AIOps Challenge | 2023 CCF国际AIOps挑战赛决赛
暨“大模型时代的AIOps”研讨会

Time-LLM: Time Series Forecasting by Reprogramming Large Language Models

Qingsong Wen

Head of AI Research & Chief Scientist, Squirrel AI (松鼠AI)

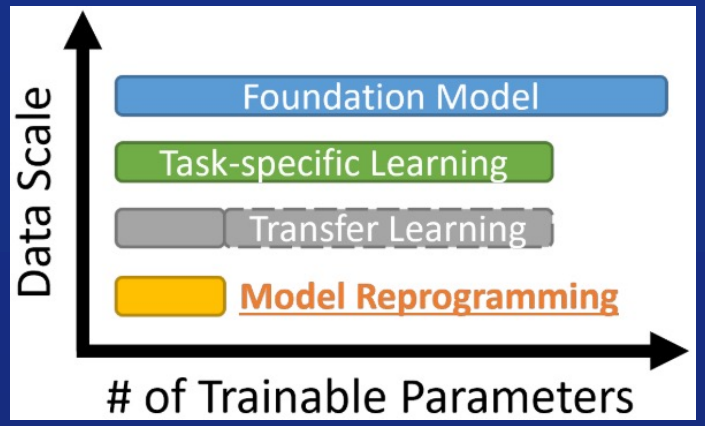
This work is collaborated with Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, from Monash University, Ant Group, Griffith University, Alibaba Group, IBM, and HKUST

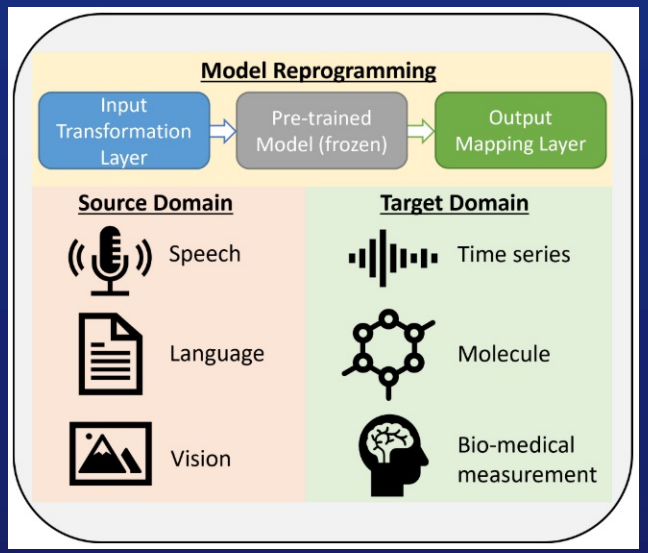
主办单位：中国计算机学会（CCF）、清华大学、中国建设银行股份有限公司、南开大学

