ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202505.00180

Can large language models precisely forecast the nuclear reactor states under never-before-seen transients?

Authors: Lin, Dr. Weiqing, Miao, Prof. Xiren, Dr. Jing Chen, Duan, Mr. Pengbin, Chen, Mr. Zening, Xu, Prof. Yong, Dr. Xinyu Liu, Jiang, Prof. Hao, Dr. Jing Chen

Date: 2025-05-14T18:27:10+00:00

Abstract

Forecasting the nuclear reactor states can provide critical support for intelligent maintenance in nuclear power plants. However, given economic and security purposes, collecting sufficient transient data from real-world nuclear reactors is extremely difficult, leading to poor forecasting performance of traditional data-driven models under never-before-seen transients. To address this, we leverage the reasoning ability of large language models (LLM) and customize a nuclear reactor-oriented LLM (NRO-LLM) for state predictions. Specifically, multi-domain public datasets are collected and decomposed into trend and residual components to characterize universal patterns and dynamic disturbances. Inspired by the human mind's habit of recalling past experiences for better decisions, an adaptive prompt mechanism is designed to retrieve highly related soft prompts from a prompt pool, which are merged with hard prompts to generate a refined prompt set for guiding the LLM. With time series and prompts as bimodal input, partial weights within the LLM are activated for fine-tuning. Experiments are conducted on two types of transients from a pressurized water reactor, containing real grid peak shaving and simulated accidents. The results confirm that the fine-tuned LLMs can accurately predict reactor states under transients through cross-domain knowledge extraction and bimodal feature fusion. This work provides deep insights into the application of LLMs in reasoning about unusual reactor states.

Full Text

Preamble

Can large language models precisely forecast nuclear reactor states under never-before-seen transients?∗

Weiqing Lin,[1] Xiren Miao,[1] Jing Chen,[1],† Pengbin Duan,[2] Zening Chen,[1] Yong Xu,[3] Xinyu Liu,[1] and Hao Jiang[1]
[1]College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China
[2]China Nuclear Power Technology Research Institute Company Limited, Shenzhen 518000, China
[3]China National Nuclear Power Operation Maintenance Technology Company Limited, Hangzhou 311200, China

Forecasting nuclear reactor states can provide critical support for intelligent maintenance in nuclear power plants. However, given economic and security constraints, collecting sufficient transient data from real-world nuclear reactors is extremely difficult, leading to poor forecasting performance of traditional data-driven models under never-before-seen transients. To address this, we leverage the reasoning ability of large language models (LLMs) and customize a nuclear reactor-oriented LLM (NRO-LLM) for state predictions. Specifically, multi-domain public datasets are collected and decomposed into trend and residual components to characterize universal patterns and dynamic disturbances. Inspired by the human mind's habit of recalling past experiences for better decisions, an adaptive prompt mechanism is designed to retrieve highly related soft prompts from a prompt pool, which are merged with hard prompts to generate a refined prompt set for guiding the LLM. With time series and prompts as bimodal input, partial weights within the LLM are activated for fine-tuning. Experiments are conducted on two types of transients from a pressurized water reactor, containing real grid peak shaving and simulated accidents. The results confirm that the fine-tuned LLMs can accurately predict reactor states under transients through cross-domain knowledge extraction and bimodal feature fusion. This work provides deep insights into the application of LLMs in reasoning about unusual reactor states.

Keywords: Nuclear power plant (NPP), pressurized water reactor (PWR), large language model (LLM), DeepSeek, prompt engineering

Introduction

Nuclear power plays an essential role in global energy policies and sustainability strategies. Compared with intermittent wind and solar energies, it exhibits stable power output and high energy density [1]. Generally, nuclear power plants (NPPs), as the base-load supply, operate in full power mode. However, as thermal power units are gradually phased out, nuclear power peak regulation becomes vital for adopting new energy penetration, easing peak demand, and addressing overcapacity [2]. This shift poses significant challenges to the safety and reliability of nuclear energy systems.

Transients in NPPs can be divided into two categories: normal operations and abnormal accidents. The former include power adjustments, shutdowns, and startups, while the latter involve emergencies caused by coolant leaks, equipment failures, etc. These transients cause sharp changes in various operating parameters that characterize nuclear reactor states [3]. The concept of "unmanned surveillance, few people on duty" depicts the vision for future NPP automation [4]. A pivotal technology for fulfilling this vision is implementing precise prediction for nuclear reactor states, especially under transients, to meet the demands for early anomaly warning and intelligent maintenance in NPPs [5].

Currently, classical data-driven models (such as multi-layer perceptron (MLP), long short-term memory (LSTM), and Transformer) are frequently developed for state prognosis through forecasting various operating parameters, including in-core power [6], flow speed [7], and coolant temperatures [8]. For example, Song et al. [9] employed RELAP5-based thermal-hydraulic models and data-driven gated recurrent unit (GRU) models for forecasting various operating parameters, comprehensively analyzing their performance. Their study concluded that data-driven models exhibit better predictive accuracy and efficiency. Kim et al. utilized a bidirectional LSTM with a conditional variational autoencoder to forecast coolant temperatures and provide uncertainty bounds during loss-of-coolant accidents. Fu et al. [11] combined a GRU with physical modeling to build an interpretable prediction framework for severe accident trends, demonstrating effectiveness across various accidents. In general, data-driven models with nonlinear mapping capability perform well on large-scale and full-coverage training sets. However, their predictive performance significantly declines when facing never-before-seen scenarios due to unusual evolutionary trends of reactor states.

Given economic and safety concerns, frequent trials in real-world NPPs may cause uncontrollable incidents and equipment damage. In some emerging nuclear reactors (e.g., high-temperature gas-cooled reactors, lead-cooled fast reactors) and newly built NPPs, data availability is limited due to strict privacy and confidentiality. Thus, collecting sufficient transient data is extremely difficult, making accurate prediction challenging. To address this, scholars have developed various system-level simulators to generate simulated data for training and verification, such as RELAP5, TRACE, PCTRAN, CASMO5, and PANGU [12, 13]. Yang et al. [14] used a thermal-hydraulic simulator to generate simulated data and developed a hybrid model combining LSTM and graph convolutional networks to predict coolant temperature, pressure, and flow rate in the primary circuit. Miao et al. [15] employed a full-scope simulator to simulate loss-of-coolant accidents and proposed a lightweight spatial-temporal model for sectional power prediction and early anomaly warning in NPPs. Li et al. [16] utilized simulated data from the full-scope simulator of the Qinshan NPP to design a hybrid model that effectively predicts steam flow and water temperature during transients for the next 5 seconds.

