prowess
  - LLM-as-Predictors: utilizing their extensive internal knowledge and emerging reasoning abilities to benefit a range of downstream tasks, e.g., forecasting
  - LLM-as-Agents: transcending conventional roles to actively engage in and transform spatio-temporal data mining

Categories of LLM-as-Predictors



(a) Tuning-Based

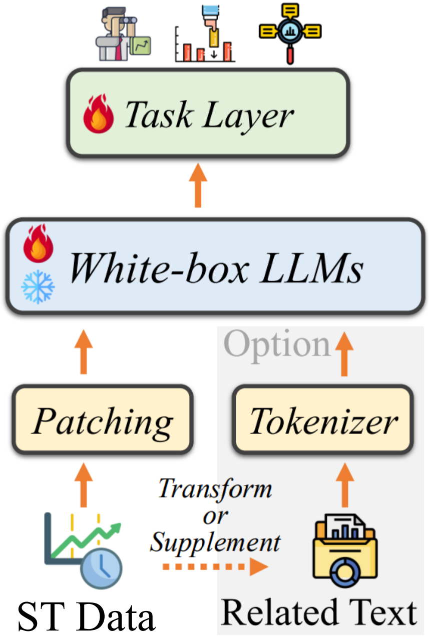


(b) Non-Tuning-Based

Tuning-based LLM Predictors

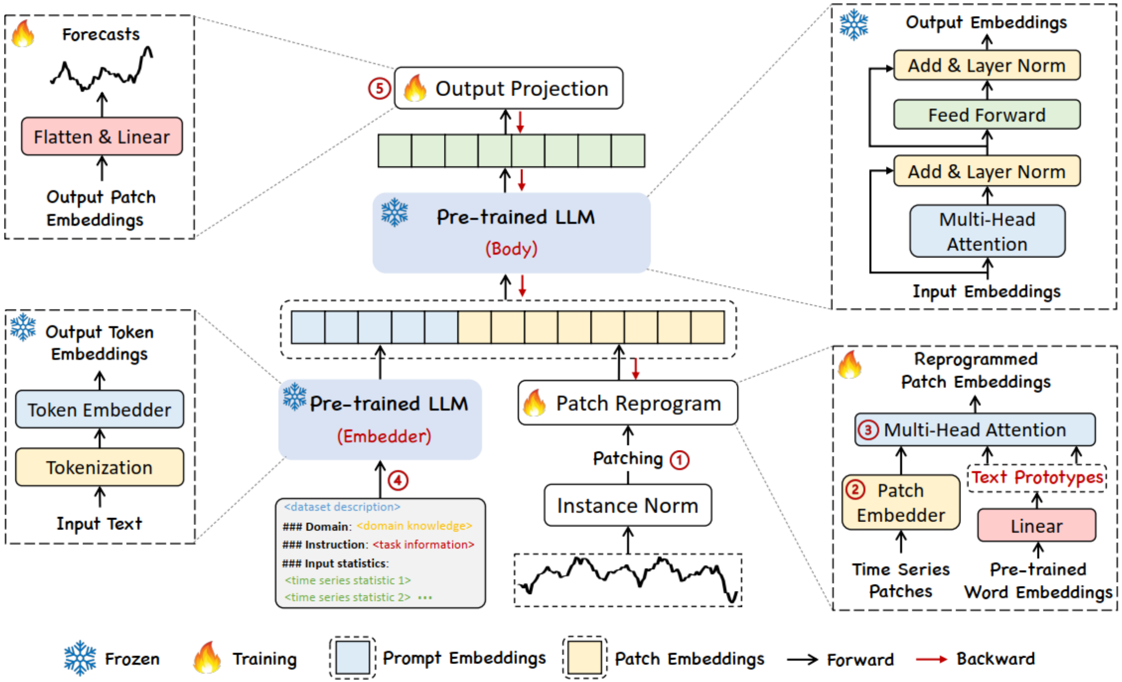
- Tuning-based predictors use accessible LLM parameters, typically involving **patching** and **tokenizing** numerical signals and related text data, followed by fine-tuning for ST data

- Examples
  - Time-LLM
  - Time-VLM



Time-LLM for Time Series Analysis

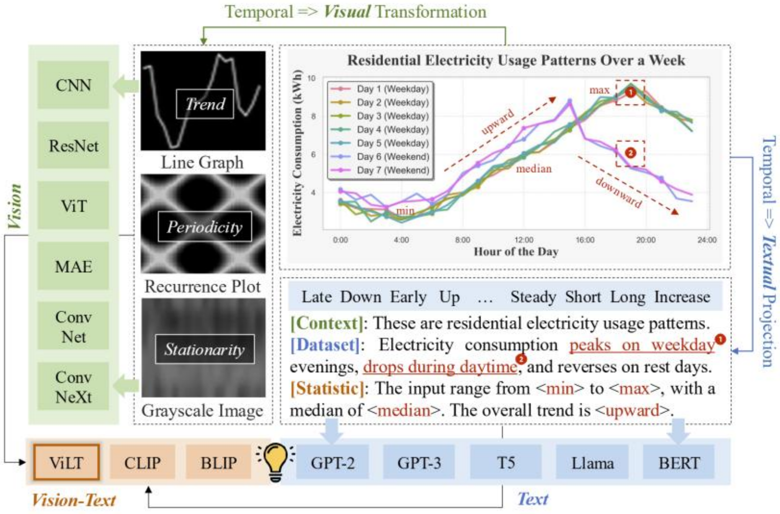
- Key question: **How to enable LLMs to understand time series?**

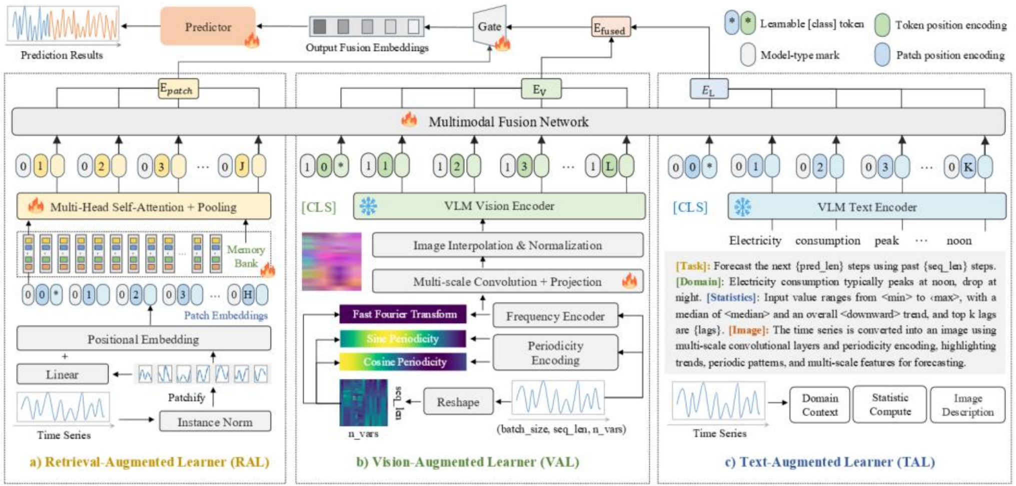


M. Jin et al., Time-LLM: Time Series Forecasting by Reprogramming Large Language Models. **ICLR** 2024.

Time-VLM for Multimodal Learning

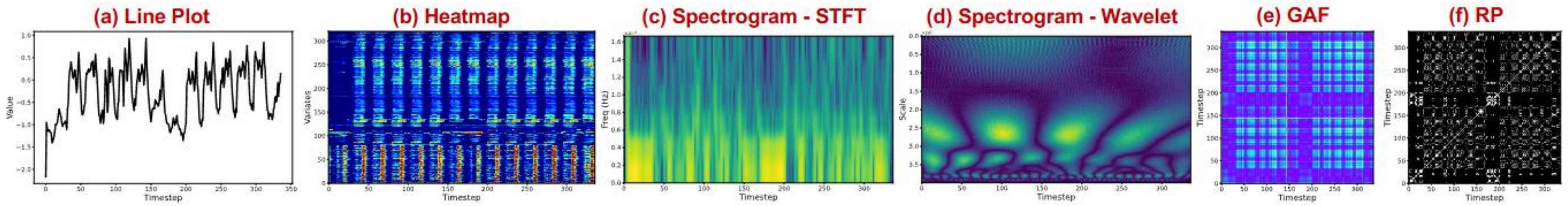
- It presents the first attempt to leverages **pretrained Vision-Language Models** (VLMs) to bridge **temporal**, **visual**, and **textual** modalities for time series analysis