承办单位：中国计算机学会互联网专委会、清华大学计算机科学与技术系、中国建设银行股份有限公司运营数据中心、南开大学软件学院、北京必示科技有限公司

赞助单位：华为技术有限公司、国网宁夏电力有限公司电力科学研究院、软通动力信息技术（集团）股份有限公司

1. Background





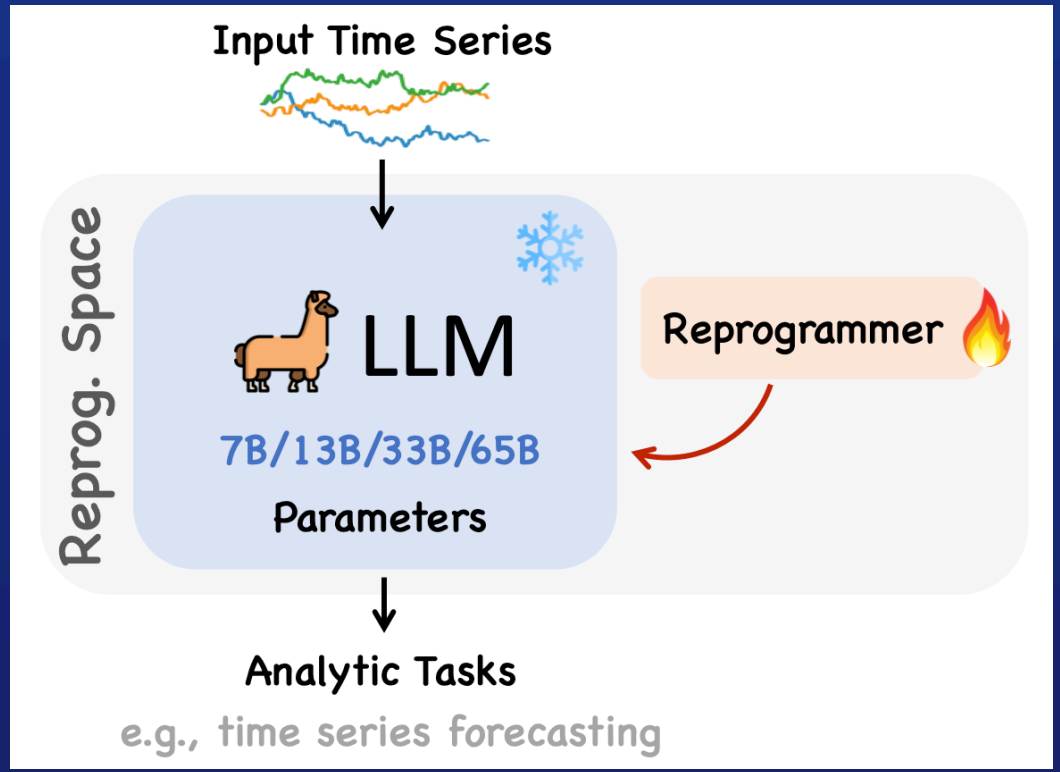
- **Task-Specific Learning**: Training a specific ML model from the scratch by minimizing the task-specific loss
- **Transfer Learning**: A common practice for in-domain knowledge transfer. One notable limitation is that in some target domains there may lack adequate pre-trained models from similar domains for effective finetuning
- **Foundation Model**: It features task-agnostic pre-training (often on large-scale datasets) and efficient finetuning to downstream tasks
- **Model Reprogramming**: Only requires training the inserted input transformation and output mapping layers while keeping the source pre-trained model intact

Chen, P. Y. (2022). Model reprogramming: Resource-efficient cross-domain machine learning. arXiv preprint arXiv:2202.10629.

2. Motivation

AIOps Challenge

- Reprogramming makes LLMs instantly ready for time series tasks



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

⚙ **Reprogramming ≈ Adaptation + Alignment**

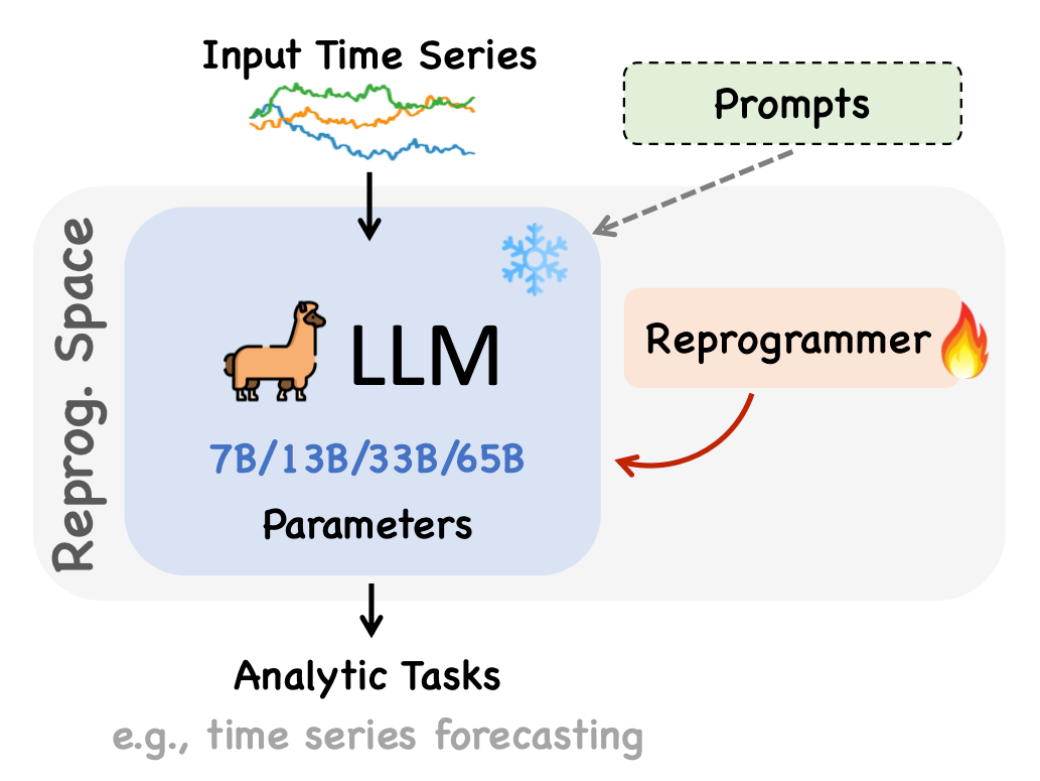
Adaptation makes LLMs to understand how to process the input time series data → Breaking domain isolation and enabling knowledge sharing