Although existing simulators provide high-fidelity physical modeling, systematic discrepancies remain between simulated and real data in terms of operating noise, numerical distributions, and dynamic behavior [5]. The fundamental reason is that mathematical equations in simulators are simplified from real systems and cannot perfectly emulate nuanced physical processes [15, 17]. This domain gap presents a "zero-shot" forecasting challenge: models trained on simulated data struggle to generalize to real-world data, decreasing predictive accuracy under actual transients. The challenge becomes more critical during never-before-seen reactor incidents, where increased complexity expands the model's blind zones and limits its ability to predict rare trend evolution.

As artificial general intelligence (AGI) continues to evolve, large language models (LLMs) stand out for their ability to autonomously reason and process natural language, such as ChatGPT, Llama, and Qwen [18]. Through chain-of-thought prompts and weight tuning, LLMs have the potential to generalize well in zero-shot and few-shot tasks, such as fault diagnosis [19], battery health management [20], and agricultural queries [21]. Qi et al. [22] utilized an LLM to capture transient knowledge in prompt texts for accident identification of NPPs. Zhou et al. [23] solely tuned the positional encoding and normalization layers in GPT-2 and achieved solid performance on various tasks, including anomaly detection, imputation, classification, and prediction. To enhance reasoning in wind speed prediction, Wu et al. [24] devised temporal and spatial prompts with autoregressive fine-tuning, yielding satisfying results. The impressive performance of LLMs across disciplines can be attributed to their large-scale parameters, enabling what are known as surprising abilities (called emergent abilities [25]). Even in highly specialized domains with inadequate training data, properly fine-tuned LLMs still exhibit extrapolation ability.

In summary, existing studies on customizing LLMs can be broadly classified into three categories: (1) applying pretrained LLMs to generate specific outputs using prompt settings; (2) focusing on fine-tuning partial weights in LLMs; and (3) combining task-specific prompts with partial weight tuning in LLMs. Despite significant advancements, current studies have not explored the capabilities of LLMs for forecasting nuclear reactor states, especially under unusual operations and unexpected accidents. The challenges posed by complex systems and extremely limited transient data in NPP scenarios demand further exploration.

To answer the question "Can large language models precisely forecast nuclear reactor states under never-before-seen transients?" this paper devises a nuclear reactor-oriented LLM (NRO-LLM) framework, which incorporates four sub-modules: temporal decomposition, reversible instance normalization, adaptive prompt mechanism, and LLM fine-tuning. The zero-shot forecasting performance is validated under two types of transient conditions: real grid peak shaving and simulated abnormal accidents. The main contributions are summarized as follows:

1. Unlike conventional data-driven modeling that relies on a sufficient specific dataset for training, this study explores the potential of LLMs by leveraging multi-domain public datasets for fine-tuning, enabling the extraction of cross-domain knowledge and enhancing extrapolation in zero-shot transient forecasting for NPPs.
2. Inspired by human reasoning of recalling past experiences for better decisions, the adaptive prompt mechanism within the NRO-LLM is designed to retrieve highly relevant soft prompts from a prompt pool based on input data, which are integrated with hard prompts to construct a refined prompt set for guiding LLMs.
3. A heterogeneous fusion architecture is established to endow diverse LLMs with bimodal encoding capability. The joint representation of time-series data and prompt information enhances state prediction under never-before-seen transients in nuclear reactors.

The remainder of this article is organized as follows: Section II introduces the key operating parameters that characterize nuclear reactor states. Section III introduces the proposed NRO-LLM framework in detail. In Section IV, the precision and superiority of the proposed method are validated by comparative experiments. Section V concludes this article.

Preliminaries

Fig. 1 [Figure 1: see original paper] (a) illustrates the power generation process of a commercial NPP, comprising the primary circuit, the secondary circuit, and the tertiary circuit. The primary circuit, which consists of the nuclear reactor, pressure vessel, main pump, pressurizer, and steam generator, handles the transfer of nuclear fission heat via steam generator tubes. The secondary circuit performs the thermal-to-mechanical energy conversion: water vaporizes in the steam generator, drives the turbine, and powers the generator. The tertiary circuit condenses exhaust steam and dissipates heat via cooling towers or seawater systems.

The nuclear reactor is the vital component of energy conversion. Therein, in-core operating parameters—such as sectional power, axial power deviation, outlet temperature, and inlet temperature—reveal the reactor's state from different perspectives. Sectional power indicates internal power distribution; axial power deviation shows axial uniformity; outlet and inlet temperatures reflect coolant thermal dynamics. Transient conditions in nuclear reactors can be categorized as either normal operations (e.g., power adjustment, shutdown, startup) or abnormal accidents (e.g., coolant leakage, equipment failure). As presented in Fig. 1 (b), in-core physical parameters exhibit spatial heterogeneity [26]. Improper operation or coolant leakage may cause abrupt changes in the reactor's physical state, resulting in temperature field distortion and power imbalance [27]. As a complex multi-physics coupled system, different accident types can lead to sharp fluctuations in critical physical variables, varying in scope and intensity. This presents a difficulty for traditional data-driven modeling due to limited coverage of training data. To address this, this study leverages reasoning capabilities contained in LLMs to infer reactor behavior under unknown conditions.

Methodology

This study proposes a transient prediction method for nuclear reactor states using the NRO-LLM framework. As depicted in Fig. 2 [Figure 2: see original paper], the framework consists of several modules: temporal decomposition, reversible instance normalization (RevIN), adaptive prompt mechanism, and LLM fine-tuning.

Specifically, multi-domain public datasets (including nuclear power, electricity load, weather, traffic, exchange rate, and transformer temperature) are used as input for fine-tuning the LLM. Temporal decomposition and RevIN are applied to characterize trend-noise features and mitigate scale discrepancies within diverse data. According to input data features, the adaptive prompt mechanism retrieves similar soft prompts from a prompt pool, which are combined with hard prompts to establish the refined prompt set. To fine-tune the pretrained LLM, gradient-based optimization and low-rank adaptation are utilized for updating partial weights in the DeepSeek model. Finally, transient data from NPPs and prompt information are fed into the fine-tuned DeepSeek model, enabling zero-shot prediction under unknown conditions and preemptive understanding of complex nuclear system evolution.