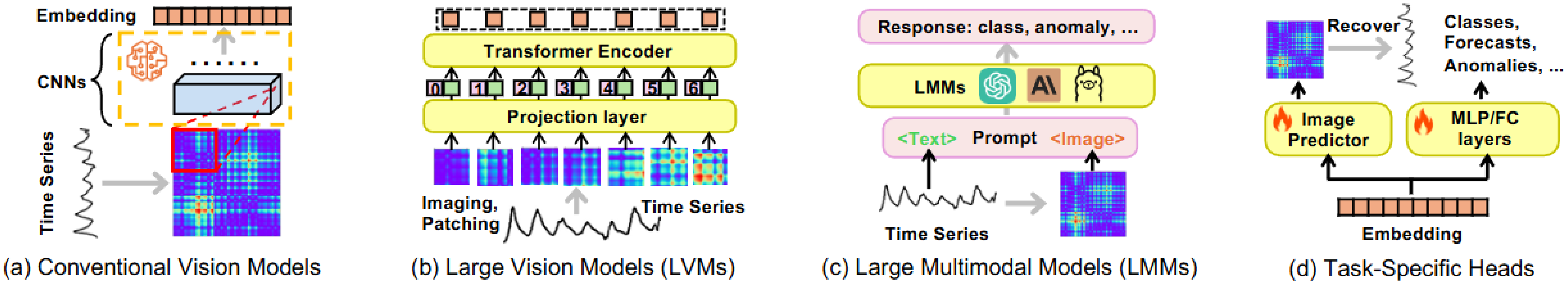




S. Zhong et al., Time-VLM: Exploring Multimodal Vision-Language Models for Augmented Time Series Forecasting. **ICML** 2025.

Supplementary on Vision-Based TS Modeling

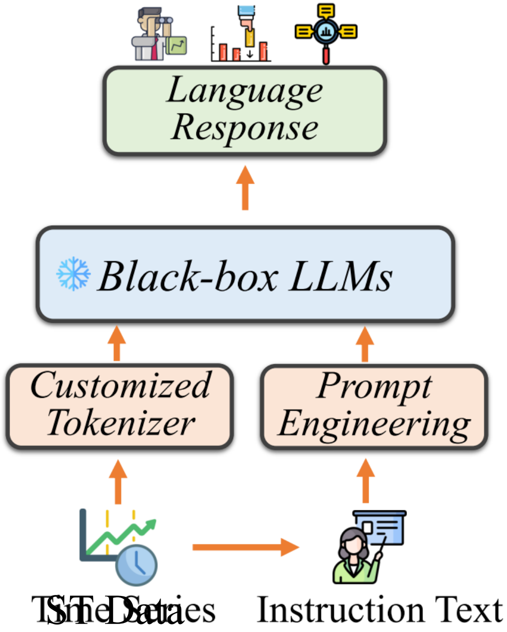




J. Ni et al., Harnessing Vision Models for Time Series Analysis: A Survey. arXiv 2025.

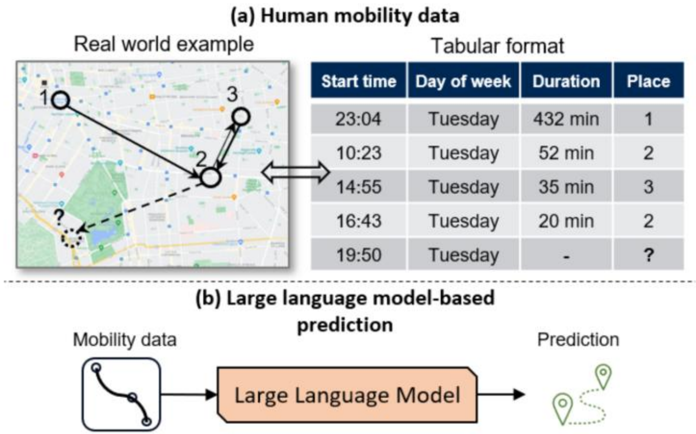
Non-Tuning-based LLM Predictors

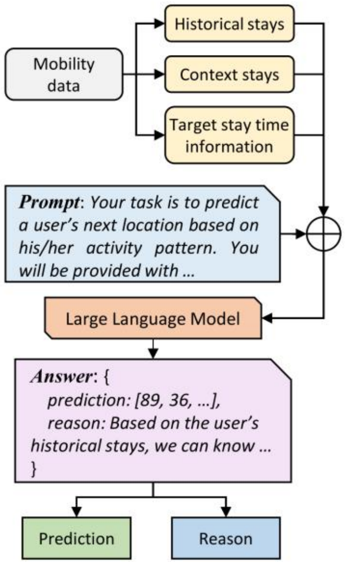
- Non-tuning-based predictors, suitable for closed-source models, involve preprocessing ST data to fit LLM input spaces
  - Tokenizer
  - Prompt design
  - In-context learning
- Example
  - LLM-Mob for human mobility



Example: LLM-Mob

- LLM-Mob leverages the language understanding and reasoning capabilities of LLMs for analyzing human mobility data





Instruction | Specify the task Your task is to predict a user's next location based on his/her activity pattern.

Data | Describe the data You will be provided with <history> which is a list containing this user's historical stays, then <context> which provide contextual information about where and when this user has been to recently. Stays in both <history> and <context> are in chronological order. Each stay takes on such form as (start_time, day_of_week, duration, place_id). The detailed explanation of each element is as follows:
start_time: the start time of the stay in 12h clock format.
day_of_week: indicating the day of the week.
duration: an integer indicating the duration (in minute) of each stay. Note that this will be None in the <target_stay> introduced later.
place_id: an integer representing the unique place ID, which indicates where the stay is.
Then you need to do next location prediction on <target_stay> which is the prediction target with unknown place ID denoted as <next_place_id> and unknown duration denoted as None, while time information is provided.

Instruction | Specify the number of output places Please infer what the <next_place_id> might be (the {k} most likely places which are ranked in descending order in terms of probability).

Instruction | Guide the model to "think" Please consider the following aspects:
1. the activity pattern of this user that you learned from <history>, e.g., repeated visits to certain places during certain times;
2. the context stays in <context>, which provide more recent activities of this user;
3. the temporal information (i.e., start_time and day_of_week) of target stay, which is important because people's activity varies during different time (e.g., nighttime versus daytime) and on different days (e.g., weekday versus weekend).

Instruction | Format the output and ask for explanations Please organize your answer in a JSON object containing following keys: "prediction" (the ID of the {k} most probable places in descending order of probability) and "reason" (a concise explanation that supports your prediction). Do not include line breaks in your output.

Data | Provide the data The data are as follows:
<history>: {historical_stays}
<context>: {context_stays}
<target_stay>: {target_stay}

X. Wang et al., Where Would I Go Next? Large Language Models as Human Mobility Predictors. arXiv 2024.

LLM-as-Agents

- Our standpoint is that LLMs can serve as the central hub for understanding and advancing ST Data Science in three principal ways
  - LLM-as-Enhancers: augmenting ST data and existing approaches with enhanced external knowledge and analytical prowess
  - LLM-as-Predictors: utilizing their extensive internal knowledge and emerging reasoning abilities to benefit a range of downstream tasks, e.g., forecasting
  - LLM-as-Agents: transcending conventional roles to actively engage in and transform spatio-temporal data mining

Limitations of LLM-as-Predictors

- **Tuning-based LLM-as-Predictors** utilize LLMs as robust model checkpoints, attempting to adjust certain parameters for specific domain applications.
  - However, this approach often sacrifices the interactive capabilities of LLMs and may not fully exploit the benefits offered by LLMs, such as in-context learning or chain-of-thought.
- **Non-tuning-based LLM-as-Predictors**, integrating ST data into textual formats or developing specialized tokenizers
  - Facing limitations due to LLMs' primary training on linguistic data, hindering their comprehension of complex ST knowledge and patterns not easily captured in language

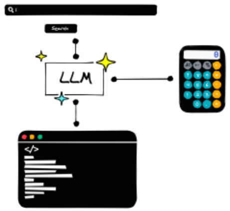
A new promising paradigm rises:

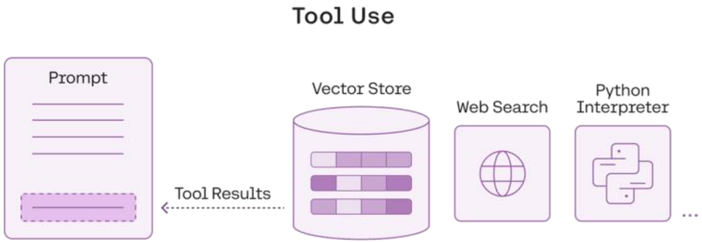
LLM-as-Agents!

