AH-SpanBERT: Fine-Tuning SpanBERT with Archerfish Optimization for Power Grid Dispatcher Training

Liao Binjie[1], Zhang Hongtu[1], He Zhihua[1], Zhang Peng[1], Yang Tingtian[1], Liu Shudi[2], Luo Rongsen[2], Liu Feng[3*]

[1]State Grid Sichuan Electric Power Company, Sichuan, 610041, China

[2]State Grid Sichuan Electric Power Research Institute, Sichuan, 610041, China

[3]BeiJing Tsintergy Technology Co., Ltd., BeiJing, 100085, China

E-mail: liufeng8764@outlook.com, feiliu562@outlook.com, shengkaiwang2@outlook.com, jindu12@outlook.com, erlanxie123@outlook.com, weixiang234@outlook.com, fengzuoguo22@outlook.com, xueyanliu2407@outlook.com

Keywords: power grid dispatchers, large language models (LLMs), black start scenarios, decision-making, archerfish hunting fine-tuned span bidirectional encoder representations from Transformers (AH-SpanBERT)

Recieved: June 2, 2025

Power grid dispatchers play a critical role in maintaining the stability and efficiency of electrical networks. As power systems grow in complexity, traditional training methods struggle to equip dispatchers with the necessary skills for rapid decision-making and human-machine collaboration. This research explores the application of fine-tuning general large language models (LLMs) to enhance internal training processes for power grid dispatchers. The research proposed Archerfish Hunting Fine-tuned Span Bidirectional Encoder Representations from Transformers (AH-SpanBERT), a model that integrates the SpanBERT architecture with Archerfish Hunting (AH) optimization to improve decision-making and operational efficiency in power system management. To fine-tune the model, a comprehensive dataset of 1,000 simulated power grid operational records was created, covering scenarios such as equipment failures, grid fluctuations, and emergency responses. The data was preprocessed using domain-specific tokenization and term normalization to ensure consistency and contextual relevance. The AH-SpanBERT model was trained using this dataset, with specific prompt strategies designed to simulate real-world dispatch scenarios and foster interactive, scenario-based learning. The model's performance was evaluated across multiple key metrics, including factuality, logicality, stability, and security. Results show significant improvements in factuality (8.48 in operation monitoring), logicality (9.74 in general scenarios), stability (9.15 in black start procedures), and security (9.62 in black start procedures). The AH-SpanBERT model outperforms existing LLMs such as GPT-4 and GAIA-70B in these areas, demonstrating its potential to enhance dispatcher decision-making and human-machine collaboration in critical power grid operations. This research highlights the effectiveness of fine-tuning general LLMs with domain-specific data to improve dispatcher training and operational performance in power grid management.

Povzetek: Članek obravnava usposabljanje dispečerjev energetskih omrežij za obvladovanje naraščajoče kompleksnosti omrežja in neenotnih podatkov. Predlaga AH-SpanBERT, tj. SpanBERT fino uglašen z optimizacijo Archerfish Hunting, treniran na 1000 simuliranih dispečerskih zapisih. Metoda izboljša faktualnost, logičnost, stabilnost in varnost odločanja v scenarijih nad GPT-4 in GAIA-70B.

1 Introduction

A persistent power connection device requires the realization of data connectivity and business collaboration. As the power grid's intelligence increases, the power sector has to manage the rapidly expanding amount of information. The intelligent evolution of the power grid greatly depends on the significant amount of unstructured data that was gathered during the energy grid's construction and development. The information is a significant component of big data [1]. The evolution of Internet technologies has been continuously supported by the advancement of information technology. Numerous data resources, such as pertinent standard specifications, technical and product documentation, management documents, fault resolution records, etc., remain to be acquired by the energy sector [2]. Enhancing power grid operating safety and stability for incorporating the long-term forecasting of alternative energies into power grid dispatching is necessary [3]. Energy is an alternative resource with great potential and a rapidly expanding global capacity. One of the main applications of energy is power generation. Considering that energy can be inconsistent and volatile, electrical grid stability and safety are of important concern [4]. The power system is an essential component of infrastructure for advancing the social and economic development of the nation. The present grid composition remains more complex, and the globalization of a local disturbance impact is becoming more prominent as ultra-high-voltage grids, extensive grid interactions, and grid coupling continue to be established constantly [5]. A significant quantity of power dispatch communication has been obtained as a consequence of the rapid technological advances and the comprehensive development of smart power grids. Designing a power dispatch knowledge visualization system and acquiring valuable information from relevant sources are essential for increasing team efficiency and assisting professionals in formulating decisions on power dispatch [6]. Energy sources, particularly solar and wind, have been actively developed in recent years, and installed capacity has increased quickly. However, solar power's intermittency and unpredictability make power grid scheduling more challenging, while wind power output's uncertainty and instantaneous volatility have a significant impact on the quality of energy and the power grid's stable operation [7]. Inadequate optimization could affect the model's overall performance, hyperparameter adjustments are complex, and the network's training time is delayed. The majority of the research presently in production applies incorrect data-processing techniques for important factors or fails to adequately take into consideration the factors influencing the immediate demand for regional power grids [8].

