EAI Endorsed Transactions
on Energy Web

Research Article **EAI.EU**

Intelligent Question-Answering System and Inference Mechanism for Grid Dispatching with Energy Storage Access Scenarios Based on Semantic Models

Lin Yang[1]*, Fan Yang[1], Yongqing Feng[1], Cheng Zeng[1], Peng Zhou[1], Hongliang Gao[1]

[1]China Southern Power Dispatching & Control Center, Guangzhou, 510000, China

Abstract

INTRODUCTION: In the power grid energy storage dispatch system, after the large-scale integration of new energy storage systems, the demand for inquiries regarding the operation and maintenance of storage equipment as well as the coordinated power regulation has significantly increased. The intelligent question-answering function is of vital importance in enhancing the efficiency of dispatchers.
OBJECTIVES:The research focuses on the field of scheduling, integrating expertise in power dispatching and energy storage grid connection operation with natural language processing technology, and has constructed an efficient question-answering generation framework. A semantic model optimized by dynamic masking and large-batch training is combined with a dispatching knowledge graph to enhance semantic understanding.
METHODS: The model achieves an intent classification accuracy of 0.97 and an F1 score of 98.1%. The accuracy of terminology disambiguation and complex sentence analysis reaches 92.5% and 84.8%, respectively. Based on this framework and a reasoning mechanism, an intelligent question answering system is developed using multi-algorithm collaborative reasoning to support answer generation and optimization.
RESULTS: Experimental results show that the system maintains an accuracy of 91.3% even at difficulty level 5. The violation rate of dispatching procedures is as low as 22.5%, with a maximum memory usage of 805 MB and a response time of 63.2 seconds.
CONCLUSION: This system can automatically generate effective responses, optimize the accuracy and rationality of question-and-answer interactions, and precisely output compliant control plans for abnormal handling issues related to energy storage grid connection. It demonstrates significant advantages in enhancing response efficiency and accuracy.

Keywords: Power grid dispatching; Reasoning mechanism; Intelligent question answering; Robustly optimized BERT pretraining approach; Translation embedding; Energy storage and grid connection

Received on 01 July 2026, accepted on 01 September 2026, published on 21 September 2026

doi: 10.4108/ew.13834

*Corresponding author. Email: yl251223@163.com

1. Introduction

In recent years, with the large-scale deployment of grid-connected energy storage systems, the power system has entered a new stage characterized by energy storage–grid coordination and flexible regulation. With the digital transformation of energy storage dispatching operations, there is a growing need for intelligent question answering systems that can understand technical terms, complex operational logic, and relationships among equipment to support decision-making efficiently [1]. In real-world dispatching, when dealing with equipment faults or multi-procedure coordination tasks, traditional question answering methods often fail due to shallow semantic understanding and insufficient knowledge association, resulting in term misjudgment and logical errors [2]. Existing approaches mainly rely on general-purpose pretrained models such as Bidirectional Encoder Representations from Transformers (BERT) or simple knowledge retrieval methods to answer dispatching questions [3]. While these models show some success in basic semantic analysis, they often lack domain knowledge, which leads to poor adaptability, weak reasoning ability, and unstructured responses. These limitations prevent them from meeting the strict requirements for accuracy, expertise, and logical consistency in power dispatching [4]. The Robustly Optimized BERT Pretraining Approach (RoBERTa) enhances semantic modeling through dynamic masking and large-batch training. Translation Embedding (TransE) is effective in mining relationships among entities in knowledge graphs. Multi-algorithm collaborative reasoning integrates rule-based, case-based, and logic-based reasoning approaches to support accurate response generation [5]. As large-scale energy storage gradually integrates into the grid dispatch system, new inquiries such as energy storage charging and discharging control, and equipment anomaly consultation have further raised the professional analysis requirements for dispatching questions [6-7]. This study combines improved RoBERTa for semantic understanding, TransE for enhanced knowledge retrieval, and collaborative reasoning for generating coherent and professional answers. A full-process intelligent question-answering system is built, offering precise semantic parsing, deep knowledge linkage, self-consistent reasoning, and standardized response generation. The innovation lies in constructing a domain-adaptive loop of semantics, knowledge, and reasoning, providing a professional and intelligent solution to support efficient dispatching decisions.

2. Related works

RoBERTa, as a key model for enhancing semantic understanding, was optimized through strategies such as dynamic masking and large-batch training. It deeply explored semantic associations and long-range dependencies in text, improving both the accuracy and generalization of semantic modeling. This model attracted widespread attention in natural language processing and related fields from scholars worldwide [8]. For example, Özkurt C addressed the need for performance evaluation of text-based question answering systems in natural language processing. Using the Stanford Question Answering Dataset as a benchmark, the study adopted RoBERTa and showed that the model delivered strong performance, offering useful references for the development and application of question answering systems in both research and industry [9]. Cheruku R et al. focused on improving sentiment analysis accuracy on social media platforms. They modified the RoBERTa model to extract more relevant contextual information and conducted comparative experiments on a Twitter comment dataset. The results demonstrated significant performance improvements, with the highest accuracy reaching 84.6% [10]. Madichetty S et al. tackled the challenge of identifying multimodal information tweets during disasters. They proposed a method combining pretrained RoBERTa and the VGG-16 model. Experiments on multiple disaster-related datasets showed that their approach outperformed existing baselines across various evaluation metrics [11]. Kumar B V et al. studied sentiment classification of tweets on social media using both BERT and RoBERTa models as well as their fusion architecture. Their findings revealed that both models achieved excellent performance in terms of F1 score, recall, and accuracy, enabling high-precision emotion classification for tweets [12].