LLMs Open New Opportunities

- LLMs are good at **processing multi-modal data**
  - Multi-modal data understanding
    - e.g., vision, texts, time series
  - Multi-format data parsing
    - tabular, json, images
  - Example: Qwen-VL

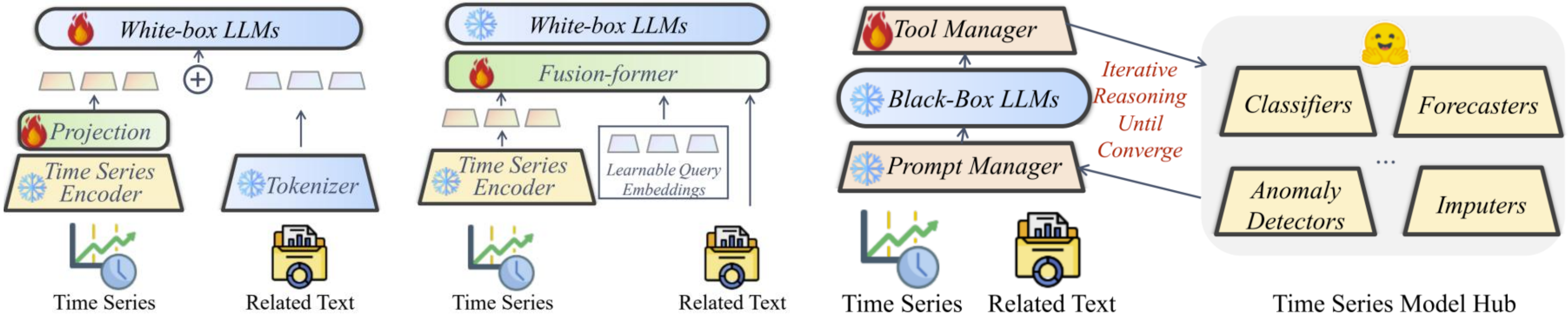
- **Tool calling capability** (LLM Agents)





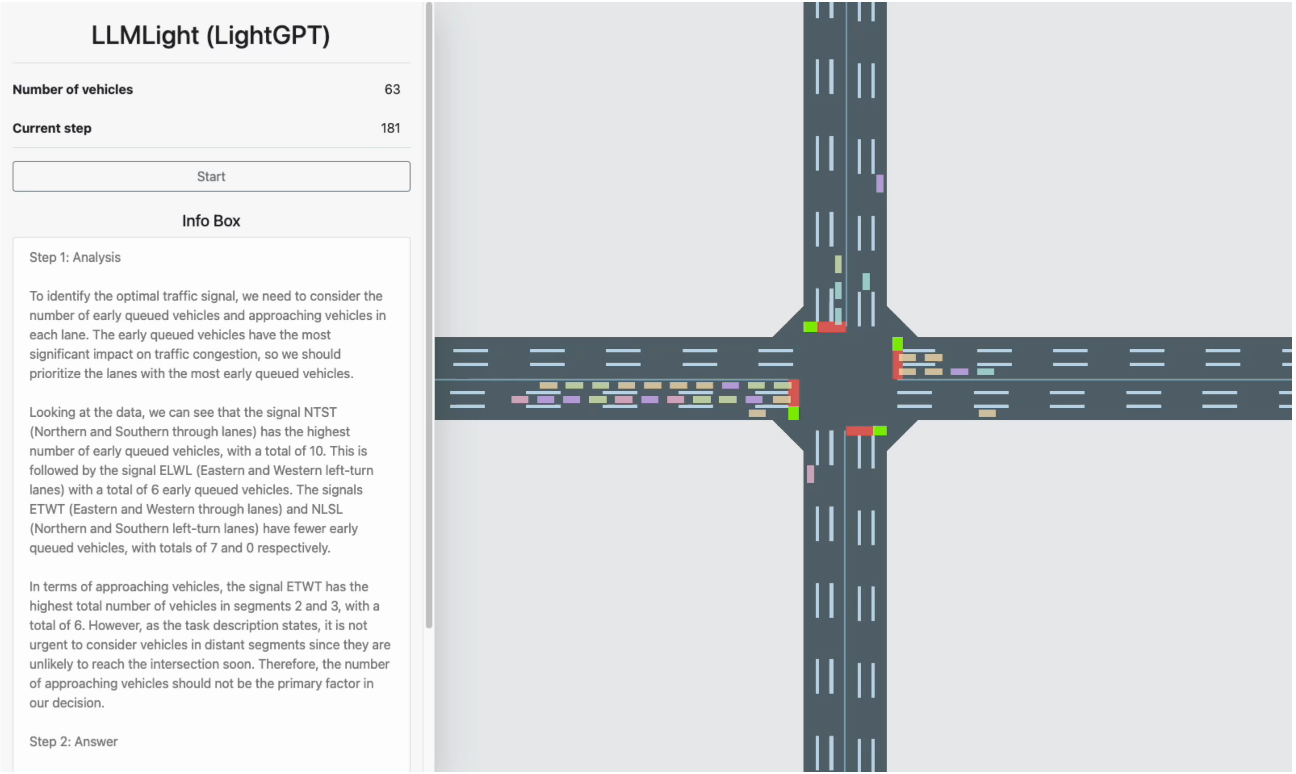
LLM-as-Agents

- Categories
  - **Problem-solving agents**
  - **Simulation-based agents**
- Key Capabilities of LLM Agents



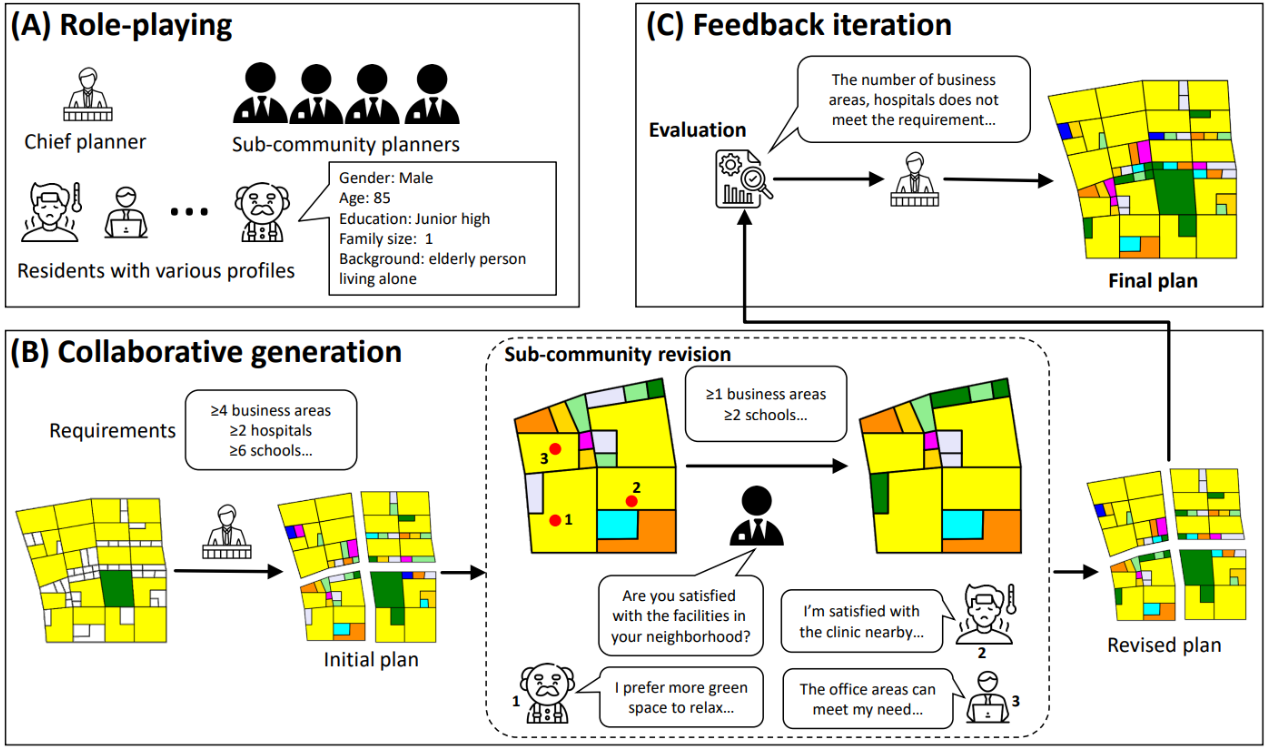
(a) Aligning (b) Fusion (c) Using External Tools