1.1 Problem statement

The increasing complexity of power grids challenges traditional methods for dispatchers, who struggle to handle vast, unstructured data and dynamic operational demands. Existing tools lack effective integration of domain-specific knowledge, leading to limited decision-making support and preparedness. This gap reduces dispatch efficiency and risks grid stability, especially during emergencies. There is an urgent need for advanced training solutions using fine-tuned large language models to improve dispatcher skills, enhance real-time decision accuracy, and facilitate better human-machine collaboration in power grid management.

1.2 Objective and contributions of this research

Investigating the use of optimized Large Language Models (LLMs) to improve power grid dispatchers' internal training is the intention of this research. Through domain-specific data adaptation of the Archerfish Hunting Fine-tuned Span Bidirectional Encoder Representations from Transformers (AH-SpanBERT) model, it enhances dispatcher decision-making and operational training. The objective of the research is to improve dispatchers' scenario-based learning by assessing essential instructional factors like accuracy, consistency, safety, and adaptability.

- To improve internal training procedures for power grid dispatchers, this research explores the use of domain-specific data to improve generalized LLMs.
- The power grid dispatcher operations dataset is to facilitate the creation and optimization of LLMs developed specifically for power grid dispatchers' internal training.
- To preprocess the obtained data, tokenization and domain-specific term normalization techniques are employed to provide consistency and contextual relevance.
- An extensive range of power system operational activities and decision-making scenarios, including operation adjustment, operation monitoring, and black start procedures, are supported by the proposed AH-SpanBERT model.
- The proposed method provides superior performance in the application of general large model fine-tuning technology of natural language in the internal training of power grid dispatchers.

2 Related work

The impact of the adoption of electric vehicles (EVs) on Italy's national power grid, with a particular emphasis on distribution and transmission systems, was described [9]. It concludes that curtailment of renewable energy, dispatching expenses, and grid breaches can all be considerably decreased by smart charging. However, there are drawbacks, such as the influence of localized distribution networks and the unpredictability of EV charging behavior. To develop a medical waste plasma hybrid peak load system for coal-fired power units, integrating syngas production, gas turbines, and renewable energy sources was described [10]. It finds that the system enhances energy efficiency and reduces operational costs, with a 6.20-year payback period. Limitations include the system's relatively low energy and exergy efficiency (37.38% and 36.19%, respectively). An optimization framework for a PV-grid-integrated EVCS with battery storage and peer-to-peer charging strategies was determined [11]. The model focuses on minimizing operational costs, ensuring reliability, and enhancing profitability. Simulation results show a reduction in energy demand costs and a reduction in maximum demand. However, the model is sensitive to daily varying weather and load conditions, which may affect long-term performance. A low-carbon economic dispatching strategy using a feasible region (FR) model to manage the interaction between wind power (WP), energy storage (ES), and carbon capture power plants (CCPP) was proposed [12]. The model reduces carbon emissions by promoting WP consumption while addressing its uncertainty. Simulation results on IEEE 39-bus and 118-bus test cases confirm its effectiveness. However, the model's reliance on robust optimization and the complexity of the column constraint generation algorithm may limit scalability.