With the increasing integration of large-scale grid-connected energy storage systems, power system operation has gradually evolved from traditional power dispatching toward energy storage–oriented dispatching and coordinated regulation. In this context, many studies have focused on improving the intelligence, safety, and efficiency of energy storage dispatching systems. For example, Yang L et al. addressed the problem of poor alignment performance in multi-entity knowledge graphs within power grid energy storage dispatching scenarios. They proposed a collective entity alignment model that learns entity and relation embeddings through an attention network and integrates structural, attribute, and relational similarities. Experimental results demonstrated that the proposed method outperforms existing approaches in knowledge fusion tasks for energy storage dispatching systems [13]. Li X et al. focused on the instability of power flow prediction in energy storage dispatching environments caused by the lack of long-term forecasting. They proposed a Conformer-RLpatching optimization algorithm, which integrates Transformer-based long-term prediction, confidence estimation for error reduction, and a dispatching evaluation mechanism. Experimental results showed that the method achieves higher safety scores compared with baseline algorithms in energy storage–assisted dispatching scenarios [14]. Fan S and Guo J investigated large-scale energy storage integrated power grid dispatching and control systems. They discussed the architecture, key technologies, and application scenarios of human–machine hybrid intelligent systems, aiming to enhance the intelligence level of energy storage dispatching through human–machine collaboration [15]. Ji Z et al. addressed the inefficiency of manual fault handling in next-generation energy storage dispatching systems. They proposed a joint entity–relation extraction method based on pretrained models, constructed a knowledge graph for energy storage dispatching scenarios, and developed an intelligent decision-support system to improve fault handling efficiency [16].

Overall, existing studies have made significant progress in RoBERTa-based semantic modeling and its applications in power system domains. However, most existing approaches still focus on general power dispatching scenarios and lack sufficient modeling of energy storage-specific operational logic, such as charging/discharging coordination, state-of-charge constraints, and grid interaction mechanisms. Therefore, this study further adapts the RoBERTa model to energy storage dispatching scenarios and integrates knowledge graph reasoning and multi-algorithm collaborative inference to construct a semantic understanding module, thereby improving semantic parsing accuracy and knowledge association capability in intelligent energy storage dispatching question-answering systems.

3. Intelligent question-answering system for energy storage dispatching

3.1 Layered construction of question answering generation framework in dispatching

In the field of energy storage dispatching, intelligent question answering systems need to accurately parse natural language questions from dispatchers. These questions often contain specialized terminology, complex operational logic, and topological relationships among equipment, which place high demands on the professionalism and precision of semantic understanding. However, traditional BERT mainly relies on general-domain corpora during pretraining, lacking professional knowledge in the power dispatching domain. As a result, it performs poorly in capturing specialized meanings, often leading to term misinterpretation and broken contextual logic [17]. In contrast, RoBERTa introduces dynamic masking, adopts larger batch sizes, and removes the next sentence prediction task. These improvements allow it to better learn deep semantic features of text. Especially when fine-tuned with domain-specific corpora, RoBERTa shows superior performance in representing specialized terms and modeling contextual associations [18]. Therefore, this study adopts the improved RoBERTa as the core algorithm model of the semantic understanding module. The structure and the semantic parsing process are shown in Figure 1.

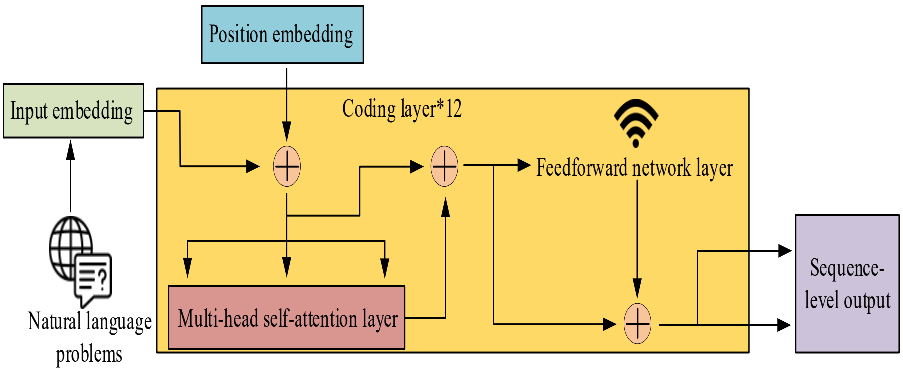

Figure 1. Module composition and semantic analysis flow diagram (Icon source from: https://www.freepik.com/)

As shown in Figure 1, the semantic understanding module is built on the improved RoBERTa architecture. It takes natural language questions from the power dispatching domain as input. The input embedding layer maps professional terms into low-dimensional semantic vectors, and the position embedding layer captures the sequential relationships among words. These are passed into a 12-layer encoder composed of multi-head self-attention and feed-forward networks. The attention mechanism extracts semantic associations within the text, and the feed-forward layer enhances the nonlinear representation of features. The module finally outputs sequence-level and sentence-level semantic vectors, which serve as precise semantic representations for the downstream knowledge retrieval and matching modules, laying the foundation for structured analysis of dispatching questions. Model adaptation expands the semantic encoding requirements for energy storage entities such as energy storage power systems and energy storage stations. The training loss of the masked language model is defined in Equation (1).

