Toward unified and accurate description of multidimensional nuclear observables via a prior-informed large language model

Shijie Guo(郭世杰)[1], Shouyu Wang(王守宇)[1,2*], Enhong Wang(王恩宏)[1,2], Zhongming Niu(牛中明)[3] and Yiming Ding(丁一鸣)[1]

[1]*Shandong Provincial Key Laboratory of Nuclear Science, Nuclear Energy Technology and Comprehensive Utilization, Weihai Frontier Innovation Institute of Nuclear Technology, School of Nuclear Science, Energy and Power Engineering, Shandong University, Jinan 250061, People's Republic of China*

[2]*WeiHai Research Institute of Industrial Technology, Shandong University, Weihai 264209, People's Republic of China*

[3]*School of Physics, Anhui University, Hefei 230601, People's Republic of China*

A prior–informed large language model (LLM) driven multi–task learning framework is proposed for the unified and accurate description of multiple nuclear observables. By fine–tuning the pre–trained DeepSeek-R1-1.5B model with Low–Rank Adaptation (LoRA), lightweight adapters are introduced while preserving general pre–trained parameters. Under a causal language modeling paradigm, the model is trained autoregressively on deviations between experimental and theoretical values. Significant accuracy improvements are achieved across seven observables, including charge radii, masses, binding energies, separation energies, and decay energies, with the training loss decreasing by over 98% across all tasks. This demonstrates that the LLM–based framework, through structured prior embedding, offers an efficient and shared approach for multi–task regression in fundamental nuclear properties.

Key words : Nuclear masses and charge radii; Multi-task; Large language model; Low-rank adaptation

The precise description of the fundamental properties of atomic nuclei has long been one of the central driving forces in nuclear research. Nuclear observables not only serve as important labels on the nuclear chart but also provide insight into the complex interactions between nucleons [1–4]. These observables collectively form an intricate and inseparable constraint network. In strongly interacting nuclear systems, a unified, self–consistent, and high–precision theoretical description of these properties requires a proper treatment of complex many–body correlations. Achieving an accurate description of these observables remains challenging for existing models [5–9].

In recent years, data–driven machine learning (ML) methods have opened new gates for the study of nuclear properties [10,11]. Tree models [12–14], Bayesian analysis [13,15–18,21,22], neural networks [16,19–21,23,24] and kernel ridge regression [25–28] have been widely applied to the prediction of nuclear properties. These include decay lifetimes [29–31], excitation energies [32,33], reaction cross sections [34–37], charge radii [13,21,38–40], etc. These advancements fully demonstrate the unique advantages of machine learning in capturing complex nonlinear correlations within nuclear structure data.

High–precision fitting of a single observable does not necessarily imply the self–consistent unification of multi–dimensional physical properties. When the same model framework is applied to the study of multi–dimensional properties [41], the sensitive regions of model parameters for different observations may not be mutually compatible. As a result, their loss functions may generate opposing gradient directions in the parameter space, making it difficult for a single model to simultaneously converge to the optimal regions for each objective. To solve this problem, researchers have introduced multi–objective optimization strategies to find optimal balance solutions in multi–dimensional space [42]. These efforts have produced a series of valuable exploratory results [43–49]. Nevertheless, existing multi–objective research still has limitations in predictive accuracy, objective count, feature simplicity, and computational efficiency. Therefore, the optimization of model architectures and the enhancement of model expressiveness have become key issues.

Since the advent of large language models (LLMs), the transfer value of hundreds of billions or even trillions of pre–trained parameters has increasingly been explored in the scientific domain [50–53]. LLMs possess the potential to capture more hidden correlation patterns in nuclear structure data with their vast number of parameters providing substantial model complexity [54]. This capability is particularly crucial in multi–objective training processes, where a deeper parameter space can provide a stronger capacity for multi–task learning. Furthermore, no study has yet been conducted on the fundamental nuclear properties utilizing LLM. Based on this, the present work proposes a LLM–driven multi–objective research framework that integrates prior information. Specifically, we fine–tune a pre–trained LLM (DeepSeek-R1-1.5B) [55], and couple it with physical prior information. This enables the framework to predict a higher–dimensional target outputs by using simpler features, thereby achieving a high–precision unified description of multiple nuclear properties.