According to the programming, the investigation was limited in carrying out the operation in an average layer, as it lacks autonomous ability. To recognize nested named entities in the power dispatching domain, a Robustly Optimized BERT (RoBERTa)-Attention-FL model was suggested [13]. The findings showed that the RoBERTa-Attention-FL model, with a higher accuracy rate, enhanced the recognition performance when compared to the baseline model. The suggested method needs to recognize named entities and assess the relationship between named entity identification and relationship extraction for optimal use of potential knowledge in the field of grid dispatching. The two-stage, data-driven deep learning approach for ultra-short-term photovoltaic (PV) electricity forecasting was presented [14]. The Bidirectional Gating Recurrent Unit (BiGRU) and the skip connection were utilized in a bidirectional recurrent neural network for many historical states that captured the long- and short-time sequences of PV sequences. The attention mechanism enabled the neural network to contribute adaptive importance to more relevant historical states. The results showed that the developed approach was capable of predicting PV power reasonably well for short-range immediate forecasts. Short-term consistency and long-term frequency were removed with the unsuitable skip connection, along with a suitable approach that could not effectively train the linear regularity. A microgrid scheduling model that addresses the economic and environmental costs of microgrid schedules utilizing accurate forecasting of photovoltaic (PV) power generation was presented [15]. The proposed model incorporates a combined Sparrow Search Algorithm (SSA)-Convolutional Neural Network (CNN)-Bidirectional Long Short-Term Memory (Bi-LSTM) prediction model with attention mechanisms, and a more enhanced Quantum Particle Swarm Optimization (QPSO) algorithm to optimize dispatch. Together, the model produces high prediction accuracy and maintains stabilization of the microgrid. To improve the flexibility and stability of the complicated and frequently large-scale power systems to control properly, the Distributed AI (DAI) framework was used [16]. The results highlighted the significance and possible advantages of the suggested framework in maintaining the reliable and efficient performance of power systems. Using the system manager's monitoring, the suggested technique needs to provide smooth power distribution optimization throughout the nano-grid that greatly improves the smart grid management systems. A fault tracing method using data partition hybrid sampling and multiple incremental regression tree algorithms to improve power grid fault detection was proposed [17]. By combining anomaly detection, clustering, and information difference models, the method achieves high precision and efficiency, enhancing grid maintenance and safety. A system based on LLMs that was presented for developing a domain-specific language for urban power grid architecture was established [18]. The creation of semantically intelligent systems for smart urban power grid design was supported by expert validation that demonstrated an accuracy rate of 89.3%. The results represented a significant practical application value. Semantic depth knowledge problems were highlighted, such as the current approach of assessing alternatives, which mostly relies on word vector similarity and manual tests. To detect abnormal users in smart grids by combining Bidirectional Long Short-Term Memory (BiLSTM) and Convolutional Neural Networks (CNN) for feature extraction, followed by Adaptive Boosting (AdaBoost) for classification, aimed [19]. The dataset consists of power consumption data from a small substation. The method outperforms individual models but faces challenges in scalability and real-time deployment across larger grids. The Dynamic Black Hole-driven Deep Convolutional Generative Adversarial Network (DBH-DCGAN) to address limitations in traditional power system monitoring, enhancing real-time equipment status and operational adaptability, was proposed [20]. The method utilizes dynamic adjustments to improve model stability and flexibility. A large set of pre-processed power equipment images was used for evaluation, showing significant improvements in monitoring accuracy across various operating conditions. The DBH-DCGAN method achieved high recall, accuracy, and F1-score, demonstrating its effectiveness in power plant monitoring and advancing intelligent grid management. However, challenges remain in adapting the method to highly diverse real-world conditions and scaling for large systems. A framework for classifying renewable energy sources using a freely available multivariate time-series dataset, with data on solar, wind, and hydro, was proposed [21]. The research analyzed a range of models: Logistic Regression (LR), Support Vector Machine (SVM), XGBoost, Artificial Neural Networks (ANN), and 1D Convolutional

Neural Networks (1D-CNN). A hybrid model that incorporates an attention mechanism with the 1D-CNN to not only improve feature extraction but also concentrate on temporal patterns of interest. The attention-attuned model proved to be the highest performer in its classification ability, with elevated metrics. However, further work is required to improve model generalisation across different energy sources and operational conditions. A comparative summary table showing dataset sizes, tasks, metrics used, and results of prior models is shown in Table 1.

Table 1: Summary table

Model	Methodology	Dataset/Scope	Key Task	Performance Results	Limitations
EV-Grid Integration Evaluation [9]	Smart Charging Impact on Distribution and Transmission Systems	EV adoption in Italy's national grid	Evaluate EV's effect on the distribution network and grid dispatching	Reduces curtailment of renewable energy, dispatching expenses, and grid breaches	Influence of localized networks and unpredictability in EV charging behavior
Medical Waste Plasma Hybrid Peak Load System [10]	Hybrid system integrating syngas production, gas turbines, and renewable energy sources	Coal-fired power units	Energy efficiency and operational cost reduction	Enhances energy efficiency, 6.20-year payback period	Low energy and exergy efficiency (37.38% and 36.19%)
PV-Grid-Integrated EVCS with Battery Storage [11]	Optimization framework for minimizing operational costs	PV-grid integrated EV charging stations	Minimize operational costs, ensure reliability, and enhance profitability	Reduction in energy demand costs and maximum demand	Sensitive to daily weather and load conditions, which affect long-term performance
Low-Carbon Economic Dispatching [12]	Feasible region model to manage wind power, energy storage, and carbon capture	IEEE 39-bus and 118-bus test cases	Economic dispatching with wind power and carbon capture	Reduces carbon emissions, confirms effectiveness in test cases	Complex column constraint generation, limiting scalability
RoBERTa-Attention-FL for Named Entity Recognition [13]	RoBERTa with Attention-FL for nested named entity recognition	Power dispatch domain	Named entity recognition	Improved accuracy in entity recognition	Needs better relationship extraction and entity identification
BiGRU + Attention Mechanism for PV Forecasting [14]	BiGRU with attention mechanism for ultra-short-term PV forecasting	PV data for short-term forecasting	PV power forecasting for short-range immediate forecast	An accurate short-term forecast captures long and short time sequences	Issues with consistency and long-term frequency in predictions
Microgrid Scheduling with SSA-CNN-Bi-LSTM [15]	SSA-CNN-Bi-LSTM with QPSO for optimizing microgrid dispatching	Microgrid with PV power generation	Economic and environmental cost optimization	High prediction accuracy, stable microgrid performance	Needs more robust optimization techniques for varying conditions