$$L_{MLM} = -\sum_{i \in masked} \log P\left(\omega_i \middle| Context\right) \quad (1)$$

In Equation (1), $i \in masked$ represents the set of masked token positions, ω_i refers to the true token at position i, and $P(\omega_i | Context)$ is the predicted probability that token ω_i appears at position i based on context. The scaled dot-product attention in the multi-head self-attention layer is shown in Equation (2).

$$Attention(Q,K,V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

In Equation (2), Q represents query vectors, K stands for key vectors, V represents the value vector containing input information representations, and $\sqrt{d_k}$ is the scaling factor. After RoBERTa completes semantic parsing, implicit associations among devices, rules, and operations in the power dispatching knowledge graph still need to be mined. General TransE reasoning lacks domain constraints and tends to produce results that violate operational logic [19]. This study adapts TransE to dispatching scenarios by constructing negative samples that violate dispatching rules and assigning logical weights to distance metrics, thus enabling domain-enhanced reasoning, as shown in Figure 2.

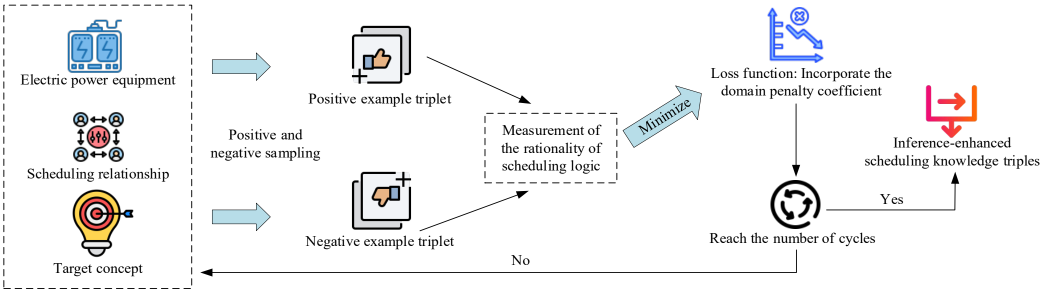

Figure 2. Schematic diagram of the power dispatch scenario adaptation process (Icon source from: https://www.freepik.com/)

In Figure 2, the process starts by constructing positive triples that follow dispatching regulations and negative triples that violate them. TransE evaluates the logical validity of each triple and incorporates domain-specific penalty coefficients into the loss function. After iterative optimization, it outputs reasoning-enhanced knowledge triples with confidence scores, effectively revealing implicit associations not explicitly stored in the graph. To measure alignment, TransE uses either the L1 or L2 norm to calculate distances. The L1 norm is defined in Equation (3) [20].

$$\text{distance}_{L1}(h+r,t) = \sum_{i=1}^{d} \left|(h+r)_i - t_i\right| \quad (3)$$

In Equation (3), h, r, and t represent the low-dimensional embeddings of the head entity, relation, and tail entity, respectively. t_i is the i-th dimension of the tail entity vector, and d represents the total number of dimensions. The L2 norm is calculated in Equation (4).

$$\text{distance}_{L2}(h+r,t) = \sqrt{\sum_{i=1}^{d} \left[(h+r)_i - t_i\right]^2} \quad (4)$$

In Equation (4), $\left[(h+r)_i - t_i\right]^2$ denotes the squared difference of the i-th dimension, emphasizing violations of strict dispatching constraints. Through knowledge retrieval and reasoning, the system obtains regulation clauses, implicit knowledge, and logical constraints. However, core knowledge is often fragmented. Directly generating responses from such fragments may result in colloquial terminology or disordered operation steps [21]. These fragmented inputs fail to meet the strict requirements for professionalism and logical clarity in power dispatching and may even introduce ambiguity into dispatching instructions. Therefore, a systematic approach is needed to integrate and optimize the knowledge. This study builds a three-level response generation mechanism—knowledge-enhanced generation, domain-specific validation, and logic optimization—to address these issues, as shown in Figure 3.

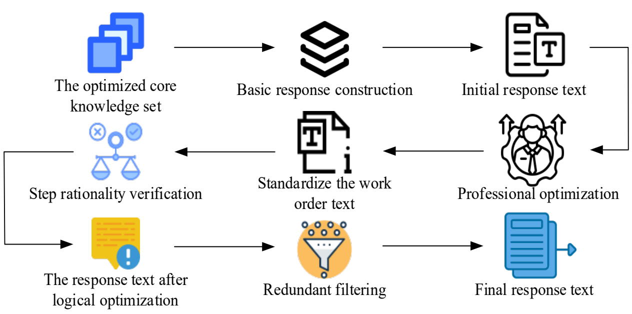

Figure 3. Schematic diagram of the process of integrating and optimizing knowledge (Icon source from: https://www.freepik.com/)

As shown in Figure 3, the module uses the optimized core knowledge set as input. First, a generation model fine-tuned on dispatching corpora using RoBERTa produces initial responses. Then, with support from a dispatch terminology library and TransE-based regulation references, the response is refined into a standardized work order. Next, irrelevant knowledge is removed via redundancy filtering, and step consistency is validated based on operational process rules. Finally, the output is a logically coherent, professionally formatted response that complies with dispatching regulations. The generation loss is defined in Equation (5).

$$L_{gen} = -\sum_{i=1}^{N} y_i \log\left(p_i\right) \qquad (5)$$

In Equation (5), N is the token length of the generated text, y_i is the actual token at position i in the reference answer, and p_i is the predicted probability of token y_i at position i. Overall, this framework integrates the semantic understanding layer, knowledge retrieval and matching layer, and answer generation layer to ensure the accuracy, professionalism, and logical consistency of question answering. It provides reliable support for dispatchers to quickly obtain valid information and improve work efficiency.