Theoretical framework. Unlike the traditional regression paradigm that directly models the numerical space, LLM processes data within a generative framework based on natural language sequence generation. Taking N and Z as input

*Corresponding authors. Email: sywang@sdu.edu.cn

© 2026 Chinese Physical Society and IOP Publishing Ltd

features, Q_α, Q_β, *mass*, *BE* as output targets as an example, the original numerical data must first be formatted into a JSON–structured file to adapt to the input sequence parsing specifications of DeepSeek-R1-1.5B:

$$\begin{aligned}\mathbf{x} &= \mathrm{concat}\left(t_{\text{im_start}}, t_{\text{sys}}, \mathbf{s}, t_{\text{im_end}}, t_{\text{im_start}}, t_{\text{user}}, \mathbf{u}, t_{\text{im_end}}, t_{\text{im_start}}, t_{\text{assist}}, \mathbf{v}, t_{\text{im_end}}, t_{\text{eos}}\right) \\ &= \langle|\text{im_start}|\rangle\ \text{system}\ \mathbf{s}\ \langle|\text{im_end}|\rangle\ \langle|\text{im_start}|\rangle\ \text{user}\ \mathbf{u}\ \langle|\text{im_end}|\rangle\ \langle|\text{im_start}|\rangle\ \text{assistant}\ \mathbf{v}\ \langle|\text{im_end}|\rangle\ \langle|\text{endoftext}|\rangle,\end{aligned} \tag{1}$$

where $\mathbf{s}$ is system instructions that embed domain–specific prior knowledge, $\mathbf{u}$ encodes the query content including N and Z, and $\mathbf{v}$ represents the target response placeholder to be predicted during autoregression. The special tokens $\langle|\text{im_start}|\rangle$ and $\langle|\text{im_end}|\rangle$ demarcate the structural boundaries of each dialogue turn, and $\langle|\text{endoftext}|\rangle$ marks the end of the sequence. After parsing:

$$\begin{aligned}\mathbf{x} = {} & \langle|\text{im_start}|\rangle\text{system You are .... } \langle|\text{im_end}|\rangle \\ & \langle|\text{im_start}|\rangle\text{user The neutron number is } N\text{, the proton number is } Z\text{, result is? } \langle|\text{im_end}|\rangle \\ & \langle|\text{im_start}|\rangle\text{assistant The } \alpha\text{-decay energy is } Q_\alpha\text{, the } \beta\text{-decay energy is } Q_\beta\text{, ... } \langle|\text{im_end}|\rangle\ \langle|\text{endoftext}|\rangle.\end{aligned} \tag{2}$$

The input sequence $\mathbf{x}$ constructed above produces a discrete token series of length T, denoted as:

$$\mathbf{x} = [x_1, x_2, \ldots, x_T], \tag{3}$$

where each element $x_t \in \mathcal{V}$ is an integer token identifier from the model's vocabulary, $x_1 =$ `<|im_start|>`, x_2 corresponds to the role token `system`, and $x_T =$ `<|endoftext|>` marks the boundary of the sequence. Under the causal language modeling paradigm, the model is trained to predict the subsequent token based on all preceding context. Correspondingly, the target sequence $\mathbf{y}$ is constructed by shifting the input sequence one position to the right:

$$\mathbf{y} = [x_2, x_3, \ldots, x_T]. \tag{4}$$

It contains $T-1$ prediction targets. For any position $t \in \{1, 2, \ldots, T-1\}$, given the prefix context $[x_1, x_2, \ldots, x_t]$, the model is optimized to maximize the probability of generating the true token $y_t = x_{t+1}$. This autoregressive form enables the model to learn the underlying conditional distribution of token sequences, thereby capturing the syntactic structure of prompt templates and the physical laws embedded within structured data in the kernel.

The input sequence $\mathbf{x}$ of length T is first mapped to a sequence of high–dimensional embedding vectors via a frozen embedding layer:

$$\mathbf{e}_t = \mathrm{Embedding}(x_t) \in \mathbb{R}^d, \quad t = 1, \ldots, T. \tag{5}$$

Therein, the model dimension is $d = 1536$, corresponding to the adopted DeepSeek-R1-1.5B architecture. The model stacks $L = 28$ identical transformer layers. Each layer employs grouped-query attention (GQA) with $H = 12$ query heads and $H_{kv} = 2$ key/value heads, RMSNorm for normalization, a SwiGLU feed-forward network with intermediate size $d_{ff} = 8960$, and rotary position embedding (RoPE) applied to queries and keys.