Alignment further eliminates domain boundary to facilitate knowledge acquiring

2. Motivation

AIOps Challenge

- Reprogramming makes LLMs more powerful for time series tasks



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

⚙ **Reprogramming ≈ Adaptation + Alignment**

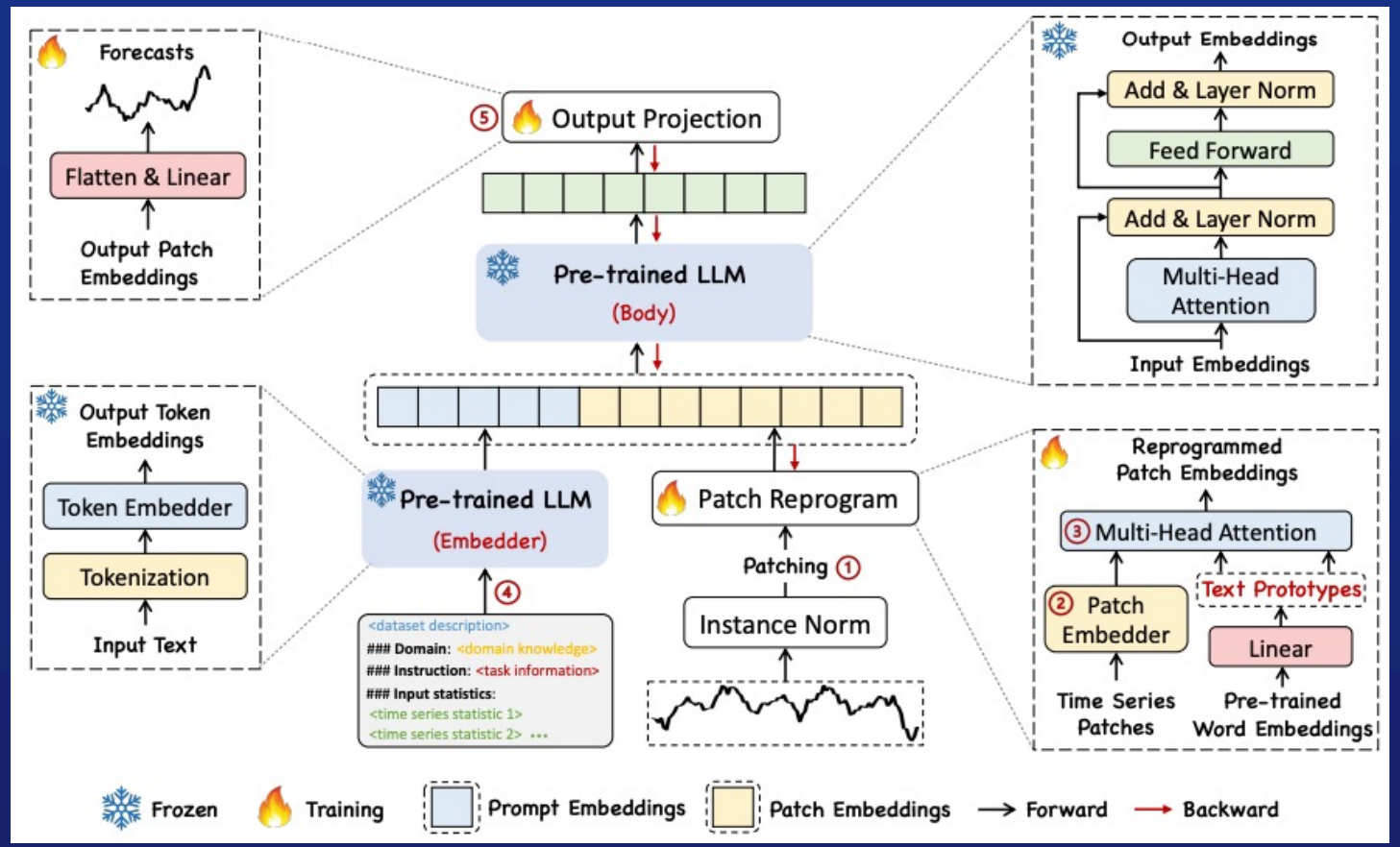
Adaptation makes LLMs to understand how to process the input time series data → Breaking domain isolation and enabling knowledge sharing