3.2 Intelligent question answering system based on reasoning mechanism and framework

Although the RoBERTa semantic model and TransE reasoning support the transformation from natural language to standardized responses, static knowledge retrieval alone is insufficient for deep requirements such as hidden knowledge associations and complex scenario adaptation. To address this, the study introduces a multi-algorithm collaborative reasoning mechanism that combines rule-based, case-based, and logic-based reasoning. This mechanism systematically resolves knowledge gaps, logical inconsistencies, and scenario mismatches in dispatching QA. The reasoning flow is shown in Figure 4.

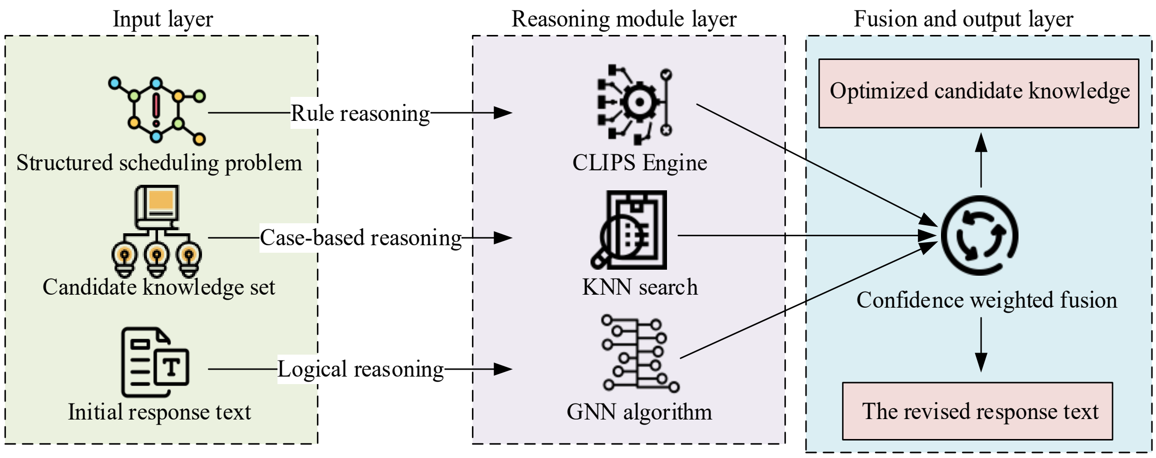

Figure 4. Schematic diagram of the reasoning mechanism of multi-algorithm collaboration (Icon source from: https://www.freepik.com/)

As shown in Figure 4, the mechanism takes structured questions, candidate knowledge sets, and initial responses as input. First, the CLIPS engine performs rule-based reasoning to verify whether the candidate knowledge complies with dispatching regulations. In parallel, KNN-based case reasoning searches historical dispatching cases for similar scenarios and adjusts to current equipment parameters to provide references. Then, the GNN algorithm performs logic reasoning by building topological relationships and operational step chains among devices to check and fix logical conflicts in the initial answer. Finally, a confidence-weighted fusion of reasoning results produces optimized candidate knowledge and a revised response. The rules and case database are synchronously inputted to store the relevant regulations and historical fault cases related to energy storage grid connection management [22]. In the case-based reasoning module, the number of retrieved neighbors in KNN was empirically set to 5 according to the validation performance on the dispatching case database, which provided a balance between retrieval accuracy and computational efficiency. The logical reasoning module adopted a graph convolutional network (GCN) structure to model the topology and operation dependencies among dispatching entities. During reasoning fusion, the outputs from rule-based reasoning, case-based reasoning, and logical reasoning were normalized and combined through confidence-weighted aggregation. The final confidence score was calculated by considering the reliability of each reasoning source, where rule-based reasoning was assigned higher priority for regulatory constraints, while case-based and logical reasoning contributed complementary scenario information.

In rule-based reasoning, conflict indicators and penalty coefficients are used to check for regulation violations, as defined in Equation (6) [23].

$$L_{rule} = \lambda \cdot \sum_{g \in \Re} \square\left(\mathrm{Conflict}(k,g)\right) \tag{6}$$

In Equation (6), k and g represent a candidate knowledge item and a single rule from the rule base. λ and $\Re$ are the penalty weights and the full set of dispatching regulations, respectively. To support multi-featured dispatching scenarios, a weighted cosine similarity function is used to retrieve similar historical cases, as defined in Equation (7).

$$\mathrm{Sim}(q,c) = \frac{\sum_{i=1}^{m} \omega_i \cdot q_i \cdot c_i}{\sqrt{\sum_{i=1}^{m} \omega_i \cdot q_i^2}\sqrt{\sum_{i=1}^{m} \omega_i \cdot c_i^2}} \tag{7}$$

In Equation (7), q and c are feature vectors of the current and historical problems, ω_i is the weight of the i-th feature, and m is the total number of dimensions. To further improve accuracy, logical consistency, and system adaptability, the study builds a reasoning fusion system based on the layered QA framework and multi-algorithm reasoning. This system is named EDIQAR. The overall structure of EDIQAR is shown in Figure 5.

