Energy Engineering

Tech Science Press

Doi:10.32604/ee.2025.073567

REVIEW

Applications of Large Language Model in HVDC Systems: Concepts, Development, and Perspectives

Xing Wen[1], Huan Chen[1], Ning Wang[1,*], Yu Song[1], Zhuqiao Qiao[2] and Bin Zhang[1]

[1]China Southern Power Grid Extra High Voltage Power Transmission Company, Guangzhou, 510663, China
[2]China Southern Power Grid Extra High Voltage Power Transmission Company Kunming Bureau, Kunming, 650217, China

*Corresponding Author: Ning Wang. Email: wangning6782025@outlook.com

Received: 21 September 2025; Accepted: 16 December 2025; Published: 12 July 2026

ABSTRACT: High voltage direct current (HVDC) systems play a pivotal role in long-distance, high-capacity, and cross-regional power transmission. However, their complex structure, wide-ranging impact of faults, and stringent safety requirements pose significant challenges to operational stability. Conventional model-based and data-driven methods for tasks such as text classification, fault diagnosis, and operation and maintenance support suffer from limited scalability and interpretability. Recent advances in large language model (LLM) provide new opportunities to address these issues. This paper provides a systematic review of LLM applications in HVDC systems. Firstly, it introduces the core architecture and training mechanisms of LLM. Subsequently, a comprehensive evaluation framework is established for LLM's application in HVDC systems across four dimensions: accuracy, generalization capability, interpretability, and complexity. Secondly, this paper examines the applications of LLM in three key domains: text classification, fault diagnosis, and question answering. It further extends the discussion to additional application areas while evaluating the performance of various methodological approaches. Specifically, in text classification, LLM enhance the semantic interpretation of operation and maintenance reports. In fault diagnosis, LLM achieve accurate fault identification through multi-source data fusion and reasoning. In question answering applications, LLM provide intelligent decision support through retrieval-augmented and multimodal frameworks. Finally, from the perspectives of real-time performance, interpretability, and applicability, this paper presents three key conclusions and discusses two critical aspects. It further outlines three forward-looking research directions focusing on multimodal data fusion and real-time decision-making. The goal is to provide an up-to-date and comprehensive reference for researchers exploring the integration of LLM into HVDC system analysis, operation, and management.

KEYWORDS: Large language model; high voltage direct current; text classification; fault diagnosis; question answering

1 Introduction

The emergence of large language model (LLM) marks a revolutionary advancement in the field of artificial intelligence [1]. Relying on massive datasets and largescale neural network architectures, LLM demonstrate exceptional capabilities in language understanding and text generation, including reasoning, few shot learning, and the integration of extensive world knowledge from pre-trained models [2–4]. These advances have accelerated progress in diverse application domains, such as natural language processing, machine translation, creative content generation, and chatbot development, but has also triggered profound reflections within the industry on the transformative power of LLM and their potential to drive advancements in artificial intelligence. In particular, the emergence of models such as ChatGPT has further intensified attention and discussion in this field.

High voltage direct current (HVDC) systems, as a cornerstone of modern power transmission [5–7], play a vital role in enabling long-distance power delivery and interconnection between asynchronous grids, as illustrated in Fig. 1 [8,9]. However, their intrinsic complexity, coupled with diverse and fast-evolving fault mechanisms, places stringent demands on accurate diagnosis and intelligent operation [10–12]. In the actual operation and maintenance of HVDC, the following three main challenges are encountered [13]. First, HVDC systems generate massive amounts of unstructured data, such as inspection logs, maintenance reports, and fault analyses, yet lack effective tools for semantic understanding and information extraction, leading to underutilization of valuable diagnostic knowledge. Second, fault mechanisms in HVDC systems are highly coupled, with subtle precursors and rapid propagation, making manual or rule-based diagnosis difficult to scale and prone to delay or misjudgment. Third, operation still depends heavily on a small number of domain experts whose tacit knowledge is difficult to formalize and inherit, resulting in fragmented expertise and inconsistent decision-making. In text-classification tasks, traditional techniques represent documents with term frequency-inverse document frequency and then apply machine-learning algorithms, yielding limited feature expressiveness. By contrast, LLM can directly model and classify large-scale operation logs and reports, providing richer inputs for subsequent monitoring and decision-making. Traditional fault diagnosis approaches for HVDC systems, encompassing model-driven, data-driven, and knowledge-driven methods, each method exhibits their own limitations and challenges [14]. Mathematical model- or physics model-based fault diagnosis approaches are prone to the "curse of dimensionality" when dealing with complex systems [15]. For instance, the parity equation model-based approach proposed in [16], while avoiding a large number of parameter adjustments, did not significantly improve the "curse of dimensionality." Data-driven methods mostly rely on simulation models, and their applicability to real-world engineering remains to be verified [17]. Reference [18] introduced the Sparse Counterfactual Generation Network (SCF-Net), which was tested in two real-world case studies within the oil and gas industry and yielded promising results. However, it has yet to be implemented in actual engineering applications. Knowledge-based approaches, due to long fault causality chains, generally cannot determine the root cause through a single inspection. For example, the fault diagnosis method based on knowledge-enhanced domain adversarial learning proposed in reference [19], while demonstrating strong adaptability to various operating conditions and unknown fault types, still has limitations in terms of computational complexity and training efficiency. However, LLM can integrate multiple diagnostic paradigms and process intricate fault data to deliver accurate diagnostic results. Moreover, as society advances, power-sector users exhibit growing demand for specialized information, yet conventional retrieval entails considerable difficulty. To overcome this bottleneck, a question answering system is urgently needed to furnish operators and engineers with real-time technical support, thereby enhancing operational efficiency and response speed. Therefore, investigating the applications in these three domains can demonstrate the potential of LLM in enhancing the diagnostic capabilities, text processing, and information retrieval of HVDC systems, which is of significant importance for improving the overall performance and reliability of the systems.

In current research on the application of LLM, various scholars and research teams have conducted in-depth explorations and summaries from multiple dimensions. To clearly illustrate the literature screening process, Fig. 2a provides a detailed flowchart. Statistical data on research in this field over the past decade were collected from platforms such as Web of Science, Engineering Index, and Wiley Online Library, and Fig. 2b presents the publication trends.

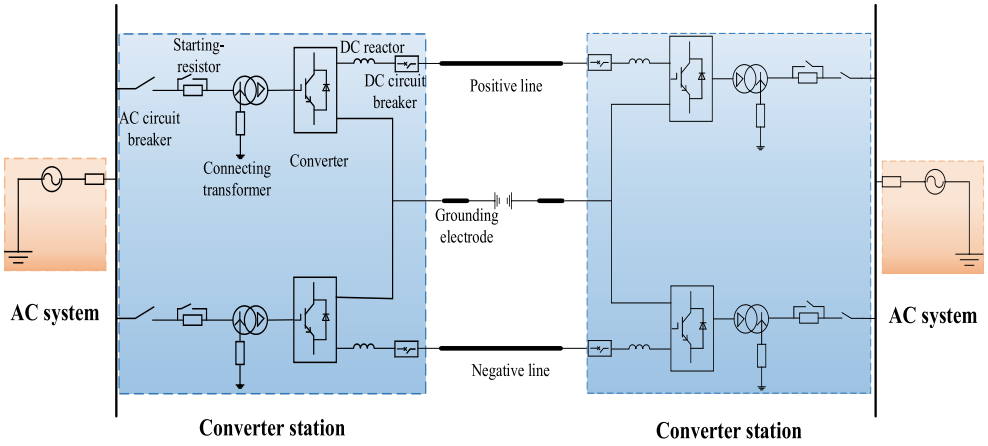

Figure 1: Fundamental architecture of HVDC systems

Figure 2: Review methodology of related reference. (**a**) Execution procedure. (**b**) Research statistics

Notably, although these studies focus on different aspects, they all demonstrate a profound awareness of the potential risks and challenges inherent in the application of LLM, as well as a shared pursuit of overcoming these obstacles through technological innovation. Reference [20] thoroughly investigated data privacy in LLM, covering leakage, attacks, and protection technologies across inference stages, yet omitted considerations for user interaction and the scalability of protections. Literature [21] reviewed LLM in metaheuristic optimization, proposing a role-based taxonomy, but its limited literature scope undermined the comprehensiveness and generalizability of its findings. Reference [22] provided a comprehensive review of advanced LLM fine-tuning techniques and their impacts but may lack coverage of the latest advancements and specific implementation guidance. Although literature [23] explored the integration of LLM with traditional metaheuristics, it lacked detailed implementation steps, limiting its practical application value. Study [24] systematically reviewed the application of Transformers and LLM in the energy sector but noted that concrete evidence for foundation models in this domain remains limited. Existing studies collectively reveal the immense potential and broad prospects of LLM in cross-domain applications. To address the gaps in existing research and highlight the contributions of this paper, Table 1 systematically compares an overview of this work with that of five key literature sources, alongside their main shortcomings. This effort establishes a clear reference framework for subsequent research on large language models for HVDC systems.

Table 1: Overview and analysis of shortcomings in recent survey research on key technical areas of LLM

Focused area	Contribution summary	Limitation
LLM and proxy data privacy	Analyze privacy threats arising from LLM and their use of proxy data.	Lacks specific guidance on the practical implementation.
The role of LLMs in meta-inspired optimization [20]	Evaluate the effectiveness of existing privacy protection mechanisms within LLM.	
Advanced fine-tuning techniques for LLMs Generative AI and LLMs in network security [21]	Analyze the functional roles of LLM within meta-inspired optimization processes.	Incomplete research sample and geographical representation.
	Propose a novel role-based classification method, categorizing LLM as Consultant, Enhancer, and Innovator.	
Transformer models and LLMs in energy	Provide a comprehensive review and analysis of advanced fine-tuning techniques for LLM.	The comparison of the experimental data summarized in the review lacks rigor.
Management LLM and proxy data privacy	Introduce a structured classification method, organizing methods into six distinct categories.	
The role of LLMs in meta-inspired optimization [22]	Offer insights and guidance through experimental evaluation across various tasks.	
Advanced fine-tuning techniques for LLMs	Explore the applications of Generative AI and LLMs in the field of network security.	It fails to adequately analyze the energy consumption, inference latency, and hardware requirements of large-scale LLMs in real-time cybersecurity scenarios.

(Continued)

Table 1 (continued)

Focused area	Contribution summary	Limitation
Generative AI and LLMs in network security [23]	Evaluate the performance of 42 LLM models in terms of their network security knowledge and hardware security understanding.	
Transformer models and LLMs in energy management [24]	Review the application of Transformer Models and LLMs in energy forecasting, efficiency optimization, and smart grid management.	Limited in concrete evidence of foundation models' application in the energy domain.
	Introduce the concept of digital twins for proxy systems, discussing their role in enhancing situational awareness and autonomy.	
Applications of LLM in HVDC systems (This Work)	Delve into the comprehensive application capabilities of LLM in text classification, fault diagnosis, and question answering systems.	Limited to the three main application directions of LLMs in HVDC systems, without detailed consideration of other more specific uses.
	Outline potential future research directions for LLM in HVDC systems, including multimodal data fusion, real time fault prediction, adaptive operation and maintenance strategies.	

However, despite the promising potential demonstrated by LLM across multiple domains, its roles, challenges, and potential benefits within HVDC systems remain insufficiently defined. To bridge this research gap, this paper aims to address the following key questions:

(1) What are the potential application scenarios and key technical challenges of LLM into HVDC systems, particularly for text classification, fault diagnosis, and question answering?
(2) What are the potential application scenarios and key technical challenges of integrating LLM into HVDC systems, particularly for text classification, fault diagnosis, and question answering?
(3) What future directions should be pursued to enable multimodal data fusion, adaptive reasoning, and real-time decision-making in LLM-driven HVDC systems?

To address these questions, this paper systematically reviews the emerging literature on LLM in HVDC systems, with particular focus on three representative areas: text classification, fault diagnosis and question answering. The main contributions are as follows:

(1) Systematic literature review of LLM research in HVDC system. This paper, for the first time, systematically presents the first structured review of existing studies, providing a comprehensive reference framework for subsequent research.
(2) Comprehensive application capability evaluation and optimization. This paper delves into the comprehensive application capabilities of LLM in text classification, fault diagnosis, and question answering systems.
(3) Proposal of future research directions. Based on the current research status, this paper outlines potential future research directions for LLM in HVDC systems, including multimodal data fusion, real time fault prediction, adaptive operation and maintenance strategies, etc., offering valuable references for subsequent research.

The remainder of this paper is organized as follows: Section 2 introduces the LLM architecture in detail, laying the foundation for subsequent discussions; Section 3 presents a comprehensive evaluation framework for LLM applications in HVDC systems; Section 4 discusses the applications of LLM in HVDC systems targeting text classification, fault diagnosis, and question answering systems; Section 5 summarizes the main findings of this paper and Section 6 proposes future application and research directions for LLM in HVDC systems.

2 Large Language Model

2.1 Principles of Large Language Model

LLM was initially applied in the field of natural language processing (NLP), serving two primary functions: (1) evaluating whether the outputs of language generation tasks conform to human linguistic patterns, thereby aiding natural language understanding; (2) performing statistical analysis on large-scale corpora to provide linguistic knowledge for language models. Early language models were mainly categorized into statistical language model (SLM) and neural network language model (NNLM), with current mainstream LLM evolving from the latter.

2.1.1 Statistical Language Model

SLM originated from research in natural language processing systems based on statistical methods. It is used to represent the probability distribution of basic linguistic units and describe the statistical rules underlying language generation. Typically, SLM decompose the probability of a sentence into the product of the conditional probabilities of individual words [25]. The probability of a sentence is expressed as:

$$P(S) = P(\omega_1, \omega_2, \cdots, \omega_n) = \prod_{i=1}^{n} P\left(\omega_i \mid \omega_1^{i-1}\right) \tag{1}$$

where S denotes; n is the length of the sentence; and ω_i represents the i-th word in S.

Due to the problem of excessive parameter size in the original model, the N-gram model based on the Markov assumption was proposed. The N-gram model assumes that each predicted word depends only on the preceding $N-1$ words as its context, and is represented as:

$$P(S) = \prod_{i=1}^{n} P\left(\omega_i \mid \omega_{i-N+1}^{i-1}\right) \tag{2}$$

where N is the order of the model, which determines both its accuracy and complexity.

The main focus of the N-gram model is to achieve data smoothing, and the language it models differs significantly from natural language. Limited by data availability and scale, statistical language models struggle to integrate massive data into coherent linguistic knowledge.

2.1.2 Neural Network Language Model

With the development of deep neural network learning techniques, the use of neural networks for natural language processing has developed rapidly. Bengio published a pioneering study that used neural networks to estimate the probability distribution of N-gram model, thereby proposing the NNLM [26]. The objective function optimized by NNLM is defined as:

$$F = \sum_{D} P\left(\omega_i \mid \omega_{i-(n-1)}, \omega_{i-(n-2)}, \cdots, \omega_{i-1}\right) \tag{3}$$

where D denotes the collection of text sequences derived from S, where each element in D is a text sequence of length n.

Due to the fixed-length context limitation of NNLM, its ability to capture long-range dependencies in historical text is limited. To address this, Mikolov proposed training language models using recurrent neural network (RNN) with directed loops. However, RNN suffer from the long-term dependency problem: when processing complex linguistic contexts, it tends to rely primarily on the most recent inputs and struggle to retain earlier information. To overcome this limitation, long short-term memory (LSTM) introduces a gating mechanism that regulates the retention and forgetting of information. Experimental results have shown that LSTM achieve better performance as the size of the training dataset increases.

To enhance the expressive capability of NNLM, embedding and SoftMax layers are incorporated at both ends of the network. The embedding layer digitizes natural language symbols by converting words into dense vector representations. Popular models for training embedding include Word2Vec and GloVe. The SoftMax layer transforms the network's output into a probability distribution over the vocabulary.

2.1.3 Large Language Model

LLM is a language training model based on deep neural networks. It learns complex linguistic patterns and semantic representations through unsupervised pretraining on massive volumes of unlabeled text data. By leveraging hundreds of billions of parameters and extremely large-scale training datasets, LLM exhibits high scalability, enabling strong contextual understanding and effective long-text processing capabilities [27]. LLM is further enhanced through key techniques such as instruction tuning and alignment, ultimately achieving multi-domain language capabilities including text generation, logical reasoning, and common-sense dialogue. At present, LLM has been widely applied in power systems and have achieved remarkable results [28].

Principle of Transformer

The transformer, a self-attention-based feature extractor proposed by Google, has become the foundational neural network architecture for most LLM. The transformer architecture consists primarily of encoder and decoder. The encoder integrates features from the input sequence through a stack of identical layers, transforming the input text into high-dimensional representations that facilitate hierarchical feature extraction. The decoder incorporates masked multi-head attention (MHA) to model dependencies between the input sequence and the partially generated output, thereby enabling accurate sequence generation [29]. The architecture of the Transformer is illustrated in Fig. 3.

The transformer primarily consists of two sublayers: MHA and a feed-forward network (FFN). Both MHA and FFN are equipped with residual connections and layer normalization to enhance the model's trainability, as illustrated below.

$$\begin{cases} z_l^{'} = \mathrm{MHA}\left(\mathrm{LN}\left(z_{l-1}\right)\right) + z_{l-1} \\ z_l = \mathrm{FFN}\left(\mathrm{LN}\left(z_l^{'}\right)\right) + z_l^{'} \end{cases} \tag{4}$$

where z_{l-1} and $z_l^{'}$ denote the input features to the MHA and FFN at layer l, respectively, while z_l represents the output features of layer l.