Alignment further eliminates domain boundary to facilitate knowledge acquiring

3. Time-LLM

AIOps Challenge

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



Unlocking the LLM's ability for time series

Cross-modal Adaptation: ① ② ⑤

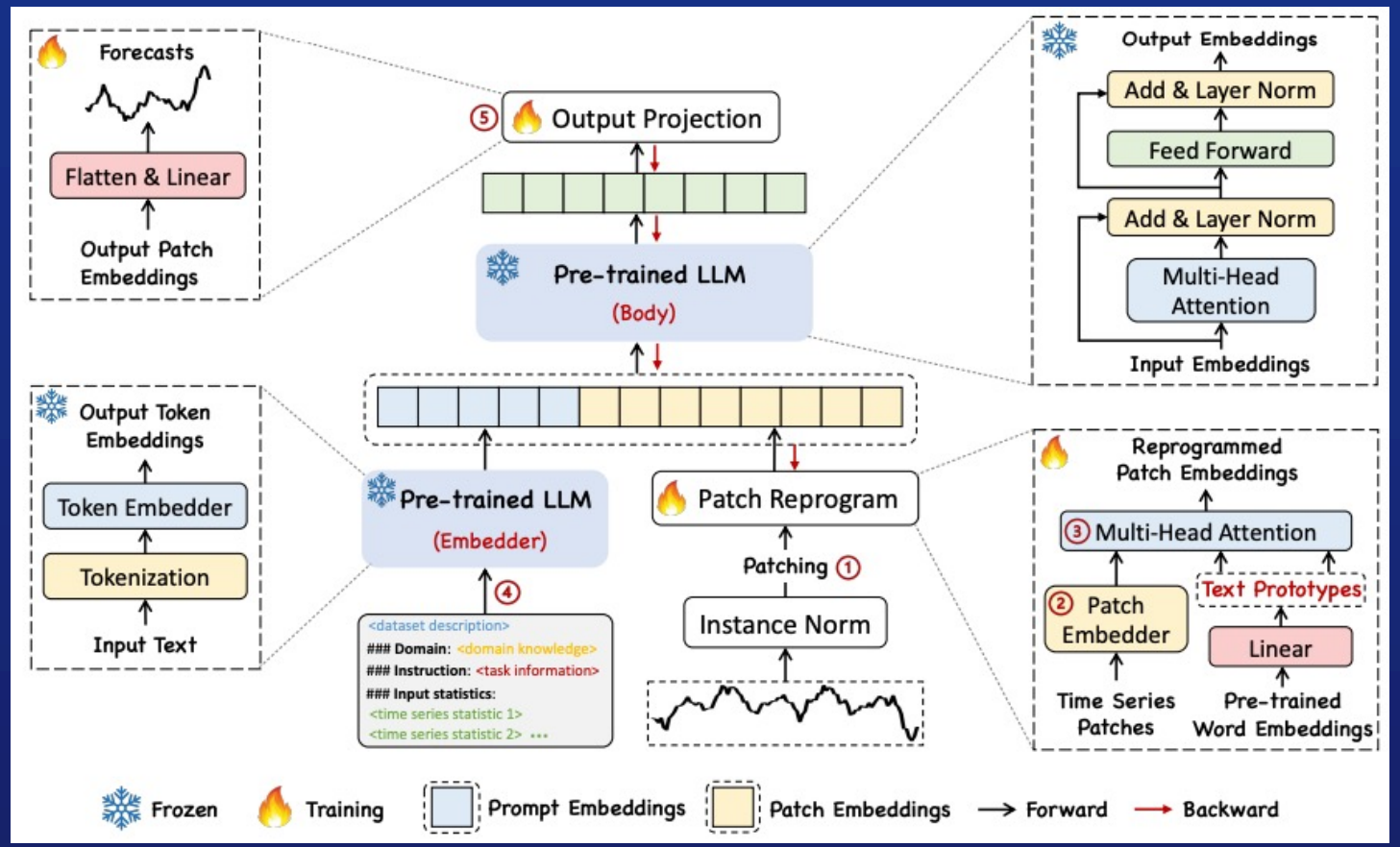
Cross-modal Alignment: ③ ④

- **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.