A. Problem Formulation

To forecast the state evolution of nuclear reactors under never-before-seen transients, this study leverages LLMs to explore zero-shot prediction for in-core key operating parameters. Zero-shot prediction refers to the model's ability to forecast future values in a different dataset without being trained on the target dataset, leveraging prior knowledge or auxiliary information. The LLM is fine-tuned with public datasets to learn cross-domain time-series features, including patterns, trends, and dynamics. The model is then tested on unseen transient conditions from an NPP.

The raw dataset (X ^(N×T)) is split into multiple subsequences (, X^(i−1), X^(i), X^(i+1),). Various time-series data (X^(i) ^(N×Tx)) and prompt tokens (P) are input into the model to forecast evolution of the target parameters (Y^(i) ^(N×Ty)) over the next Ty steps. The predictive process of the ith subsequence (X^(i)) is expressed as:

Y^(i) = F_{forecast}(X^(i), P)

where T and N denote the total length of the data and the number of target parameters. Tx and Ty are input and predictive lengths, respectively.

B. Temporal Decomposition

Given the significant trend exhibited by multi-domain data, along with interference from system or environmental noise, we introduce a simple average pooling layer for temporal decomposition at the initial terminal of the model. Unlike traditional variational mode decomposition or empirical mode decomposition, the input data (X^(i)) is split into trend and residual components, allowing the model to learn dominant variations and dynamic perturbations separately. The decomposition is expressed as follows:

X^(i) = X^(i)*{Tre} + X^(i)*{Res}

where X_{Tre} and X_{Res} are the trend and residual components, respectively.

C. Reversible Instance Normalization (RevIN)

Owing to various distributional discrepancies across diverse datasets, we adopt the RevIN technique. By combining instance-level normalization with a learnable affine transform, RevIN mitigates variation across domains while preserving the ability to reconstruct the original distribution.

For target parameters (x^(i)_n ^T) from a subsequence (X^(i) = {x$^{(i)n\}(n=1)}$N), RevIN proceeds in three steps: computing instance statistics, normalization, and inverse normalization. The formulas are as follows [28]:

$^{(i)n\ =\ (1/Tx)\ \ (t=1)}${Tx} x^(i)_n,t σ$^{(i)n\ =\ \surd((1/Tx)\ \ (t=1)}${Tx} (x^(i)_n,t − ^(i)_n)^2 +) x̃^(i)_n,t = γ_n (x^(i)_n,t − ^(i)_n)/σ^(i)_n + β_n
ỹ^(i)_n,t = (ŷ^(i)_n,t − β_n) • σ^(i)_n/γ_n + ^(i)_n

where x̃^(i)_n,t and ỹ^(i)_n,t are the normalized and inverse normalization values for the nth parameter at time t. ^(i)_n and σ^(i)_n are the mean and standard deviation of the nth parameter. is a stability constant. γ_n and β_n are trainable affine matrices that tune the normalized output.

After temporal decomposition and RevIN, the data is segmented into different patches to characterize local temporal patterns. The number of patches (Mp) can be calculated by:

Mp = (T − l)/S + 2

where l is patch size and S is the stride. • represents the truncation operation.

D. Adaptive Prompt Mechanism

Prompts are essential instructions in LLM interactions, where different prompts can guide LLMs to produce distinct outputs. As illustrated in Fig. 3 [Figure 3: see original paper], users can input hard prompts via dialogue to guide the model in forecasting in-core sectional power over the next 10 minutes. Well-crafted prompts enhance the precision, coherence, and overall quality of the generated content, tailoring them to user expectations [29].

A common approach is to employ hard prompts for guiding models [22]. However, this mode may be subjective, inflexible, and limited in semantic scope, leading to suboptimal effect. When encountering unfamiliar problems, humans tend to recall past experiences for better reasoning and decision-making [30]. Inspired by this habit, we devise an adaptive prompt mechanism, which retrieves highly relevant soft prompts from a prompt pool based on input data to combine them with hard prompts, enabling a well-crafted prompt set. Specifically, the universal variation patterns of multi-domain datasets could be encoded into a prompt pool, in which soft prompts denote temporal features of trend or residual components. This design allows the LLM to leverage related past experiences. Even in unseen NPP scenarios, the LLM can dynamically pair features of operating parameters with historically relevant soft prompts to combine with hard prompts, enabling precise predictions.

As illustrated in Fig. 4 [Figure 4: see original paper], hard prompt templates are designed separately for the public training dataset and the NPP test dataset, then processed through a pretrained embedding layer. The prompt pool can be viewed as a shared memory space stored as distinct key-value pairs. According to features of input data, the adaptive prompt mechanism uses similarity matching to retrieve relevant key-value pairs from the prompt pool. The prompt pool (V) is defined as follows:

V = {V^1, V^2, , V^M}

where V^1, V^2, , V^M ^(Lp×D) denotes different soft prompts. Lp and D are the word length and embedding dimension.

For trend component (X^(i)*{Tre}) and its hard prompt (P^(i)*{Tre}), learnable key-value pairs are established for each soft prompt: {(k^1, V^1), (k^2, V^2), , (k^M, V^M)} where M is the length of the prompt pool. k^M ^(LE) and V^M ^(LP×LE) denote the single key and value. LP and LE are the word length and embedding dimension.

The scoring function (cosine distance, $\gamma(\cdot)$) expressed by Eq. (9) measures the similarity between the input data and each key-value pair, retrieving the top-K most relevant soft prompts, which are combined with hard prompts to obtain an adaptive prompt set. This process can be defined in Eq. (10).

γ(P^(i)_{Tre}, k^M) = P$^{(i)\{Tre\}\cdot k\hat{}M\ /\ (\|P\hat{}(i)\{Tre\}\|\cdot\|k}$M‖)

P^(i)_{Tre} = [V^1, , V^K; P^(i)_{Tre}]

Both trend and residual components follow the same selection process.

E. Fine-tuning LLM

The proposed NRO-LLM framework adapts to diverse LLMs for fine-tuning. Given numerical and logical inferences in the forecasting task for reactor states, we employ the open-source DeepSeek-R1-Distill-Qwen-1.5B (called the DeepSeek model) as the pretrained backbone. Developed by the DeepSeek team, the model distills knowledge from DeepSeek-R1 (671 billion) into lightweight Qwen2.5-Math-1.5B, a math-oriented student model with 1.78 billion parameters. To improve logical inference, prompts are merged with multi-domain time-series data to input the DeepSeek model. Inside the original LLM, we update the weights in the root-mean-square normalization layer using the gradient update method from [23], and incrementally train the query and value matrices in the multi-head attention mechanism using low-rank adaptation (LoRA) from [31].