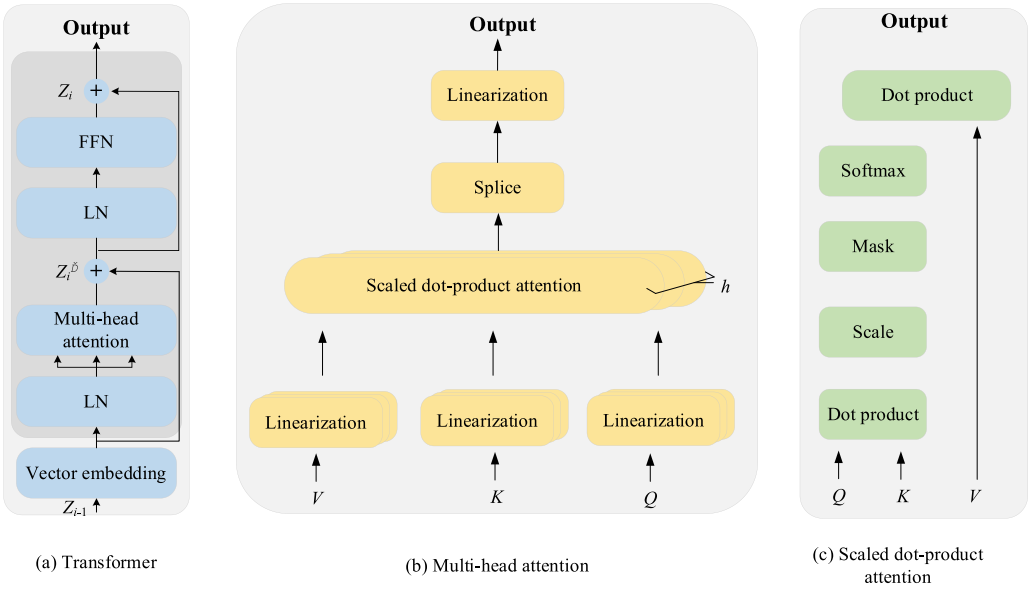

Figure 3: Transformer framework schematic diagram

MHA divides the input sequence into multiple sub-sequences and applies the attention mechanism to each sub-sequence independently. Since the attention weight parameters of each head H_h are learned independently, MHA can capture diverse features in parallel and concatenate them for integration, thereby enhancing the large-scale parallel computing capabilities of LLM.

$$\begin{cases} \text{MHA}\left(Q, K, V\right) = \text{Concat}\left(H_1, H_i, \cdots, H_h\right) W^O \\ H_i = \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right) i = 1, 2, \cdots, h \end{cases} \tag{5}$$

where W^O, W_i^Q, W_i^K and W_i^V are linear projection matrices, and h denotes the number of parallel attention heads.

The input to the self-attention mechanism consists of the query matrix, the key matrix, and the value matrix. The output attention weights are computed using the SoftMax function. The self-attention mechanism is formally defined as follows:

$$\text{Attention}\left(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}\right) = softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{V} \tag{6}$$

where d denotes the dimensionality of the $\boldsymbol{Q}$ and $\boldsymbol{K}$.

The Transformer consists of the encoder and the decoder. The encoder encodes the input sequence, while the decoder decodes the encoded information to generate the output. Different LLM adopt varying transformer architectures: models such as GPT, ERNIE Bot, and DeepSeek use the decoder-only structure; BERT and RoBERT employ the encoder-only structure; and T5 adopts the encoder-decoder structure. Currently, the most widely used underlying architecture in LLM is the decoder, as it is well suited for text generation tasks.

Training Process of Large Language Model

The training process of LLM is typically divided into two stages: pretraining and tuning. The pretraining stage is a critical phase in the development of LLM capabilities. It involves learning from massive volumes of unlabeled data to capture linguistic patterns, structures, and features. The amount of data used during

pretraining often reaches the scale of hundreds of billions of tokens, encompassing sources such as texts, web pages, and books, with language modeling as the training objective. Given a text sequence $S = \{s_1, s_2, \cdots, s_n\}$, the model predicts the target token s_i autoregressively based on the preceding tokens $s_1, s_2, \cdots, s_{i-1}$ [30]. The objective function $L(s)$ is defined as a maximum likelihood function, as follows.

$$L(s) = \sum_{i=1}^{n} \lg P(s_i \mid s_1, s_2, \cdots, s_{i-1}) \tag{7}$$

where n denotes the number of text samples, and $P(\cdot)$ represents the likelihood function.

After pretraining, the model must be fine-tuned to adapt to specific downstream tasks. The fine-tuning phase primarily transfers the model from general language knowledge to domain-specific tasks, thereby improving its performance in applications such as document classification and domain-specific question answering. Currently, reinforcement learning from human feedback (RLHF) is widely used in the fine-tuning stage of LLM. RLHF typically consists of three main stages: supervised fine-tuning (SFT), reward modeling (RM), and proximal policy optimization (PPO).

During the SFT stage, the pretrained model is fine-tuned using human-annotated data to learn basic task execution capabilities. Given a training sample pair (x_i, y_i) the objective of SFT is to minimize the cross-entropy loss function:

$$L_{\mathrm{SFT}} = -\sum_{i=1}^{N} \log \pi_\theta(y_i \mid x_i) \tag{8}$$

where x_i denotes the input text, y_i represents the human-annotated target output, $\pi_\theta(y_i \mid x_i)$ is the probability assigned by the LLM with parameters θ to the reference response y_i, and N is the number of training samples.

In the RM stage, the model primarily learns the human-preferred ranking of different responses. Specifically, for a given input text x_i, humans may provide multiple candidate responses. The reward model $r_\varnothing(x, y)$ takes an input-output pair as input and produces a scalar score as output. The objective of RM is to maximize the probability of assigning higher scores to human-preferred responses.

$$L_{\mathrm{RM}} = -\sum_{i=1}^{N} \log \sigma\left(r_\phi\left(x_i, y_i^A\right) - r_\phi\left(x_i, y_i^B\right)\right) \tag{9}$$

where $\sigma(\cdot)$ denotes the sigmoid function, and $r_\phi(x_i, y_i)$ is the score assigned to the output y by the reward model with paraments $\varnothing$.

PPO uses the reward scores from the RM as feedback to guide the LLM in generating responses that better align with human preferences. During the PPO stage, the LLM is treated as a policy π_θ that generates the response y_i given the input x_i, and receives feedback based on the reward function $r_\phi(x_i, y_i)$. The objective function is defined as follows:

$$L_{\mathrm{PPO}}(\theta) = \mathbb{E}_{(x,y)\sim\pi_\theta}\left[\min\left(\rho_t A_t,\ \mathrm{clip}\left(\rho_t, 1-\epsilon, 1+\epsilon\right) A_t\right)\right] \tag{10}$$

where ρ_t denotes the ratio between the new and old policies, A_t is the advantage function, which evaluates how favorable the current output is compared to the average, ϵ controls the magnitude of policy updates, and $\mathrm{clip}(\cdot)$ restricts the ratio within the range $[1-\epsilon, 1+\epsilon]$.

2.2 Evolution and Applications of Large Language Model

2.2.1 Evolution of Large Language Models

LLM can be traced back to early sequence-based language models. Based on technological iterations, their evolution can be categorized into three stages: the emergence of LLM, the advancement of large-scale language models, and the evolution toward general-purpose LLM.

Emergence of LLM: A milestone in the rise of LLM was the introduction of the Transformer architecture in 2017. By eliminating the limitations of sequential computation in traditional models and leveraging GPU-based parallelism, the Transformer significantly improved training efficiency and enabled model scale-up. This shift led to the mainstream adoption of pretraining on large-scale unlabeled data, giving rise to two major architectures: BERT and GPT. BERT excels in language understanding through masked word prediction, while GPT is specialized in text generation using autoregressive modeling. Both models achieved breakthroughs in parameter scale and generalization ability, redefining NLP benchmarks and laying the foundation for modern LLM.

Advancement of large-scale models: The release of GPT-3 in 2020 marked the beginning of the trillion-parameter era. This leap in parameter scale led to qualitative improvements in model capabilities, particularly in few-shot and zero-shot learning. For example, GPT-3 demonstrated a strong correlation between parameter count and generalization performance. With 175 billion parameters, it could complete various generation tasks using only a few or even no examples. This breakthrough reduced reliance on massive labeled datasets and spurred a global race toward developing even larger models, which have since shown remarkable performance in knowledge understanding and contextual learning.

Evolution toward general-purpose LLM: Since 2022, LLM research has focused on alignment with human intent and controllability. Models such as ChatGPT and GPT-4 integrated RLHF, significantly improving coherence, controllability, and alignment with human preferences. Technologies such as instruction tuning, multitask training, and multimodal fusion have advanced LLM toward personalization. Meanwhile, open-source LLM like DeepSeek and ChatGPT, along with efficient fine-tuning techniques, are accelerating industrial and enterprise-level deployment [31].

2.2.2 Applications of LLM in HVDC

As digitalization and intelligent transformation advance in modern power systems, the practical applications of LLM in the power sector are beginning to show promising results.

Major energy enterprises such as State Grid Corporation of China and China Southern Power Grid have entered the LLM development arena, aiming to accelerate the digital transformation of power grids. As a critical component of power transmission, HVDC significantly influence the stability, efficiency, and economic performance of the grid. LLM have shown great potential in HVDC applications, including fault detection, line diagnostics, knowledge graph construction for operation and maintenance, and intelligent human-machine interaction [32].

DaWatt, the first large-scale LLM platform tailored for power systems, has introduced various HVDC-specific models. For instance, ultra-high voltage subsidiaries of China Southern Power Grid developed DC operation and safety models by integrating edge-cloud-terminal architectures and video recognition technology, enabling real-time staff behavior identification and warnings—boosting operational efficiency by over 125%. Additionally, DaWatt has driven the construction of digital production and safety control platforms for HVDC, facilitating digital twins, intelligent inspections, fault alarms, and maintenance decision support [33].

Guangming Power LLM developed by State Grid features trillion-parameter scale and multimodal processing capabilities. It supports cross-domain and multi-business operations and has been applied in HVDC maintenance scenarios. Specifically, it enables intelligent health assessment, risk path identification, and maintenance suggestion generation based on multimodal analysis of operation and maintenance texts, work orders, and image data from key HVDC equipment like converter and substation facilities. Furthermore, the model can generate natural language explanations and emergency handling strategies in complex scenarios such as power flow reversal or short-term system fluctuations, reducing the cognitive load on dispatchers [34].

In addition, LLM developed by companies such as SenseTime, Baidu Wenxin, and Huawei Pangu support a wide range of HVDC-related tasks, including fault diagnosis, grid inspection, and professional document analysis and response. A representative timeline of LLM development is shown in Fig. 4.

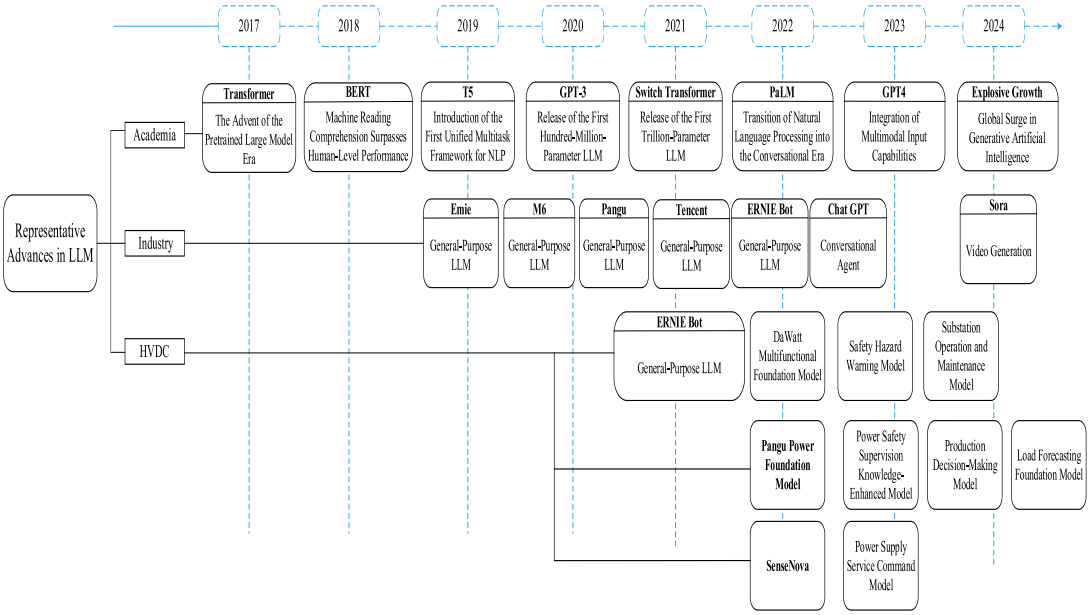

Figure 4: Development of representative large language models

3 Comprehensive Evaluation Framework for LLM Applications in HVDC Systems

To systematically assess and compare the comprehensive performance of LLM across different application tasks in HVDC systems, this section proposes a unified quantitative evaluation framework. This framework aims to move beyond single accuracy metrics and, based on the practical requirements for deployment in industrial systems, provides multi-dimensional decision-making basis for method selection and optimization.

3.1 Evaluation Dimensions and Quantitative Metrics

To comprehensively and objectively evaluate the application performance of LLM in HVDC systems, this section constructs a multi-dimensional comprehensive evaluation framework. This framework abandons the limitation of focusing solely on a single performance metric, instead considering four core dimensions: accuracy, generalization capability, interpretability, and complexity. Each dimension is defined with quantifiable scoring functions, whose output values are normalized to the interval $[0, 1]$, where a higher value indicates better performance.

3.1.1 Accuracy

Accuracy is the benchmark performance metric that measures the model's ability to complete its core task. Given that the success criteria differ across tasks, this framework employs task-specific core evaluation metrics for quantification. Its scoring function is defined as follows:

$$S_{acc} = M_{core} \tag{11}$$

where, M_{core} represents the recognized core evaluation metric for a specific task: Text classification tasks use the Macro-F1 score to properly handle scenarios with imbalanced class distributions; Fault diagnosis tasks use diagnostic accuracy, i.e., the proportion of correctly diagnosed fault cases to the total cases; Question answering systems use the normalized value of the human-evaluated accuracy score, rated by domain experts based on the correctness of the answers.

3.1.2 Generalization Capability

Generalization capability reflects the model's robustness when facing data distribution shifts, noise interference, and unknown scenarios, and is key to assessing the model's applicability in real-world industrial environments. This dimension is measured by the relative performance degradation of the model on a benchmark test set vs. a challenging test set:

$$S_{\text{gen}} = \max\left(0, 1 - \frac{M_{\text{core}}^{\text{clean}} - M_{\text{core}}^{\text{challenge}}}{M_{\text{core}}^{\text{clean}}}\right) \tag{12}$$

where, $M_{\text{core}}^{\text{clean}}$ is the model's core metric score on a clean, in-distribution test set, and $M_{\text{core}}^{\text{challenge}}$ is the model's score on a challenging test set containing noise, missing data, or out-of-distribution samples. A value closer to 1 indicates less performance degradation and stronger generalization capability.

3.1.3 Interpretability

Interpretability measures the extent to which the model's decision-making process can be understood and trusted by human experts. This dimension is crucial for safety-critical systems like HVDC. This framework assigns a graded score based on the intrinsic characteristics of the model's methodology:

$$S_{\text{int}} = \begin{cases} 1.0, & \text{Method based on rules or symbolic knowledge} \\ 0.8, & \text{Method providing explicit attention weights traceable to input features} \\ 0.5, & \text{Method relying on post-hoc explanation techniques} \\ 0.2, & \text{End-to-end black-box model} \end{cases} \tag{13}$$

This scoring is predefined based on the inherent potential and convenience of different methodologies in providing intuitive and reliable explanations.

3.1.4 Complexity

Complexity assesses the model's demand for computational resources and time costs during deployment and runtime, directly impacting its feasibility in real-time or resource-constrained HVDC industrial systems.

This framework quantifies this dimension using relative efficiency compared to a baseline model, with its scoring function defined as follows:

$$S_{\text{eff}} = \min\left(1, \frac{1}{2}\left(\alpha \cdot \frac{N_{\text{base}}}{N_{\text{model}}} + (1-\alpha) \cdot \frac{T_{\text{base}}}{T_{\text{model}}}\right)\right) \tag{14}$$

where, N_{model} and N_{base} represent the number of trainable parameters of the model under evaluation and the baseline model, respectively; T_{model} and T_{base} represent the average inference time required to process a single sample for the model under evaluation and the baseline model, respectively, under identical hardware configurations; α is a weighting coefficient used to balance the importance between model size (parameter count) and inference speed, with a value range of [0, 1]. In the absence of specific preference, $\alpha = 0.5$ can be set to indicate equal importance.

The core of this formula is to calculate the comprehensive efficiency improvement of the model under evaluation relative to the baseline model. $\frac{N_{\text{base}}}{N_{\text{model}}} > 1$ indicates a lighter model, and $\frac{T_{\text{base}}}{T_{\text{model}}} > 1$ indicates a faster model. Taking the minimum of the weighted sum of the two metrics and 1 aims to obtain a robust, normalized efficiency score. A higher S_{eff} value indicates better performance in terms of complexity and efficiency.

3.2 Weighted Comprehensive Scoring Formula

As different tasks and application scenarios have varying priorities, this paper defines a comprehensive scoring formula with adjustable weights:

$$S = w_1 \cdot S_{\text{acc}} + w_2 \cdot S_{\text{gen}} + w_3 \cdot S_{\text{int}} + w_4 \cdot S_{\text{eff}} \tag{15}$$

where $w_1 + w_2 + w_3 + w_4 = 1$. The weight assignment reflects the application's priority requirements. To provide a more intuitive comparison of the performance of different methods, this paper classifies the comprehensive scores into distinct tiers: * for $S \in [0, 0.6]$, ** for $S \in (0.6, 0.7]$, *** for $S \in (0.7, 0.8]$, **** for $S \in (0.8, 0.9]$, and ***** for $S \in (0.9, 1]$.

3.3 Recommended Weight Configurations for Different Tasks

To provide more specific guidance, this paper recommends different weight configurations for the three tasks in HVDC systems in Table 2, reflecting their distinct requirements in engineering practice.

Table 2: Recommended evaluation weight configurations for different LLM application tasks

Application task	w_1	w_2	w_3	w_4	Rationale and scenarios for weight configuration
Text classification	0.4	0.3	0.2	0.1	Scenario: Real-time alarm filtering and command classification. Rationale: Accuracy is critical, as false positives/negatives pose direct risks. Strong generalization capability is required to handle diverse alarm texts. Moderate interpretability is necessary for operator judgment. Complexity requirements are relatively lenient since these are typically non-extreme real-time scenarios.

(Continued)

Table 2 (continued)

Application task	w_1	w_2	w_3	w_4	Rationale and scenarios for weight configuration
Fault diagnosis	0.5	0.2	0.2	0.1	Scenario: Complex fault root cause analysis. Rationale: Accuracy is the highest priority, as incorrect diagnoses can lead to severe consequences. Interpretability is equally important as accuracy, as experts must trust and understand the diagnostic basis. Generalization capability and complexity are important but can be partially sacrificed for higher accuracy and credibility.
Question answering	0.3	0.2	0.3	0.2	Scenario: Operator knowledge query and decision support. Rationale: Interpretability carries the highest weight, as the system must provide answer sources or reasoning processes (e.g., via RAG) to be trusted. Accuracy is fundamental. Complexity has a relatively high weight due to the need for rapid responses to interactive queries.

4 Applications and Technical Discussions of LLM in HVDC Systems

4.1 Types of Large Language Model Training Datasets for HVDC Systems

The training of LLM relies on diverse and multimodal datasets to support tasks such as text classification, fault diagnosis, and question answering. Based on data structure, source, and task characteristics, the datasets can be categorized into the following five types:

(1) Unstructured text datasets: These datasets contain complex and irregular semantic structures, including domain terminology, abbreviations, and mixed expressions, such as operation logs, maintenance records, inspection reports, fault analysis documents, and dispatch instructions. Such data are primarily used to train LLM for semantic understanding and text classification, enabling them to extract equipment conditions, fault information, and operational knowledge from large volumes of unstructured text.