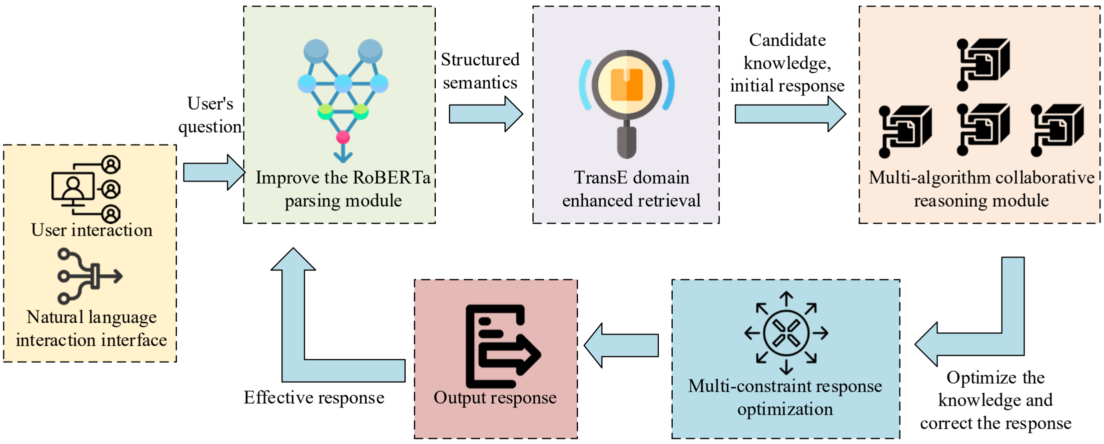

Figure 5. Schematic diagram of EDIQAR framework (Icon source from: https://www.freepik.com/)

In Figure 5, EDIQAR first receives natural language questions via a user interface. The improved RoBERTa module parses the input into structured semantics. Then, the TransE-based retrieval module fetches candidate knowledge and an initial response. The multi-algorithm reasoning module further processes and refines the results. After that, the multi-constraint optimization module polishes the response and outputs the final result. Finally, the validated answer is fed back to the RoBERTa module, enabling iterative optimization and completing the intelligent QA process. This provides accurate and professional QA support for dispatching scenarios. In RoBERTa, the input text first passes through word embedding and position encoding, then enters the Transformer layers. Word embeddings map tokens into a low-dimensional space, while position encodings capture sequential relationships, as shown in Equation (8).

$$X = \mathrm{LayerNorm}\left(E_{word} + E_{pos}\right) \tag{8}$$

In Equation (8), E_{word} is the word embedding matrix, E_{pos} is the position encoding matrix, and LayerNorm represents the layer normalization operation. The TransE model is based on the translation assumption that the head entity plus the relation approximates the tail entity. It learns vector representations of entities and relations in the knowledge graph, as defined in Equation (9).

$$d(h+r,t) = \left\|(h+r)-t\right\|_{L1/L2} \tag{9}$$

In Equation (9), $d(\cdot)$ is the distance function (L1 or L2 norm), used to evaluate the semantic fit of triples. Smaller distances indicate higher plausibility.

4. Performance verification of the intelligent question answering system and reasoning mechanism for power dispatching

4.1 Performance of the RoBERTa-based semantic model improved from BERT

The baseline models were implemented based on their original architectures and fine-tuned using the same power dispatching dataset under identical experimental conditions. BERT was adopted as the representative bidirectional language representation model, LightGPT was selected as a lightweight generative pretrained model, and GAIA was used as a domain-oriented power dispatching language model for comparison. All baseline models and the proposed RoBERTa-based model were trained with the same training/test data partition, optimization strategy, and hardware environment. The main training settings, including learning framework, computing platform, and dataset configuration, were kept consistent to ensure a fair comparison. The dataset was organized according to eight dispatching intent categories, including equipment faults, parameter queries, operational procedures, charging and discharging control, abnormal condition handling, and other related scenarios. The samples were randomly divided into training and test sets with a ratio of 70% and 30%, respectively, while maintaining a similar distribution of different intent categories in both subsets. The annotation process was conducted by three domain experts with practical experience in power dispatching, and disagreements were resolved through discussion and consensus. This procedure ensured the consistency and reliability of the labeled dataset. After manual annotation of entities, intents, and relations, the data were structured into a labeled dataset. The dataset includes some samples of grid-connected energy storage scheduling inquiries for special validation purposes. The models were trained in a PyTorch environment using 70% of the data as the training set and 30% as the test set. Detailed hardware configurations and parameter settings are shown in Table 1.

Table 1. Hardware configuration and parameter setting

Item	Parameter
Dataset	Power dispatching corpus
The number of intents	8
The number of samples	2200
Operating system	Linux Ubuntu 20.04 LTS
Experiment condition	AMD Ryzen 9 7950X, 64GB DDR5, NVIDIA RTX 4090 24GB
Deep learning framework	PyTorch 2.1.0
Random seed	42

To assess how accurately the models recognized domain-specific entities, operational intents, and entity relations in dispatch questions, the study compared the intent classification accuracy and F1 scores of the four models. The results are shown in Figure 6.