Distributed AI Framework for Power Systems [16]	Distributed AI for large-scale power systems control	Smart grid and large-scale systems	Power system control and management	Improves flexibility and stability in power system control	Lack of real-time autonomous decision-making
Fault Detection with Hybrid Sampling [17]	Data partition, hybrid sampling, and incremental regression trees	Power grid fault data	Fault detection and grid maintenance	High precision and efficiency in fault detection	Complexity in handling large-scale datasets
LLMs for Urban Power Grid Architecture [18]	Large Language Models (LLMs) for developing a domain-specific language	Urban power grid design	Design of intelligent systems for urban grid architecture	89.3% accuracy in expert validation	Relies on word vector similarity and manual testing for semantic depth
GPT-4 [23]	Variable (General)	General language understanding, text generation	Factuality, Logicality, Stability	Factuality: 7.05, Logicality: 9.71, Stability: 7.52	Limited domain adaptation, no specific optimization for power grid tasks
GAIA-70B [23]	Power Grid Specific	Power grid management, load forecasting	Factuality, Logicality, Security	Factuality: 7.79, Logicality: 7.79, Stability: 8.64	Lack of real-time scenario adaptability, limited domain-specific vocabulary

2.1 Research gap:

Current models for grid optimization, energy forecasting, and AI-driven power dispatching, such as smart charging for EVs and RoBERTa-Attention-FL, face significant challenges in real-time adaptability, scalability, and semantic understanding. Existing solutions struggle to address localized network behaviors and the unpredictability of charging patterns, limiting their effectiveness in dynamic environments. Additionally, short-term forecasting models like BiGRU + Attention Mechanism lack consistency over the long term, while AI models show promise in entity recognition but fall short in relationship extraction and complex semantic processing. AH-SpanBERT overcomes these issues by leveraging advanced span-based entity recognition to handle power-specific vocabularies and dynamic relationships in grid dispatching. Its ability to recognize named entities and relationship extraction allows for accurate, scalable, and real-time decision-making. By combining DL techniques like BiGRU, CNN-BiGRU Hybrid, and QPSO, AH-SpanBERT enhances forecasting accuracy, grid stability, and autonomous operations in large-scale energy systems.

2.2 Research methodology

To improve the internal training procedures for power grid dispatchers, this research investigates the use of domain-specific knowledge to improve generic LLMs. For effective performance, the research explores the data collection process, preprocessing techniques like tokenization, along with domain-specific term normalization, and the proposed method applications more comprehensively. Figure 1 shows the process of research methodology.