Fig. 5 [Figure 5: see original paper] illustrates the principle of LoRA, which freezes original weights within the pretrained model and uses low-rank matrices (A and B) to approximate weight updates (ΔW). This greatly reduces the number of trainable weights and enhances fine-tuning efficiency. For the weight matrix (W) in the original model, the forward propagation and objective function of the LoRA are described by:

$$h = W_0x + \Delta Wx = W_0x + BAx$$

$$L\_\Phi(\Theta) = \ \_(u,v)\ Z\ lg(P\_\Phi_0+\Delta\Phi(\Theta)(v\_t \mid u,\ v<t))$$

where h is the fine-tuned output. Θ is a smaller-sized set of weights. The above objective function aims to find optimal Θ. Z denotes the training set for fine-tuning. u and v are the input and output of the LoRA. Θ_0 represents the initial weights of the LLM.

The DeepSeek model outputs high-dimensional features for both trend and residual components (Z^(i)*{Res} ^(N×P×Ty)), which are then transformed through a fully connected layer into Ŷ*{Res} ^(N×Ty). Finally, predictive values (Y^(i)) are obtained using the inverse normalization in Eq. (4) and Eq. (13).

Y^(i) = Y^(i)*{Tre} + Y^(i)*{Res}

where Y^(i)*{Tre} and Y^(i)*{Res} are predictive values of trend and residual components, respectively.

Experiments

This section conducts transient prediction experiments of the nuclear reactor using real normal operation and simulated abnormal accidents from a commercial pressurized water reactor (PWR). In-core key operating parameters include sectional power at 24 locations (P1 P24, % FP), axial power deviation across 4 channels (ΔP1 ΔP4, % FP), and outlet/inlet temperatures at 3 positions (T1 T3 and T4 T6, °C). Therein, % FP stands for the percent of full power in the nuclear reactor. P1 P24 and ΔP1 ΔP4 reflect in-core power distribution states, while T1 T3 and T4 T6 correspond to thermotechnical-related behaviors.

A. Experimental Settings

In traditional prediction tasks, a single dataset is typically divided into training and test sets. To evaluate the NRO-LLM model's predictive capabilities under never-before-seen transient conditions in NPPs, we follow the data isolation principle [23, 32], fine-tuning the LLM with a mixture of training data from multi-domain public datasets, including nuclear power, electricity load[1], weather[2], traffic[3], exchange rate[4], and transformer temperature[5], with sampling frequencies ranging from minutes to hours. To avoid undue bias and ensure fair representation of each domain's data in the mixed training data, we select similar proportions of training samples from each domain's dataset to concatenate. Then, the model is tested on unseen private datasets from an NPP. This ensures objective evaluation on unseen datasets.

Notably, as illustrated in Fig. 6 [Figure 6: see original paper], the public nuclear power dataset is generated using the open-source PCTRAN (personal computer transient analysis for nuclear reactors) simulator [33], covering 18 normal and abnormal operating scenarios. We utilize the nuclear reactor-related accidents as training sets. 30% of the nuclear power dataset is randomly selected as the validation set for enhancing feature extraction of NPP behaviors. This setup allows the LLM to learn specific temporal patterns and prior knowledge of abnormal system behaviors in nuclear power plants during fine-tuning.

The test dataset consists of two parts: (1) Real data collected from a PWR, including 40,000 samples (10-second intervals) under normal transients such as power adjustment, shutdown, and startup during grid peak shaving. (2) Simulated data generated by a full-scope simulator, which simulates small-break loss-of-coolant accidents (SBLOCAs), i.e., pipe breaks in the primary circuit. The simulated dataset covers 18 abnormal cases across three loops with breaks at hot, cold, and transition legs, totaling 129,600 samples (1-second intervals). As shown in Fig. 7 [Figure 7: see original paper], the high-fidelity simulator replicates the control room of an NPP and consists of a simulation control room, computing system, and instructor interface. It can accurately simulate a wide range of transient processes and operating conditions, both normal and abnormal. Simulated time progresses at the same rate as real-world time, with a one-to-one correspondence per second.

B. Evaluation Indicators and Comparative Models

We evaluate the forecasting performance for the next 60 time steps (i.e., the next 10 and 1 minutes in real and simulated scenarios, respectively). Experiments are run on a workstation with an Intel(R) Xeon(R) w5-2465X 16-Core CPU, an NVIDIA GeForce RTX 4090 GPU, CUDA 12.2, and CUDNN 8.9.3. All models are developed with Python 3.8 and PyTorch 2.0.1+cu118. Three precision indicators—root mean square error ($\delta\{RMSE\}$*), symmetric mean absolute percentage error (*$\delta\{\text{SMAPE}\}$), and relative absolute error ($\delta\{RAE\}$*)—are utilized to evaluate forecasting performance. The smaller the* $\delta\{\text{RMSE}\}$, $\delta\{SMAPE\}$*, and* $\delta\{\text{RAE}\}$ indicators obtained, the higher the prediction accuracy manifested. To mitigate randomness in testing, each experiment is repeated five times, with the mean and standard deviation of all indicators reported.

We compare the NRO-LLM with 11 representative and state-of-the-art forecasting models, including GRU, sequence to sequence (Seq2Seq), Transformer, Informer [34], decomposed linear (DLinear) [35], PatchTST [36], segmentation recurrent neural network (SegRNN) [37], ESTformer [38], fine-tuning T5 for time series (T5-4TS) [23], fine-tuning GPT-2 for time series (GPT-2-4TS) [23], and Time-LLM [39]. Hyperparameters of all comparative models are optimized through grid search on the validation set. Table 1 lists the hyperparameters of the proposed NRO-LLM.

Table 1 . Model parameter settings of the NRO-LLM model.