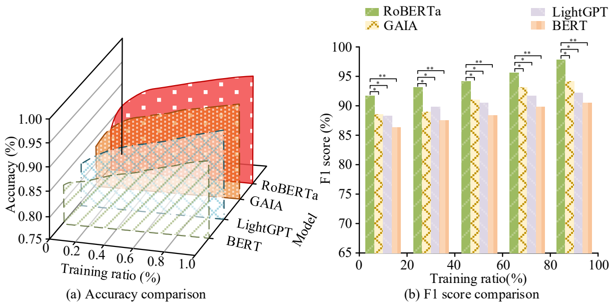

Figure 6. Comparison of Intent classification accuracy and F1 scores

Note: Comparison of semantic understanding performance among different models. Statistical significance was evaluated using paired statistical tests, where * indicates $p < 0.05$ and ** indicates $p < 0.01$ compared with the proposed RoBERTa-based model.

As shown in Figure 6, RoBERTa consistently outperformed GAIA, LightGPT, and BERT in both intent classification accuracy and F1-score under different training proportions. With the increase in training data, RoBERTa maintained high classification accuracy, reaching 0.97 when using 100% of the training data, while its F1-score peaked at 98.1%. Statistical significance analysis was further conducted to evaluate the reliability of the performance improvements. Compared with GAIA, LightGPT, and BERT, RoBERTa achieved statistically significant improvements in both accuracy and F1-score ($p < 0.05$). The results demonstrate that RoBERTa not only provides higher classification performance but also exhibits stronger stability under different training conditions. This advantage is attributed to optimization strategies such as dynamic masking and large-batch training, which enhance semantic representation capability and long-range dependency modeling. Consequently, RoBERTa can better capture professional semantics and complex associations in power dispatch texts, leading to more accurate intent classification. To further evaluate the models' capabilities in handling specific semantic tasks in power dispatch, the study conducted a comparative analysis on two core tasks: dispatch terminology disambiguation and complex sentence structure parsing. The performance results are presented in Figure 7.

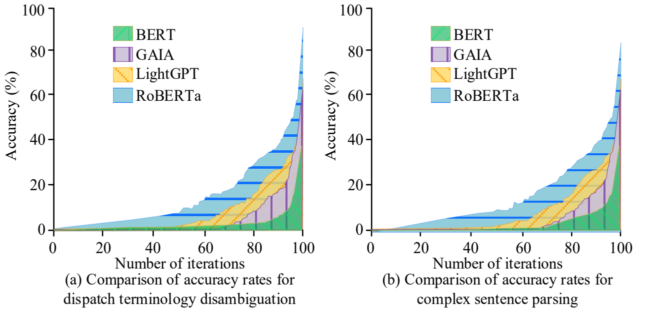

Figure 7. Accuracy of dispatch terminology disambiguation and complex sentence parsing

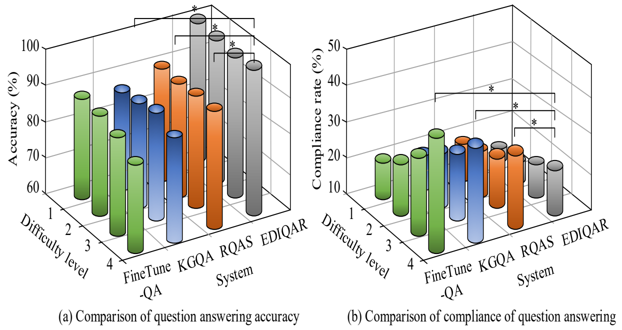

Figure 8. Comparison of answer accuracy and regulation compliance rate

As shown in Figure 7(a), RoBERTa consistently achieved higher accuracy in dispatch terminology disambiguation throughout the training iterations compared to the other models. Its accuracy rose quickly with more iterations, reaching 92.5% after 100 iterations. This demonstrated RoBERTa's superior ability to distinguish between ambiguous dispatch terms such as load shedding and AGC adjustment. With dynamic masking and large-batch strategies, it effectively leveraged the professional semantics embedded in the knowledge graph, reducing interference from general semantics and enabling more accurate interpretation of context-specific meanings. As shown in Figure 7(b), RoBERTa also outperformed other models in parsing complex dispatch sentence structures involving multiple conditional relations. Its accuracy reached 84.8% after 100 iterations. This improvement was attributed to its enhanced ability to model long-distance semantic dependencies. By leveraging entity relations and regulation constraints from the power dispatch knowledge graph, RoBERTa was able to produce more comprehensive and logically coherent parsing results.

4.2 Application performance of the intelligent reasoning fusion system EDIQAR

To validate the effectiveness of the proposed intelligent question-answering and reasoning system EDIQAR in professional QA tasks, the study selected three representative benchmark systems for comparison: the Rule-based Dispatch Question Answering System (RQAS), the Knowledge Graph-based Question Answering System (KGQA), and the Domain-fine-tuned Question Answering System (FineTune-QA). The evaluation focused on core performance metrics to highlight EDIQAR's strengths. First, the study classified power dispatch question answering tasks into four difficulty levels and randomly selected 100 questions from each level. It then compared the answer accuracy and regulation compliance rate of each system on the test set. The results are shown in Figure 8.