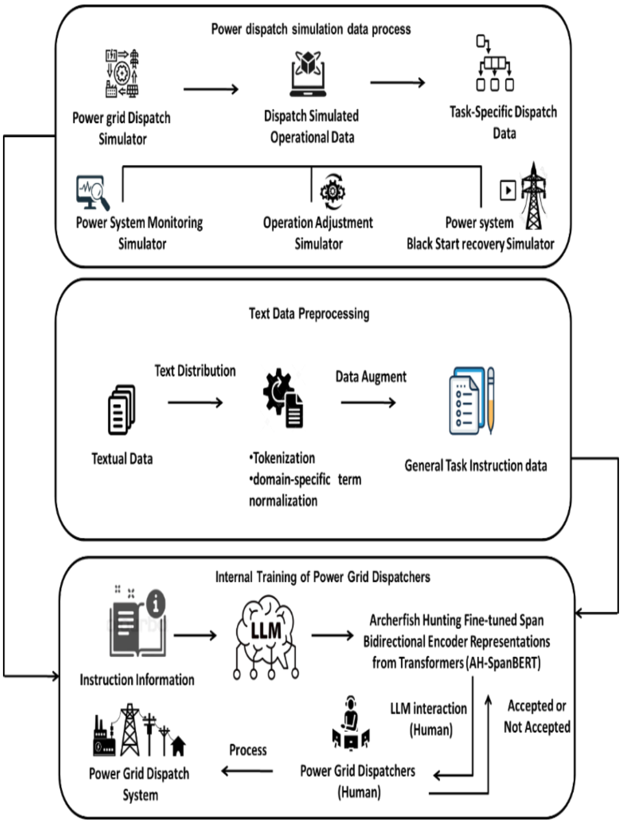

Figure 1: Methodology process involved with natural language in the internal training of power grid dispatchers

2.3 Data collection

The power grid dispatcher operations dataset is obtained from the open-source Kaggle [22]. The purpose of the dataset is to facilitate the creation and optimization of LLMs developed specifically for power grid dispatchers' internal training. The dataset contains 1,000 simulated functional records that span real-world situations like normal monitoring, emergency responses, equipment failures, and grid fluctuations. This dataset is beneficial for scenario-based LLM training, interactive dispatcher learning simulations, decision-making analysis under various complexity and urgency levels, along with collaborative human-machine modeling.

The data is split into training, validation, and test sets, in which 80% of the data will be used for training, 10% for validation, and 10% for testing. This split allows for proper evaluation of the model and prevents overfitting by ensuring that there is separate data to validate and test the model.

2.4 Data preprocessing through tokenization and domain-specific term normalization

Initial measures to clean and prepare unprocessed textual information for efficient performance are known as data pre-processing. There are two preprocessing techniques, such as tokenization and domain-specific term normalization, that are employed to preprocess the obtained data for the application of general large model fine-tuning technology of natural language in the internal training of power grid dispatchers.

2.4.1 Tokenization

Textual information is divided into meaningful components through proper tokenization, facilitating advanced modeling and computational processes. Text can be efficiently processed and comprehended by NLP algorithms by breaking the textual data into tokens, which leads to more precise and perceptive outcomes across a wide range of language processing applications. Modifying tokenization for domain-specific needs, including the needs of social media data, legal documentation, or research papers, is known as domain-specific tokenization.

2.4.2 Domain-specific term normalization

Impurity assessment is enhanced with the normalization step to distinguish between various document lengths and domain corpora sizes. The normalizing step and the impurity measure, as demonstrated by the research, make the technique more accurate in characterizing words for the application of general large-scale model fine-tuning technology of natural language.

2.5 Stemming: Stemming is a text preprocessing technique in NLP that reduces words to their root form by removing prefixes and suffixes. This process helps in simplifying words to a common base, making them more consistent and improving model accuracy by treating different forms of a word as equivalent. Algorithms like the Porter Stemmer and Snowball Stemmer are commonly used for this purpose. While stemming enhances data consistency and model performance, it may occasionally result in non-standard words, such as reducing "better" to "bet." Despite this, stemming is essential for improving tasks like text classification and information retrieval in NLP.

2.6 Power system operational tasks and decision-making scenarios development through archerfish hunting fine-tuned span bidirectional encoder representations from transformers (AH-SpanBERT)

The suggested Archerfish Hunting Fine-tuned Span Bidirectional Encoder Representations from Transformers (AH-SpanBERT) model supports a wide range of power system operational tasks and decision-making scenarios that integrate the Span Bidirectional Encoder Representations from Transformers (SpanBERT) and Archerfish Hunting (AH) optimization that enhances the internal training processes for power grid dispatchers. AH-SpanBERT is described in Figure 2**.**