(2) Semi-structured text datasets: These datasets originate mainly from technical specifications, manufacturer manuals, protection device event messages, and operational work orders, featuring partial formatting and hierarchical structure. They enhance the ability of LLM to extract knowledge from semi-structured content and serve as essential training sources for question answering and knowledge graph construction.

(3) Structured datasets: These include equipment parameter tables, operational indicators, monitoring sequences, protection device signals, and online monitoring data. Structured data can be enriched through textual descriptions or state labels, supporting causal inference, event interpretation, and state recognition in LLM-based fault diagnosis tasks.
(4) Multi-source heterogeneous and cross-modal datasets: These include real-time waveforms, missing trigger-pulse logs, and image inspection data. Through textual transformation, feature representation, or RAG-based retrieval enhancement, these data sources can serve as auxiliary corpora for LLM reasoning chains or knowledge graph inference, providing contextual reinforcement and cross-validation for multimodal fault diagnosis.
(5) Retrieval-augmented and knowledge graph datasets: These include equipment topology, fault-causality chains, and operation and maintenance knowledge. These datasets are not suitable for direct pretraining but serve as augmented resources for SFT, RAG, and KGQA stages. They enhance the question-answering capabilities of LLM while improving logical consistency and interpretability.

4.2 Applications of Large Language Model in Text Classification

4.2.1 Text Classification Tasks

Text classification refers to the task of assigning textual data to predefined categories. Common applications include sentiment detection, news sorting, and thematic identification. Recent studies demonstrate that numerous natural language understanding (NLU) problems can be reformulated as text-classification tasks by enabling deep-learning models to process text pairs as input. This section outlines five representative tasks: three conventional text-classification problems and two NLU tasks frequently approached through text classification in modern deep-learning research [35].

(1) Sentiment analysis: This task determines the polarity or attitude expressed in text. It can be framed as a binary decision—positive vs. negative—or as a multi-class prediction involving fine-grained labels or graded intensity levels [36].
(2) News categorization: News articles remain a vital information source. Automated news classifiers assist users by detecting emerging topics or recommending content aligned with personal interests, thus providing timely, relevant information.
(3) Topic analysis: Also called topic classification, this task identifies the primary subject of a document, such as discerning whether a product review focuses on 'customer support' or 'ease of use'.
(4) Question answering: Two major forms exist: extractive and generative. Extractive question answering can be treated as a classification problem, where the system judges candidate text spans as correct answers to a question (e.g., in SQuAD datasets). Generative question answering, in contrast, requires producing answers directly and is considered a text-generation task. This paper addresses only the extractive type [37].
(5) Natural language inference (NLI): Also known as textual entailment, NLI predicts whether the meaning of one statement follows from another, assigning labels such as entailment, contradiction, or neutral. A related variant is paraphrase detection, which evaluates whether two sentences convey the same semantic content [38].

4.2.2 Methods for Text Classification

Feed-Forward Neural Networks

The feedforward neural network (FFNN) is one of the earliest DL models applied to document classification and is particularly suitable for scenarios involving relatively simple semantic structures or

coarse-grained textual data. FFNN treats the input text as a bag of words and maps individual words into vector representations $\{x_1, x_2, \ldots, x_n\} \in \boldsymbol{R}^d$ using embedding models such as Word2Vec or Glove. Then, the model performs an averaging operation over all word vectors to obtain a dense vector representation v of the texts [39].

$$v = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{16}$$

The text vector is fed into the hidden layer of the FFNN to perform nonlinear feature transformation. Taking a single hidden layer as an example, the transformation can be expressed as:

$$h = \sigma\left(W_1 v + b_1\right) \tag{17}$$

where $W_1 \in \boldsymbol{R}^{m\times d}$ is the weight matrix of the hidden layer, $b_1 \in \boldsymbol{R}^m$ is the bias vector, $\sigma(\cdot)$ denotes the activation function and h represents the hidden feature representation.

h after nonlinear transformation, is passed through the output layer for category prediction, where the Softmax function is applied to generate a probability distribution:

$$\hat{y} = softmax\left(W_2 h + b_2\right) \tag{18}$$

where $W_2 \in \boldsymbol{R}^{C\times m}$ is the output weight matrix, $b_2 \in \boldsymbol{R}^C$ is the bias vector, C denotes the number of classes, and $\hat{y} \in \boldsymbol{R}^C$ represents the predicted probability for each class.

During training, the model optimizes its parameters by minimizing the cross-entropy loss between the predicted probability distribution and the true labels. The loss function is defined as follows:

$$L = -\sum_{j=1}^{C} y_j \log\left(\hat{y}_j\right) \tag{19}$$

Long Short Term Memory

The LSTM is an improved version of the RNN that incorporates a gating mechanism. It is widely used for text classification tasks involving strong contextual dependencies. The input text is first represented as a sequence of word embeddings $\{x_1, x_2, \ldots, x_n\} \in \boldsymbol{R}^d$, and the LSTM processes this sequence while retaining long-range dependencies. At each time step t, the LSTM maintains the current word vector x_t, the previous hidden state h_{t-1}, and the previous cell state c_{t-1}, and updates its internal states according to the following equations [40].

$$\begin{cases} f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \\ i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \\ \tilde{c}_t = \tanh\left(W_c \cdot \left[h_{t-1}, x_t\right] + b_c\right) \\ c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\ o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \\ h_t = o_t \odot \tanh\left(c_t\right) \end{cases} \tag{20}$$

where the forget gate f_t determines how much information from the previous cell state is retained, input gate i_t and the candidate memory cell $\tilde{c}_t$ control the extent to which the current input is used for updating the memory, the output gate o_t regulates the information output from the current hidden state h_t, $\odot$ represents element-wise multiplication, W and b denote the weight and bias parameters of each gate, respectively.

After processing the input text sequence, the LSTM can either use the hidden state at the final time step h_t, or apply a pooling operation over all hidden states to generate a fixed-length text representation vector h. This vector h is then fed into a fully connected layer followed by the Softmax activation function for classification.

Convolutional Neural Network

Convolutional neural network (CNN) extracts local semantic patterns from text using a local receptive field mechanism and obtains the most salient global semantic representation through max-pooling operations. This makes them particularly suitable for short text classification tasks such as sentiment analysis and topic identification. For a given text input, pre-trained embedding methods such as Word2Vec or GloVe are used to convert words into vectors, resulting in a matrix representation of the entire text [41]. A convolution operation is then applied using sliding windows over the embedding matrix to extract local features:

$$c_i = f\left(\langle \boldsymbol{W}, \boldsymbol{X}_{i:i+h-1} \rangle + b\right) \tag{21}$$

where $\boldsymbol{W} \in \boldsymbol{R}^{h \times d}$ denotes the convolutional kernel weights, h is the window size, b is the bias term, and $f(\cdot)$ typically represents the ReLU activation function. By sliding the convolutional window across the text sequence, a one-dimensional feature map $\boldsymbol{c} = [c_1, c_2, c_{n-h+1}]$ is generated, which captures the significance of local structures at different positions within the text.

To better aggregate features and mitigate the impact of long text inputs, max-pooling is typically applied to each feature map to extract the most salient semantic features, as follows:

$$\hat{c} = \max_i \left(c_i\right) \tag{22}$$

Since the presence of key semantic units is often more informative than their exact positions, max-pooling preserves the strongest activation in each feature channel. The outputs from multi-scale convolutional kernels are pooled and concatenated into a global feature vector $\boldsymbol{Z}$, enabling the integration of diverse linguistic patterns. Finally, the feature vector is projected into the category space through a fully connected layer.

Robustly Optimized BERT Pretraining Approach

Robustly optimized BERT pretraining approach (RoBERTa) is a language representation model proposed by Facebook AI in 2019 as an improved version of BERT. While retaining the original BERT architecture, RoBERTa enhances pretraining performance by removing the next sentence prediction task, increasing batch size and training data, and employing a dynamic masking strategy. In document classification tasks, RoBERTa is commonly used as a text encoder to convert entire documents into context-aware vector representations, which are then used for category prediction.

The input document is tokenized into a sequence of tokens, with special symbols [CLS] (for classification tasks) and [SEP] (for sentence separation) added accordingly:

$$\text{Input} = [[\text{CLS}], t_1, t_2, \cdots, t_n, [\text{SEP}]] \tag{23}$$

RoBERTa employs a standard multi-layer Transformer encoder, which is primarily composed of multi-head self-attention mechanisms and position-wise feedforward networks. The process is formally expressed as follows.

For each token at position i, the Query, Key, and Value vectors are computed as follows:

$$\begin{aligned} \boldsymbol{Q} &= \boldsymbol{X}W^{Q} \\ \boldsymbol{K} &= \boldsymbol{X}W^{K} \\ \boldsymbol{V} &= \boldsymbol{X}W^{V} \end{aligned} \tag{24}$$

The bullish form is:

$$\text{MultiHead}\left(\boldsymbol{X}\right) = \left[\text{head}_1; \ldots; \text{head}_h\right] W^{O} \tag{25}$$

Each layer passes through a feedforward network:

$$\text{FFN}\left(x\right) = \text{GELU}\left(xW_1 + b_1\right) W_2 + b_2 \tag{26}$$

The hidden state vector corresponding to the [CLS] token is ultimately used as the semantic representation of the entire document:

$$\boldsymbol{z} = \boldsymbol{h}_{[\text{CLS}]} \tag{27}$$

Table 3 presents a comparative analysis of the advantages, limitations, complexity, and applicability of four commonly used methods for text classification in HVDC systems.

Table 3: Comparison of advantages and limitations of various existing HVDC text classification methods

Methods of text classification	Advantages of HVDC applications	Limitations of HVDC applications	Complexity	Applicability	Score
FFNN [38]	Simple architecture, suitable for high-real-time alert filtering or basic text classification Low computational load and minimal hardware requirements	Lacks temporal and contextual understanding Making complex fault causality or cross-paragraph semantics hard to detect Low accuracy when processing control commands or equipment status records	Low complexity: simple text feature extraction and few parameters	Suitable for simple text archiving and single fault alarm classification scenarios	**
LSTM [39]	Strong long-sequence dependency modeling, enabling cross-temporal event chain analysis Strong contextual understanding, suitable for processing maintenance logs and control commands	Slow inference speed, unsuitable for online tasks requiring second-level response Many training parameters, with moderate scalability for long texts and large datasets	High complexity: requiring fine-tuning and long training time	Suitable for fault evolution analysis scenarios High computational demand with complex engineering tuning	***

(Continued)

Table 3 (continued)

Methods of text classification	Advantages of HVDC applications	Limitations of HVDC applications	Complexity	Applicability	Score
CNN [40]	Excels at extracting local key information, effectively identifying short-text warnings and event trigger words High computational parallelism and fast inference, suitable for real-time warning	Insufficient long-sequence dependency modeling, unable to handle cross-document causal relationships Requiring multi-scale convolution kernels to capture HVDC semantic patterns	Moderate complexity: training and inference are easily parallelized.	Suitable for short-sentence alarm classification and rapid topic identification applications Moderate hardware requirements, enabling rapid migration and deployment	****
RoBERTa [42]	Strong contextual understanding, capable of capturing global semantics and cross-domain relationships in complex fault chains Suitable for processing multimodal text and integrating with knowledge graph technologies	High computational and storage costs Fine-tuning requires extensive labeled data, and building HVDC-specific corpora is costly	High complexity: requiring GPU and distributed training Relatively long inference latency	Suitable for complex fault diagnosis, cross-system operation and maintenance question answering, and intelligent scheduling decision support	****

Note: **denotes $S \in (0.6, 0.7]$, ***denotes $S \in (0.7, 0.8]$, ****denotes $S \in (0.8, 0.9]$, where S is the normalized composite score derived from the respective evaluation criteria; the detailed scoring protocol is provided in Section 3, "Comprehensive Evaluation Framework for LLM Applications in HVDC Systems".

4.2.3 Applications of Large Language Model to HVDC Text Classification

Inverter Attack Detection Model Based on Volt/VAR Command Text Classification

Intelligent inverters are widely deployed in HVDC systems, where they regulate system voltage through Volt/VAR curves. However, these Volt/VAR curves are vulnerable to stealthy cyber-attacks, which may cause voltage violations and other security concerns. Traditional attack detection methods rely excessively on numerical data, which provide insufficient contextual information for effectively identifying cyber threats. Transforming numerical Volt/VAR control commands into textual representations can mask numerical details, improve data interpretability, and enhance system security. This approach makes full use of the accumulated data in HVDC systems, enabling intuitive detection of abnormal conditions while strengthening the system's resilience against cyber-attacks. Reference [43] proposes converting variations in control command values into text sequences and employing LLM for classification to detect potential cyber-attacks.

The detection architecture illustrated in Fig. 5 consists of three interrelated stages. First, numerical commands from the Volt/VAR system are transformed into text sequences. The model is designed to process these sequences, which may contain attack vectors, and the classifier is trained to identify any suspicious command. Second, the classified commands are validated to ensure that they are free from potential malicious attacks. Finally, a LLM based on BERT is employed to capture dependencies among different parts of a command sequence, while DistilBERT is adopted to meet the low-latency requirements of real-time detection across various attack scenarios. Once a potential attack is detected, operators are alerted and the

corresponding command is blocked. Otherwise, the command is converted back into its numerical form and executed by the inverter.

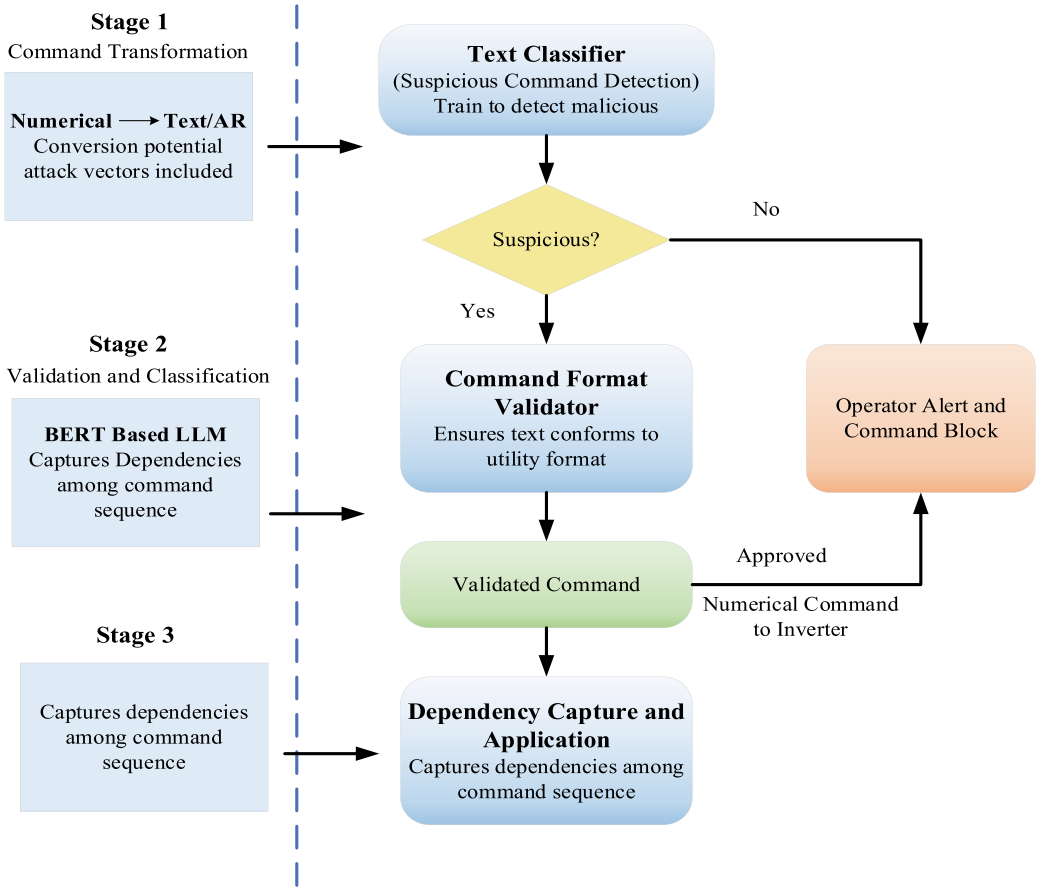

Figure 5: Inverter attack detection model based on volt/var command text classification

BERT-Based Dual-Channel Power Equipment Defect Text Assessment Model

The massive volume of textual records generated during the routine operation and maintenance of HVDC equipment serves as a critical basis for evaluating equipment health and formulating maintenance strategies. However, due to the highly specialized nature and irregular expression of such texts, risk-level assessments of equipment records still heavily rely on manual interpretation, which suffers from low efficiency and strong subjectivity. Work [44] proposes a dual-channel text feature extraction model based on BERT, enabling the coordinated capture of both global semantics and local key information within the text.

First, a Chinese dataset of power equipment defect texts was constructed, and data augmentation techniques were employed to effectively expand the dataset and mitigate the problem of class imbalance. Subsequently, the pre-trained BERT model was used to deeply capture semantic representations, providing the Bi-LSTM channel with high-quality, semantically rich features. The Bi-LSTM was then employed to model the long-term contextual dependencies of the overall information sequence, converting them into more sequence-aware features. Meanwhile, a CNN was applied to extract features related to key defect phrases within the text, which are crucial for determining defect severity levels. Finally, the feature vectors extracted by the Bi-LSTM and CNN channels were concatenated and fed into a fully connected layer for fusion, completing the final classification.

Time-series classification models such as GRU, TCN, and lightweight Transformer architectures are generally more compact and computationally efficient. The GRU model achieves higher efficiency but may lose fine-grained contextual precision when handling long fault reports. The TCN model offers excellent inference speed and parallelism, making it suitable for real-time waveform-based fault detection; however, it has limited capability in perceiving complex semantic dependencies within unstructured textual data. Lightweight Transformers possess strong global representational power but require additional domain adaptation before being effectively applied to HVDC systems. In contrast, work [36] adopts a Bi-LSTM within a dual-channel framework to capture the inherent semantic dependencies in HVDC fault texts and preserve semantic coherence, while hyperparameter optimization further reduces computational complexity. Future research on real-time HVDC fault monitoring should explore hybrid architectures that integrate deep contextual semantic understanding with lightweight and efficient sequential models.

For non-English manuals, LLM exhibit strong scalability and can be effectively applied to real HVDC systems that involve multilingual texts and real-time data. Owing to its multilingual pretraining capabilities, LLM can be fine-tuned and adapted to the HVDC domain at a low additional training cost, enabling efficient processing of non-English technical documentation. Furthermore, LLM can be integrated with online data-processing pipelines, where operational logs, equipment data, and other real-time information are preprocessed, aggregated, and subsequently transmitted to the reasoning layer. By leveraging lightweight retrieval modules or tokenization mechanisms, LLM is capable of functioning efficiently under near-real-time conditions. In addition, their robustness can be further enhanced through few-shot learning, transfer learning, and data augmentation. Moreover, by employing retrieval augmented generation (RAG), LLM can incorporate external documents to enrich their training corpus, thereby improving their comprehension of non-English contexts and strengthening domain-specific semantic understanding in HVDC applications.