Note: * indicates $p < 0.05$ compared with the proposed EDIQAR system.

As shown in Figure 8(a), despite fluctuations, EDIQAR maintained a high level of answer accuracy across all difficulty levels. Its accuracy reached 98.5% at level 1 and remained at 91.3% even at level 4. As shown in Figure 8(b), its violation rate of dispatch regulations was also relatively low, measured at 12.0% for level 1 and 22.5% for level 4. In contrast, the lowest violation rate among the benchmark systems was 30.5%. These results indicated that EDIQAR provided more accurate and regulation-compliant answers, with better performance stability ($p < 0.05$). Its advantage stemmed from the integration of multiple collaborative reasoning mechanisms. Rule verification ensured compliance, the knowledge graph supplemented related knowledge, and logical reasoning corrected procedural logic, all of which enabled precise interpretation of professional semantics and complex operation relations in power dispatch. Despite the complex challenges posed by the energy storage joint regulation system, it can still maintain a relatively high level of compliance response.

Finally, to assess the response efficiency of the four systems, the study compared their memory usage and response time under increasing data volumes. The results are shown in Figure 9.

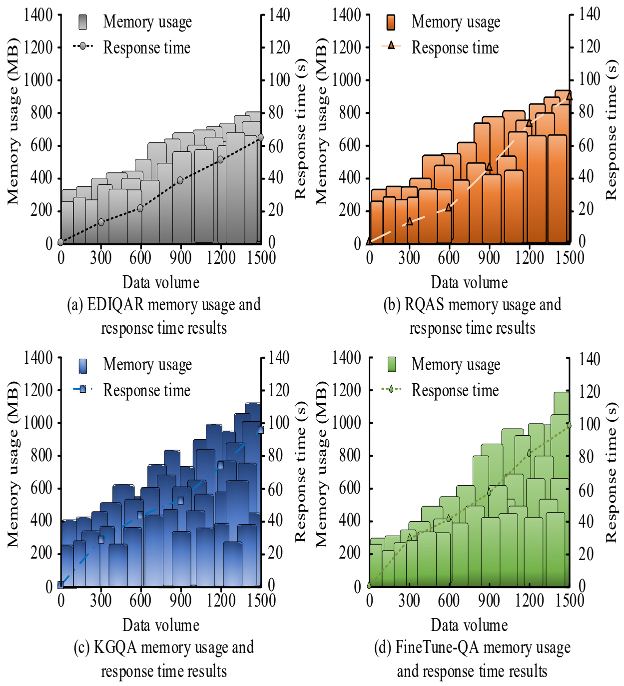

Figure 9. Comparison of memory usage and response time

In Figure 9(a), EDIQAR's memory usage increased slowly with data volume and peaked at 805MB. When processing 1,500 records, its response time was 63.2 seconds. As shown in Figure 9(b), RQAS recorded a peak memory usage of 875.3MB and the longest response time of 91.2 seconds. According to Figures 9(c) and 9(d), the peak memory usage of KGQA and FineTune-QA reached 1,184.6MB and 1,201.3MB, respectively, with maximum response times of 97.5 seconds and 100.3 seconds. In summary, EDIQAR demonstrated lower memory usage and shorter response time compared to the other systems, indicating higher operational efficiency and stability. These advantages were attributed to its lightweight knowledge retrieval and reasoning optimization strategies, including compressed storage structures for the knowledge graph and task scheduling optimization in the multi-algorithm reasoning module, which reduced redundant computations.

5. Conclusion

With the rapid development of grid-connected energy storage systems, intelligent dispatching and operation support for energy storage-grid coordination scenarios has become increasingly important. To address the issues of inaccurate semantic parsing, poor domain adaptability, and low operational efficiency in traditional intelligent question answering for power dispatch, this study put forward an intelligent question answering and reasoning fusion system named EDIQAR, which is based on an improved RoBERTa semantic model. Experimental results showed that RoBERTa enhanced its ability to resolve ambiguous technical terms and parse complex sentence structures by applying dynamic masking and large-batch training strategies, combined with a power dispatch knowledge graph. The model achieved a maximum intent classification accuracy of 0.97 and an F1 score of 98.1%. The accuracy of terminology disambiguation and complex sentence parsing reached 92.5% and 84.8%, respectively. The EDIQAR system integrated multiple collaborative reasoning algorithms. Its question answering accuracy remained as high as 91.3% even at difficulty level 5, while the violation rate of dispatch regulations was as low as 22.5%. In addition, its maximum memory usage was 805MB and the response time was 63.2 seconds, both significantly better than those of the benchmark systems. The results demonstrated that the proposed model and system offered clear advantages in semantic parsing accuracy, answer compliance, and operational efficiency, confirming its practical value in engineering applications. In addition, this system, relying on the expanded energy storage dispatch knowledge base, can efficiently support the implementation of dispatching services for grid connection of energy storage and handling of energy storage faults. However, the system's adaptability to extremely complex fault scenarios still needs improvement. In the future, it can enhance its decision-making ability in generalized scenarios by reinforcing dynamic collaborative reasoning across multiple regulation sets.