Within each layer, the attention score[50] is computed by the scaled dot–product:

$$\begin{aligned}\mathbf{h}^{(l)} &= \mathbf{h}^{(l-1)} + \mathrm{ATT}(\mathrm{RM}(\mathbf{h}^{(l-1)})), \\ \mathbf{h}^{(l)} &= \mathbf{h}^{(l)} + \mathrm{FFN}(\mathrm{RM}(\mathbf{h}^{(l)})), \\ \mathrm{ATT}(\mathbf{h}^{(l)}) &= \left(\sum_{j=1}^{T} \alpha_{t,j}\,(\mathbf{h}_j^{(l)} W_v)\right) W_o, \\ \alpha_{t,j}^{(h)} &= \mathrm{softmax}\left(\frac{(\mathrm{RoPE}(\mathbf{h}_t^{(l)} W_q^{(h)}))(\mathrm{RoPE}(\mathbf{h}_j^{(l)} W_k^{(g(h))}))^T}{\sqrt{d_k}} + M_{t,j}\right), \\ \mathrm{FFN}(\mathbf{h}_t^{(l)}) &= \left(\mathrm{SiLU}(\mathbf{h}_t^{(l)} W_{\text{gate}}) \odot (\mathbf{h}_t^{(l)} W_{\text{up}})\right) W_{\text{down}},\end{aligned} \tag{6}$$

where, $\mathbf{W} \in \mathbb{R}^{d \times d_k}$ are projection matrices, RM represents RMS normalization (RMSNorm), $\mathrm{SiLU}(x) = x \cdot \sigma(x)$ is the Sigmoid Linear Unit (Swish) activation used in the SwiGLU variant, $\odot$ denotes element-wise multiplication, $g(h)$ maps query head h to its corresponding KV head in GQA, and M is employed to restrict information sharing between the current position t and position j:

$$M_{t,j} = \begin{cases} 0, & t \geq j, \\ -\infty, & t < j. \end{cases} \tag{7}$$

All projection matrices are in a low–rank adaptation (LoRA) manner[56,57]. Unlike traditional parameter update methods, the projection matrix is reparameterized as:

$$\mathbf{W}_q^{\text{final}} = \mathbf{W}_q^{\text{pretrained}} + \mathbf{B}_q \mathbf{A}_q, \tag{8}$$

among these, the pre–trained weights $\mathbf{W}_q^{\text{pretrained}}$ are kept frozen, and only the low–rank matrices $\mathbf{B}_q \in \mathbb{R}^{d \times r}$ and $\mathbf{A}_q \in \mathbb{R}^{r \times d_k}$ (rank $r \ll d$) participate in training.

After propagation through all L layers, the resulting hidden layer $\mathbf{h}_t^{\text{final},L}$ is passed to the output layer. The final hidden state at each position is mapped to a probability distribution over the vocabulary $\mathcal{V}$ via an output projection matrix $\mathbf{W}_{\text{head}}$:

$$\begin{aligned} \mathbf{o}_t &= \mathbf{h}_t^{\text{final},L}\mathbf{W}_{\text{head}} \in \mathbb{R}^{|\mathcal{V}|}, \\ \mathbf{p}_t &= \text{softmax}(\mathbf{o}_t), \\ p_t(k) &= \frac{\exp(o_{t,k})}{\sum_{j=1}^{|\mathcal{V}|}\exp(o_{t,j})}. \end{aligned} \tag{9}$$

The loss function for a single sequence is defined as the average negative log–likelihood over $T-1$ predicted positions:

$$\mathcal{L}(\mathbf{\Theta}) = -\frac{1}{T-1}\sum_{t=1}^{T-1}\log p_t(y_t), \tag{10}$$

where $p_t(y_t)$ is the predicted probability for the ground truth label y_t. The total loss on the training set $\mathcal{D}$ is the empirical average:

$$\mathcal{L}_{\text{total}}(\mathbf{\Theta}) = \frac{1}{|\mathcal{D}|}\sum_{k=1}^{|\mathcal{D}|}\mathcal{L}_k(\mathbf{\Theta}). \tag{11}$$

Trainable parameters $\mathbf{\Theta}$ only contain LoRA matrices $\{\mathbf{B}_q, \mathbf{A}_q, \mathbf{B}_v, \mathbf{A}_v, \mathbf{B}_o, \mathbf{A}_o, \ldots\}$ in all layers. These parameters are determined by minimizing a loss function via stochastic gradient descent. The gradients of the loss with respect to the low–rank matrices (e.g., $\mathbf{A}_q$) are computed via backpropagation using the chain rule.

The essence of this method lies in introducing lightweight adapters trained on a specific domain, while preserving the frozen pre–trained weights for general language and reasoning capabilities. By reconstructing and learning multiple related target quantities, the model can leverage the deeper implicit shared structural information within the data, thereby enhancing the generalization ability for all targets.

Data preparation and model design. Data sources[9,58,59] and statistics of preprocessed data are shown in Table 1. The residuals between experimental values and theoretical values are taken as the training target by the model. In the theoretical model construction, Q_α was predicted using the Random Forest algorithm, while Q_β was predicted based on the LightGBM algorithm. The evaluation results on the testing set indicate that the MAE and RMSE of ML_{RF} are 0.2259 MeV and 0.3852 MeV, respectively, while the corresponding errors for $\text{ML}_{\text{LightGBM}}$ are 0.1447 MeV and 0.3754 MeV. Additionally, r is given by a semi–empirical formula, and the other four parameters are derived from the mass model WS4.