Model parameter	Value	Model parameter	Value
Input length (Tx)	60	Output length (Ty)	60
Batch size	32	Learning rate	0.001
Epoch	50	Dropout	0.1
Patch size (l)	12	Patch stride (S)	6
Pool length (M)	100	Number of soft prompts (K)	3
Low rank (r)	8	Zoom factor	2
Pretrained model	DeepSeek-R1-Distill-Qwen-1.5B		

C. Prediction and Early Warning

To answer the question "Can large language models precisely forecast nuclear reactor states under never-before-seen transients?" we select the vanilla Transformer as the typical data-driven model to compare with the proposed NRO-LLM. Unlike the NRO-LLM, the test data is standardized by conventional Z-score to input the Transformer. Fig. 8 [Figure 8: see original paper] presents the zero-shot transient predictions of sectional power (P1 P6) and axial power deviation (ΔP1). During normal operation of the NPP, reactor transients occurred in response to grid load adjustments over holidays, experiencing power adjustment (100 75% FP), shutdown (75 0% FP), and startup (0 100% FP). Since Transformer is trained on multi-domain public datasets, it suffers from numerical distribution discrepancies and exhibits larger forecasting deviations. Although the NRO-LLM model is not fine-tuned on the test dataset, it effectively transfers cross-domain knowledge and integrates bimodal features (operating parameters and prompt information) to achieve precise predictions. Even with distinct dynamic behaviors between sectional power and axial deviation, the NRO-LLM model demonstrates a reliable ability to track and anticipate their evolution.

Fig. 9 [Figure 9: see original paper] illustrates a simulated extreme accident involving a hot-leg pipe break in the reactor's primary circuit. The break causes coolant leakage, impairs in-core heat removal, and triggers a rapid rise in upper-core temperature. Unlike the variation patterns in Fig. 8, local power distributions become irregular, with P6 increasing significantly and ΔP1 exceeding the alarm limit. Under the never-before-seen accident, the predictive performance of Transformer further declines because unexpected accidents cause sharp changes in reactor state and expand the model's blind zones. In the task of forecasting axial power deviation, the curve predicted by Transformer is significantly lagging, causing delayed warning. In contrast, the NRO-LLM's prediction reveals that only minor predictive deviations appear in sectional power curves around the 5500th point. As a whole, the NRO-LLM model correctly forecasts the abnormal reactor state evolution and provides an early warning for ΔP1, exceeding the limit 29 seconds in advance. In summary, the above cases preliminarily verify that the fine-tuned LLM effectively forecasts reactor states under unseen transients compared to conventional data-driven models.

D. Comparison with Various Predictive Models

To further evaluate the superiority of zero-shot prediction, we compare the NRO-LLM against 11 representative and state-of-the-art models: GRU, Seq2Seq, Transformer, Informer, DLinear, PatchTST, SegRNN, ESTformer, T5-4TS, GPT-2-4TS, and Time-LLM. In this experiment, all models are improved by the RevIN module to mitigate numerical distribution discrepancies among multiple-domain datasets.

Table 2 lists prediction errors on four operating parameters. GRU, Seq2Seq, and Transformer exhibit large errors, limited by their baseline architectures. Informer, PatchTST, and ESTformer, as enhanced Transformers, leverage specialized temporal-aware modules that enable more effective feature representation. DLinear and SegRNN, though structurally simple, perform modestly in zero-shot tasks. Notably, T5-4TS and GPT-2-4TS activate pretrained position encoding and specific weights, achieving improved results in both real operations and simulated accidents. Time-LLM reformulates prediction tasks of reactor states as a "language task" using abundant prompts and a reprogramming framework, further boosting predictive accuracy. Overall, four LLM-based models (T5-4TS, GPT-2-4TS, Time-LLM, and NRO-LLM) outperform conventional data-driven models, highlighting better generalization across varying reactor states. Moreover, the proposed NRO-LLM model consistently exhibits optimal precision across all operating parameters.

Table 2. Comparison with various models for forecasting operating parameters.

Parameter	Indicator	GRU	Seq2Seq	Transformer	Informer	DLinear	PatchTST	SegRNN	ETSformer	T5-4TS	GPT-2-4TS	Time-LLM	NRO-LLM
Sectional power	δ_{RMSE} (10^{-1}%FP)	4.45±1.07	3.38±1.41	3.83±0.60	3.37±0.49	2.64±0.41	2.17±0.39	1.55±0.29	1.67±0.36	1.35±0.30			
	δ_{SMAPE} (10^{-2})	4.50±0.44	4.11±0.68	4.04±0.38	3.27±0.30	3.13±0.54	2.34±0.25	1.85±0.23	1.65±0.23	1.47±0.18			
	δ_{RAE} (10^{-2})	8.51±1.51	8.61±1.73	8.52±2.19	6.88±1.64	4.19±0.68	2.51±0.45	3.06±0.96	1.94±0.47	1.45±0.28			
Axial power deviation	δ_{RMSE} (10^{-1}%FP)	4.96±0.44	4.11±0.47	3.01±0.22	2.78±0.19	2.18±0.12	2.15±0.10	2.02±0.08	1.68±0.06	1.57±0.06			
	δ_{SMAPE} (10^{-2})	11.98±1.10	14.6±0.98	11.40±0.97	13.5±0.89	12.3±0.54	9.30±0.23	11.6±0.56	8.14±0.10	7.21±0.2			
	δ_{RAE} (10^{-2})	5.34±0.26	5.28±0.19	4.63±0.28	4.80±0.16	4.86±0.26	4.74±0.15	4.77±0.21	4.46±0.13	4.68±0.22			
Outlet temperature	δ_{RMSE}	0.71±0.24	0.77±0.21	0.66±0.11	0.54±0.08	0.38±0.02	0.36±0.06	0.26±0.03	0.26±0.06	0.27±0.07			
	δ_{SMAPE} (10^{-1}%)	4.11±0.02	3.50±0.01	1.97±0.04	1.56±0.03	0.88±0.01	0.55±0.02	0.69±0.02	0.39±0.03	0.34±0.03			
	δ_{RAE} (10^{-2})	0.88±0.05	0.84±0.06	0.70±0.07	0.52±0.06	0.25±0.02	0.13±0.01	0.24±0.03	0.11±0.01	0.09±0.01			
Inlet temperature	δ_{RMSE}	1.58±0.31	1.35±0.25	1.13±0.31	1.05±0.29	0.59±0.16	0.33±0.07	0.42±0.02	0.23±0.04	0.22±0.01			
	δ_{SMAPE} (10^{-1}%)	3.54±0.10	3.82±0.09	3.06±0.07	2.86±0.09	1.38±0.03	0.65±0.01	1.14±0.01	0.42±0.01	0.38±0.01			
	δ_{RAE} (10^{-2})	1.21±0.8	1.21±0.05	1.02±0.09	0.96±0.08	0.48±0.04	0.29±0.02	0.31±0.04	0.19±0.02	0.18±0.02			

To visually demonstrate the superiority, Fig. 10 [Figure 10: see original paper] presents local prediction curves of axial power deviation (ΔP1) during actual grid peak shaving. Also, we plot the x-line for quantifying predictive deviations. During this period, power distributions in the nuclear reactor respond dynamically, causing significant non-stationary fluctuations. Although traditional models like GRU and Transformer are combined with the RevIN module for mitigating discrepancies among cross-domain data, they still show larger prediction errors. Advanced models such as DLinear, PatchTST, and SegRNN reduce these errors but still struggle to accurately predict local trends. In contrast, the finesse of the NRO-LLM model involves adopting temporal decomposition to characterize dominant variations and dynamic perturbations, in addition to utilizing the adaptive prompt mechanism to enhance LLM reasoning. This conjoint design mitigates inherent modality gaps between time-series data and prompt texts, achieving stable and accurate predictions across varying states in nuclear reactors.