3. Time-LLM

AIOps Challenge

TL;DR Domain 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

- **Prompt-as-Prefix:** natural language-based prompts (e.g., domain knowledge & task instructions) can act as prefixes to enrich the input context and guide the transformation of reprogrammed TS patches

3. Time-LLM

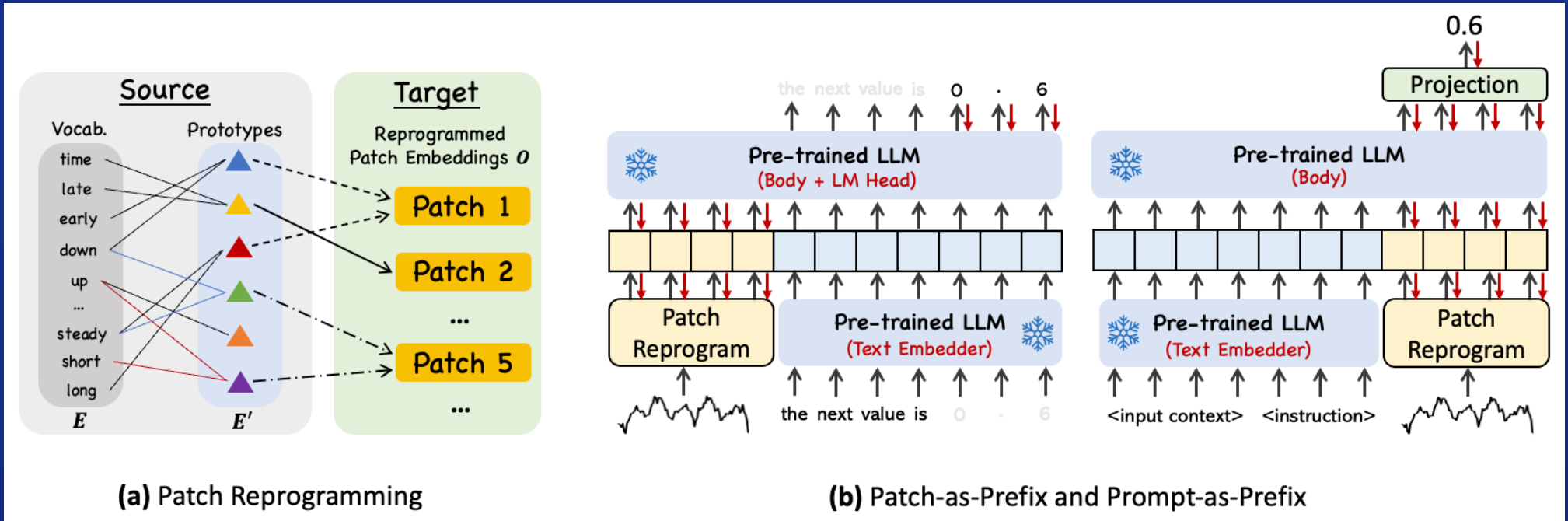


Illustration of **(a)** patch reprogramming and **(b)** Patch-as-Prefix versus Prompt-as-Prefix

- **Patch reprogramming:** Text prototypes learn connecting language cues, e.g., “short up” (red lines) and “steady down” (blue lines), which are then combined to represent the local patch information (e.g., “short up then down steadily” for characterizing Patch 5)