Acknowledgments

The research work in this paper was supported by the Research and Development of Scheduling Think Tank and Intelligent Writing System Technology Based on Big Data and Artificial Intelligence [Project No.: ZDKJXM20240064], for which we express our gratitude.

References

[1] Ugboke P, Ebimaro J, Olanrewaju P, et al. The impact of digitization towards technological revolution in the power industry[J]. NIPES-Journal of Energy Technology and Environment, 2024, 6(2): 193-200.

[2] Biancofiore G M, Deldjoo Y, Noia T D, et al. Interactive question answering systems: Literature review[J]. ACM Computing Surveys, 2024, 56(9): 32-38.

[3] Kim K, Park S. Aobert: All-modalities-in-one BERT for multimodal sentiment analysis[J]. Information Fusion, 2023, 92: 37-45.

[4] Laurer M, Van Atteveldt W, Casas A, Welbers K. Less annotating, more classifying: Addressing the data scarcity issue of supervised machine learning with deep transfer learning and BERT-NLI[J]. Political Analysis, 2024, 32(1): 84-100.

[5] Xu J, Guo Y. TransE-KCB: An improved knowledge graph representation method for negative sample sampling[J]. Computer Applications and Software, 2024, 41(8): 345-350.
[6] Jiang T, Shen D, Zhang Z, Liu H, Zhao G, Wang Y, Chen W. Battery technologies for grid-scale energy storage[J]. Nature Reviews Clean Technology, 2025, 1(7): 474-492.
[7] Chang X, Zhao Y M, Yuan B, Fan M, Meng Q, Guo Y G, Wan L J. Solid-state lithium-ion batteries for grid energy storage: opportunities and challenges. Science China Chemistry, 2024, 67(1): 43-66.
[8] Sobhanam H, Prakash J. Analysis of fine tuning the hyper parameters in RoBERTa model using genetic algorithm for text classification[J]. International Journal of Information Technology, 2023, 15(7): 3669-3677.
[9] Özkurt C. Comparative analysis of state-of-the-art Q&A models: BERT, RoBERTa, DistilBERT, and ALBERT on SQuAD v2 dataset[J]. Chaos and Fractals, 2024, 1(1): 19-30.
[10] Cheruku R, Hussain K, Kavati I, et al. Sentiment classification with modified RoBERTa and recurrent neural networks[J]. Multimedia Tools and Applications, 2024, 83(10): 29399-29417.
[11] Madichetty S, M S, Madisetty S. A RoBERTa based model for identifying the multi-modal informative tweets during disaster[J]. Multimedia Tools and Applications, 2023, 82(24): 37615-37633.
[12] Kumar B V, Sadanandam M. A fusion architecture of BERT and RoBERTa for enhanced performance of sentiment analysis of social media platforms[J]. International Journal of Computing and Digital Systems, 2024, 15(1): 51-66.
[13] Yang L, Lv C, Wang X, et al. Collective entity alignment for knowledge fusion of power grid dispatching knowledge graphs[J]. IEEE/CAA Journal of Automatica Sinica, 2022, 9(11): 1990-2004.
[14] Li X, Yang N, Li Z, et al. Confidence estimation transformer for long-term renewable energy forecasting in reinforcement learning-based power grid dispatching[J]. CSEE Journal of Power and Energy Systems, 2022, 10(4): 1502-1513.
[15] Fan S, Guo J, Ma S, et al. Framework and key technologies of human-machine hybrid-augmented intelligence system for large-scale power grid dispatching and control[J]. CSEE Journal of Power and Energy Systems, 2023, 10(1): 5-12.
[16] Ji Z, Wang X, Zhang J, Wu D. Construction and application of knowledge graph for grid dispatch fault handling based on pre-trained model[J]. Global Energy Interconnection, 2023, 6(4): 493-504.
[17] Garrido-Merchan E C, Gozalo-Brizuela R, Gonzalez-Carvajal S. Comparing BERT against traditional machine learning models in text classification[J]. Journal of Computational and Cognitive Engineering, 2023, 2(4): 352-356.
[18] Kim J H, Park S W, Kim J Y, et al. RoBERTa-CoA: RoBERTa-based effective finetuning method using co-attention[J]. IEEE Access, 2023, 11: 120292-120303.
[19] Meng D, Ma B, Shang Z. A knowledge set recommendation method for online education in universities based on DV-TransE model and social networks[J]. International Journal of Networking and Virtual Organisations, 2024, 30(1): 44-56.
[20] Chen L, Zhou Y, Wei Q, et al. ELPGPT: Large language models enhancing link prediction in electrical knowledge graph[J]. International Journal of High Speed Electronics and Systems, 2025, 34(01): 2540115.
[21] Kang M, Lee S, Baek J, et al. Knowledge-augmented reasoning distillation for small language models in knowledge-intensive tasks[C]//Advances in Neural Information Processing Systems. 2023, 36: 48573-48602.
[22] Irham A, Roslan M F, Jern K P, et al. Hydrogen energy storage integrated grid: A bibliometric analysis for sustainable energy production[J]. International Journal of Hydrogen Energy, 2024, 63: 1044-1087.
[23] Zhou N, Xu Y, Cho S, Wee C T. A systematic review for switchgear asset management in power grids: Condition monitoring, health assessment, and maintenance strategy[J]. IEEE Transactions on Power Delivery, 2023, 38(5): 3296-3311.