E. Comparison with Various LLMs

The above qualitative and quantitative experiments have demonstrated the advantages of forecasting nuclear reactor states under unknown transients using fine-tuned LLMs. Notably, the proposed NRO-LLM model employs the open-source DeepSeek-R1-Distill-Qwen-1.5B model as its pretrained backbone. The model, developed by the DeepSeek team, distills mathematical reasoning from DeepSeek-R1 (671 billion) into lightweight Qwen2.5-Math-1.5B (1.78 billion). Given the wide variety of open-source LLMs available, this raises another question: would fine-tuning other LLMs achieve similar performance? Thus, five open-source LLMs with similar parameter quantities are utilized to replace the backbone in NRO-LLM to analyze the impact on prediction performance. Table 3 lists the basic information of various LLMs, including bidirectional encoder representations from Transformers (BERT), large language model Meta AI (Llama), Qianwen (Qwen), text-to-text transfer Transformer (T5), and generative pretrained Transformer (GPT). All models are fine-tuned using the similar activation strategy in Section III.E for fair comparison.

Table 3. Basic information of various LLMs.

Version	Model	Weight
BERT-base-uncased[1]	BERT	0.11 billion
Llama-3.2-1B[2]	Llama	1.24 billion
Qwen2.5-1.5B[3]	Qwen	1.54 billion
T5-base[4]	T5	0.22 billion
GPT-2[5]	GPT	0.13 billion
DeepSeek-R1-Distill-Qwen-1.5B[6]	DeepSeek	1.78 billion

[1] BERT: https://huggingface.co/google-bert/bert-base-uncased

[2] Llama: https://huggingface.co/meta-llama/Llama-3.2-1B

[3] Qwen: https://huggingface.co/Qwen/Qwen2.5-1.5B

[4] T5: https://huggingface.co/google-t5/t5-base

[5] GPT: https://huggingface.co/openai-community/gpt2

[6] DeepSeek: https://huggingface.co/deepseek-ai/DeepSeek-R1-Distill-Qwen-1.5B

Fig. 11 [Figure 11: see original paper] shows predictive error comparisons of different LLMs. GPT achieves lower errors in forecasting sectional power and axial power deviation, reflecting stronger scalability in deep temporal modeling. As a whole, comparing the indicators between Table 2 and Fig. 11 comprehensively, Qwen, T5, and GPT as pretrained backbones within our NRO-LLM framework exhibit better effect than conventional data-driven models (ESTformer, SegRNN, PatchTST, etc.). Notably, DeepSeek-R1-Distill-Qwen-1.5B specializes in solving complex mathematical problems and logical reasoning, inheriting knowledge distilled from the larger DeepSeek-R1 model. This enables effective learning of cross-domain temporal features and supports accurate extrapolation under unseen transients in NPPs. The experiment indicates that DeepSeek-R1-Distill-Qwen-1.5B exhibits an unfathomable potential as a powerful foundation for intelligent maintenance and real-world deployment in NPPs with limited computing power.

F. Module Contribution Verification

To delve deeper into the efficacy of each module in the fine-tuning process, a leave-one-out experiment is conducted on NRO-LLM and compared with five variant models: (1) Variant 1: Remove the RevIN module. (2) Variant 2: Disable fine-tuning for the LLM. (3) Variant 3: Exclude the adaptive prompt mechanism. (4) Variant 4: Omit hard prompt information. (5) Variant 5: Cancel the temporal decomposition.

Fig. 12 [Figure 12: see original paper] shows the effectiveness validation of each module (employing the δ_{RMSE} metric). Removing the RevIN module (variant 1) leads to a significant drop in accuracy due to data scale discrepancies across multi-domain datasets, making fine-tuning difficult to converge. In Variant 2, freezing all LLM weights limits the model's ability to integrate time-series features and prompt information, impacting predictive performance. Variants 3 and 4 achieve moderate improvements, but their prompt content remains insufficient. This indicates that incorporating semantic information as discrete indicators within language models can more effectively fuse domain knowledge, particularly in tasks demanding high accuracy and relevance. Without temporal decomposition, Variant 5 struggles to extract inherent patterns and dynamic perturbations within in-core parameters, resulting in slight degradation. Quantitative analysis confirms that each module contributes positively to NRO-LLM's performance, highlighting the significance of proper LLM fine-tuning and refined prompt design for zero-shot transient prediction in NPPs.

G. Fine-tuning Sensitivity Analysis

This study fine-tunes the multi-head attention module in the DeepSeek model using Low-Rank Adaptation (LoRA), aiming to extract cross-domain temporal features and integrate bimodal information. Specifically, LoRA approximates weight updates through a low-rank matrix determined by a low rank (r), significantly reducing the quantity of trainable weights and improving computational efficiency. The rank (r) controls the dimensionality of the low-rank matrices. The smaller the r set, the fewer model weights trained, and the lower computing power resources required. In NPPs with limited computing power, the trade-offs between model capacity and hardware resource usage are critical.

To explore the influence of rank size on fine-tuning, a gradient experiment is conducted with r {1, 2, 4, 8, 16, 32}, with all zoom factors set to twice the rank. As shown in Fig. 13 [Figure 13: see original paper], the NRO-LLM model maintains stable performance for r between 2 and 16. When r = 1, computing power usage is minimal, but model accuracy deteriorates due to underfitting. This indicates that fine-tuning with LoRA benefits from moderate ranks and is sensitive to extreme values, revealing a nuanced relationship between model hyperparameter and generalization.