Table 1. Data description.

Feature name	Description	Experimental value	Theoretical value	Dataset size
r	nuclear charge radius	[58]	$1.1 \times A^{1/3}$	799
$mass$	nuclear mass			
BE	binding energy	[9]	WS4	3231
S_p	single-proton separation energy			
S_n	single-neutron separation energy			
Q_α	α-decay energy	[59]	ML_{RF}	3413
Q_β	β-decay energy		$\text{ML}_{\text{LightGBM}}$	3166

To achieve multi–dimensional comparative analysis, five distinct training paradigms were constructed, with the feature combinations for each paradigm defined as shown in Table 2. Across all task categories, the dataset was divided into training and testing sets with a ratio of 8:2. The data distributions of the training and testing sets for Type4 and Type5 are shown in Figure 1. The testing set was not involved in model training at any stage and was solely utilized for evaluating generalization performance post–training, thereby preventing data leakage from compromising result fidelity. Two evaluation metrics, MAE and RMSE, were employed to quantify the performance of the LLM for comparison with other reference results.

Table 2. Model design.

Paradigm	Input features	Output features	Size
Type1	N, Z, M_p, M_n, δ	r	799
Type2	N, Z	r	799
Type3	N, Z	$mass$, BE, S_p, S_n	3231
Type4	N, Z	$mass$, BE, Q_α, Q_β	2986
Type5	N, Z, M_p, M_n, δ	Q_α, Q_β, BE, S_p, S_n	2864

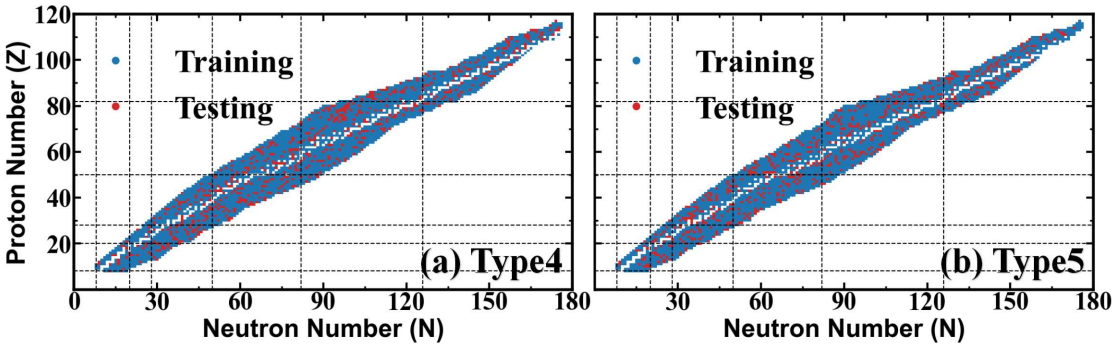

Fig. 1. The data distributions of the training and testing sets in Type4 (a) and Type5 (b). The dashed lines in the figure indicate the positions of magic numbers.

Results and discussion. To reveal the characteristics of the LLM in processing regression tasks, a single–task learning framework is first employed to predict the charge radius. Type1 takes N, Z, M_p, M_n and δ as input features, where M_p (M_n) denotes the distance between the proton (neutron) number and the nearest magic number, and $\delta = \frac{1}{2}[(-1)^Z + (-1)^N]$ for characterizing the nucleon pairing effect. After training, across the entire dataset, the model achieved MAE and RMSE of 0.0086 fm and 0.0167 fm on the training set, and 0.0136 fm and 0.0221 fm on the testing set, respectively. In the $Z \geq 8$ region, the RMSE for the training and testing sets further decreased to 0.0102 fm and 0.0195 fm, exhibiting a slight accuracy advantage compared with Refs. [38,39]. The fitting curve and outlier distribution are shown in Figure 2 (a).

Unlike traditional ML, which relies on manually constructed features, the LLM theoretically possesses the capability to automatically extract effective representations from raw inputs. In this regard, Type2 analyzes r using only N and Z as input. Under this approach, the model achieves MAE and RMSE of 0.0139 fm and 0.0467 fm on the training set for all regions, and 0.0198 fm and 0.0515 fm on the testing set. Compared to Type1, the RMSE increase exceeds a factor of 2. In the region $Z \geq 8$, the dual–feature scheme achieves RMSE values of 0.0154 fm and 0.0204 fm on the training and testing sets, respectively, which are at almost the same level as Refs. [38,39] and Type1, effectively demonstrating the applicability of the LLM for regression tasks. Figure 2 (b) also illustrates the fitting curves and outlier distribution.