Problem-Solving Agents: Traffic Light Control



S. Lai et al., LLMLight: Large Language Models as Traffic Signal Control Agents. **KDD** 2025.

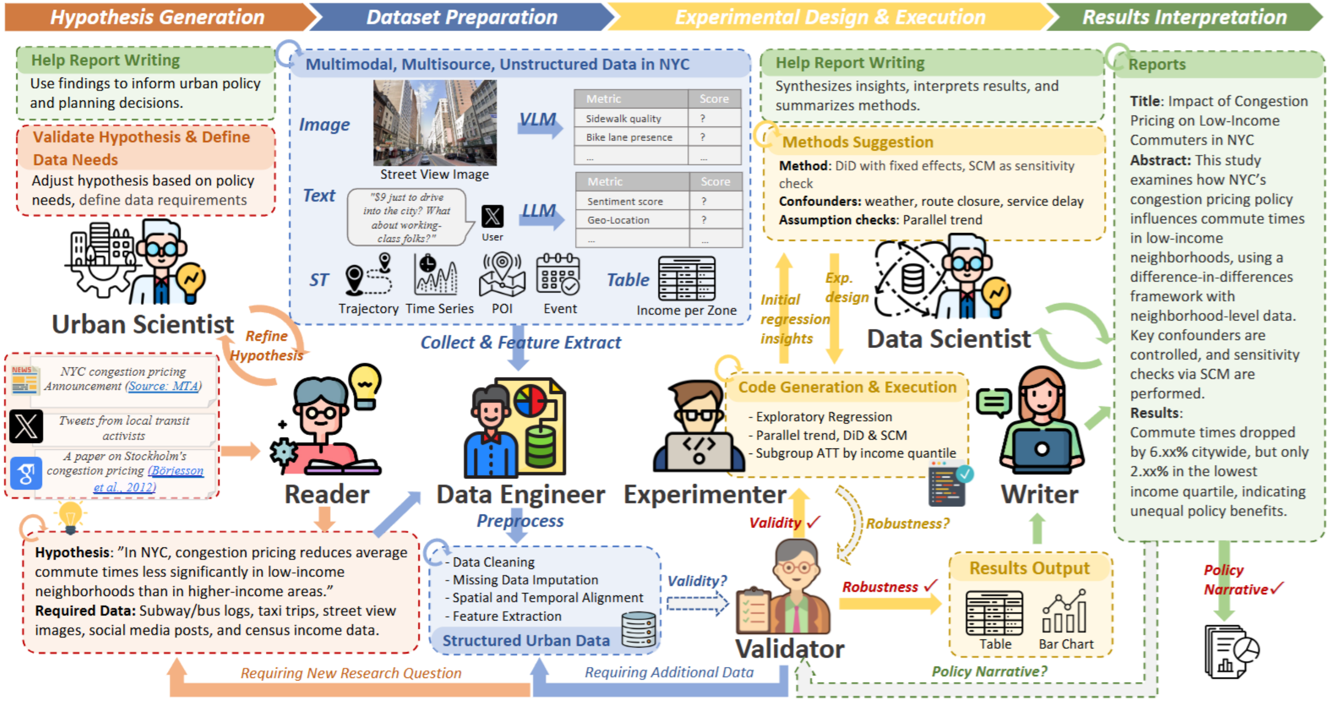
Simulation-based Agents: Urban Planning



Z. Zhou et al., Large language model empowered participatory urban planning. arXiv 2024.

Simulation-based Agents: Urban Causal Inference

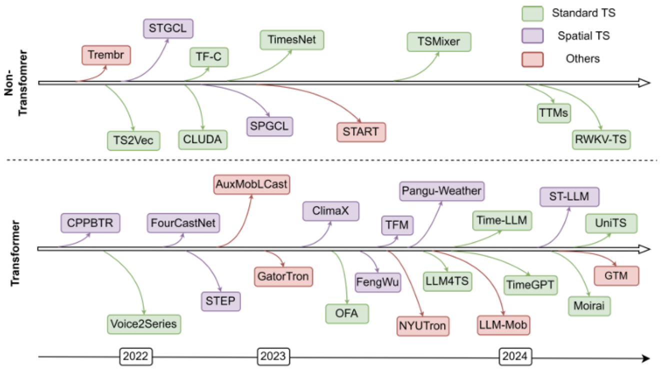
- What if we create **a team of agents** to simulate an **urban research lab**?



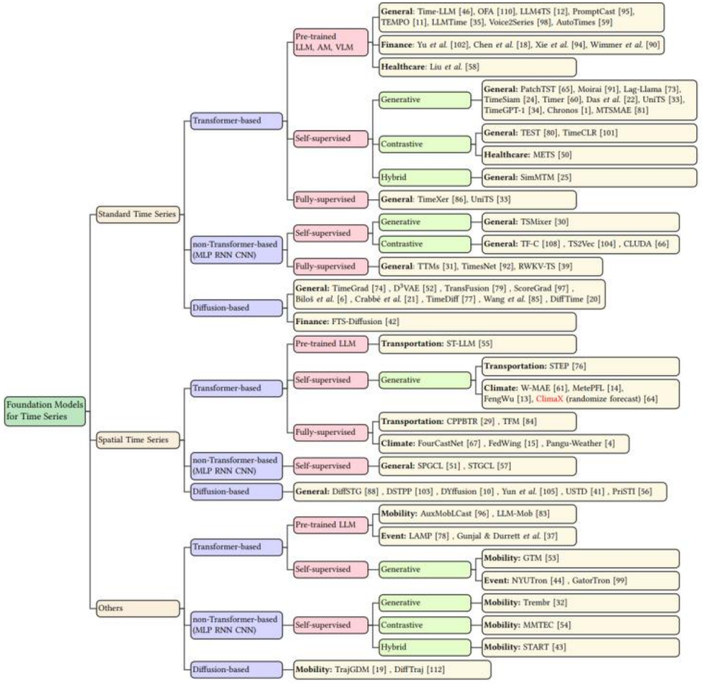
Y. Xia_et al., Reimagining Urban Science: Scaling Causal Inference with Large Language Models. arXiv 2025

PFMs for ST Data Science

Survey on PFMs



[KDD'24] & [arXiv'25]

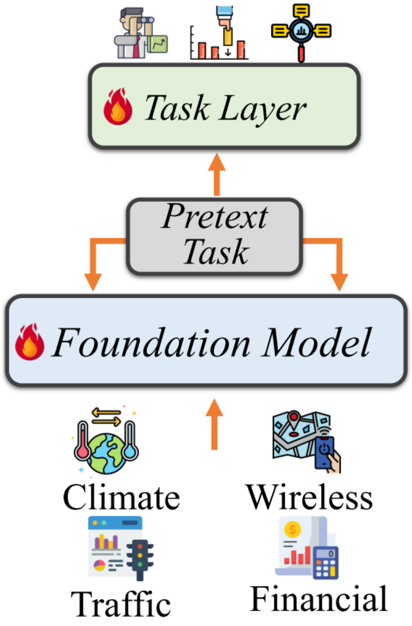


Y. Liang et al. Foundation Models for Time Series Analysis: A Tutorial and Survey. **KDD** 2024

Y. Liang et al. Foundation Models for Spatio-Temporal Data Science: A Tutorial and Survey. **KDD** 2025

Pretrained Foundation Models (PFM)

- Beyond the previously LLM-based methods, another significant approach in STFM involves building foundation models from scratch



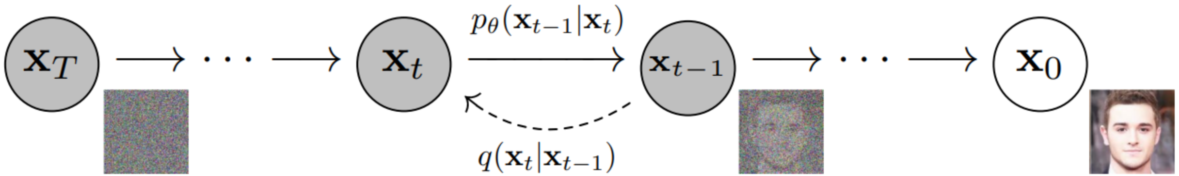
A major challenge:
Large-scale datasets!

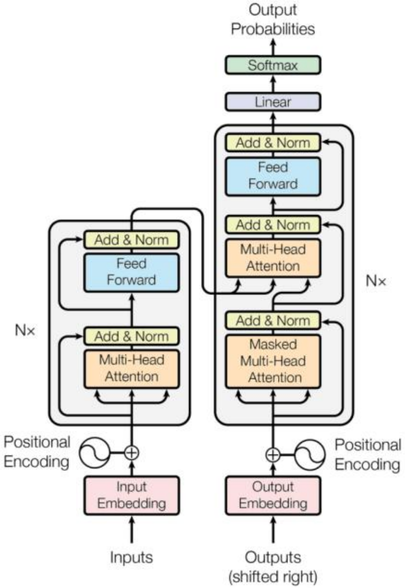
Table 14. Datasets and key properties from the Monash Time Series Forecasting Repository.