4.2.4 Semantic Relation Extraction Process of LLM

In the semantic understanding framework of LLM, the extraction of semantic relations relies on the integration of multi-head contextual attention and probabilistic generative modeling to ensure semantic consistency within the vector space. For a given text sequence S, Eq. (6) is used to assign weights to dependencies across arbitrary spans. This mechanism enables the model to assign semantic dependency weights globally across all positions, thereby capturing long-range semantic relations, such as the implicit logic embedded in the equipment-fault-handling semantic chain.

Subsequently, to map word-level semantics onto the entity and relational levels, LLM typically employ joint semantic modeling. Given a text containing an entity set ε and a latent relation set R, the joint objective is formulated as the minimization of a composite loss function:

$$\mathcal{L} = \mathcal{L}_{\text{ent}} + \mathcal{L}_{\text{rel}} + \lambda \mathcal{L}_{\text{cons}} \tag{28}$$

where, $\mathcal{L}_{\text{ent}}$ and $\mathcal{L}_{\text{rel}}$ represent the cross-entropy terms for entity recognition and relation classification, respectively, while $\mathcal{L}_{\text{cons}}$ denotes a consistency constraint term designed to ensure that the triplets extracted from unstructured text comply with semantic dependency constraints.

To further enhance the explicit representation of cross-sentence semantic relations, graph-enhanced semantic modeling can be introduced at the structural level. In this approach, the extracted entities and relations are organized into a semantic graph, where a graph attention propagation layer enables the structured aggregation of semantic contexts. Furthermore, this framework enables the capture of semantic dependencies and causal pathways across multiple scales, within sentences, between sentences, and across documents, thereby providing structured semantic support for subsequent reasoning and question-answering tasks.

In addition, when handling long texts or multi-document scenarios, LLM typically incorporate retrieval-augmented generation (RAG) or chain-of-thought mechanisms. This generation-reasoning architecture substantially enhances the model's ability to capture causal and logical relations, and its joint probabilistic modeling can be expressed as follows:

$$p(y|x) = \sum_{s} p(y|x,s)\, p(s|x) \tag{29}$$

where, s denotes the intermediate reasoning steps or the set of retrieved documents.

4.3 Applications of Large Language Model in Fault Diagnosis

4.3.1 Classification of HVDC Fault Diagnosis

AC-Side Faults

AC-side faults refer to short-circuit or grounding events occurring on the converter station's AC bus, AC filter bus, or incoming line sections, and commonly include single-phase-to-ground, phase-to-phase, three-phase faults, and composite faults accompanied by line tripping [45,46]. At the instant of fault, the AC voltage magnitude drops sharply, its phase shifts abruptly, and the negative-sequence component surges, resulting in severe distortion of the converter-bus voltage waveform. Meanwhile, the DC power plunges rapidly because the firing angle is forced to advance; if the fault is not cleared in time, successive commutation failures are easily triggered and eventually lead to DC blocking. In addition, the fault disrupts the reactive-power balance between sending and receiving ends, the converter transformer suffers DC bias caused by distorted excitation current, and the accumulated effect shortens core life [47]. The AC system frequency also declines due to the power deficit, causing under-frequency load-shedding relays to operate and, in multi-infeed regions, may set off cascading events that enlarge the disturbance.

DC-Side Faults

DC-side faults mainly appear on the pole conductors, neutral bus, electrode lines, and DC filter branches, and take the forms of ground flashover, line breakage, pole-to-pole short circuits, or high-resistance grounding [48]. Lightning strikes, pollution flashover, tree contact, or insulator ageing can all degrade line insulation, producing a sudden rise in DC current and a steep collapse of DC voltage [49]. A broken conductor introduces resonant over-voltages on the DC side, which may damage smoothing reactors and DC voltage dividers. The fault-current magnitude decays exponentially with fault distance and grounding resistance [50]; high-resistance faults far from the station are weak and easily missed by conventional relays. If protection fails to detect and block within a few milliseconds, sustained over-current will puncture thyristor stacks, saturate smoothing-reactor cores, and rupture DC-filter capacitor banks [51], ultimately resulting in severe primary-equipment destruction.

Converter Valve Faults

Converter-valve faults cover valve short-circuit, false firing, and misfiring. Valve short-circuits often originate from thyristor chip defects, damping-circuit open circuits, or trigger-unit failures [52], generating surge currents within microseconds and causing valve-component temperatures to jump by more than one hundred degrees, leading to cracked ceramic insulation and coolant leakage. False or missed firing, caused by distorted trigger pulses, ageing opto-converters, or valve-control board faults [53], manifests itself as distorted valve-voltage waveforms, drifting firing angles, and amplified harmonics; long-term operation degrades valve life and may escalate into system-wide commutation failures when the AC grid is disturbed [54]. Because valve current is not directly measured in practice, diagnostics must integrate indirect

valve-voltage and current measurements, trigger-pulse feedback, and valve-control self-check signals using multi-source data fusion techniques for early warning [55,56].

Commutation Failure Faults

Commutation failure is a control-type fault unique to inverter stations, whose essence is that the valve fails to regain blocking capability during the reverse-voltage period, so the valve that should turn off continues to conduct. AC-voltage sag, loss of trigger pulses, insufficient extinction angle, or valve-control delays can all initiate the failure. Once it occurs, DC current surges, the AC side experiences significant transient over-voltage, and fundamental-frequency components intrude into the DC circuit, causing an instantaneous drop in DC power [57]. Repeated commutation failures force the DC system into emergency blocking, creating large power deficits and frequency swings. Moreover, commutation failures amplify DC-side harmonics, excite resonances between DC filters and lines, and further deteriorate system stability. In multi-infeed receiving networks, a single commutation failure can, through voltage coupling, induce successive lockouts of adjacent DC links, leading to massive power transfers and voltage collapse [58].

4.3.2 Existing HVDC Fault Diagnosis Methods

Model-Driven Fault Diagnosis Methods

Existing methods for power line fault diagnosis primarily include approaches based on analytical models, expert systems, data-driven techniques, and multi-technology integration [59–61]. The earliest fault diagnosis methods relied on physical or mathematical models, achieving fault detection and localization through theoretical analysis of power grids combined with traditional mathematical tools such as equivalent circuits, traveling wave theory, and wavelet analysis [62,63]. In study [64], a method was proposed that utilizes a high-voltage DC power supply and insulated gate bipolar transistors (IGBTs) to periodically charge transformers while measuring oscillatory wave responses via winding bushings to assess transformer condition. This approach integrates physical and mathematical models, enabling fault diagnosis by analyzing changes in the transformer's equivalent circuit parameters. Given the structural and operational similarities between high-voltage DC transformers and AC transformers, this method can theoretically be extended to fault detection in HVDC transformers, particularly for insulation and winding condition assessment. Building on transformer condition monitoring, reference [65] shifted focus to line-level protection by introducing a novel fault detection and localization method based on phase let transform and continuous wavelet transform (CWT). This approach effectively suppresses undesired fluctuations in HVDC line currents, reducing fault localization error to below 1% while achieving 98.5% accuracy. The method's versatility is demonstrated through its applicability to both monopolar and bipolar HVDC systems, covering various fault types including ground and line faults. Complementing these system-level approaches, work [66] proposed a single-ended fault diagnosis method based on traveling wave theory and redundant discrete wavelet transform (RDWT). This solution enables comprehensive fault detection, classification, and localization using only single-ended voltage signals from HVDC lines, offering a cost-effective alternative to conventional double-ended measurement systems while maintaining high diagnostic accuracy.

Although model-based methods offer clear principles and strong interpretability, their low diagnostic efficiency and high computational demands make them unsuitable for modern power systems with massive data volumes and real-time requirements. With the increasing prevalence of power electronic devices and monitoring data, there is an urgent need for more efficient and adaptive intelligent diagnostic technologies.

Data-Driven Fault Diagnosis Methods

Data-driven methods for power line fault diagnosis have become a current research hotspot [67–70]. These methods primarily employ machine learning algorithms to automatically learn fault characteristics

from actual operational data or simulation data, thereby establishing diagnostic models [59]. Compared to traditional model-based approaches, data-driven methods exhibit better fault tolerance and adaptability [71], effectively addressing nonlinear problems and uncertainties in power systems. However, these methods also face inherent challenges such as poor model interpretability and the need for large amounts of training data [72]. In terms of specific research, literature [73] proposed a hybrid diagnostic framework that integrates power system domain knowledge with machine learning. This framework constructs 21 novel physical features based on sequence components, phase angle dynamics, and power characteristics, and combines recursive feature elimination techniques to screen key variables, enabling the XGBoost model to achieve high-precision classification of 11 types of faults in three-phase transmission lines. Building on feature engineering research, literature [74] further proposed an optimization method based on an adaptive neuro-fuzzy inference system. This approach innovatively combines signal processing techniques with intelligent optimization algorithms, performing feature extraction through principal component analysis and discrete wavelet transform, while using intelligent optimization algorithms to optimize membership function parameters, significantly improving the accuracy of fault diagnosis in HVDC transmission systems. Following the developmental trajectory of intelligent diagnostic technologies, literature [75] pioneered a novel method for converting signal spectrum information into feature images. This solution achieves visual representation of signal features through fast Fourier transform and Gramian angular field techniques, providing more effective input features for deep learning models and enabling rapid and accurate fault identification.

Despite significant progress in data-driven methods for power line fault diagnosis, several pressing issues remain to be addressed. First, the lack of model interpretability limits their application in critical power facilities. Second, the dependence on high-quality training data restricts the generalization ability of models in practical engineering. Additionally, the adaptability of existing methods to complex grid topologies and multi-fault coupling scenarios requires further validation. Future research needs to focus on resolving these bottlenecks to promote the practical application of data-driven methods in power system fault diagnosis.

Knowledge-Driven Fault Diagnosis Methods

Early knowledge-driven methods primarily relied on expert systems [76], which consist of two main components: a knowledge base and an inference engine. The knowledge base includes a rule base and a database, where the rule base stores judgment rules for power line faults, and the database store fault characteristic information. The inference engine utilizes knowledge from the knowledge base to analyze newly acquired fault signals, identifying faults by logically matching fault information with the expert knowledge base. In specific research, reference [77] proposed an innovative approach that combines information theory with expert systems, transforming traditional rule-based matching expert systems into numerical calculation methods. This approach not only eliminates the influence of signal timing but also quantitatively describes uncertainties in information transmission, effectively addressing issues of information ambiguity and communication uncertainty. While expert systems can simulate expert decision-making processes and handle complex professional problems, their construction and maintenance require significant expert involvement, and the systems themselves lack flexibility.

In recent years, graph model-based methods have become a major research direction in fault diagnosis. These methods leverage structured graphical representations and logical reasoning capabilities to effectively address fault localization and analysis in complex systems [78]. However, accurately modeling certain complex processes remains challenging [79]. The introduction of knowledge graph technology has opened new possibilities for fault diagnosis, significantly improving engineering knowledge utilization through pattern recognition and correlation analysis of entity and relationship data. Existing knowledge graph-based fault diagnosis methods primarily employ techniques such as knowledge reasoning, querying, question-answering mechanisms, and recommendation algorithms.

Nevertheless, current knowledge graph-based fault diagnosis methods still have limitations. In scenarios involving complex large-scale industrial equipment, due to the extensive scope of investigation and lengthy fault causality chains, it is often difficult to determine the root cause through a single investigation. This highlights the urgent need for more flexible and targeted dynamic fault investigation and diagnosis methods. This demand has driven the transition of fault diagnosis methods from static knowledge matching to dynamic intelligent reasoning, pointing the way for the development of next-generation intelligent diagnosis systems.

Table 4 presents a visual comparison of the advantages and limitations of various existing HVDC fault diagnosis methods, based on the summary and analysis above. The score for each type of method is the average of the scores from the corresponding references.

Table 4: Comparison of advantages and limitations of various existing HVDC fault diagnosis methods

Methods of fault diagnosis	Advantages of HVDC applications	Limitations of HVDC applications	Complexity	Applicability	Score
Model-Driven [64–66]	Clear principles Strong interpretability	Low diagnostic efficiency High computational requirements Unsuitable for modern power systems	High complexity: requiring strong electrical theory knowledge and system modeling expertise High requirements for model debugging and computational resources	Suitable for critical equipment condition assessment and structural stability of traditional HVDC lines	****
Data-Driven [73–75]	High fault tolerance Strong adaptability Effectively handles nonlinear issues and uncertainties	Poor model interpretability Requiring a large amount of training data	Moderate complexity: requiring feature engineering, data cleaning, and model training	Suitable for real-time online fault detection, large-scale monitoring data analysis, and complex multi-fault coupling scenarios	****
Knowledge-Driven [77–79]	Simulates expert decision-making Processes Handles complex professional issues	Requires significant expert involvement in construction and maintenance The system lacks flexibility	High complexity: requiring multidisciplinary experts for knowledge modeling and ongoing maintenance, along with support for graph construction and reasoning platforms	Suitable for cross-site fault chain analysis, intelligent operation and maintenance question answering, and complex causal inference	*****

Note: ****denotes $S \in (0.8, 0.9]$, and *****denotes $S \in (0.9, 1]$, where S is the normalized composite score derived from the respective evaluation criteria; the detailed scoring protocol is provided in Section 3, "Comprehensive Evaluation Framework for LLM Applications in HVDC Systems".

4.3.3 Hybrid LLM-Enhanced HVDC Fault Diagnosis Method Technical Implementation Path

Focusing on HVDC transmission systems, this paper presents a concept for an LLM-based HVDC fault-diagnosis method. With the large language model as its core, the approach handles both training and deployment of the HVDC fault-diagnosis model, enabling a closed-loop upgrade of the entire HVDC fault-diagnosis workflow from manual rules to semantic-driven automation. Owing to its ability to unlock the latent potential of metadata, the knowledge graph has increasingly captured the attention of practitioners in the fault-diagnosis field; therefore, the following adopts the knowledge graph as the HVDC fault-diagnosis approach and explores the feasibility of an LLM-optimized HVDC fault-diagnosis method. Fig. 6 illustrates a side-by-side comparison of the technical routes for traditional knowledge-graph construction vs. LLM-based knowledge-graph construction.

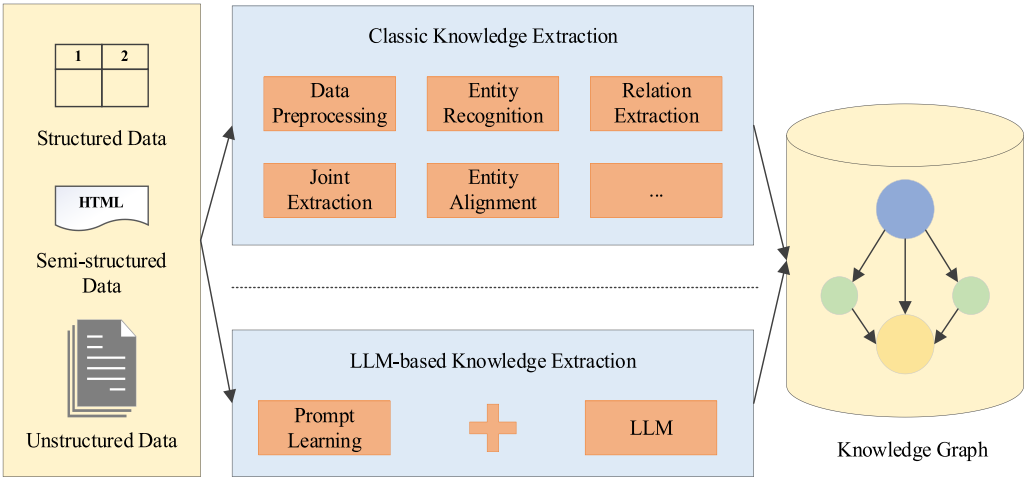

Figure 6: Comparison of the technical routes for classic knowledge-graph construction vs. LLM-based knowledge-graph construction

Prompt Tailored for Constructing an HVDC Fault Knowledge Graph

The extraction-oriented prompt for knowledge-graph construction comprises four sequentially progressive modules. First, the background and historical input module furnishes the language model with a coherent domain context: commencing with the topological structure of the HVDC system, it subsequently incorporates the fault-classification taxonomy and typical protection configurations, and ultimately links to pre-compiled entity, relation, and attribute constraints. Through this continuous information chain, the model operates within a unified context and the probability of overlooking domain-specific terminology is markedly reduced. Next, the source text description module continues from the preceding context and orderly presents the domain metadata-such as fault-record narratives, dispatch logs, or maintenance reports-allowing the model to interface directly with the raw material without additional contextual supplementation. Thereafter, the task description module refines the operational directives on the basis of the source text: it begins by articulating the overarching requirement of enumerating core concepts, proceeds to the more granular tasks of identifying entities or relations, and concludes with the specific instruction to generate class definitions, thereby establishing a progressive command sequence from macroscopic to microscopic levels. Finally, the structured output specification module, anchored to the three prior components, mandates the required output format-such as JSON or RDF-ensuring seamless linkage with preceding modules and

consistency for subsequent parsing, storage, and reasoning stages. Fig. 7 illustrates the structure of the aforementioned knowledge-extraction prompt.

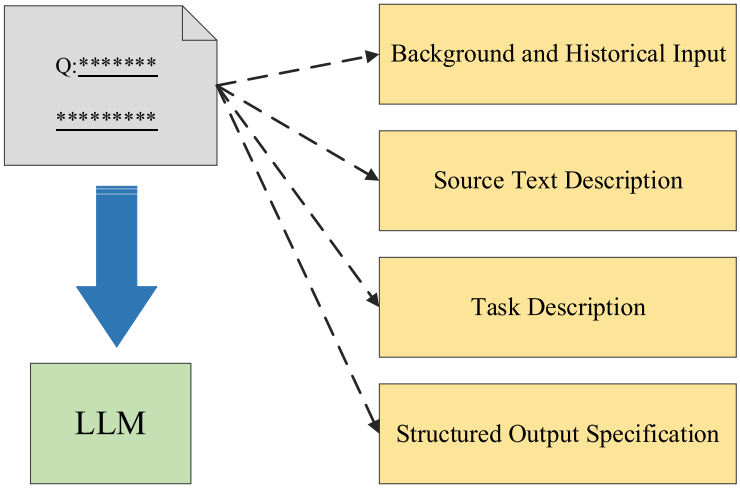

Figure 7: The structure of knowledge-extraction prompt (*denotes the textual input, i.e., the posed question)

Within the HVDC fault-operation-and-maintenance knowledge-graph pipeline, the domain-concept prompt, as the initial step, naturally inherits the aforementioned four-part structure. Specifically, the background information first demarcates the fault-operation scenario and lays the contextual foundation for subsequent steps. The task description then, within this contextual framework, requires the model to systematically generate the core concepts, thereby achieving a smooth transition from scenario to concept. Subsequently, the structured-output specification reiterates the task description by obliging the model to define concepts and their attributes at the class level according to the prescribed syntax. Consequently, the three stages interlock to provide a unified and coherent semantic framework for the instantiation of entities and relations.