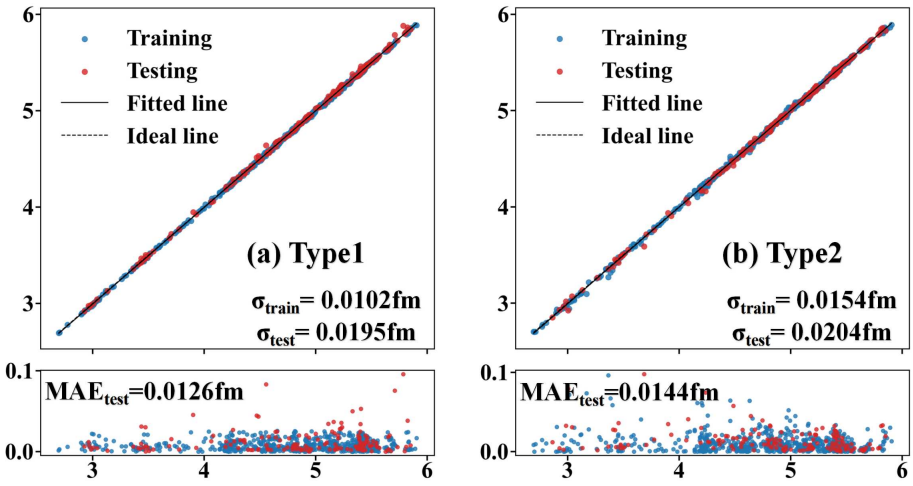

Fig. 2. Comparison between experimental and predicted r in the region $Z \geq 8$. (a) Results from Type 1 with five input features, showing the fitting curve and outlier distribution. (b) Results from Type 2 with only N and Z as input features. The model accurately captures the main trend of charge radii even with minimal input, though precision degrades slightly in the light–nucleus region compared to Type 1.

Using only N and Z, the LLM can capture the main trend of r variation. This stems from the autonomous feature extraction capability brought by the deeper and more complex architectural advantages of the LLM: even for simple input features, the model can autonomously extract the underlying deep correlations. However, for both Type1 and Type2, the accuracy in the $Z \geq 8$ region is higher than that in the entire region, revealing that the light nuclear region, due to the complexity of its few–body correlation and cluster structure, has always been a more challenging domain for

charge radius prediction. This phenomenon also indirectly indicates the necessity of more prior information: the prior information extractable from N and Z alone has inherent limitations in the light nuclear region. When physical features, including magic numbers and pairing effects, are explicitly introduced, the model's prediction error can be effectively suppressed.

To evaluate the framework's stability and uncertainty, the trained model is used for 20 predictions. The predicted values fluctuated by no more than 0.7%.

Given that simple input features did not significantly negatively impact accuracy in the single–objective task, N and Z are directly used for training in the multi–objective task: Type3 simultaneously predicted $mass$, BE, S_p and S_n. The MAE values on the training and testing sets were between 0.20 MeV and 0.30 MeV, and the RMSE was between 0.30 MeV and 0.55 MeV. The numerical accuracy was slightly insufficient, but it is noteworthy that the prediction accuracy of $mass$ and BE improved almost synchronously, showing that the model outputs maintained the built-in algebraic consistency between these two quantities. When the output targets were replaced from $mass$ and BE to Q_α and Q_β, the RMSE remained between 0.25 MeV and 0.45 MeV, showing no significant improvement. This indicates that the presence of implicit equality constraints between target tasks is not the dominant factor determining the accuracy level.

In contrast, Type4 targets $mass$, BE, Q_α, Q_β, with the precision of each physical quantity presented in Table 3. The MAE for Q_α and Q_β in the testing set reached approximately 0.09 MeV, with RMSE within 0.13 MeV, while $mass$ corresponded to 0.16 MeV and 0.20 MeV, respectively, all exhibiting consistent performance improvements [44–46]. The errors between predicted and true values for all data are plotted in Figure 3. The model's systematic underestimation for outputs S_p and S_n stems from the data structure of these two physical quantities and nuclear many–body effects.

Table 3. The RMSE(MeV) for Each Multi–task Objective.