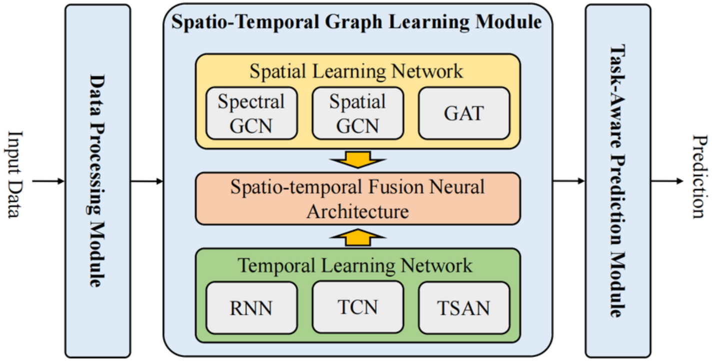
Dataset	Domain	Frequency	# Time Series	# Targets	# Past Covariates	# Obs.
London Smart Meters	Energy	30T	5,520	1	0	166,238,880
Wind Farms	Energy	T	337	1	0	172,165,370
Wind Power	Energy	4S	1	1	0	7,397,147
Solar Power	Energy	4S	1	1	0	7,397,222
Oikolab Weather	Climate	H	8	1	0	800,456
Elecdemand	Energy	30T	1	1	0	17,520
Covid Mobility	Transport	D	362	1	0	148,602
Kaggle Web Traffic Weekly	Web	W	145,063	1	0	16,537,182
Extended Web Traffic	Web	D	145,063	1	0	370,926,091
M1 Yearly	Econ/Fin	Y	106	1	0	3,136
M1 Quarterly	Econ/Fin	Q	198	1	0	9,854
M1 Monthly	Econ/Fin	M	617	1	0	44,892
M3 Yearly	Econ/Fin	Y	645	1	0	18,319
M3 Quarterly	Econ/Fin	Q	756	1	0	37,004
M3 Monthly	Econ/Fin	M	1,428	1	0	141,858
M3 Other	Econ/Fin	Q	174	1	0	11,933
M4 Yearly	Econ/Fin	Y	22,739	1	0	840,644
M4 Quarterly	Econ/Fin	Q	24,000	1	0	2,214,108
M4 Monthly	Econ/Fin	M	48,000	1	0	10,382,411
M4 Weekly	Econ/Fin	W	359	1	0	366,912
M4 Hourly	Econ/Fin	H	414	1	0	353,500
M4 Daily	Econ/Fin	D	4,227	1	0	9,964,658
NN5 Daily	Econ/Fin	D	111	1	0	81,585
NN5 Weekly	Econ/Fin	W	111	1	0	11,655
Tourism Yearly	Econ/Fin	Y	419	1	0	11,198
Tourism Quarterly	Econ/Fin	Q	427	1	0	39,128
Tourism Monthly	Econ/Fin	M	366	1	0	100,496
CIF 2016	Econ/Fin	M	72	1	0	6,334
Traffic Weekly	Transport	W	862	1	0	82,752
Traffic Hourly	Transport	H	862	1	0	14,978,112
Australian Electricity Demand	Energy	30T	5	1	0	1,153,584
Rideshare	Transport	H	2,304	1	0	859,392
Saugeen	Nature	D	1	1	0	23,711
Sunspot	Nature	D	1	1	0	73,894
Temperature Rain	Nature	D	32,072	1	0	22,290,040
Vehicle Trips	Transport	D	329	1	0	32,512
Weather	Climate	D	3,010	1	0	42,941,700
Car Parts	Sales	M	2,674	1	0	104,286
FRED MD	Econ/Fin	M	107	1	0	76,612
Pedestrian Counts	Transport	H	66	1	0	3,130,762
Hospital	Healthcare	M	767	1	0	55,224
COVID Deaths	Healthcare	D	266	1	0	48,412
KDD Cup 2018	Energy	H	270	1	0	2,897,004
Bitcoin	Econ/Fin	D	18	1	0	74,824
US Births	Healthcare	D	1	1	0	7,275

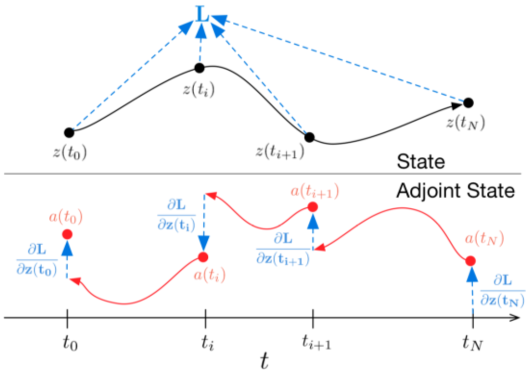
Neural Architectures

- Transformer-based
- Diffusion-based
- Graph-based
- Others, e.g., ODE-based, SSM-based