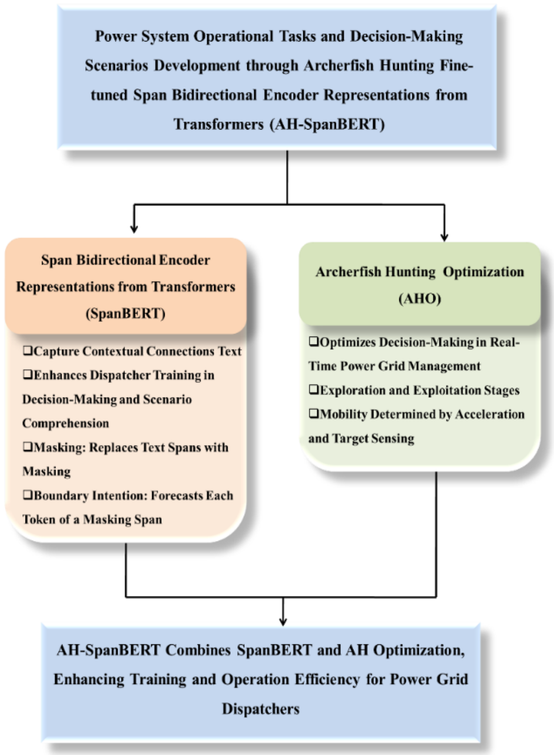

Figure 2: Flowchart of AH-SpanBERT

2.6.1 Span bidirectional encoder representations from transformers (SpanBERT)

A transformer-based model called SpanBERT predicts text spans and performs an exceptional task at capturing contextual connections in text. The technique is designed to comprehend domain-specific language, enhancing dispatcher training in decision-making and scenario comprehension. To improve the representation and prediction of text spans, a self-supervised pre-training technique called SpanBERT was introduced. One extended textual segment is sampled by SpanBERT for every training instance. The categories of SpanBERT processes, such as masking and boundary intention, are discussed.

Masking: This model iteratively samples text spans until the masking resource is exhausted, selecting a subset of tokens $B \subseteq A$ from a series of tokens $A = (a_1, a_2, \dots, a_N)$. The method initiates iteration by sampling a span length (the number of words) from a geometrical distribution$\ell \sim GEO(q)$, which tends to benefit from shorter spans. Next, uniformly and randomly chooses an initial point for the masking of the span. Instead of using subword tokens, always samples a sequence of whole words, and the initial position should be the first character of a single word. The technique replaces all of the tokens in a span with masking, and it is performed at the span level.

Boundary Intention: The span selection approaches engage with a span's boundary tokens to provide a fixed-length representation of the span. The representations at the end of the span can ideally compress the majority of the inside span information. To accomplish the function, SpanBERT introduces a span boundaries target that uses the representations of the observation tokens at the boundaries to forecast each token of a masking span.

Representing the transformer encoder's output in each word of the sequence is represented by $a_1, \dots, a_N$. Using the output encodings of the outer boundary tokens a_{t-1} anda_{f+1}, along with the position encoding of the target token q_{j-t+1}, it represents each token a_j in the masked span of tokens $(a_t, \dots, a_f) \in B$, where (t, f) denotes its beginning and the end positions in Eq. (1). It allows for a real-time operational scenario to be built, and helps dispatchers make informed decisions about system performance, such as response to grid changes, equipment failures, and energy demand changes, improving training and operation efficiency.

$$b_j = f\left(a_{t-1}, a_{f+1}, q_{j-t+1}\right) \tag{1}$$

Where the masked words' positional relationships to the left boundary word a_{t-1} are indicated by positional embeddings$q_1, q_1, \dots, etc.$ For every token a_j in the masked span $(a_t, \dots, a_f)$, SpanBERT adds the loss from the span border and the regular masked language model desired outcomes, utilizing the input embedded for the target tokens described in Eq. (2).

$$\mathcal{L}(a_j) = \mathcal{L}_1(a_j) + \mathcal{L}_2(a_j) = -\log Q\left(a_j \middle| \boldsymbol{a}_j\right) - \log Q\left(a_j \middle| \boldsymbol{b}_j\right) \tag{2}$$

Where, $\mathcal{L}(a_j)$ measures the prediction error of a_j based on its history, and $\mathcal{L}_2(a_j)$measure the error based on the related variable $\boldsymbol{b}_j$. The likelihood $\left(a_j \middle| \boldsymbol{a}_j\right)$ and $Q\left(a_j \middle| \boldsymbol{b}_j\right)$ assess prediction accuracy for both self-prediction and cross-variable relationships. This loss function could be applied to minimize the difference between actual and predicted outcomes, such as grid behaviors or operational decisions, and to increase the accuracy of the models for both the historical context and future predictions.

SpanBERT loss function Equation 2 is indeed the probability of predicting a_j given itself; this term would ideally be zero, as the model would predict a_j perfectly, making the likelihood zero abd contributing nothing to the loss. This is not consistent with the typical loss components used with SpanBERT-type models. The normal SpanBERT loss function is the combination of two objectives: the Masked Language Modeling (MLM) loss, which is concerned with predicting masked tokens, and the Span Boundary Objective (SBO) loss, which ensures span boundaries are predicted better. The equation in question does not capture either of these two components or the appropriate math form for either. A fix would involve properly combining MLM loss and SBO loss to be consistent with established formulations, to include both span prediction and token prediction objectives in the loss function. To pre-train span representations, SpanBERT uses a geometric distribution-based masking approach that masks full-word spans and a single-sequence data flow to optimize a supporting span-boundary achievement.