The data-extraction prompt likewise adheres to the four-stage progressive logic. The background-and-historical-input segment first reuses the fault-operation RDF Schema generated by the domain-concept prompt, thus seamlessly inheriting the prior context. The source-text segment then continues from this schema by explicitly introducing the fault cases or industrial-equipment data tables to be processed, effecting a natural transition from template to instances. Following this, the task-description segment refines the extraction requirements on the basis of the source text and further specifies detailed constraints. Ultimately, the structured-output specification connects to the task description by requiring the model to render each identified entity, entity-attribute pair, and triple in RDF syntax and to output a complete RDF graph, thereby forming a full-chain closed loop from input to output.

To minimize the occurrence of "hallucinated facts" in LLM outputs while ensuring result accuracy and interpretability, a prompt engineering template based on LLM has been designed. The template is structured as follows:

$$prompt = identity + description + examples + input \tag{30}$$

where, *identity* is used to define the role the model should assume in the given context; *description* provides a specific explanation of the task; *examples* demonstrate instances of task implementation; and *input* refers to the text to be processed.

Based on the above discussion, this paper presents a prompt template tailored for constructing an HVDC fault-diagnosis knowledge graph:

Prompt-1: Concept-Layer Generation

Context & Historical Input The objective is to construct a machine-interpretable HVDC fault-diagnosis knowledge graph that encompasses device attributes, operational fault phenomena, fault causes, and maintenance strategies.

Task Description Enumerate the core concept categories required for fault diagnosis and maintenance (e.g., device type, fault type, temporal attribute, spatial attribute). Do not list concrete instances. For each category, define two disjoint attribute kinds (datatype and object properties) with maximal granularity.

Output Format Organize the resulting concepts under two top-level classes: Industrial Equipment and Fault Operation Event. Provide Turtle-syntax RDF Schema declarations.

Prompt-2: Data Extraction

Context & Historical Input At this stage, the established RDF Schema for HVDC fault diagnosis is reused; the objective is to extract entities, attributes, and relations from concrete fault cases or equipment data tables and to serialize the result into an RDF-compliant file.

Source Text The input takes the form of either narrative fault-case descriptions or structured equipment data tables.

Task Description Guided by the established HVDC fault-diagnosis RDF Schema, perform end-to-end extraction on the supplied narrative or tabular data. First, identify all industrial-equipment and fault-event entities; next, extract for each entity the complete set of attributes—such as fault time, location, device status, and type—and capture causal or associative relations between devices and faults as well as among faults themselves, ultimately producing an entity-attribute-relation triple set that can be directly serialized into RDF.

Output Format Assign a unique URI to each entity and bind it to the corresponding RDF: class; express every entity-attribute pair and every entity → relation → entity triple in RDF statements, and emit a complete RDF graph.

HVDC Fault Diagnosis Method Knowledge Graph Based on LLM

Building on the two fault-operation knowledge-graph prompts designed in the previous section, this paper proposes an LLM-driven two-stage framework: RDF schema generation followed by RDF data-layer population. First, the domain-concept prompt is fed into the large language model to obtain a complete RDF schema. Next, the raw fault-operation data are supplied as source text and, together with the generated schema, are passed into the data-extraction prompt, enabling the model to extract entities, attributes, and relationships in real time, thereby populating the RDF data layer. This sequential workflow is illustrated in Fig. 8.

To effectively handle updates and ensure consistency over time, the LLM-based HVDC fault diagnosis method employs the following mechanisms: First, the dynamic update mechanism of the knowledge graph is realized through a real-time data bridging channel, which can promptly reflect new fault data and maintenance information to ensure the timeliness of information during fault diagnosis. Meanwhile, the multi-source data fusion technology integrates various data sources, including valve voltage and current measurements, trigger pulse feedback, and valve control self-check signals, not only enhancing the accuracy of diagnosis but also continuously updating the knowledge graph to maintain its consistency. In addition, the continuous learning capability of the LLM enables it to update its internal knowledge representation by receiving new fault cases and maintenance data, better adapting to new fault patterns and diagnostic

requirements. The system also includes a feedback mechanism that feeds new diagnostic results and experience back into the model to further optimize model performance and knowledge base content. When constructing the knowledge graph, a structured output specification is adopted to ensure the consistency and interpretability of the knowledge graph's format and content at different time points. The flexibility of the two-stage framework allows for independent updates of the RDF schema and data layer, thus enabling flexible handling of new fault types and diagnostic requirements without affecting the overall architecture. These mechanisms collectively ensure the system's efficiency and adaptability, providing strong support for the stable operation of HVDC systems.

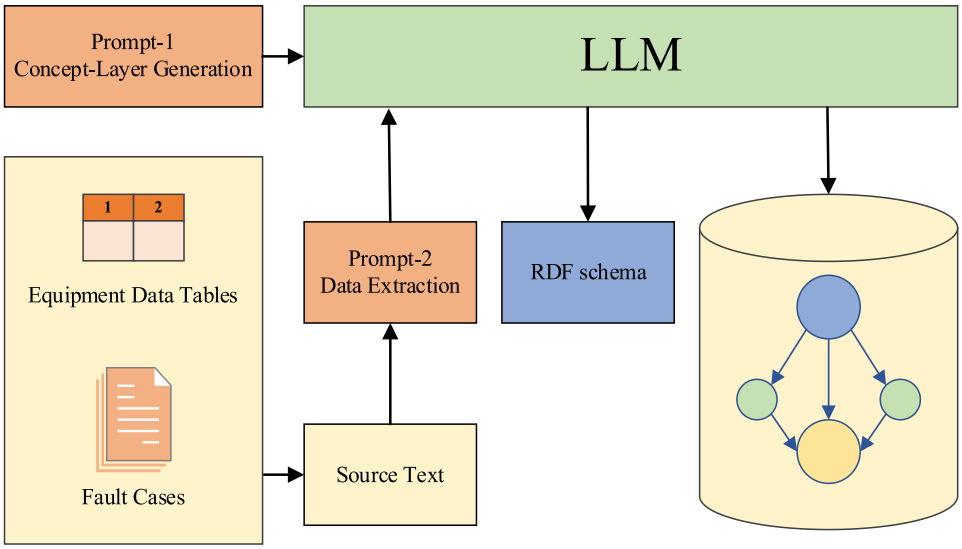

Figure 8: LLM-based knowledge graph for HVDC fault diagnosis constructing framework

4.4 Applications of Large Language Model in Question Answering

4.4.1 Classification of Question Answering Systems

Existing question answering systems commonly employ retrieval mechanisms that rely on a pre-established knowledge base containing potential answers and their source information. Based on different knowledge base structures and retrieval techniques, these systems can be categorized into four types: frequently asked question answering, knowledge graph-based question answering, table-based question answering, and machine reading comprehension. The following sections provide detailed descriptions of these four types of question answering systems.

(1) Frequently asked question-based question answering

Frequently asked question systems are retrieval systems based on predefined question-answer pairs. These systems typically include a pre-built question-answer database, where text-matching techniques are employed to identify user queries and search for the most similar question-answer pairs in the database [80]. The operational principle of such systems resembles that of search engines but focuses specifically on matching question-answer pairs. To enhance retrieval accuracy and efficiency, a two-stage framework involving recall and ranking may be adopted. First, the system retrieves candidate answers from the predefined question-answer pairs that are relevant to the user's query. Subsequently, these candidate answers are ranked based on relevance scores, and the highest-scoring answer is ultimately returned to the user.

(2) Knowledge graph-based question answering

Knowledge graph-based question answering systems utilize structured data from knowledge graphs to interpret and respond to user queries [81]. Such systems initially perform semantic parsing on user questions to convert them into executable queries within the knowledge graph. Subsequently, relevant information is retrieved through entities and relationships in the knowledge graph to generate answers. This approach relies on the quality and coverage of the knowledge graph to provide accurate and comprehensive responses. Additionally, such systems can further enrich the knowledge graph by extracting information from unstructured data using information extraction techniques.

(3) Table-based question answering

Table-based question answering systems are specifically designed to handle structured tabular data. The objective of such systems is to convert natural language questions posed by users into structured query statements, enabling the execution of queries within tables and the retrieval of accurate answers [82]. This process typically involves natural language processing techniques, such as semantic parsing and entity recognition, to interpret the meaning of questions and map them to corresponding fields within tables [83]. Table-based question answering systems are particularly useful in application scenarios involving large volumes of structured information, such as database queries and data report analysis.

(4) Machine reading comprehension based question answering

Machine reading comprehension systems simulate human reading comprehension capabilities by extracting information from large volumes of text to answer user questions [84] Such systems typically consist of two main modules: passage retrieval and reading comprehension [85]. First, the system retrieves passages relevant to the question from a large-scale document collection. Then, a reading comprehension model analyzes these passages to extract or generate answers. This approach relies on advanced natural language processing techniques, such as DL and attention mechanisms, to understand the deeper meaning and context of the text. Machine reading comprehension systems excel in handling open-domain question answering and scenarios requiring in-depth understanding of textual content.

4.4.2 Mainstream Domain-Specific Question Answering System Technologies

In the current field of question answering systems, methods such as fine-tuning, prompt learning, knowledge graphs (KG), and RAG have been demonstrated to play indispensable roles in constructing efficient and intelligent question answering systems. These technologies each contribute from different perspectives, providing substantial support for enhancing the performance of question answering systems. The following sections will elaborate in detail on the research status of these four mainstream questions answering system technologies, aiming to offer valuable references for researchers in related fields.

(1) Fine-tuning based question answering

Fine-tuning methods build upon pre-trained models by further training them with task-specific data from target objectives. During this process, model parameters are adjusted and optimized according to the characteristics and requirements of the target task, enabling the model to transition from learning general knowledge to adapting to specific applications. A common practice involves keeping the parameters of most layers frozen or updating only a subset of layers. This approach leverages the knowledge encoded in the pre-trained model while mitigating the risk of overfitting when training on limited datasets.

Current fine-tuning approaches can be classified into four main categories: full-parameter fine-tuning, partial-parameter fine-tuning, additional-parameter fine-tuning, and parameter-free fine-tuning. Full-parameter fine-tuning updates all model parameters and generally yields optimal performance, though it requires substantial computational resources. To mitigate this limitation, memory-efficient full-parameter

methods such as Diff pruning [86] and MeZO [87] have been proposed. Partial-parameter fine-tuning updates only specific components, such as biases or selected weights [88], and can be further divided into heuristic and self-learning types. While this approach has shown effectiveness as a resource-efficient alternative to full-parameter updating [89,90], it still involves computing full gradients across all parameters, which remains computationally intensive [91]. Additional-parameter fine-tuning, exemplified by Adapter-based methods [92], LoRA-based techniques, and soft prompting, introduces a limited number of external parameters to adapt models efficiently. This strategy addresses key challenges associated with full-parameter and partial-parameter methods, including large model size [93] and high computational demand [94]. Parameter-free fine-tuning relies exclusively on prompt design and external retrieval mechanisms [95], avoiding any updates to model parameters. Although this method significantly reduces computational costs for task adaptation, it faces limitations such as restricted context length, limited robustness, scalability constraints, and efficiency concerns [96]. Each fine-tuning strategy presents distinct trade-offs, and the selection among them should be guided by available computational resources, inference latency requirements, and specific task objectives in practical applications. In the context of parameter-efficient fine-tuning methodologies for LLM, reference [97] conducted a systematic investigation into instruction fine-tuning techniques applied to compact LLM-including Mistral-7B, Llama3-8B, and Phi3-mini—demonstrating substantial performance improvements in financial text classification tasks. Building upon task-specific adaptation, reference [98] proposed the Low-Rank Gradient Estimator, an efficient fine-tuning framework designed to enhance model capability across diverse applications, with demonstrated effectiveness particularly in question answering systems. Complementing gradient-based optimization approaches, reference [99] introduced the LOMO optimizer, which enables memory-efficient full-parameter fine-tuning of large language models under constrained resources, maintaining competitive downstream performance while significantly reducing memory requirements. Further expanding the parameter-efficient paradigm, reference [100] developed a prompt tuning approach that learns adaptive soft prompts to tailor frozen language models for specific downstream applications, achieving balanced performance while maximizing parameter efficiency.

(2) Prompt learning based question answering

Although large models possess extensive knowledge coverage and powerful text generation capabilities, demonstrating excellent performance in general question-answering tasks, their accuracy in generating answers often falls short when addressing domain-specific or specialized tasks. To address this limitation, prompt learning technology designs specialized prompt templates for different domains and professional scenarios, significantly enhancing model performance in specialized question-answering tasks. Specifically, as a guidance mechanism based on pre-trained models, prompt learning effectively improves the model's understanding of professional questions by integrating domain-specific prompt information with pre-trained models, enabling question-answering systems to produce more accurate and professional responses.

From a technical implementation perspective, prompt learning methods systematically guide language models to generate target outputs by embedding predefined text fragments into input data. This mechanism is particularly suitable for question-answering applications, as large models, despite their excellent text generation fluency and contextual understanding capabilities enabling multi-turn dialogue interactions, can further enhance their domain adaptability through carefully designed prompt templates. Notably, recent studies have developed several innovative prompt learning frameworks with specialized applications. In cross-modal question answering, the PEG framework [101] employs prompt learning to generate contextual descriptions for images, integrates prefix tuning to extract background knowledge from pre-trained models, and combines vision-language joint encoding with contrastive learning, achieving substantial performance

gains on WebQA and MultimodalQA datasets. Advancing beyond supervised cross-modal settings, the ZPV-QA framework [102] introduces a zero-shot prompt learning strategy that enables large language models to first generate image descriptions relevant to queries without training examples, then synthesize question-answer pairs, attaining breakthrough results on VQAv2 and OK-VQA benchmarks. Further expanding prompt-based representation learning, the PGCL method [103] guides textual and visual representation learning through attribute prompt templates, incorporates self-supervised contrastive learning to enhance multimodal fusion, and demonstrates exceptional performance on the ScienceQA dataset. Complementing these specialized frameworks, the ChatLD approach [104] leverages the GPT-4o model with chain-of-thought prompting to achieve high-accuracy, training-free sample classification via reasoning chains, while validating method scalability across multiple scenarios. These advances collectively demonstrate that carefully architected prompt learning strategies can effectively unlock the application potential of large models in specialized domains.

(3) KG based question answering

A KG serves as a graph-structured data organization format that achieves symbolic representation and structured storage of real-world knowledge through networked connections of "entity-relation-entity" triples and their associated attribute values. This unique semantic database structure provides a solid foundation for knowledge graph question answering (KGQA) systems, which parse users' natural language questions, retrieve relevant entities from the knowledge graph, and return accurate answers. Current mainstream KGQA methods are primarily divided into two technical approaches: the first is based on semantic parsing, which converts questions into logical form representations and executes queries directly on the knowledge graph to obtain answers; the second is based on information retrieval, which extracts relevant subgraphs from the knowledge graph, performs reasoning on the subgraphs, and selects highly ranked entities as final answers.

It is noteworthy that although large language models pre-trained on large-scale corpora demonstrate exceptional language understanding capabilities and broad knowledge coverage in natural language processing tasks, they still exhibit significant limitations. Specifically, these models struggle to accurately capture factual information, often generating factually incorrect statements-a phenomenon known as hallucination [105]. Additionally, due to a lack of domain-specific knowledge and training data, models trained on general corpora face challenges in effectively generalizing to specialized domains. To address these challenges, researchers have developed several innovative solutions. Reference [106] proposed a semantic-oriented fusion model for KGQA that optimizes topic entity recognition through deep transition RNNs, while innovatively designing a dynamic candidate path generation algorithm and quadruple similarity computation model. Building on semantic modeling approaches, literature [107] introduced a contrastive learning enhancement framework that employs a sampling strategy to construct semantically consistent positive pairs and inconsistent negative pairs. By jointly optimizing contrastive and generative objectives, this framework effectively mitigates exposure bias in encoder-decoder models during teacher-forced training. Extending to long-tail knowledge scenarios, study [108] constructed the LTGen benchmark comprising both single-turn QA and multi-turn dialogue tasks, systematically evaluating how non-parametric knowledge enhances LLM performance in long-tail QA. Experimental results demonstrated that KG triple prompts significantly improve accuracy while reducing hallucination. Further advancing the integration paradigm, reference [109] developed the LLaMaKG system, which bridges LLMs' natural language capabilities with KGs' structured knowledge through instruction fine-tuning, providing an effective solution to mitigate hallucination issues in large models. Collectively, these advances facilitate deeper integration of knowledge graphs with large language models, establishing new pathways for developing more reliable domain-specific question answering systems.

(4) RAG based question answering

RAG is a technology that integrates information retrieval with generative models, with its core objective being to enhance the performance of natural language processing tasks. Specifically, this technology inputs retrieved relevant content along with user queries into large language models, leveraging external knowledge to enable the models to generate more accurate and higher-quality responses [110]. From a technical advantage perspective, RAG primarily excels in two key aspects: First, by organically combining information retrieval with text generation processes, the technology significantly improves the accuracy and reliability of responses. Unlike traditional generative models that rely solely on training data, RAG retrieves relevant information from large-scale knowledge bases or document collections and generates responses based on this information, effectively avoiding the production of false information. Second, thanks to the design of its information retrieval module, RAG can access updated databases or internet resources in real-time, ensuring the timeliness of provided information-a feature that effectively compensates for the inherent shortcomings of pure generative models in terms of information updates.

It is noteworthy that although large language models demonstrate exceptional capabilities in language understanding and generation, they still face several technical bottlenecks. For instance, models are prone to hallucination phenomena and often underperform in retrieving key information. The root cause lies in the fact that the pre-training process does not explicitly teach models how to effectively utilize retrieved texts to accomplish generation tasks, especially when dealing with lengthy and complex retrieved content, where models often struggle to accurately extract key information. To address these challenges, RAG technology enriches the generative capabilities of large language models by integrating external data sources, enabling them to dynamically retrieve relevant content from external databases while processing queries or generating text.

In advancing RAG applications, several domain-specific frameworks have been developed. Reference [111] designed a RAG-based question-answering system for nutrigenetics, integrating vector database retrieval of scientific literature with large language models to significantly enhance response accuracy in specialized domains while avoiding the computational costs and data privacy risks associated with conventional fine-tuning. Expanding beyond biomedical domains, document [112] systematically investigated how OCR quality affects RAG performance and introduced an innovative method that converts OCR outputs into structured table formats, substantially improving information extraction accuracy and system reliability. Addressing educational applications, work [113] proposed the PLRTQA framework, which combines RAG with transfer learning to achieve notable performance in textbook-based question-answering, particularly in handling long-text comprehension and cross-disciplinary knowledge retrieval.