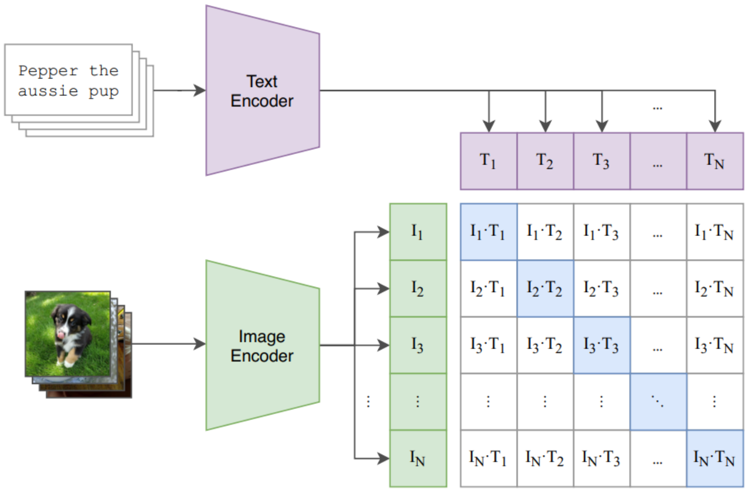




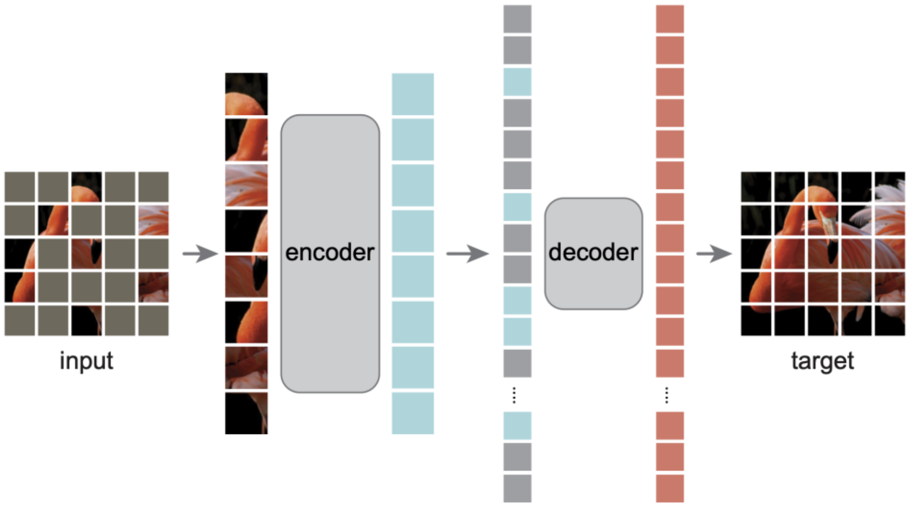


Pretraining Schemes

- Contrastive Learning

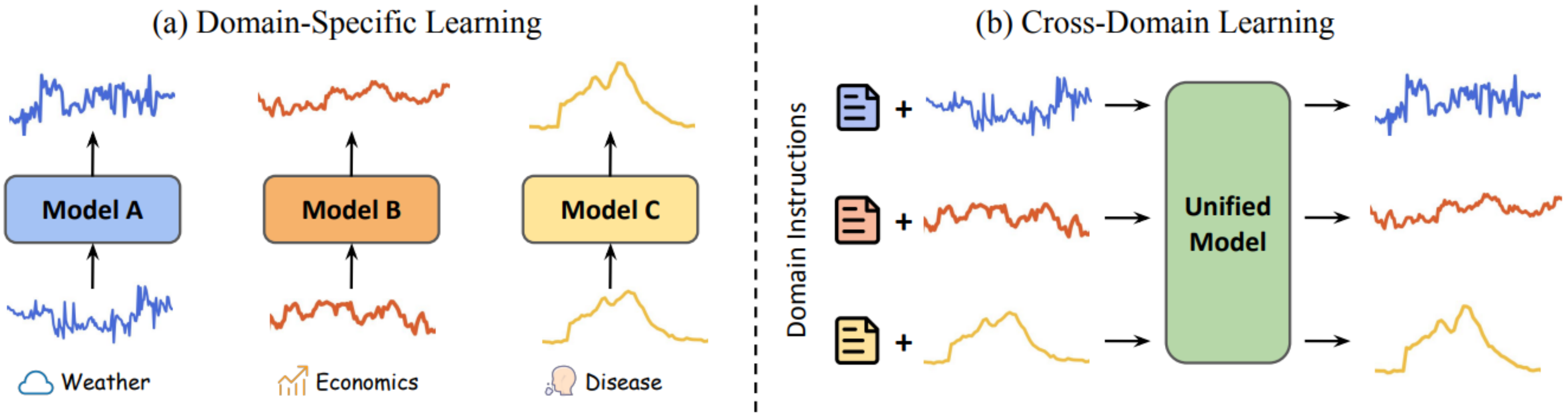


- Generative Learning



UniTime: Building Time Series Foundation Models

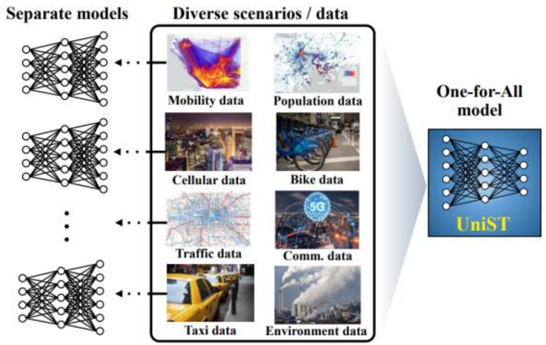
- The prerequisite of training a **Foundation Model for time series** is training a model on cross-domain time series all at once

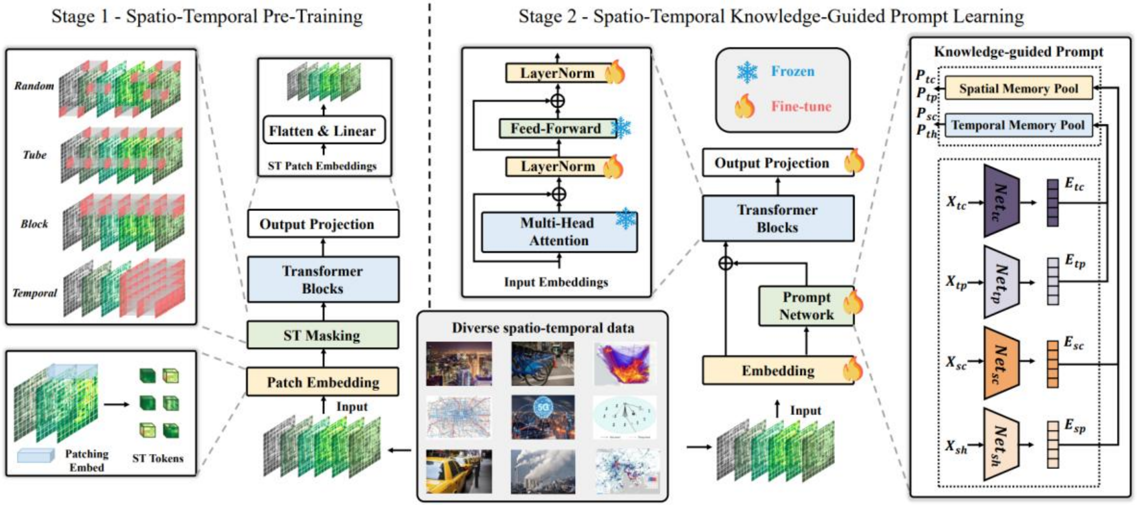


X. Liu , Y. Liang* et al., UniTime: A Language-Empowered Unified Model for Cross-Domain Time Series Forecasting. **WWW** 2024

UniST: A Spatio-Temporal Foundation Model

- Integrate ST correlations into UniTime while designing spatio-temporal prompts to enhance out-of-domain generalization

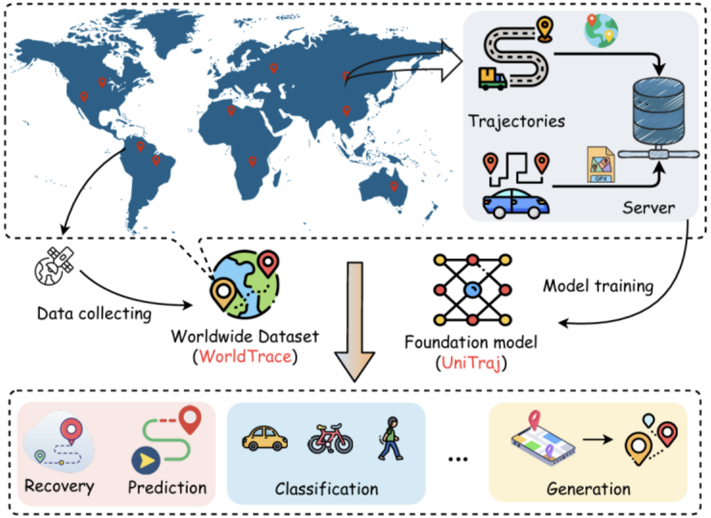




Y Yuan et al., UniST: A Prompt-Empowered Universal Model for Urban Spatio-Temporal Prediction. **KDD** 2024

UniTraj: A Trajectory Foundation Model

- The **first** trajectory foundation model across the world!



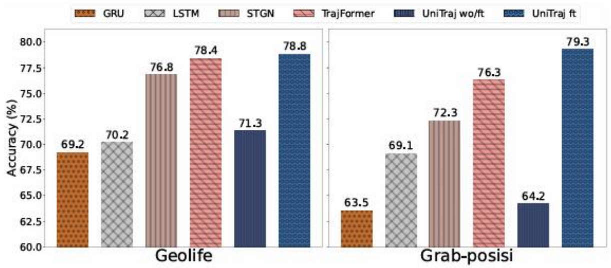
Recovery

Methods	WorldTrace		Chengdu		Xi'an		GeoLife		Grab-Posisi		Porto	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Linear	427.68	516.15	205.74	258.52	176.49	220.87	196.85	249.76	507.41	617.28	396.61	482.39
DHTR	220.35	302.47	75.19	98.68	62.85	83.43	80.04	168.25	351.20	415.16	194.37	232.59
Transformer	130.82	147.62	55.23	62.85	45.85	51.96	94.68	113.77	136.58	163.29	104.36	126.96
DeepMove	51.16	62.29	29.32	39.02	27.31	35.67	86.38	107.78	126.93	168.07	136.66	174.96
TrajBERT	58.13	70.14	26.48	33.83	19.45	25.13	34.53	43.24	112.68	136.24	78.77	99.23
TrajFM	47.64	58.92	19.10	25.09	18.86	24.13	59.34	64.24	107.64	130.69	71.15	92.96
UniTraj (zero-shot)	10.22	13.56	11.98	20.94	8.93	13.83	37.21	63.89	114.07	167.01	78.28	100.14
Improvement(%)	↑78.55	↑76.99	↑37.28	↑16.54	↑52.65	↑42.69	↓7.76	↓47.46	↓5.97	↓27.79	↓10.02	↓7.72
UniTraj (fine-tune)	**6.94**	**9.67**	**6.92**	**10.41**	**6.50**	**9.93**	**23.23**	**34.70**	**48.95**	**69.23**	**60.18**	**79.76**
Improvement(%)	↑85.43	↑83.59	↑63.77	↑58.51	↑65.54	↑58.85	↑32.73	↑19.75	↑54.52	↑47.03	↑15.42	↑14.20