2.6.2 Archerfish hunting (AH) optimization

The AH optimization enhances LLM training for operational responsibilities in complicated, real-time power grid management by optimizing decision-making through the simulation of environment-specific problem-solving. The proposed AH's exploration and exploitation stages depend on the hunting and jumping behaviors of archerfish. Any optimization issue can be resolved using AH, provided that the power grid management is properly formulated. The AH optimization was selected for its biologically inspired exploration–exploitation balance and suitability for nonlinear, high-dimensional scenarios like power grid dispatcher simulation. AH demonstrated improved convergence behavior and enhanced training dynamics in the AH-SpanBERT model with standard optimizers such as Adam and Particle Swarm Optimization (PSO) under identical training conditions. This will help quantify AH's performance advantage in terms of model accuracy, convergence rate, and decision quality in scenario-based training tasks. Assume that there are many archerfish in a search area of dimensions D. Archerfish location j at iteration s is as follows (Eq. (3)), and the population size is$n$.

$$A^{j,s} = (a_1, a_2, \dots, a_D) \tag{3}$$

Using Eq. (4), the point $A^{\langle j,0\rangle}$ is initialized at random with repetitions$s = 0$.

$$A^{\langle j,0\rangle} = \left(\alpha_1 \times \left(a_1^{maxi} - a_1^{mini}\right) + a_1^{mini}, \dots, \alpha_D \times \left(a_D^{maxi} - a_D^{mini}\right) + a_D^{mini}\right) \tag{4}$$

$A^{\langle j,0\rangle}$ is the initial position vector, α_1 is the random multiplier, and a uniform distribution. a_k^{maxi} and a_k^{mini} are upper and lower bounds, D defining the complexity of the optimization problem. Equation (4) initializes agent positions in the AH optimization algorithm and creates different places to begin within the defined bounds. The issues of starting position support the search space for optimization of energy grids, particularly for supporting task schedules and the overall efficiency of operations, ensuring a reachable solution as it relates to decision-making in active or dynamic grid management. AH initialization formula raises a valid concern regarding the ambiguity of the notation. The variables $(a_1, a_2, \dots, a_D)$ are described as "uniformly distributed random integers between o and 1," but the formula implies these values might be sequential components of the single random vector, which could lead to confusion. In standard optimization algorithms, each dimension is typically initialized with independent random values. Additionally, the notation a_1^{maxi} and a_D^{mini}, with the subscript i, is unclear regarding whether the bounds are shared across all dimensions or specific to each one.

Where the uniformly distributed random integers between 0 and 1 are denoted by$\alpha_1, \dots, \alpha_D$. An archerfish uses Eq. (5) to travel in the direction of a target when it identifies the generation of the vibration.

$$A^{\langle j,s+1\rangle} = A^{\langle j,s\rangle} + f^{-\left\|A_{prey}^{\langle l,s\rangle} - A^{\langle j,s\rangle}\right\|^2}\left(A_{prey}^{\langle l,s\rangle} - A^{\langle j,s\rangle}\right) \tag{5}$$

The position of the prey is determined in Eq. (6). The archerfish's mobility is determined by its acceleration of gravity (g), launch speed (v), and sensing angle (θ0), while the air friction is minimal. The desired outcome is assumed to be at the peak of the motion visualization. Using Eq. (7), an archerfish moves in the direction of the target that intends to capture.

$$A_{prey}^{\langle l,s\rangle} = A^{\langle j,s\rangle} + \left(0, \dots, \frac{u^2}{2Gr} \times \sin 2\theta_0, \dots, 0\right) + \varepsilon \tag{6}$$

$A^{\langle j,s+1\rangle}$ is the updated position of the j-th archerfish, $A^{\langle j,s\rangle}$ is the current position of the jth archerfish at iteration T. $A_{prey}^{\langle l,s\rangle}$ is the position of the prey at the sth iteration for the lth archerfish. e is the scaling factor that adjusts the magnitude of the movement based on the distance between the archerfish and its prey, $A_{prey}^{\langle l,s\rangle} - A^{\langle j,s\rangle}$ is the squared Euclidean diatance between the archerfish and he prey, ε is the small random perturbation, v is the lunch speed of the archerfish, g is the acceleration due to gravity, θ_0 is the sensing angle of the archerfish. The motion of an archerfish can be characterized by these equations as part of the AH optimization algorithm. The update process has two main components. First, the prey is updated via a physics formulation (Equation 6) based on its speed, gravitational pull, and the angle of the prey based on the archerfish's three-dimensional surface imaging.