Data	Task	Q_α	Q_β	$mass$	BE	S_p	S_n
Training	Type4	0.1051	0.1060	0.1639	0.1638	–	–
	Type5	0.0778	0.0838	–	0.1495	0.1878	0.1941
Testing	Type4	0.1275	0.1226	0.2012	0.2011	–	–
	Type5	0.1051	0.1059	–	0.1693	0.2096	0.2118

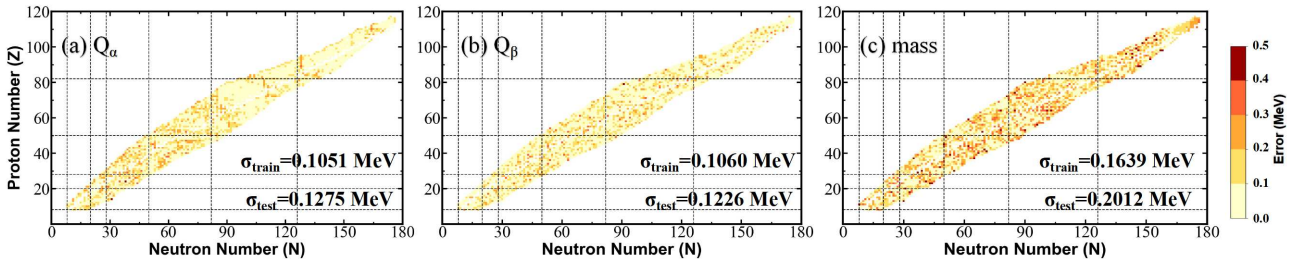

Fig. 3. Predicted versus experimental values for multi–task learning with Type 4 (input: N and Z). The deviations are shown for all four observables. The RMSE for Q_α and Q_β on the testing set is below 130 keV, while those for $mass$ and BE are around 200 keV. The dashed lines in the figure indicate the positions of magic numbers.

The separation energy is essentially a differential operation on the binding energies of adjacent nuclides. This structure amplifies the mutation of the binding energy surface near magic numbers, leading to significant local fluctuations and non–smooth features in the data distribution. When only N and Z are used as input, the model does not possess any explicit prior information, making direct learning difficult. Conversely, Q_α and Q_β, during their construction, systematically cancel out a large number of local fluctuations influenced by shell effects, making the overall function smoother and thus easier for the model to fit. This explains why, under the same input conditions, the prediction error for separation energy is higher than that for decay energy, and further confirms that the data smoothness of the target physical quantity itself plays a decisive role during training.

To compensate for this deficiency and maximize the prediction accuracy of various physical quantities as much as possible, physical quantities related to magic numbers and pairing effects, M_p, M_n and δ are explicitly introduced as additional input features. The performance of the trained Type5 on the results confirms the effectiveness of this approach, with detailed results listed in Table 3. On the training set, the MAE for Q_α and Q_β reached within 0.065 MeV, and the RMSE was around 0.080 MeV. For S_p and S_n, the prediction accuracy was also significantly improved, decreasing to within 0.150 MeV and 0.195 MeV, respectively, with BE reaching 0.116 MeV and 0.149 MeV. On the testing set, the errors for each physical quantity were highly consistent with the training set, with deviations controlled within 0.020 MeV. Specifically, the MAE for Q_α and Q_β was around 0.080 MeV, and the RMSE was around 0.105 MeV. S_p and S_n corresponded to 0.170 MeV and 0.210 MeV, respectively, with BE reaching 0.135 MeV and 0.169 MeV [44–46]. The strong consistency between training and testing performance under random train-test splitting stems from the

model's parameter-sharing mechanism, which effectively encodes local neighborhood correlations on the (N, Z) nuclear chart. In stricter extrapolation validation, the prediction accuracy for each fundamental nuclear observable degrades by approximately 50–70 keV.

To further verify the architectural advantage of the LLM-LoRA framework, Figure 4 plots the test-set RMSE of different Transformer architectures on Type 4 and Type 5. The results demonstrate that the LLM-LoRA framework can tap into the representational capacity of a far larger base model with substantially lower training computational cost, and achieves higher fitting accuracy across both tasks.

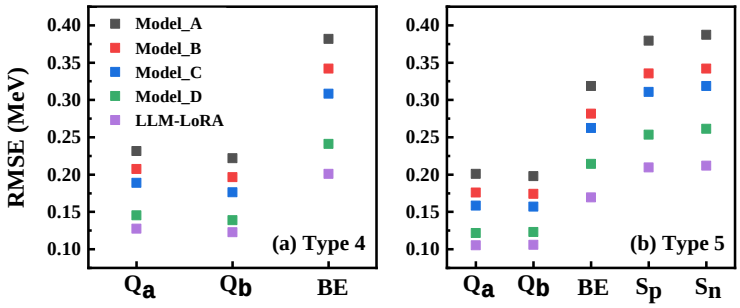

Fig. 4. Test-set RMSE performance of different Transformer architectures on Type 4 and Type 5. Model_A denotes the standard Transformer with 20 million trainable parameters, which is fully trained from scratch. Model_B, Model_C and Model_D denote Transformers sharing identical architecture with the LLM-LoRA base model, with 20 million, 40 million, and the compressed effective parameter size of this base model, respectively. All of them are randomly initialized and fully trained from scratch without LoRA adaptation.