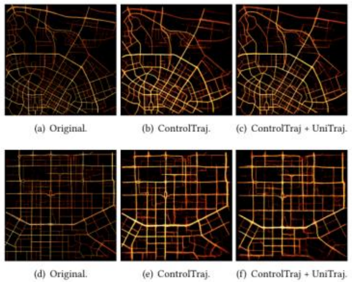
Prediction

Methods	WorldTrace		Chengdu		GeoLife	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
Linear	153.12	159.65	156.85	164.58	189.02	201.34
DHTR	146.48	151.63	123.47	129.73	180.32	187.59
Transformer	114.25	117.07	67.38	70.86	165.02	170.84
DeepMove	55.69	58.67	36.31	39.10	116.46	123.20
BERT	80.57	86.36	64.73	68.92	113.68	121.18
TrajFM	75.45	81.32	77.82	80.48	121.94	128.16
UniTraj (zero-shot)	49.85	55.02	42.75	45.93	108.35	133.60
Improvement(%)	↑10.49	↑6.22	↓17.74	↓17.46	↑4.69	↓10.25
UniTraj (fine-tune)	**30.10**	**34.46**	**28.78**	**32.44**	**90.97**	**102.88**
Improvement(%)	↑45.95	↑41.27	↑20.74	↑17.03	↑19.98	↑15.10

Classification



Generation



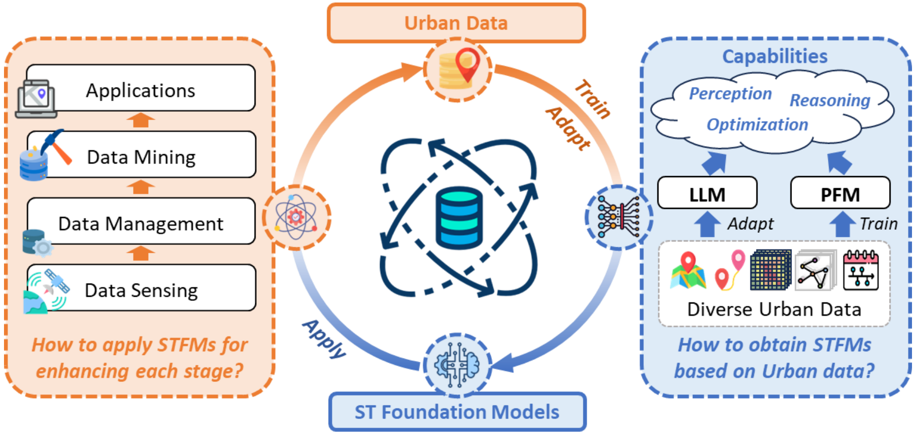
Y. Zhu et al., UniTraj: Learning a Universal Trajectory Foundation Model from Billion-Scale Worldwide Traces. arXiv 2024

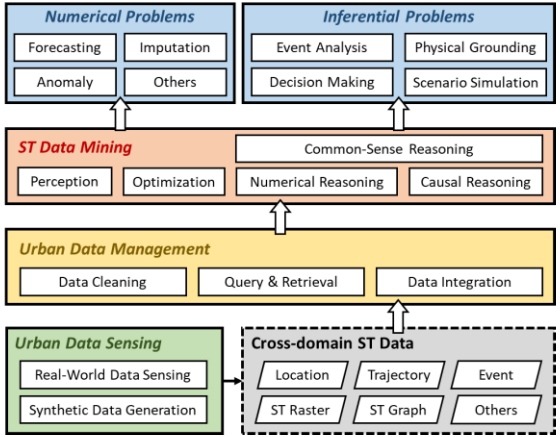
Comparison between LLMs and PFMs

Capabilities	Large Language Models (LLMs)	Pretrained Foundation Models (PFMs)
Perception	✓ Strong perception for unstructured modalities, such as texts; ▲ Limited native ST perception	✓ Strong perception for structured ST data (e.g., trajectories, ST graphs); ▲ Weak for free-form text & images
Optimization	✓ Agent-based reasoning for decision-making; relies on prompting and heuristics	▲ Limited; lacks interaction and decision-making ability for control and planning
Common-sense Reasoning	✓ Strong via pretraining on vast textual data; can be enhanced with fine-tuning	▲ Limited; relies on structured ST data rather than broad world knowledge
Numerical Reasoning	▲ Handles arithmetic but struggles with structured ST computations	✓ Designed for numerical problems, e.g., forecasting, anomaly detection
Causal Reasoning	✓ Can infer causal relations from texts; ▲ Lacks causality modeling in structured ST data	✓ Built-in graph-based and ST causal modeling

Summary

- When ST Data Science Meets Large Foundation Models
  - **How to leverage STFMs for benefiting each stage of ST Data Science?**
  - **How to obtain STFMs based on Urban Data?**
    - ➢ Utilizing LLMs in zero-shot or few-shot ways
    - ➢ Pretraining STFMs from scratch based on large-scale urban data

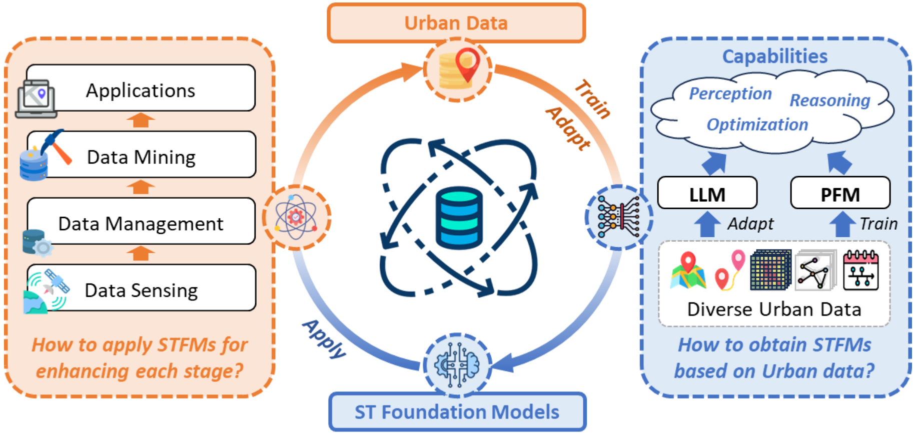


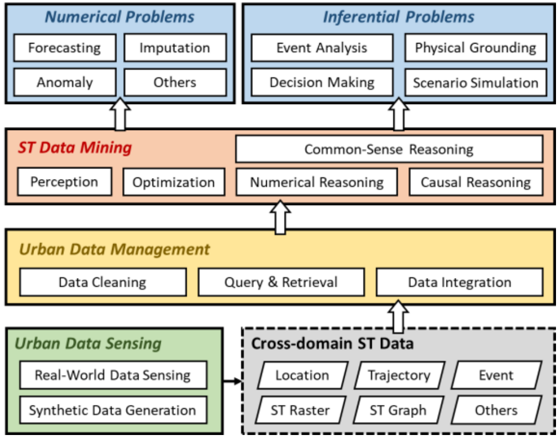


Limitations & Future Opportunities

- The curse of accuracy against interpretability
- Large foundation models are all we need?
- One-fit-all FMs across the full workflow of ST Data Science
- Integrating STFMs with multimodal understanding →

ACM multimedia
Dublin, Ireland 27-31.10.2025





Main Reference

Foundation Models for Spatio-Temporal Data Science: A Tutorial and Survey

Yuxuan Liang[1], Haomin Wen[2,1], Yutong Xia[3], Ming Jin[4], Bin Yang[5], Flora Salim[6], Qingsong Wen[7], Shirui Pan[4], Gao Cong[8]
[1]The Hong Kong University of Science and Technology (Guangzhou) [2]Carnegie Mellon University
[3]National University of Singapore [4]Griffith University [5]East China Normal University
[6]University of New South Wales [7]Squirrel Ai Learning, USA [8]Nanyang Technology University
{yuxliang,yutong.x}@outlook.com,{wenhaomin.whm,mingjinedu,qingsongedu}@gmail.com
flora.salim@unsw.edu.au,byang@dase.ecnu.edu.cn,s.pan@griffith.edu.au,gaocong@ntu.edu.sg

Abstract

Spatio-Temporal (ST) data science, which includes sensing, managing, and mining large-scale data across space and time, is fundamental to understanding complex systems in domains such as urban computing, climate science, and intelligent transportation. Traditional deep learning approaches have significantly advanced this field, particularly in the stage of ST data mining. However, these models remain task-specific and often require extensive labeled data. Inspired by the success of Foundation Models (FM), especially large language models, researchers have begun exploring the concept of Spatio-Temporal Foundation Models (STFMs) to enhance adaptability and generalization across diverse ST tasks. Unlike prior architectures, STFMs empower the entire workflow of ST data science, ranging from data sensing, management, to mining, thereby offering a more holistic and scalable approach. Despite

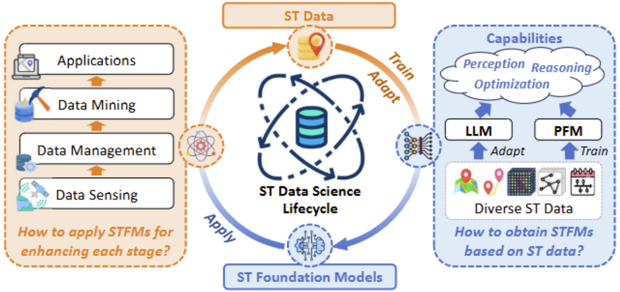


Figure 1: ST Foundation Models (STFM), which include LLM and PFM, are pretrained with or applied to diverse ST data, with the abilities of perception, optimization, and reasoning. STFMs can, in turn, enhance each stage of ST data science.