3. Time-LLM

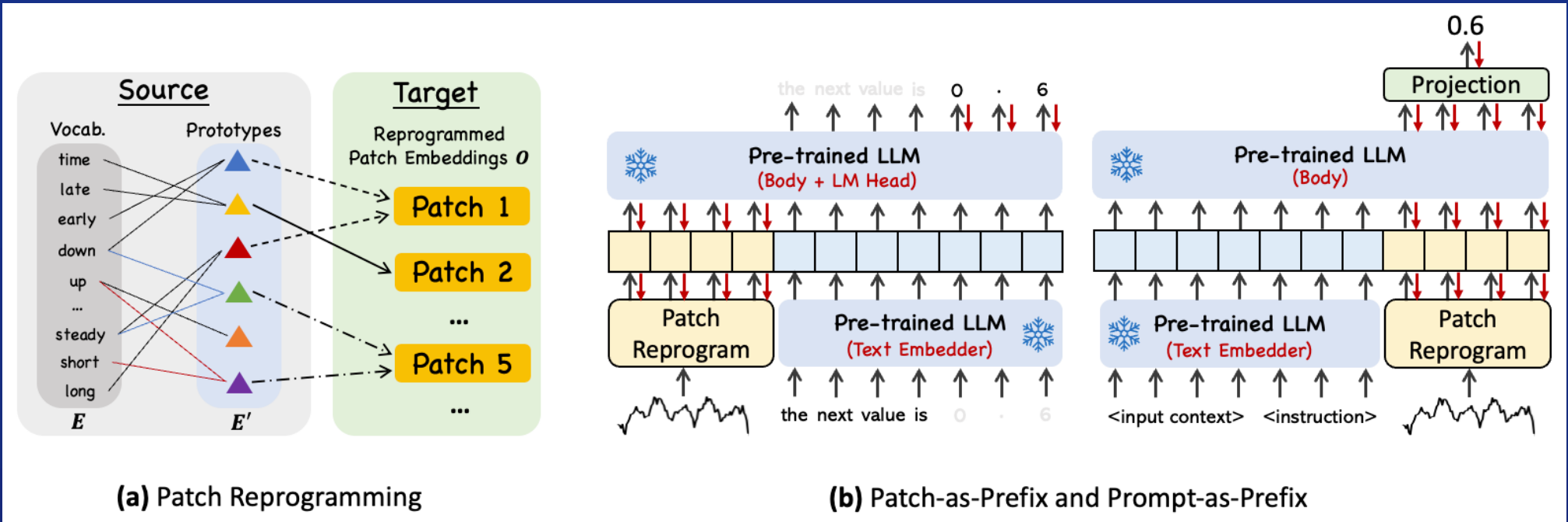


Illustration of **(a)** patch reprogramming and **(b)** Patch-as-Prefix versus Prompt-as-Prefix

- **Prompt-as-Prefix** is proposed to enrich the input context and guide the transformation of reprogrammed time series patches
- **Prompt-as-Prefix** is more desired in time series forecasting compared to **Patch-as-Prefix**

4. Brief Results

AIOps Challenge

Table 1: Long-term forecasting results. All results are averaged from four different forecasting horizons: $H \in \{24, 36, 48, 60\}$ for ILI and $\{96, 192, 336, 720\}$ for the others. A lower value indicates better performance. **Red**: the best, Blue: the second best. Our full results are in Appendix D.

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	**0.408**	**0.423**	0.465	0.455	0.422	0.437	0.413	0.430	0.458	0.450	0.440	0.460	0.496	0.487	0.570	0.537	0.542	0.510	0.491	0.479	1.040	0.795	1.029	0.805
ETTh2	0.334	0.383	0.381	0.412	0.431	0.446	**0.330**	**0.379**	0.414	0.427	0.437	0.449	0.450	0.459	0.526	0.516	0.439	0.452	0.602	0.543	4.431	1.729	6.736	2.191
ETTm1	**0.329**	**0.372**	0.388	0.403	0.357	0.378	0.351	0.380	0.400	0.406	0.448	0.452	0.588	0.517	0.481	0.456	0.429	0.425	0.435	0.437	0.961	0.734	0.799	0.671
ETTm2	**0.251**	**0.313**	0.284	0.339	0.267	0.333	0.255	0.315	0.291	0.333	0.305	0.349	0.327	0.371	0.306	0.347	0.293	0.342	0.409	0.436	1.410	0.810	1.479	0.915
Weather	**0.225**	**0.257**	0.237	0.270	0.248	0.300	**0.225**	0.264	0.259	0.287	0.309	0.360	0.338	0.382	0.288	0.314	0.271	0.334	0.261	0.312	0.634	0.548	0.803	0.656
ECL	**0.158**	**0.252**	0.167	0.263	0.166	0.263	0.161	**0.252**	0.192	0.295	0.214	0.327	0.227	0.338	0.193	0.296	0.208	0.323	0.229	0.329	0.311	0.397	0.338	0.422
Traffic	**0.388**	0.264	0.414	0.294	0.433	0.295	0.390	**0.263**	0.620	0.336	0.610	0.376	0.628	0.379	0.624	0.340	0.621	0.396	0.622	0.392	0.764	0.416	0.741	0.422
ILI	**1.435**	0.801	1.925	0.903	2.169	1.041	1.443	**0.797**	2.139	0.931	2.847	1.144	3.006	1.161	2.077	0.914	2.497	1.004	7.382	2.003	5.137	1.544	4.724	1.445
1st Count	7		0		0		5		0		0		0		0		0		0		0		0	