The loss (including both syntactic patterns and numerical precision) variation trends and proportions for various tasks are summarized in Table 4. During the training process, the loss value decreased significantly: the single–objective task decreased by over 99%, and the optimization range for the multi–objective task remained stable above 98%. When the input features contain more prior physical information (magic numbers, pairing effects), the model exhibits lower initial loss and a higher optimization proportion. This also validates the positive role of prior information in the LLM framework, but its improvement is more limited compared to ML. These results indicate that the LLM-based framework offers another effective route for nuclear property prediction.

Table 4. Loss variation values across tasks.

Task	Initial loss	Final loss	Improvement
Type1	9.5360	0.0459	99.52%
Type2	10.1834	0.0682	99.33%
Type4	8.8967	0.1775	98.13%
Type5	8.6367	0.1105	98.72%

In summary, a large language model–driven multi–objective research framework integrating prior information is proposed for the study of nuclear observables. By adopting LoRA technology, this framework uses lightweight adapters tailored for specific nuclear property tasks while preserving the general pre–trained parameters. Under the causal language modeling paradigm, the model undergoes autoregressive training on the residuals between experimental data and theoretical models.

In the single–objective task, predicting r based on five input features, the MAE and RMSE on the testing set in the region where $Z \geq 8$ were 0.0126 fm and 0.0195 fm, respectively. In the multi–objective task, using only N and Z as input features to train Q_α, Q_β, $mass$ and BE, the model achieved RMSE below 130 keV for Q_α and Q_β on the testing set, around 200 keV for $mass$ and BE, and effectively captured the intrinsic correlation between $mass$ and BE. When the input features were expanded to five, and the output features were Q_α, Q_β, BE, S_p and S_n, the RMSE for Q_α and Q_β in the testing set reached approximately 100 keV, BE improved to within 170 keV, while S_p and S_n were around 200 keV. Compared with some traditional ML methods, the framework demonstrated consistent performance improvements across various tasks. Meanwhile, during the training process, the loss value showed a significant downward trend: the decrease exceeded 99% in the single–objective task and remained stable above 98% in the multi–objective task, validating the robustness and effectiveness of the framework.

The aforementioned results indicate that by incorporating structured prior embeddings, the LLM–based framework establishes a novel and efficient approach for multi–objective regression tasks in fundamental nuclear properties.

Although LLMs can achieve decent accuracy across various tasks, they also have certain limitations. More complex architectures, while improving performance, tend to enhance model interpretability challenges. Meanwhile, the high computational cost restricts their application scenarios. This is one of the directions that requires discussion and exploration.

Acknowledgements. This work is partly supported by the National Natural Science Foundation of China (Grants No. 12225504 and No. 1232100). The numerical calculations in this work have been done on the super computing system in Shandong University, Weihai.