References

- M. Jin, Y. Liang*, Q. Wen* et al., Position: What Can Large Language Models Tell Us about Time Series Analysis. **ICML** 2024.
- M. Jin, Y. Liang, Q. Wen* et al., Time-LLM: Time Series Forecasting by Reprogramming Large Language Models. **ICLR** 2024.
- Y. Liang, Q. Wen* et al. Foundation Models for Time Series Analysis: A Tutorial and Survey. **KDD** 2024.
- M. Jin, Q. Wen, Y. Liang et al., Large Models for Time Series and Spatio-Temporal Data: A Survey and Outlook. Under review. 2024.
- Y. Liang, Q. Wen et al. Foundation Models for Spatio-Temporal Data Science: A Tutorial and Survey. **KDD** 2025.

Hot!!! Total citations > 1.2K since 2024

[ICML'24]
Position Paper

[ICLR'24]
Time-LLM

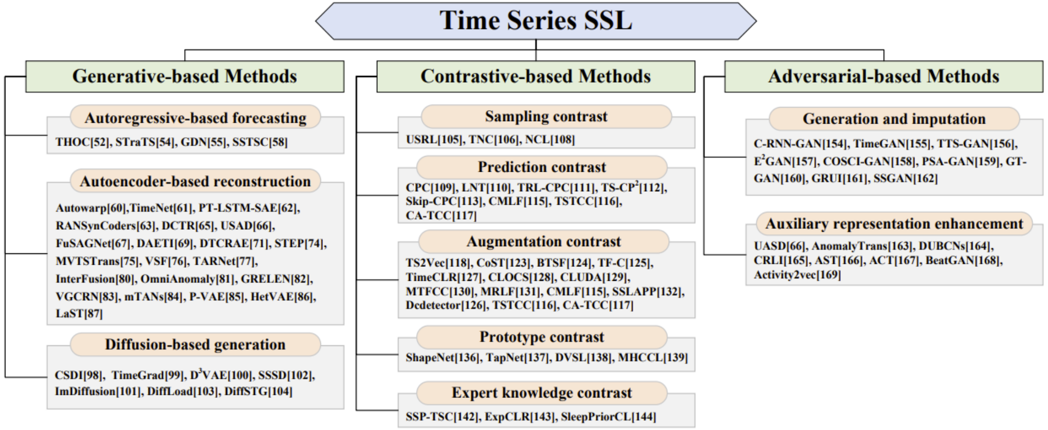
[KDD'24] Survey -TS
Foundation Models

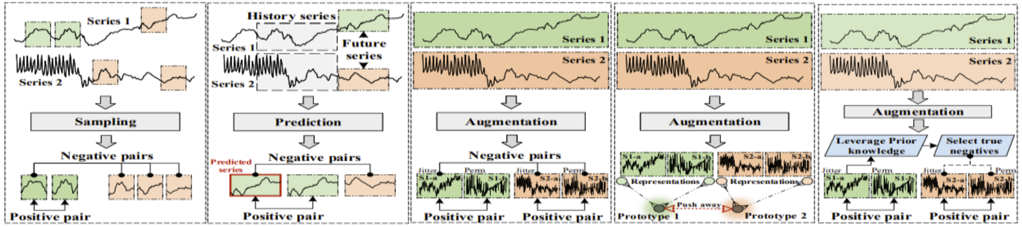
[arXiv'23] Survey –ST/TS
Large Models

[KDD'25] This Talk

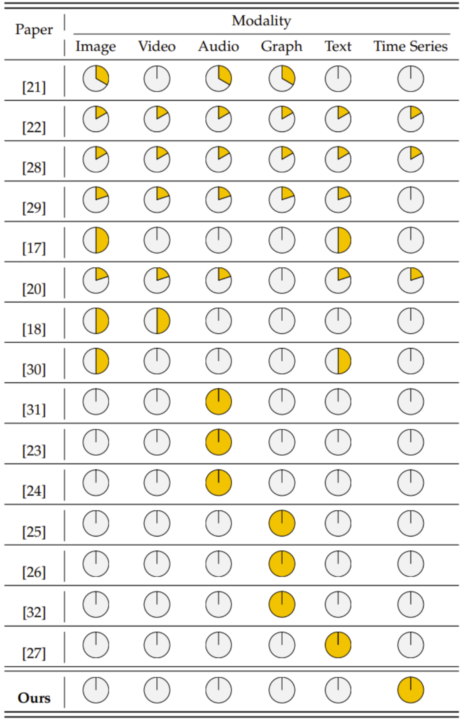
Other Related Surveys

- **[TPAMI'24]** Zhang et al. Self-Supervised Learning for Time Series Analysis



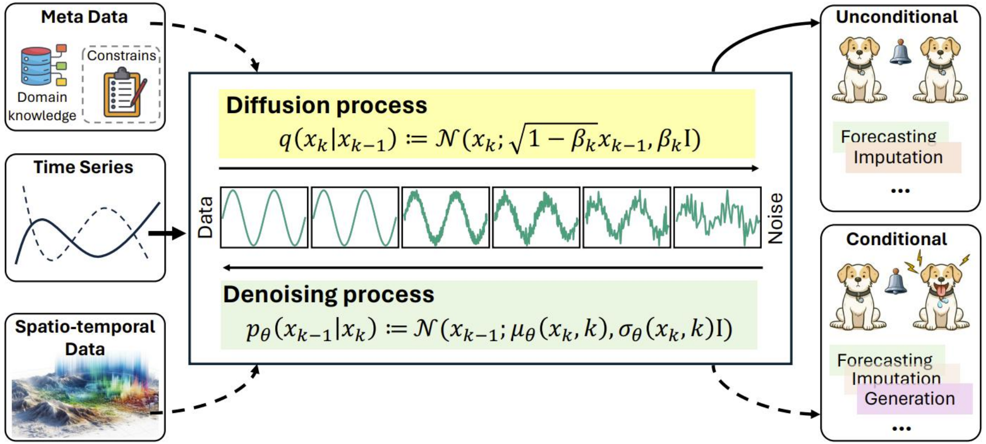


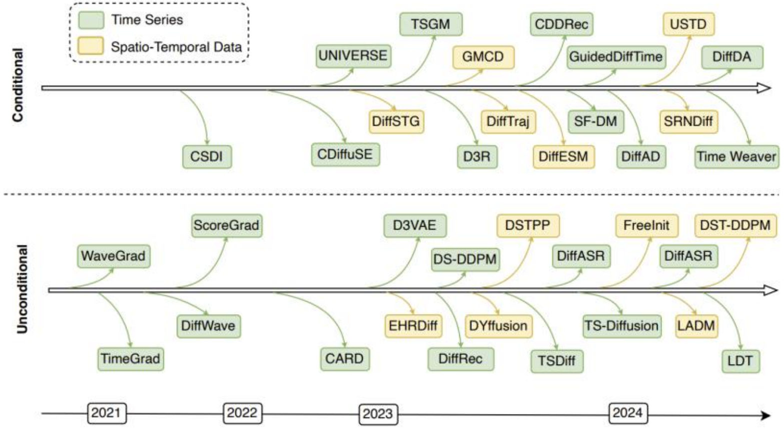
(a) Sampling contrast (b) Prediction contrast (c) Augmentation contrast (d) Prototype contrast (e) Expert knowledge contrast



Other Related Surveys

- [arXiv] **Diffusion Models** for Time Series and Spatio-Temporal Data





Our Related Workshops & Tutorials at **KDD'25**

- We just organized **Workshop** on Urban Computing (Urbcomp) at KDD 2025!
- We just organized **Workshop** on Mining and Learning from Time Series at KDD 2025!
- We just organized **Tutorial** on Deep Learning in the Frequency Domain: Advances, Challenges, and Applications for Time Series Analysis at KDD 2025!

Q&A

Yuxuan Liang

Assistant Professor, HKUST(GZ)

yuxliang@outlook.com

Thanks!

Qingsong Wen

Head of AI, Squirrel Ai Learning

qingsongedu@gmail.com