Parallel progress has been made in industrial applications. Study [114] developed the KG-RAG framework, which standardizes Failure Mode and Effects Analysis (FMEA) data into a knowledge graph and integrates vector retrieval with graph query techniques to enhance both accuracy and interpretability in risk analysis. The framework incorporates a depth-first search-based embedding generation algorithm and a dual-path retrieval mechanism, enabling effective processing of semantically ambiguous queries and precise numerical reasoning. Advancing multi-document processing, reference [115] proposed the IIER framework, which transforms multi-document content into a unified Chunk-Interaction Graph structure. By integrating structural, semantic, and keyword-based relationships across documents, the framework effectively mitigates contextual fragmentation in multi-document QA and offers a novel technical pathway for enhancing LLM performance in knowledge-intensive tasks. Collectively, these developments demonstrate that continued optimization of RAG technology can effectively overcome inherent limitations of large language models, advancing natural language processing toward more specialized and reliable applications.

Table 5 presents a visual comparison of the advantages and limitations of various QA system methods, based on the summary and analysis above. The score for each type of method is the average of the scores from the corresponding references.

Table 5: Comparison of advantages and limitations of various question answering System methods

Methods of question answering system	Advantages of HVDC applications	Limitations of HVDC applications	Complexity	Applicability	Score
Fine-tuning based question answering [97–100]	Leverages pre-trained knowledge, mitigates overfitting	Requires task-specific data, computational resources	High complexity: requiring full processes of data preparation, parameter tuning, and training deployment	Suitable for fixed question answering on standard operating procedures and historical fault review	***
Prompt learning based question answering [101–104]	Enhances model understanding of professional questions	Requires domain-specific prompt design	Low complexity: mainly involving prompt template iteration and validation	Suitable for rapid-iteration OM, question answering and routine scheduling consultation	***
KG based question answering [105–109]	Provides structured knowledge representation	Struggles with factual accuracy, domain-specific knowledge	High complexity: requiring knowledge modeling, graph database, and reasoning platform support Complex to update and maintain	Suitable for cross-site fault tracing, expert-system-level question answering, and OM decision support	***
RAG based question answering [112–115]	Improves accuracy and reliability, accesses updated information	Prone to hallucination, struggles with complex retrieved content	Moderate complexity: requiring vector retrieval, index management, and integration with generative models	Suitable for real-time fault question answering and dynamic OM knowledge updates	****

Note: ***denotes $S \in (0.7, 0.8]$, ****denotes $S \in (0.8, 0.9]$, where S is the normalized composite score derived from the respective evaluation criteria; the detailed scoring protocol is provided in Section 3, "Comprehensive Evaluation Framework for LLM Applications in HVDC Systems".

4.4.3 HVDC Question Answering System Based on LLM

Building upon the technical approach for vertical domain adaptation using LLM, the following construction scheme for an HVDC question answering system based on LLM can be designed.

(1) Interface architecture design

The HVDC question answering system adopts an innovative three-layer architecture, with each layer possessing distinct technical characteristics. The top layer is the natural language interaction layer, equipped with an advanced bidirectional semantic parsing engine that employs DL methods to accurately identify professional terminology and convert it into executable API instructions for underlying devices, while simultaneously transforming device responses into explanatory text comprehensible to technicians. The middle layer serves as the dynamic context management layer, utilizing a conversation state tracking module with

attention mechanisms to maintain coherence in multi-turn dialogues, ensuring the system can understand contextually related queries. The bottom layer functions as the domain protocol adaptation layer, specifically handling HVDC-specific protocol messages to achieve seamless integration with industrial control systems such as SCADA and PMU. This architectural design preserves the reliability of traditional power system applications while incorporating the interactive intelligence unique to question answering systems.

(2) Knowledge system construction

The HVDC question answering system requires the establishment of a multidimensional knowledge network framework. Initially, a question-knowledge mapping matrix is created, employing semantic similarity algorithms to intelligently associate typical interrogative sentence patterns with knowledge nodes. Subsequently, an explanatory knowledge extraction module is developed, embedding not only structured data such as equipment parameters but also technical principles derivation chains within the knowledge graph. Finally, a real-time data bridging channel is established to dynamically update the state of knowledge entities, ensuring the timeliness of system responses. This knowledge system supports multi-level question answering requirements ranging from basic parameter queries to complex fault diagnostics.

Given the complex structure and interconnected nature of knowledge in the field of fault diagnosis, a hyperbolic space distance metric is introduced for similarity measurement during retrieval. The local knowledge base contains i files, denoted as $D_i\ (i = 1, 2, \cdots, n)$. First, the textual content of the files in the knowledge base is segmented into text chunks $B_{ij}\ (i = 1, 2, \cdots, n; j = 1, 2, \cdots, m)$, where B_{ij} represents the j-th text chunk of the i-th file. Then, through hyperbolic embedding techniques, each text chunk is mapped into hyperbolic space to obtain a vector matrix, and a vector index $I_i\ (i = 1, 2, \cdots, n \times m)$ is established. During retrieval, the query text is first vectorized and mapped into hyperbolic space. In this space, the hyperbolic distance function is used to calculate the distance between the query vector and the precomputed text chunk vectors in the local knowledge base. Based on the results of the hyperbolic distance function, the text vectors in the database are ranked. The top k most similar text chunk vectors are then selected, and the corresponding textual information is retrieved from the knowledge base.

The hyperbolic distance $D_h\ (E, F)$ between two points $E\ (E_1, E_2, \cdots, E_n)$ and $F\ (F_1, F_2, \cdots, F_n)$ in hyperbolic space, representing the query and the answer, respectively, is expressed by the formula:

$$D_h\ (E, F) = \arccos\left(\sum_{i=1}^{n} (E_i - F_i)^2\right) \tag{31}$$

The distance is converted into a similarity metric function:

$$Similarity = f\ (D_h\ (E, F)) \tag{32}$$

$$f\ (x) = e^{-x} \tag{33}$$

In the above, the function $f\ (x)$ indicates that a smaller distance corresponds to a higher similarity. Finally, the retrieved text chunks are post-processed, integrated, and output to provide a professional response.

(3) Decision chain generation mechanism

The system implements a rigorous three-phase decision-making process. During the thought chain decomposition phase, relevant reasoning techniques are employed to break down complex problems into executable steps. The plugin coordination engine phase activates specialized modules: diagnostic plugins identify fault types, control plugins generate response strategies such as power reduction operations, and explanatory plugins reference relevant standard clauses to produce technical justifications. Ultimately, the system generates a traceable decision tree containing complete technical pathways, with each node annotated with data sources and theoretical foundations, significantly enhancing decision transparency and credibility.

(4) Algorithm generation differentiation strategy

Tailored to question answering system characteristics, specialized algorithm generation paradigms are developed. Regarding code generation, the objective shifts from singular control functions to "control + explanation" composite functions, such as outputting mathematical proofs for capacitor voltage balancing principles when generating MMC control algorithms. The output format combines executable code with Markdown technical documentation, allowing engineers to directly examine algorithmic principles. The verification mechanism incorporates an explanation computation dimension alongside traditional functional testing, utilizing BLEU and ROUGE metrics to quantify explanatory text quality. This approach enables the system to both accurately execute control commands and clearly articulate technical principles.

(5) Functional migration and upgrade pathway

System functionality migration follows a progressive upgrade path, achieving capability expansion while maintaining core architectural stability. Initially, a version compatibility matrix is established to ensure interface protocol consistency between old and new modules. Subsequently, a grayscale release mechanism is designed, with migration effectiveness verified through A/B testing. Finally, a rollback protection system is constructed to enable rapid service recovery during upgrade anomalies. This strategy simultaneously reduces migration risks and ensures continuous system evolution. Through systematic migration pathways, three core capabilities are elevated: knowledge transfer efficiency improvement, decision transparency enhancement, and system adaptability expansion. This upgrade model preserves the stable characteristics of the original system while endowing it with sustained evolutionary vitality.

Upon completion of the aforementioned steps, an HVDC question-answering system based on LLM is established; it receives user queries and returns accurate results rapidly, as illustrated in Fig. 9.

In terms of hardware configuration for the question answering system, the use of high-performance computing hardware, distributed computing architectures, model compression techniques, and caching mechanisms are employed to further reduce latency and meet the stringent requirements of HVDC systems.

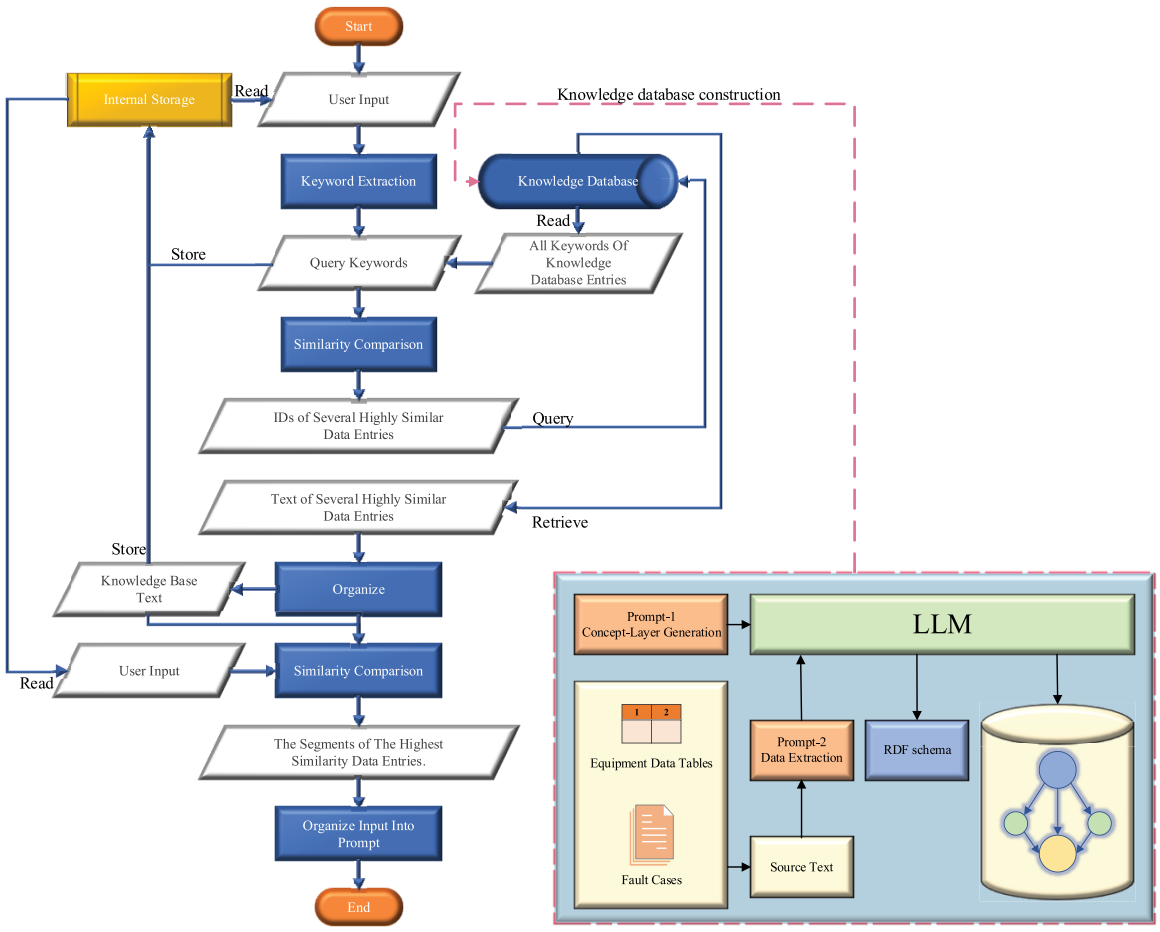

Figure 9: Knowledge question answering system retrieval process

4.5 Applications of Large Language Model in Other Tasks

From a practical engineering perspective, in addition to the three representative tasks discussed above, LLM can also support a broader range of HVDC tasks.

(1) Knowledge management and document intelligence

Knowledge management and document intelligence aim to unlock the value of large volumes of HVDC documents, including operation logs, maintenance reports, technical specifications, manufacturer manuals, and fault analyses. LLM-based text classification and information extraction can automatically organize these heterogeneous materials by equipment, fault type, and operating condition, while knowledge-graph construction techniques transform extracted entities and relations into structured HVDC operation and maintenance knowledge [116]. On this basis, semantic search and domain-specific summarization provide operators and engineers with rapid access to relevant experience and procedures, which in turn enhances the effectiveness of downstream fault diagnosis and decision support [117].

The study on the ElecBench benchmark in [118] further demonstrates how knowledge management and document intelligence can be incorporated into language models for power systems. ElecBench collects documents such as technical standards, fault records, and defect logs from power systems and reorganizes them into representative natural language processing tasks, forming a large scale evaluation suite that can serve as a foundational reference for knowledge management and document intelligence in HVDC systems.

(2) Multimodal inspection and safety supervision

Multimodal inspection and safety supervision exploit visual-language models and multimodal LLM to enhance digital inspections of converter stations and DC yards. In such scenarios, LLM can fuse image or video data from cameras, drones, or inspection robots with textual inspection records, automatically detecting abnormal equipment appearances, environmental hazards, and unsafe staff behaviors. The models can further link detected anomalies to relevant safety procedures and standards, and produce structured inspection summaries and compliance hints, thus improving both inspection quality and on-site safety [119].

Reference [120] outlines three primary implementations of multimodal detection in power systems. The first category comprises SAM-based approaches, which utilize foundation segmentation models such as SAM and its adapted variants to accelerate infrared image annotation and achieve robust segmentation of power equipment in substations. The second category involves VLM-based models, including Grid-Blip and Power-LLaVA, which integrate high-resolution visual encoders with chat-oriented LLMs fine-tuned on power-industry corpora. These models are capable of detecting abnormal conditions and generating alert messages or natural-language descriptions of hazardous scenarios. The third category consists of UAV-based solutions that leverage large models to jointly analyze drone trajectories, inspection targets, and equipment conditions, thereby enabling automated flight-path planning and the triggering of focused re-inspection or alarm mechanisms.

(3) Digital twin assisted situational awareness and simulation explanation

Digital twin assisted situational awareness and simulation explanation treat LLM as natural language interfaces to HVDC digital twins. Operators can pose "what-if" questions about power-flow reversals, commutation failures, or multi-infeed interactions in plain language, while the digital twin performs large scale simulations in the background. LLM then summarize the key results, highlight dominant constraints and risk evolution paths, and provide human-readable explanations of complex transient phenomena. This greatly reduces the cognitive load on operators and helps bridge the gap between high-dimensional simulation data and practical decision-making [121]. Work [122] proposes an LLM-driven multi-agent digital twin framework that integrates temporal data with document semantics, using LLM-based agents for diagnosis, analysis, and decision-making. This framework provides a comprehensive example of digital twin assisted situational awareness and simulation explanation for HVDC systems.

(3) Cybersecurity and security-log understanding

Cybersecurity and security-log understanding leverage LLM to analyze heterogeneous security logs. By classifying alerts, clustering similar events, and extracting attack paths and affected assets from textual descriptions, LLM can assist security analysts in rapidly identifying suspicious behaviors and configuration errors. In addition, LLM can automatically draft preliminary incident reports, which can then be reviewed and refined by human experts, improving the efficiency and consistency of cybersecurity operations in HVDC environments [123].

Literature [124] proposes the use of LLM to enhance communication security among devices in digital substations and introduces an LLM-based cybersecurity framework. The framework comprises data preprocessing for communication systems and human-in-the-loop (HITL) training, incorporating human-recommended cybersecurity guidelines.

5 Summary and Discussion

The paper provides a comprehensive review of the applications of LLM in HVDC systems, with particular emphasis on three representative domains: text clarification, fault diagnosis, and question answering. By synthesizing existing literature and recent advances, the study highlights the strengths and limitations

of current approaches and elucidates their implications for HVDC operation and research, as illustrated in Fig. 10. The main conclusions can be summarized as follows:

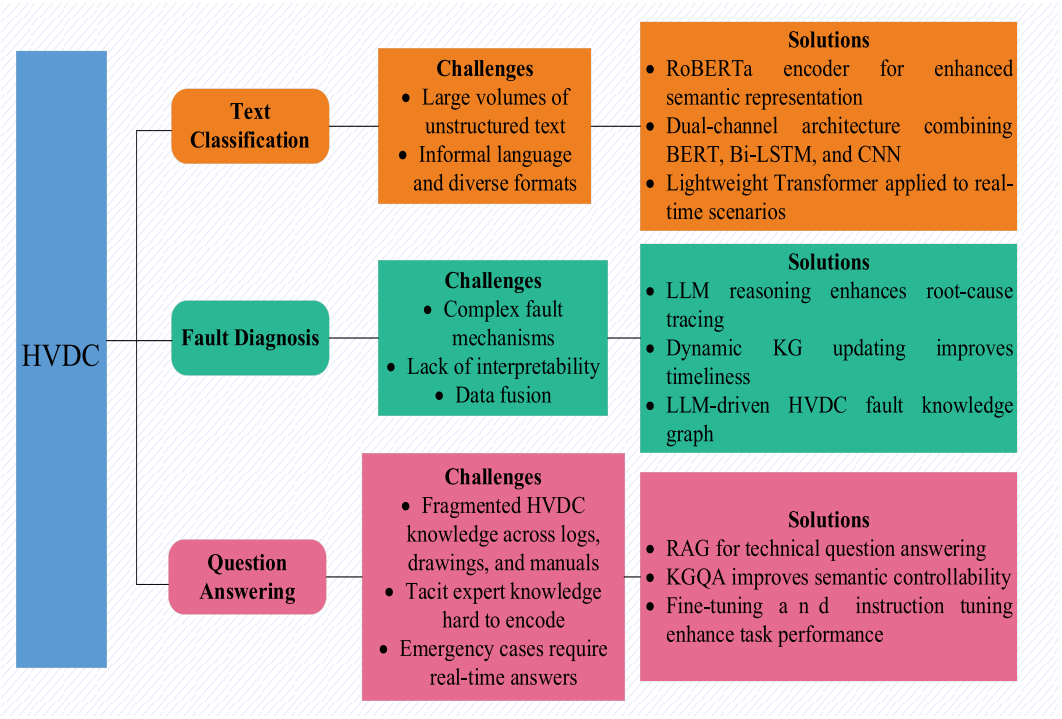

Figure 10: Summary of LLM applications in HVDC systems

Firstly, existing studies demonstrate a progression from traditional neural models toward pre-trained Transformers and emerging LLM-based solutions. While LSTM and CNN models are adequate for short and structured logs, they struggle with multi-sentence semantics and multilingual manuals. Pre-trained encoders and LLM substantially enhance representation quality and support downstream tasks such as defect-level tagging and anomaly command detection. However, through the evaluation framework, the survey identifies that text classification requires high accuracy and robustness but only moderate interpretability, making lightweight pre-trained variants more suitable than full-scale general LLM for deployment.