Table 2: Short-term time series forecasting results on M4. The forecasting horizons are in [6, 48] and the three rows provided are weighted averaged from all datasets under different sampling intervals. A lower value indicates better performance. **Red**: the best, Blue: the second best. More results are in Appendix D.

Methods		TIME-LLM (Ours)	GPT4TS (2023a)	TimesNet (2023)	PatchTST (2023)	N-HiTS (2023b)	N-BEATS (2020)	ETSformer (2022)	LightTS (2022a)	DLinear (2023)	FEDformer (2022)	Stationary (2022)	Autoformer (2021)	Informer (2021)	Reformer (2020)
Average	SMAPE	**11.983**	12.69	12.88	12.059	12.035	12.25	14.718	13.525	13.639	13.16	12.780	12.909	14.086	18.200
	MASE	**1.595**	1.808	1.836	1.623	1.625	1.698	2.408	2.111	2.095	1.775	1.756	1.771	2.718	4.223
	OWA	**0.859**	0.94	0.955	0.869	0.869	0.896	1.172	1.051	1.051	0.949	0.930	0.939	1.230	1.775

4. Brief Results

AIOps Challenge

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
$ETTm1 \rightarrow ETTh2$	**0.381**	**0.412**	0.433	0.439	0.992	0.708	0.464	0.475	0.439	0.438	0.457	0.454
$ETTm1 \rightarrow ETTm2$	**0.268**	**0.320**	0.313	0.348	1.867	0.869	0.335	0.389	0.296	0.334	0.322	0.354
$ETTm2 \rightarrow ETTh2$	**0.354**	**0.400**	0.435	0.443	0.992	0.708	0.455	0.471	0.409	0.425	0.435	0.443
$ETTm2 \rightarrow ETTm1$	**0.414**	**0.438**	0.769	0.567	1.933	0.984	0.649	0.537	0.568	0.492	0.769	0.567

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 on ETTh1 in forecasting different steps ahead.

Length	ETTh1-96			ETTh1-192			ETTh1-336			ETTh1-512		
Metric	Param. (M)	Mem. (MiB)	Speed(s/iter)	Param. (M)	Mem. (MiB)	Speed(s/iter)	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	3404.62	37988	0.632	3404.69	39004	0.697
D.2 LLama (8)	975.83	11370	0.184	975.87	12392	0.192	975.92	13188	0.203	976.11	13616	0.217
D.3 w/o LLM	6.39	3678	0.046	6.42	3812	0.087	6.48	3960	0.093	6.55	4176	0.129

5. Summary

AIOps Challenge

- Time-LLM shows promise in adapting frozen LLMs for time series forecasting by reprogramming time series data into natural language representation space more natural for LLMs and providing natural language guidance via Prompt-as-Prefix to augment reasoning
- Our evaluations demonstrate the adapted LLMs can significantly outperform many specialized expert models, indicating their potential as effective time series machines
- We provide a novel insight that time series forecasting can be cast as yet another “language” task that can be tackled by an off-the-shelf LLM to simply achieve or match SOTA performance
- We are the first to achieve “multimodal augmented time series forecasting” – We can even do more with the Prompt-as-Prefix!

Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J.Y., Shi, X., Chen, P.Y., Liang, Y., Li, Y.F., Pan, S. and Wen, Q., Time-LLM: Time series forecasting by reprogramming large language models. arXiv preprint arXiv:2310.01728, 2023

2023 CCF国际AIOps挑战赛决赛暨“大模型时代的AIOps”研讨会

THANKS