References

[1] Dobaczewski J, Michel N, Nazarewicz W *et al.* 2007 *Prog. Part. Nucl. Phys.* **59** 432
[2] Tullney K, Allmendinger F, Burghoff M *et al.* 2013 *Phys. Rev. Lett.* **111** 100801
[3] Wen P S, Holt J W, and Li M 2024 *Phys. Rev. Lett.* **133** 252501
[4] Ding M Q, Fang D Q, and Ma Y G 2024 *Nucl. Sci. Tech.* **35** 211
[5] Möller P, Sierk A J, Ichikawa T *et al.* 2016 *At. Data Nucl. Data Tables* **109** 1
[6] Weizsäcker C F v 1935 *Z. Phys.* **96** 431
[7] Geng L S, Toki H, and Meng J 2005 *Prog. Theor. Phys.* **113** 785
[8] Afanasjev A V, Agbemava S E, Ray D *et al.* 2013 *Phys. Lett. B* **726** 680
[9] Wang N, Liu M, Wu X Z *et al.* 2014 *Phys. Lett. B* **734** 215
[10] Xu X Y, Deng L, Chen A X *et al.* 2024 *Nucl. Sci. Tech.* **35** 91
[11] He W B, Li Q F, Ma Y G *et al.* 2023 *Sci. China Phys. Mech. Astron.* **66** 282001
[12] Liu G P, Wang H L, Zhang Z Z *et al.* 2025 *Phys. Rev. C* **111** 024306
[13] Li Z L, Wang Y J, Li Q F *et al.* 2025 *Phys. Rev. C* **112** 014312
[14] Liu H, Li X X, Yuan Y *et al.* 2025 *Nucl. Sci. Tech.* **36** 21
[15] Xie J Z, Wang K P, Wang C *et al.* 2024 *Phys. Rev. C* **109** 064317
[16] Niu Z M, and Liang H Z 2018 *Phys. Lett. B* **778** 48
[17] Neufcourt L, Cao Y C, Nazarewicz W *et al.* 2018 *Phys. Rev. C* **98** 034318
[18] Zhang X Y, Li W F, Fang J Y *et al.* 2024 *Nucl. Phys. A* **1043** 122820
[19] Gernoth K A, Clark J W, Prater J S *et al.* 1993 *Phys. Lett. B* **300** 1
[20] Gazula S, Clark J W, and Bohr H 1992 *Nucl. Phys. A* **540** 1
[21] Dong X X, An R, Lu J X, and Geng L S 2023 *Phys. Lett. B* **838** 137726
[22] Liu J, Tan K Z, Wang L *et al.* 2025 *Nucl. Sci. Tech.* **36** 215
[23] Cao Y Y, Guo J Y, and Zhou B 2023 *Nucl. Sci. Tech.* **34** 152
[24] Richardson K A, Trifinopoulos S, and Williams M 2025 arXiv:2508.08370
[25] Wu X H, and Zhao P W 2020 *Phys. Rev. C* **101** 051301
[26] Wu X H, and Pan C 2024 *Phys. Rev. C* **110** 034322
[27] Tian J L, Ma P F, Wu X H *et al.* 2025 *Phys. Rev. C* **112** 064306
[28] Wu X H, Pan C, Zhang K Y *et al.* 2024 *Phys. Rev. C* **109** 024310
[29] Niu Z M, Liang H Z, Sun B H *et al.* 2019 *Phys. Rev. C* **99** 064307
[30] Jalili A, Pan F, Luo Y A *et al.* 2025 *Phys. Rev. C* **111** 034321
[31] Ma N N, Bao X J, and Zhang H F 2021 *Chin. Phys. C* **45** 024105
[32] Lv B F, Li Z L, Wang Y J *et al.* 2024 *Phys. Lett. B* **857** 139013
[33] Li Z L, Lv B F, Wang Y J *et al.* 2026 *Chin. Phys. C* **50** 014107
[34] Li Z L, Gao Z P, Liu L *et al.* 2024 *Phys. Rev. C* **109** 024604
[35] Akkoyun S, Yesilkanat C M, and Bayram T 2024 *Comput. Phys. Commun.* **74** 109055
[36] Choi H, Mitra S, Brodksy J *et al.* 2025 *Phys. Rev. C* **112** 044601
[37] Amrani N, and Akkoyun S 2025 *Appl. Radiat. Isot.* **225** 112059
[38] Wang Y D, Shang T S, Xie H H *et al.* 2026 *Nucl. Sci. Tech.* **37** 93
[39] Shang T S, Xie H H, and Li J 2024 *Phys. Rev. C* **110** 014308
[40] Tang L, and Zhang Z H 2024 *Nucl. Sci. Tech.* **35** 19
[41] Caruana R 1997 *Mach. Learn.* **28** 41
[42] Zhang Y, and Yang Q 2017 *Natl. Sci. Rev.* **5** 30
[43] Meyer R, Weichselbaum M, Hauser A W *et al.* 2020 *Theory Comput.* **16** 5685
[44] Wu X H, Lu Y Y, and Zhao P W 2022 *Phys. Lett. B* **834** 137394
[45] Ming X C, Zhang H F, Xu R R *et al.* 2022 *Nucl. Sci. Tech.* **33** 48
[46] Ye W H, and Wan N 2026 *Phys. Rev. C* **113** 024304
[47] Li C Q, Tong C N, Du H J *et al.* 2022 *Phys. Rev. C* **105** 064306
[48] Bai J H, Niu Z M, Sun B Y *et al.* 2021 *Phys. Lett. B* **815** 136147
[49] Kitouni O, Nolte N, Trifinopoulos S *et al.* 2023 arXiv:2306.06099
[50] Vaswani A, Shazeer N, Parmar N *et al.* 2023 arXiv:1706.03762
[51] OpenAI 2024 arXiv:2303.08774
[52] Zhao W X, Zhou K, Li J Y *et al.* 2026 arXiv:2303.18223

[53] Lu C, Lu C, Lange R T *et al.* 2026 *Nature* **651** 914
[54] Song X Y, Li O, Lee C *et al.* 2025 arXiv:2402.14547
[55] DeepSeek-AI 2025 *Nature* **645** 633
[56] Ding N, Qin Y, Yang G *et al.* 2023 *Nat. Mach. Intell.* **5** 220
[57] Hu E J, Shen Y L, Wallis P *et al.* 2021 arXiv:2106.09685
[58] Angeli I 2004 *At. Data Nucl. Data Tables* **87** 185
[59] Wang M, Huang W W, Kondev F G *et al.* 2021 *Chin. Phys. C* **45** 030003