Second, LLM can automatically extract entities, attributes, and relationships from diverse fault datasets, thereby facilitating the construction of semantically rich knowledge bases that enable fault tracing and causal reasoning in HVDC systems. This represents a paradigm shift in HVDC fault diagnosis from traditional expert-based assessments toward adaptive reasoning frameworks. The survey additionally shows that current LLM-driven diagnosis methods form a complementary pattern with model-driven and data-driven techniques, providing both semantic abstraction and structured reasoning. However, the fault scenarios of HVDC transmission are characterized by complexity, long causal chains and high safety requirements, which pose obstacles to the application of LLM in fault diagnosis of HVDC systems. Ensuring semantic consistency, maintaining high-quality data, and embedding interpretability are essential for the reliable application of LLM in HVDC fault diagnosis.

Third, LLM-based question-answering systems for HVDC can provide operators with real-time access to technical knowledge, operational procedures, and emergency strategies. By integrating RAG and multi-modal data fusion, these systems alleviate cognitive burden and support evidence-based decision-making.

This survey further reveals that QA tasks benefit most from architectures combining retrieval accuracy, LLM reasoning, and structured knowledge constraints. Nevertheless, incomplete coverage of rare-event knowledge and insufficient mechanisms for evidence citation constrain their practical deployment in critical HVDC tasks.

Building on the above summary, the application of LLM in HVDC systems can be examined from three complementary dimensions: methodology and practice.

Firstly, compared with traditional approaches that rely on physical modeling or purely data-driven techniques, LLM is capable of integrating diverse information sources, inferring latent relationships, and generating human-interpretable explanations. However, this flexibility also introduces uncertain challenges to the security and reliability of HVDC systems. Therefore, current research should emphasize the integration of deterministic analytical models with the semantic reasoning capabilities of LLM.

Second, LLM can incorporate domain knowledge and expert experience to support routine analyses, yet engineers' validation of their outputs and interpretation of reasoning chains remain indispensable in engineering applications. Ultimately, LLM should serve as intelligent assistants that enhance engineers' situational awareness, mitigate information overload, and promote greater consistency in operational decision-making.

In conclusion, LLM hold substantial potential in HVDC systems by serving as semantic engines that bridge unstructured linguistic information with actionable signal data. At the same time, the unique requirements of HVDC systems, such as operational security, latency constraints, and interpretability, set a high threshold for adoption. Advancing this field requires not only progress in model architectures and fine-tuning strategies, but also the development of domain-specific datasets, validation protocols, and governance frameworks that ensure compliance with engineering standards and the incorporation of human-in-the-loop supervision.

6 Conclusions and Prospects

This paper presents a comprehensive review of the applications of LLM in HVDC systems. The findings indicate that, because of their powerful semantic representation and reasoning capabilities, LLM can transform large volumes of unstructured textual data into actionable knowledge, provide adaptive and interpretable support for fault diagnosis, and serve as intuitive question-answering interfaces for operational decision-making. Collectively, these functions highlight the potential of LLM to substantially enhance the efficiency, reliability, and intelligence of HVDC system operation and maintenance.

At the same time, the strong coupling characteristics, stringent real-time requirements, and strict safety constraints of HVDC systems impose demanding conditions on the application of LLM. Critical challenges remain, including the lack of domain-specific corpora, difficulties in achieving semantic consistency, and the need to guarantee low-latency and interpretable outputs remain unresolved.

Looking ahead, progress in this field will depend on three key directions. First, the construction of high-quality, large-scale corpora and evaluation benchmarks tailored to HVDC contexts is crucial for effective pre-training, fine-tuning, and reliable performance assessment of LLM. Second, the integration of LLM with knowledge graphs, RAG, and multimodal sources, e.g., operation manuals, sensor data, and structured databases, will be necessary to improve accuracy, semantic consistency and interpretability Finally, systematic testing protocols, auditing procedures, and governance frameworks must be established to ensure that LLM-based systems remain trustworthy, transparent, and compliant with the rigorous safety requirements of HVDC operations.

Acknowledgement: This work was supported by Research and Application of Key Technologies for Digitalized Transmission Network Operation and Maintenance Based on LLM.

Funding Statement: This work was supported by Research and Application of Key Technologies for Digitalized Transmission Network Operation and Maintenance Based on LLM (CGYKJXM20230156).

Author Contributions: The authors confirm their contributions to this article as follows: Funding acquisition, guidance, and research conception: Ning Wang; investigation, data collection, and writing: Xing Wen; original draft and editing: Xing Wen and Huan Chen; manuscript preparation: Yu Song, Zhuqiao Qiao and Bin Zhang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

Nomenclature

LLM	Large language model
HVDC	High voltage direct current
AC	Alternating current
DC	Direct current
NLP	Natural language processing
SLM	Statistical language model
RNN	Recurrent neural network
LSTM	Long short-term memory
MHA	Multi-head attention
FFN	Feed-forward network
RLHF	Reinforcement learning from human feedback
SFT	Supervised fine-tuning
PPO	Proximal policy optimization
NLU	Natural language understanding
NLI	Natural language inference
FFNN	Feedforward neural network
DL	Deep learning
CNN	Convolutional neural network
RoBERTa	Robustly optimized BERT pretraining approach
KGQA	Knowledge graph question answering
RAG	Retrieval-augmented generation
$\boldsymbol{U_L}$	The root-mean-square line voltage at the inverter-side converter bus, V
n	The length of the sentence
N	Order of the model
D	The collection of text sequences derived from S
S	The sentence
z_l	The output features of layer l
$z_l^{'}$	The input features to the FFN at layer l
z_{l-1}	The input features to the MHA at layer l
h	The number of parallel attention heads
$\boldsymbol{P}(\cdot)$	The likelihood function
x_i	The input text

y_i	The human-annotated target output
$\boldsymbol{\sigma}(\cdot)$	The sigmoid function
$r_\phi\,(x_i, y_i)$	The score assigned to the output y by the reward model with paraments ∅
ρ_t	The ratio between the new and old policies
A_t	The advantage function
ϵ	The magnitude of policy updates
$\boldsymbol{W_1}$	The weight matrix of the hidden layer
$\boldsymbol{W_2}$	The output weight matrix
h	The hidden feature representation
$\boldsymbol{f_t}$	The forget gate
$\boldsymbol{i_t}$	The input gate
$\tilde{c}_t$	The candidate memory cell
$\boldsymbol{o_t}$	The output gate

References

1. Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, et al. A survey of large language models. arXiv:2303.18223. 2025. doi:10.48550/arxiv.2303.18223.
2. Chen J, Liu Z, Huang X, Wu C, Liu Q, Jiang G, et al. When large language models meet personalization: perspectives of challenges and opportunities. World Wide Web. 2024;27(4):42. doi:10.1007/s11280-024-01276-1.
3. Milano S, McGrane JA, Leonelli S. Large language models challenge the future of higher education. Nat Mach Intell. 2023;5(4):333–4. doi:10.1038/s42256-023-00644-2.
4. Wang H, Fu T, Du Y, Gao W, Huang K, Liu Z, et al. Scientific discovery in the age of artificial intelligence. Nature. 2023;620(7972):47–60. doi:10.1038/s41586-023-06221-2.
5. Shafique G, Boukhenfouf J, Gruson F, Colas F, Guillaud X. DC voltage control with grid-forming capability for enhancing stability of HVDC system. J Mod Power Syst Clean Energy. 2024;13(1):66–78. doi:10.35833/mpce.2024.000822.
6. Lin H, Liang L, Yi H, Kong X. Optimal capacity planning for converter stations in sending-end MT-HVDC systems considering uncertainties of PV generation. J Mod Power Syst Clean Energy. 2024;13(6):2180–91. doi:10.35833/mpce.2024.000949.
7. Zhang Y, Ravishankar J, Fletcher J, Li R, Han M. Review of modular multilevel converter based multi-terminal HVDC systems for offshore wind power transmission. Renew Sustain Energy Rev. 2016;61:572–86. doi:10.1016/j.rser.2016.01.108.
8. Yang B, Liu B, Zhou H, Wang J, Yao W, Wu S, et al. A critical survey of technologies of large offshore wind farm integration: summary, advances, and perspectives. Prot Control Mod Power Syst. 2022;7(1):17. doi:10.1186/s41601-022-00239-w.
9. Yang B, Jiang L, Yu T, Shu HC, Zhang CK, Yao W, et al. Passive control design for multi-terminal VSC-HVDC systems via energy shaping. Int J Electr Power Energy Syst. 2018;98:496–508. doi:10.1016/j.ijepes.2017.12.028.
10. Ruan J, Liang G, Zhao J, Zhao H, Qiu J, Wen F, et al. Deep learning for cybersecurity in smart grids: review and perspectives. Energy Convers Econ. 2023;4(4):233–51. doi:10.1049/enc2.12091.
11. Valinejad J, Mili L, Yu X, van der Wal CN, Xu Y. Computational social science in smart power systems: reliability, resilience, and restoration. Energy Convers Econ. 2023;4(3):159–70. doi:10.1049/enc2.12087.
12. Patel V, Kapoor A, Sharma A, Chakrabarti S. Taxonomy of outlier detection methods for power system measurements. Energy Convers Econ. 2023;4(2):73–88. doi:10.1049/enc2.12082.
13. Hao Y, Yu L, Nie C, Li F, Luo H. Reactive power sensitivity-based novel overvoltage suppression strategy for parallel VSC-and-LCC-HVDC transmission system. Int J Electr Power Energy Syst. 2025;172:111294. doi:10.1016/j.ijepes.2025.111294.
14. Zhao Y, Song Z, Li D, Qian R, Lin S. Wind turbine gearbox fault diagnosis based on multi-sensor signals fusion. Prot Control Mod Power Syst. 2024;9(4):96–109. doi:10.23919/pcmp.2023.000241.

15. Yang B, Zheng R, Han Y, Huang J, Li M, Shu H, et al. Recent advances in fault diagnosis techniques for photovoltaic systems: a critical review. Prot Control Mod Power Syst. 2024;9(3):36–59. doi:10.23919/pcmp.2023.000583.
16. Shokoohi S, Moshtagh J. Parity equation model-based fault diagnosis method newly applied to bearing faults in induction motors. Measurement. 2025;249:117057. doi:10.1016/j.measurement.2025.117057.
17. Lu G, Tsang CW, Yim HN, Lei C, Bu S, Yung WKC, et al. Interpretable fault diagnosis for overhead lines with covered conductors: a physics-informed deep learning approach. Prot Control Mod Power Syst. 2025;10(2):25–39. doi:10.23919/pcmp.2023.000159.
18. Barraza JF, Droguett EL, Martins MR. SCF-Net: a sparse counterfactual generation network for interpretable fault diagnosis. Reliab Eng Syst Saf. 2024;250:110285. doi:10.1016/j.ress.2024.110285.
19. Shi P, Zhao Y, Xu X, Han D. A small-sample cross-domain bearing fault diagnosis method based on knowledge-enhanced domain adversarial learning. Neurocomputing. 2025;658:131699. doi:10.1016/j.neucom.2025.131699.
20. Yan B, Li K, Xu M, Dong Y, Zhang Y, Ren Z, et al. On protecting the data privacy of large language models (LLMs) and LLM agents: a literature review. High Confid Comput. 2025;5(2):100300. doi:10.1016/j.hcc.2025.100300.
21. Ghanbarzadeh R, Mirjalili S. A structured review of large language models in metaheuristic optimisation. Decis Anal J. 2025;15:100587. doi:10.1016/j.dajour.2025.100587.
22. Pratap S, Aranha AR, Kumar D, Malhotra G, Iyer APN, Shylaja SS. The fine art of fine-tuning: a structured review of advanced LLM fine-tuning techniques. Nat Lang Process J. 2025;11:100144. doi:10.1016/j.nlp.2025.100144.
23. Ferrag MA, Alwahedi F, Battah A, Cherif B, Mechri A, Tihanyi N, et al. Generative AI in cybersecurity: a comprehensive review of LLM applications and vulnerabilities. Internet Things Cyber Phys Syst. 2025;5:1–46. doi:10.1016/j.iotcps.2025.01.001.
24. Antonesi G, Cioara T, Anghel I, Michalakopoulos V, Sarmas E, Toderean L. A systematic review of transformers and large language models in the energy sector: towards agentic digital twins. Appl Energy. 2025;401:126670. doi:10.1016/j.apenergy.2025.126670.
25. Cheng Y, Zhao H, Zhou X, Zhao J, Cao Y, Yang C, et al. A large language model for advanced power dispatch. Sci Rep. 2025;15(1):8925. doi:10.1038/s41598-025-91940-x.
26. Mikolov T, Zweig G. Context dependent recurrent neural network language model. In: Proceedings of the 2012 IEEE Spoken Language Technology Workshop (SLT); 2012 Dec 2–5; Miami, FL, USA. p. 234–9. doi:10.1109/SLT.2012.6424228.
27. Mnih A, Zhang Y, Hinton G. Improving a statistical language model through non-linear prediction. Neurocomputing. 2009;72(7–9):1414–8. doi:10.1016/j.neucom.2008.12.025.
28. Ding L, Chen Y, Xiao T, Huang A, Shen C. Exploration of generative intelligent application mode for new power systems based on large language models. Autom Electr Power Syst. 2024;48(19):1–13. (In Chinese). doi:10.7500/AEPS20231217001.
29. Wu D, Nie L, Mumtaz RA, Agarwal K. A LLM-based hybrid-transformer diagnosis system in healthcare. IEEE J Biomed Health Inform. 2025;29(9):6428–39. doi:10.1109/jbhi.2024.3481412.
30. Narayanan D, Shoeybi M, Casper J, LeGresley P, Patwary M, Korthikanti V, et al. Efficient large-scale language model training on GPU clusters using megatron-LM. In: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. St. Louis, MO, USA; 2021. p. 1–15. doi:10.1145/3458817.3476209.
31. Sindhu B, Prathamesh RP, Sameera MB, Kumara Swamy S. The evolution of large language model: models, applications and challenges. In: Proceedings of the 2024 International Conference on Current Trends in Advanced Computing (ICCTAC); 2024 May 8–9; Bengaluru, India. p. 1–8. doi:10.1109/ICCTAC61556.2024.10581180.
32. Cao Y, Zhao H, Cheng Y, Shu T, Chen Y, Liu G, et al. Survey on large language model-enhanced reinforcement learning: concept, taxonomy, and methods. IEEE Trans Neural Netw Learn Syst. 2025;36(6):9737–57. doi:10.1109/tnnls.2024.3497992.
33. Sun QY, Zhang RX, Chen DY. Application and route prospect of empowering power systems with large language models. Autom Electr Power Syst. 2025;1–22. (In Chinese).
34. Liu Y, He H, Han T, Zhang X, Liu M, Tian J, et al. Understanding LLMs: a comprehensive overview from training to inference. Neurocomputing. 2025;620:129190. doi:10.1016/j.neucom.2024.129190.

35. Kuzman T, Ljubešić N. LLM teacher-student framework for text classification with no manually annotated data: a case study in IPTC news topic classification. IEEE Access. 2025;13:35621–33. doi:10.1109/ACCESS.2025.3544814.
36. Wang D, Alfred R, Obit JH, On CK. A literature review on text classification and sentiment analysis approaches. In: Computational science and technology. Singapore: Springer; 2021. p. 305–23. doi:10.1007/978-981-33-4069-5_26.
37. Mishra A, Jain SK. A survey on question answering systems with classification. J King Saud Univ Comput Inf Sci. 2016;28(3):345–61. doi:10.1016/j.jksuci.2014.10.007.
38. Richard E, Reddy B. Text classification for clinical trial operations: evaluation and comparison of natural language processing techniques. Ther Innov Regul Sci. 2021;55(2):447–53. doi:10.1007/s43441-020-00236-x.
39. Attieh J, Tekli J. Fast text classification using lean gradient descent feed forward neural network for category feature augmentation. In: Proceedings of the 2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom); 2023 Nov 1–3; Exeter, UK. p. 2341–8. doi:10.1109/TrustCom60117.2023.00330.
40. Sari WK, Rini DP, Malik RF. Text classification using long short-term memory. In: Proceedings of the 2019 International Conference on Electrical Engineering and Computer Science (ICECOS); 2019 Oct 2–3; Batam, Indonesia. p. 150–5. doi:10.1109/ICECOS47637.2019.8984558.
41. Wang R, Li Z, Cao J, Chen T, Wang L. Convolutional recurrent neural networks for text classification. In: Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN); 2019 Jul 14–19; Budapest, Hungary. p. 1–6. doi:10.1109/ijcnn.2019.8852406.
42. Rodrawangpai B, Daungjaiboon W. Improving text classification with transformers and layer normalization. Mach Learn Appl. 2022;10:100403. doi:10.1016/j.mlwa.2022.100403.
43. Selim A, Zhao J, Yang B. Large language model for smart inverter cyber-attack detection via textual analysis of volt/VAR commands. IEEE Trans Smart Grid. 2024;15(6):6179–82. doi:10.1109/TSG.2024.3453648.
44. Zhou Z, Zhang C, Liang X, Liu H, Diao M, Deng Y. BERT-based dual-channel power equipment defect text assessment model. IEEE Access. 2024;12(3):134020–6. doi:10.1109/access.2024.3444852.
45. Wei X, Li S, Shan Y, Du H, Jiang T, Zhang J, et al. A novel second harmonic disturbance suppression method for islanded hybrid AC/DC microgrid clusters under asymmetric AC side fault. Int J Electr Power Energy Syst. 2025;171:111026. doi:10.1016/j.ijepes.2025.111026.
46. Li S, Li Y, Li T. An autonomous flexible power management for hybrid AC/DC microgrid with multiple subgrids under the asymmetric AC side faults. Int J Electr Power Energy Syst. 2022;142:107985. doi:10.1016/j.ijepes.2022.107985.
47. Shu H, Ren M, Tian X, Zhao W. Strategies for improving ride-through capability of three-terminal HVDC system under AC side fault. Int J Electr Power Energy Syst. 2023;150:109074. doi:10.1016/j.ijepes.2023.109074.
48. Xu Z, Xiao H, Xiao L, Zhang Z. DC fault analysis and clearance solutions of MMC-HVDC systems. Energies. 2018;11(4):941. doi:10.3390/en11040941.
49. Zhao Y, Zhou Y, Wang S, Liu J, Liu S. Research on fault detection technology for DC side cables of inverters in power network. AIP Adv. 2025;15(7):075008. doi:10.1063/5.0269675.
50. Elserougi AA, Abdel-Khalik AS, Massoud AM, Ahmed S. A new protection scheme for HVDC converters against DC-side faults with current suppression capability. IEEE Trans Power Deliv. 2014;29(4):1569–77. doi:10.1109/TPWRD.2014.2325743.
51. Yang J, Fletcher JE, O'Reilly J. Short-circuit and ground fault analyses and location in VSC-based DC network cables. IEEE Trans Ind Electron. 2012;59(10):3827–37. doi:10.1109/tie.2011.2162712.
52. Liu Y, Mao M, Zheng Y, Chang L. DC short-circuit fault detection for MMC-HVDC-grid based on improved DBN and DC fault current statistical features. CPSS Trans Power Electron Appl. 2023;8(2):148–60. doi:10.24295/cpsstpea.2023.00015.
53. Xiao Q, Jin Y, Jia H, Tang Y, Cupertino AF, Mu Y, et al. Review of fault diagnosis and fault-tolerant control methods of the modular multilevel converter under submodule failure. IEEE Trans Power Electron. 2023;38(10):12059–77. doi:10.1109/tpel.2023.3283286.