H. Discussion on Different Activated Modules

Despite the distinct network structure of LLMs, the multi-head attention mechanism remains fundamental, involving separate query, key, and value matrices (Q, K, and V). As mentioned in Section III.E, we incrementally update weights of Q and V using LoRA during fine-tuning. Fig. 14 Figure 14: see original paper illustrates that the DeepSeek-R1-Distill-Qwen-1.5B model contains 28 attention mechanism layers. The activation of different modules introduces varying numbers of trainable parameters and affects the model's performance. Motivated by curiosity, we further discuss this influence by comparing the following variants: (1) NRO-LLM-FAM: Freeze attention mechanism. (2) NRO-LLM-Q: Activate Q matrix. (3) NRO-LLM-K: Activate K matrix. (4) NRO-LLM-V: Activate V matrix. (5) NRO-LLM-QK: Activate Q and K matrices. (6) NRO-LLM-QV*: Activate Q and V matrices. (7) NRO-LLM-KV: Activate K and V matrices. (8) NRO-LLM-QKV: Activate Q, K, and V matrices.

As illustrated in Fig. 14(b), when all attention layers are frozen, NRO-LLM-FAM struggles to fully learn shared knowledge across multi-domain datasets, leading to higher error. As the number of activated metrics increases, prediction errors decline significantly. When two matrices among Q, K, and V are activated, the performance approaches optimal. The number of trainable parameters during fine-tuning is 1,225,728, accounting for only 0.0689% of the total. However, activating all attention matrices leads to increased errors and overfitting in prediction-specific tasks, further forgetting original knowledge in the LLM and reducing predictive extrapolation under unseen conditions in reactors. It is inferred that moderate activation of the multi-head attention mechanism strikes a better balance between model adaptability and knowledge retention.

Conclusion

This work proposed a customized LLM framework (called NRO-LLM) to address the question, "Can large language models precisely forecast nuclear reactor states under never-before-seen transients?" The NRO-LLM integrates temporal decomposition, reversible instance normalization, adaptive prompt mechanism, and LLM fine-tuning.

According to the principle of data separation, the model was validated by four in-core operating parameters under two types of transients of a nuclear reactor: sectional power, axial power deviation, outlet temperature, and inlet temperature. In real-world grid peak shaving, the NRO-LLM model accurately predicted varying reactor states through cross-domain knowledge extraction and bimodal feature fusion. In simulated abnormal accidents, the model also exhibited strong extrapolation and early warning capabilities under never-before-seen conditions. Comparative experiments confirmed that LLM-based methods had generally superior performance in transient forecasting over various conventional data-driven methods. Compared to diverse LLMs with the same-scale parameter quantity, the fine-tuned architecture built on DeepSeek-R1-Distill-Qwen-1.5B exhibits an unfathomable potential as a powerful foundation for intelligent prediction. Ablation studies validated the contribution of each module in the fine-tuning process.

To the best of our knowledge, this paper is the first study to explore the potential of LLMs for forecasting nuclear reactor states. The proposed method showcases a viable pathway for advancing intelligent transformation in NPPs. Overall, the synergistic integration of LLMs with the nuclear industry is a promising frontier that can offer opportunities and warrant extensive future interdisciplinary investigations.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Data Availability

The authors do not have permission to share data.

References

[1] W. Xu, J. Li, H. Xie et al., "Conceptual design and safety characteristics of a new multi-mission high flux research reactor," *Nucl. Sci. Tech.*, vol. 34, p. 34, 2023. doi: 10.1007/s41365-023-01172-9

[2] Q. Guo, X. Wu, H. Cai et al., "Multi-power sources joint scheduling model considering optimal nuclear power peak regulation," *Energy*, vol. 293, p. 130678, 2024. doi: 10.1016/j.energy.2024.130678

[3] W. Lin, X. Miao, J. Chen et al., "ST-SAM: Spatial-temporal state adaptation model for neutron detector fault detection and isolation in nuclear power plants," *IEEE T. Ind. Inform.*, vol. 21, no. 2, pp. 1110-1119, 2025. doi: 10.1109/TII.2024.3459022

[4] D. B. Lv, R. G. Chen, X. Y. Li et al., “Preliminary study on improving the automation level of large commercial pressurized water reactor,” in *Lecture Notes in Electrical Engineering*, vol. 779, 2021, pp. 429-435. doi: 10.1007/978-981-16-3456-7-41

[5] W. Lin, X. Miao, J. Chen et al., “Simulation-to-reality transferability framework for operating-parameter forecasting in nuclear reactors using domain adaptation,” *Nucl. Sci. Tech.*, vol. 36, no. 5, 2025. doi: 10.1007/s41365-025-01643-1

[6] D. A. Ejigu, X. J. Liu, “Gradient descent-particle swarm optimization based deep neural network predictive control of pressurized water reactor power,” *Prog. Nucl. Energy*, vol. 145, p. 104108, 2022. doi: 10.1016/j.pnucene.2021.104108

[7] H. P. Nguyen, P. Baraldi, E. Zio, “Ensemble empirical mode decomposition and long short-term memory neural network for multi-step predictions of time series signals in nuclear power plants,” *Appl. Energy*, vol. 283, p. 116346, 2021. doi: 10.1016/j.apenergy.2020.116346

[8] Q. Lu, Y. Yuan, F. C. Li et al., “Prediction method for thermal-hydraulic parameters of nuclear reactor system based on deep learning algorithm,” *Appl. Therm. Eng.*, vol. 196, p. 117272, 2021. doi: 10.1016/j.applthermaleng.2021.117272

[9] H. D. Song, X. J. Liu, M. Q. Song, “Comparative study of data-driven and model-driven approaches in prediction of nuclear power plants operating parameters,” *Appl. Energ.*, vol. 341, p. 121077, 2023. doi: 10.1016/j.apenergy.2023.121077

[10] H. Kim, J. Kim, “Long-term prediction of safety parameters with uncertainty estimation in emergency situations at nuclear power plants,” *Nucl. Eng. Tech.*, vol. 55, no. 5, pp. 1630-1643, 2023. doi: 10.1016/j.net.2023.01.026

[11] Y. J. Fu, D. Z. Zhang, Y. L. Xiao et al., “An interpretable time series data prediction framework for severe accidents in nuclear power plants,” *Entropy*, vol. 25, no. 8, p. 1160, 2023. doi: 10.3390/e25081160

[12] D. She, B. Xia, J. Guo et al., “Prediction calculations for the first criticality of the HTR-PM using the PANGU code,” *Nucl. Sci. Tech.*, vol. 32, p. 90, 2021. doi: 10.1007/s41365-021-00936-5