54. Liu L, Li X, Teng Y, Luo Y, Chen K. Improved commutation failure prevention control for inter-phase short-circuit faults. Appl Sci. 2025;15(18):9972. doi:10.3390/app15189972.
55. Fang J, Yang S, Wang H, Tashakor N, Goetz SM. Reduction of MMC capacitances through parallelization of symmetrical half-bridge submodules. IEEE Trans Power Electron. 2021;36(8):8907–18. doi:10.1109/tpel.2021.3049389.
56. Wu H, Wang Y, Liu Y, Liu Y, Li Y. An improved switching method-based diagnostic strategy for IGBT open-circuit faults in hybrid modular multilevel converters. IEEE Trans Power Electron. 2025;40(2):3578–99. doi:10.1109/TPEL.2024.3492236.
57. Jiang Q, Li B, Liu T, Blaabjerg F, Wang P. Study of cyber attack's impact on LCC-HVDC system with false data injection. IEEE Trans Smart Grid. 2023;14(4):3220–31. doi:10.1109/tsg.2023.3266780.
58. Su C, Yin C, Li F, Han L. A novel recovery strategy to suppress subsequent commutation failure in an LCC-based HVDC. Prot Control Mod Power Syst. 2024;9(1):38–51. doi:10.23919/pcmp.2023.000203.
59. Wang J, Yang B, Zeng C, Chen Y, Guo Z, Li D, et al. Recent advances and summarization of fault diagnosis techniques for proton exchange membrane fuel cell systems: a critical overview. J Power Sources. 2021;500(1):229932. doi:10.1016/j.jpowsour.2021.229932.
60. Liu H, Zhang L, Li L. Progress on the fault diagnosis approach for lithium-ion battery systems: advances, challenges, and prospects. Prot Control Mod Power Syst. 2024;9(5):16–41. doi:10.23919/PCMP.2023.000213.
61. Sawant V, Zambare P. DC fast charging stations for electric vehicles: a review. Energy Convers Econ. 2024;5(1):54–71. doi:10.1049/enc2.12111.
62. Ferreira VH, Zanghi R, Fortes MZ, Sotelo GG, Silva RBM, Souza JCS, et al. A survey on intelligent system application to fault diagnosis in electric power system transmission lines. Electr Power Syst Res. 2016;136(5):135–53. doi:10.1016/j.epsr.2016.02.002.
63. Wang X, Chen H, Yin Z, Yang F, Anuchin A, Demidova G, et al. Research on fault diagnosis of asymmetric three-level T-type power converter based on switched reluctance motor. Prot Control Mod Power Syst. 2025;10(6):81–100. doi:10.23919/pcmp.2024.000377.
64. Wu Z, Zhou L, Lin T, Zhou X, Wang D, Gao S, et al. A new testing method for the diagnosis of winding faults in transformer. IEEE Trans Instrum Meas. 2020;69(11):9203–14. doi:10.1109/tim.2020.2998877.
65. Hemmati M, Ghaffarzadeh N. A novel phaselet-based approach for fault detection and location in HVDC lines. Comput Electr Eng. 2024;120:109667. doi:10.1016/j.compeleceng.2024.109667.
66. Vieira JCC, Fernandes D, Neves WLA, Lopes FV. Fault diagnosis in bipolar HVDC systems based on traveling wave theory by monitoring data from one terminal. Electr Power Syst Res. 2023;223(1):109594. doi:10.1016/j.epsr.2023.109594.
67. Wu F, Chen K, Qiu G, Ying H, Sheng H, Wang Y. Detectability based data-driven fault diagnosis method for multiple device faults of converters. IEEE Trans Power Electron. 2025;40(4):5983–98. doi:10.1109/tpel.2024.3510749.
68. Jiang Z, Yang B, Zheng R, Hou Y, Li H, Gao D, et al. Fault diagnosis of proton exchange membrane fuel cell using multiple convolutional neural networks with multi-scale attention mechanism. Inf Sci. 2025;720:122524. doi:10.1016/j.ins.2025.122524.
69. Ge L, Du T, Xu Z, Hou L, Yan J, Li Y. A fault diagnosis method for smart meters via two-layer stacking ensemble optimization and data augmentation. J Mod Power Syst Clean Energy. 2024;12(4):1272–84. doi:10.35833/MPCE.2023.000909.
70. Su X, Deng C, Shan Y, Shahnia F, Fu Y, Dong Z. Fault diagnosis based on interpretable convolutional temporal-spatial attention network for offshore wind turbines. J Mod Power Syst Clean Energy. 2024;12(5):1459–71. doi:10.35833/mpce.2023.000606.
71. Yang B, Guo Z, Wang J, Wang J, Zhu T, Shu H, et al. Solid oxide fuel cell systems fault diagnosis: critical summarization, classification, and perspectives. J Energy Storage. 2021;34:102153. doi:10.1016/j.est.2020.102153.
72. Kong L, Mao Y, Zhang T, Chen X, Wang Z, Wang X. Online data-driven diagnosis for common electrical and sensor faults in dual three-phase PMSM drives. IEEE Trans Instrum Meas. 2025;74:1–10. doi:10.1109/tim.2025.3541811.

73. Guo B, Yang B, Wang S, Shi W, Yang F, Wang D. Machine learning-based fault diagnosis and classification of three-phase transmission lines with RFE and domain knowledge. Electr Power Syst Res. 2025;247:111777. doi:10.1016/j.epsr.2025.111777.
74. Akbari E, Samady Shadlu M. An intelligent ANFIS-based fault detection, classification, and location model for VSC-HVDC systems based on hybrid PCA-DWT signal processing technique. Comput Electr Eng. 2024;120:109763. doi:10.1016/j.compeleceng.2024.109763.
75. Ding C, Wang Z, Ding Q, Yuan Z. Convolutional neural network based on fast Fourier transform and gramian angle field for fault identification of HVDC transmission line. Sustain Energy Grids Netw. 2022;32(1):100888. doi:10.1016/j.segan.2022.100888.
76. Zhang C, Zhao Y, Zhou Y, Zhang X, Li T. A real-time abnormal operation pattern detection method for building energy systems based on association rule bases. Build Simul. 2022;15(1):69–81. doi:10.1007/s12273-021-0791-x.
77. Zhou ZJ, Hu GY, Hu CH, Wen CL, Chang LL. A survey of belief rule-base expert system. IEEE Trans Syst Man Cybern Syst. 2021;51(8):4944–58. doi:10.1109/tsmc.2019.2944893.
78. Xia L, Zheng P, Li X, Gao RX, Wang L. Toward cognitive predictive maintenance: a survey of graph-based approaches. J Manuf Syst. 2022;64:107–20. doi:10.1016/j.jmsy.2022.06.002.
79. Gharahbagheri H, Imtiaz SA, Khan F. Root cause diagnosis of process fault using KPCA and Bayesian network. Ind Eng Chem Res. 2017;56(8):2054–70. doi:10.1021/acs.iecr.6b01916.
80. Tseng WT, Hsu YC, Chen B. Effective FAQ retrieval and question matching tasks with unsupervised knowledge injection. In: Text, speech, and dialogue. Cham, Switzerland: Springer; 2021. p. 124–34. doi:10.1007/978-3-030-83527-9_11.
81. Zhu R, Liu B, Tian Q, Zhang R, Zhang S, Hu Y, et al. Knowledge graph based question-answering model with subgraph retrieval optimization. Comput Oper Res. 2025;177:106995. doi:10.1016/j.cor.2025.106995.
82. Tao Y, Liu J, Li H, Cao W, Qin X, Tian Y, et al. KFEX-N: a table-text data question-answering model based on knowledge-fusion encoder and EX-N tree decoder. Neurocomputing. 2024;593:127795. doi:10.1016/j.neucom.2024.127795.
83. Nan L, Hsieh C, Mao Z, Lin XV, Verma N, Zhang R, et al. FeTaQA: free-form table question answering. Trans Assoc Comput Linguist. 2022;10:35–49. doi:10.1162/tacl_a_00446.
84. Ahmed M, Khan H, Iqbal T, Khaled Alarfaj F, Alomair A, Almusallam N. On solving textual ambiguities and semantic vagueness in MRC based question answering using generative pre-trained transformers. PeerJ Comput Sci. 2023;9:e1422. doi:10.7717/peerj-cs.1422.
85. Wang M, He X, Liu L, Fang Q, Zhang M, Chen H, et al. HCT: Chinese medical machine reading comprehension question-answering via hierarchically collaborative transformer. IEEE J Biomed Health Inform. 2024;28(5):3055–66. doi:10.1109/jbhi.2024.3368288.
86. Guo D, Rush AM, Kim Y. Parameter-efficient transfer learning with diff pruning. arXiv:2012.07463. 2020. doi:10.48550/arxiv.2012.07463.
87. Malladi S, Gao T, Nichani E, Damian A, Lee JD, Chen D, et al. Fine-tuning language models with just forward passes. arXiv:2305.17333. 2023. doi:10.48550/arxiv.2305.17333.
88. Michel P, Levy O, Neubig G. Are sixteen heads really better than one? arXiv:1905.10650. 2019. doi:10.48550/arXiv.1905.10650.
89. Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. arXiv:1808.01974. 2018. doi:10.48550/arxiv.1808.01974.
90. Kovaleva O, Romanov A, Rogers A, Rumshisky A. Revealing the dark secrets of BERT. arXiv:1908.08593. 2019. doi:10.48550/arxiv.1908.08593.
91. Zhang Q, Chen M, Bukharin A, Karampatziakis N, He P, Cheng Y, et al. AdaLoRA: adaptive budget allocation for parameter-efficient fine-tuning. arXiv:2303.10512. 2023. doi:10.48550/arxiv.2303.10512.
92. Houlsby N, Giurgiu A, Jastrzebski S, Morrone B, de Laroussilhe Q, Gesmundo A, et al. Parameter-efficient transfer learning for NLP. arXiv:1902.00751. 2019. doi:10.48550/arxiv.1902.00751.
93. Pfeiffer J, Kamath A, Rücklé A, Cho K, Gurevych I. AdapterFusion: non-destructive task composition for transfer learning. arXiv:2005.00247. 2020. doi:10.48550/arxiv.2005.00247.

94. Wang S, Zhu Y, Liu H, Zheng Z, Chen C, Li J. Knowledge editing for large language models: a survey. arXiv:2310.16218. 2023. doi:10.48550/arXiv.2310.16218.
95. Petroni F, Rocktäschel T, Riedel S, Lewis P, Bakhtin A, Wu Y, et al. Language models as knowledge bases? In: Proceedings of the 2019 Conference on Empirical Methods In Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Stroudsburg, PA, USA: ACL; 2019. p. 2463–73. doi:10.18653/v1/d19-1250.
96. Li Y. A practical survey on zero-shot prompt design for in-context learning. In: Proceedings of the Conference Recent Advances in Natural Language Processing—Large Language Models for Natural Language Processings. Shoumen, Bulgaria: INCOMA Ltd.; 2023. p. 641–7. doi:10.26615/978-954-452-092-2_069.
97. Fatemi S, Hu Y, Mousavi M. A comparative analysis of instruction fine-tuning large language models for financial text classification. ACM Trans Manage Inf Syst. 2025;16(1):1–30. doi:10.1145/3706119.
98. Zhang L, Lou Z, Ying Y, Yang C, Zhou H. Efficient fine-tuning of large language models via a low-rank gradient estimator. Appl Sci. 2025;15(1):82. doi:10.3390/app15010082.
99. Lv K, Yang Y, Liu T, Guo Q, Qiu X. Full parameter fine-tuning for large language models with limited resources. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. Vol. 1. Stroudsburg, PA, USA: ACL; 2024. p. 8187–98. doi:10.18653/v1/2024.acl-long.445.
100. Lester B, Al-Rfou R, Constant N. The power of scale for parameter-efficient prompt tuning. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: ACL; 2021. p. 3045–59. doi:10.18653/v1/2021.emnlp-main.243.
101. Cui C, Li Z. Prompt-enhanced generation for multimodal open question answering. Electronics. 2024;13(8):1434. doi:10.3390/electronics13081434.
102. Hu N, Zhang X, Zhang Q, Huo W, You S. ZPVQA: visual question answering of images based on zero-shot prompt learning. IEEE Access. 2025;13:50849–59. doi:10.1109/access.2025.3550942.
103. Gao L, Zhang H, Liu Y, Sheng N, Feng H, Xu H. PGCL: prompt guidance and self-supervised contrastive learning-based method for visual question answering. Expert Syst Appl. 2024;251:124011. doi:10.1016/j.eswa.2024.124011.
104. Pan J, Zhong R, Xia F, Huang J, Zhu L, Yang Y, et al. ChatLeafDisease: a chain-of-thought prompting approach for crop disease classification using large language models. Plant Phenomics. 2025;7(3):100094. doi:10.1016/j.plaphe.2025.100094.
105. Zheng Y, Lu L, Hu Y, Chen Y, Wang A. Thoughtful and cautious reasoning: a fine-tuned knowledge graph-based multi-hop question answering framework. Eng Appl Artif Intell. 2025;150:110479. doi:10.1016/j.engappai.2025.110479.
106. Xiong H, Wang S, Tang M, Wang L, Lin X. Knowledge graph question answering with semantic oriented fusion model. Knowl Based Syst. 2021;221:106954. doi:10.1016/j.knosys.2021.106954.
107. Du H, Zhang X, Wang M, Chen Y, Ji D, Ma J, et al. A contrastive framework for enhancing knowledge graph question answering: alleviating exposure bias. Knowl Based Syst. 2023;280:110996. doi:10.1016/j.knosys.2023.110996.
108. Huang W, Zhou G, Lapata M, Vougiouklis P, Montella S, Pan JZ. Prompting large language models with knowledge graphs for question answering involving long-tail facts. Knowl Based Syst. 2025;324:113648. doi:10.1016/j.knosys.2025.113648.
109. Xu Y. Design of question-answering and reasoning system combining large language models and knowledge graphs. Procedia Comput Sci. 2025;262:348–57. doi:10.1016/j.procs.2025.05.062.
110. Gao Y, Xiong Y, Gao X, Jia K, Pan J, Bi Y, et al. Retrieval-augmented generation for large language models: a survey. arXiv:2312.10997. 2023. doi:10.48550/arxiv.2312.10997.
111. Benfenati D, de Filippis GM, Rinaldi AM, Russo C, Tommasino C. A retrieval-augmented generation application for question-answering in nutrigenetics domain. Procedia Comput Sci. 2024;246:586–95. doi:10.1016/j.procs.2024.09.467.
112. Song M. Defining the problem: the impact of OCR quality on retrieval-augmented generation performance and strategies for improvement. Inf Process Manag. 2026;63(1):104368. doi:10.1016/j.ipm.2025.104368.

113. Bahr L, Wehner C, Wewerka J, Bittencourt J, Schmid U, Daub R. Knowledge graph enhanced retrieval-augmented generation for failure mode and effects analysis. J Ind Inf Integr. 2025;45:100807. doi:10.1016/j.jii.2025.100807.
114. Guo T, Wang C, Liu Y, Tang J, Li P, Xu S, et al. Leveraging inter-chunk interactions for enhanced retrieval in large language model-based question answering. Neurocomputing. 2025;650(1):130931. doi:10.1016/j.neucom.2025.130931.
115. Kaddour J, Harris J, Mozes M, Bradley H, Raileanu R, McHardy R. Challenges and applications of large language models. arXiv:2307.10169. 2023. doi:10.48550/arxiv.2307.10169.
116. Wang D, Raman N, Sibue M, Ma Z, Babkin P, Kaur S, et al. DocLLM: a layout-aware generative language model for multimodal document understanding. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL; 2024. p. 8529–48. doi:10.18653/v1/2024.acl-long.463.
117. Mao Z, Bai H, Hou L, Shang L, Jiang X, Liu Q, et al. Visually guided generative text-layout pre-training for document intelligence. In: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language. Stroudsburg, PA, USA: ACL; 2024. p. 4713–30. doi:10.18653/v1/2024.naacl-long.264.
118. Zhang S, Huang Q, Zhang Q, Liang X, Liu W, Gao K, et al. ElecBench: a large language model benchmark in electric power domain. Eng Appl Artif Intell. 2025;162:112310. doi:10.1016/j.engappai.2025.112310.
119. Liu J, Zhan J. Constructing knowledge graph from cyber threat intelligence using large language model. In: Proceedings of the 2023 IEEE International Conference on Big Data (BigData); 2023 Dec 15–18; Sorrento, Italy. p. 516–21. doi:10.1109/BigData59044.2023.10386611.
120. Hao J, Tao Y. Power grid inspection based on multimodal foundation models. EAI Endorsed Trans Energy Web. 2025;12:1–7. doi:10.4108/ew.9087.
121. He X, Tang Y, Ma S, Ai Q, Tao F, Qiu R. Redefinition of digital twin and its situation awareness framework designing toward fourth paradigm for energy Internet of Things. IEEE Trans Syst Man Cybern Syst. 2024;54(11):6873–88. doi:10.1109/tsmc.2024.3407061.
122. Sun Y, Zhang Q, Bao J, Lu Y, Liu S. Empowering digital twins with large language models for global temporal feature learning. J Manuf Syst. 2024;74:83–99. doi:10.1016/j.jmsy.2024.02.015.
123. Ren H, Lan K, Sun Z, Liao S. CLogLLM: a large language model enabled approach to cybersecurity log anomaly analysis. In: Proceedings of the 2024 4th International Conference on Electronic Information Engineering and Computer Communication (EIECC); 2024 Dec 27–29; Wuhan, China. p. 963–70. doi:10.1109/EIECC64539.2024.10929078.
124. Zaboli A, Choi SL, Song TJ, Hong J. ChatGPT and other large language models for cybersecurity of smart grid applications. In: Proceedings of the 2024 IEEE Power & Energy Society General Meeting (PESGM); 2024 Jul 21–25; Seattle, WA, USA. p. 1–5. doi:10.1109/PESGM51994.2024.10688863.