[13] K. M. Yassin, M. H. Hassan, M. M. Ghoneim et al., “Multiphysics simulation of VVER-1200 fuel performance during normal operating conditions,” *Nucl. Sci. Tech.*, vol. 34, p. 28, 2023. doi: 10.1007/s41365-023-01172-9

[14] L. J. Yang, Z. Miao, T. Li et al., “LSTM-GCN based experimental multidimensional prediction framework for system level parameter analysis,” *Ann. Nucl. Energy*, vol. 210, p. 110890, 2025. doi: 10.1016/j.anucene.2024.110890

[15] W. Lin, X. Miao, J. Chen et al., “SR-STM: Simulation-reality spatial-temporal model for early warning of power tilt in nuclear power plants,” *IEEE*

Internet Things, pp. 1-15, 2025. doi: 10.1109/JIOT.2025.3539864

[16] X. Li, K. Cheng, T. Huang et al., “Research on short term prediction method of thermal hydraulic transient operation parameters based on automated deep learning,” *Ann. Nucl. Energy*, vol. 165, p. 108777, 2022. doi: 10.1016/j.anucene.2021.108777

[17] X. Y. Li, K. Cheng, T. Huang et al., “Equivalence analysis of simulation data and operation data of nuclear power plant based on machine learning,” *Ann. Nucl. Energy*, vol. 163, p. 108507, 2021. doi: 10.1016/j.anucene.2021.108507

[18] Y. Hu, H. Kim, K. Ye et al., “Applying fine-tuned LLMs for reducing data needs in load profile analysis,” *Appl. Energ.*, vol. 377, p. 124666, 2025. doi: 10.1016/j.apenergy.2024.124666

[19] L. Lin, S. H. Zhang, F. Song et al., “FD-LLM: Large language model for fault diagnosis of complex equipment,” *Adv. Eng. Inform.*, vol. 65, p. 103208, 2025. doi: 10.1016/j.aei.2025.103208

[20] H. Peng, C. Y. Liu, H. Li et al., “Large-language-model-enabled health management for internet of batteries in electric vehicles,” *IEEE Internet Things*, vol. 12, no. 6, pp. 6082-6094, 2025. doi: 10.1109/JIOT.2024.3513002

[21] J. C. Jiang, L. Yan, H. F. Liu et al., “Knowledge assimilation: Implementing knowledge-guided agricultural large language model,” *Knowl-Based Syst.*, vol. 314, p. 113197, 2025. doi: 10.1016/j.knosys.2025.113197

[22] B. Qi, J. Sun, Z. Sui et al., “Multimodal learning using large language models to improve transient identification of nuclear power plants,” *Prog. Nucl. Energy*, vol. 177, p. 105421, 2024. doi: 10.1016/j.pnucene.2024.105421

[23] T. Zhou, P. S. Niu, X. Wang et al., “One fits all: Power general time series analysis by pretrained LM,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, 2023, pp. 43322-43355. doi: 10.48550/arXiv.2302.11939

[24] T. J. Wu, Q. Ling, “STELLM: Spatio-temporal enhanced pre-trained large language model for wind speed forecasting,” *Appl. Energ.*, 2024. doi: 10.1016/j.apenergy.2024.124034

[25] B. W. Jin, G. Liu, C. Han et al., “Large language models on graphs: A comprehensive survey,” *IEEE T. Knowl. Data En.*, vol. 36, no. 12, pp. 8622-8642, 2024. doi: 10.1109/TKDE.2024.3469578

[26] J. Chen, Y. Z. Lu, H. Jiang et al., “Twin model-based fault detection and tolerance approach for in-core self-powered neutron detectors,” *Nucl. Sci. Tech.*, vol. 34, p. 117, 2023. doi: 10.1007/s41365-023-01276-2

[27] S. Yang, X. B. Li, Y. S. Liu et al., “Response characteristics of PWR primary circuit under SBLOCAs considering steam bypass discharging,” *Nucl. Sci. Tech.*, vol. 35, p. 101, 2024. doi: 10.1007/s41365-024-01457-7

[28] T. Kim, J. Kim, Y. Tae et al., "Reversible instance normalization for accurate time-series forecasting against distribution shift," in *International Conference on Learning Representations*, 2021.

[29] G. Marvin, N. Hellen, D. Jjingo et al., "Prompt engineering in large language models," in *International Conference on Data Intelligence and Cognitive Informatics*, 2024, pp. 387-402. doi: 10.1007/978-981-99-7962-2_{30}

[30] Z. F. Wang, Z. Z. Zhang, C. Y. Lee et al., "Learning to prompt for continual learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 139-149. doi: 10.48550/arXiv.2112.08654

[31] E. J. Hu, Y. L. Shen, P. Wallis et al., "LoRA: Low-rank adaptation of large language models," in *International Conference on Learning Representations (ICLR)*, 2022, pp. 1-20. doi: 10.48550/arXiv.2106.09685

[32] B. N. Oreshkin, D. Carpov, N. Chapados et al., "Meta-learning framework with applications to zero-shot time-series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 10, 2021, pp. 9242-9250. doi: 10.48550/arXiv.2002.02887

[33] B. Qi, X. Y. Xiao, J. G. Liang et al., "An open time-series simulated dataset covering various accidents for nuclear power plants," *Sci. Data*, vol. 9, p. 766, 2022. doi: 10.1038/s41597-022-01889-5

[34] H. Y. Zhou, S. H. Zhang, J. Q. Peng et al., "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, 2021, pp. 11106-11115. doi: 10.48550/arXiv.2012.07436

[35] A. L. Zeng, M. X. Chen, L. Zhang et al., "Are transformers effective for time series forecasting?" in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 9, 2023, pp. 11121-11128. doi: 10.48550/arXiv.2205.13504

[36] Y. Q. Nie, N. H. Nguyen, P. Sinthong et al., "A time series is worth 64 words: Long-term forecasting with transformers," in *International Conference on Learning Representations*, 2023. doi: 10.48550/arXiv.2205.13504

[37] S. S. Lin, W. W. Lin, W. T. Wu et al., "SegRNN: Segment recurrent neural network for long-term time series forecasting," arXiv, 2308.11200, 2023.

[38] G. Woo, C. H. Liu, D. Sahoo et al., "ESTformer: Exponential smoothing transformers for time-series forecasting," arXiv, 2202.01381, 2022.

[39] M. Jin, S. Y. Wang, L. T. Ma et al., "Time-LLM: Time series forecasting by reprogramming large language models," in *International Conference on Learning Representations*, 2024. doi: 10.48550/arXiv.2310.01728

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.