$New\_pos = current\_pos + (target\_pos - current\_pos).\, e^{-||target_{pos} - current\_pos}$ was Euclidean distance. However, it is more closely representative of physics-inspired metaheuristics (e.g., Gravitational Search Algorithms) than AH's original behavior based on a jump-based ballistic mechanism. To more closely align with the AH concept, this function must incorporate: directional jumps influenced by θ_0 projectile arc displacement perturbations ϵ. Moving forward, the updated algorithm provides an updated formulation, using a hybrid of physics-based trajectory (Equation 6) and normalized directional movement.

This must also be appropriately buffered by parameter θ_0. The perceiving angle value (θ_0) makes the exploration and exploiting phases switch. Therefore, the more effectively AH will utilize the search space when the value of θ_0 is to $\frac{\pi}{2}$or$-\frac{\pi}{2}$, and vice versa. Eq. (7) is used to produce the value of θ_0 at random.

$$\theta_0 = (-1)^y \times \alpha \times \pi \tag{7}$$

θ_0 is the perceiving angle controlling exploration, y is the random variable for alternating sign, α is the scaling factor for angle magnitude. π is the mathematical constant for angle bounds. Equation (7) was used to create a method for generating the perceiving angle θ_0 randomly, which is important in AH optimization to control both exploration and exploitation. By selecting the value of the perceptive angle randomly, ensures that the archerfish will alternate between exploring the search space (θ_0 = π/2 or θ_0 = -π/2) and exploiting areas near a target (indicated by smaller values of θ_0). The concerns with the implementation of θ_0 logic and its effect on whatever exploration/exploitation is currently being done. The current implementation of θ_0, where randomly selected to be between π/2 and -π/2 is too limiting on the flipping between an exploration phase and an exploitation phase as designed from the formula in Equation (7). Instead of having a plan of action where consideration of exploration is always occurring and the exploitation logic in the $explore\_exploit$ function is not able to be reached, to modify the code to allow θ_0 more flexibility in terms of how many values it take on a variety of angles, while simultaneously providing behavior that enacts both exploration and exploitation, making use of values controlled by θ_0, as intended for AH optimization.

The current implementation always assigns $\theta_0\ as\ \pm\pi/2$, making the condition if $abs(\theta_0)\ ==\ \pi/2$ always true and rendering the else branch redundant. To address this, will introduce continuous sampling of $\theta_0 \in [-\pi/2, \pi/2]\ using\ \theta_0 = (-1)^\wedge rand \times \alpha \times \pi$, allowing dynamic control via scaling factor α. The $explore\_exploit()$ function then probabilistically switches between exploration and exploitation based on whether $|\theta_0| > \pi/4$. The model, as it is currently set, exists to log $updated\_position$, but not to affect the model training of SpanBERT. The next iteration of AH represented the positions generated as hyperparameter vectors (learning rate, masking ratio, frozen layers, batch size) to fine-tune the model. The hyperparameters are decoded and used from the function $fine\_tune\_with\_config()$. In this version, AH was able to know how to proceed conservatively with a lower learning rate, masking ratio, and batch size depending on the scenario being suited. This marks the beginning of a closed-loop system with AH insisting on the tuning of SpanBERT training creation and SpanBERT creating feedback for AH's directed search through a performance-based fitness score.

To prevent getting trapped in localized optimums, AH optimization employs a conventional approach. Algorithm 1 shows the working procedure of the proposed AH-SpanBERT model.

Algorithm 1: AH-SpanBERT

```
Import numpy as np
Import random
Class SpanBERT:
def __init__(self, model_name):
self.model_name = model_name
def preprocess(self, data):
    Return preprocess_data(data)
deffine_tune(self, data):
    Return fine_tune_model(data)
def predict(self, input_text):
    Return self.model.predict(input_text)
Class ArcherfishHunting:
def __init__(self, dimensions, population_size):
self.dimensions = dimensions
self.population_size = population_size
self.position = np.random.rand(population_size, dimensions)
defupdate_position(self, current_position, target_position): Return current_position+(target_position-current_position)*np.exp(-np.linalg.norm(target_position - current_position))
defexplore_exploit(self, position, prey_position, theta_0):
if abs(θ_0) == np.pi / 2:
     Return self.update_position(position, prey_position)
else:
     Return position
def simulate(self):
for _ in range(self.population_size):
prey_position = self.position[random.randint(0, self.population_size - 1)]
target_position = self.position[random.randint(0, self.population_size - 1)]
θ_0= random.choice([-np.pi / 2, np.pi / 2])
new_position = self.explore_exploit(self.position, prey_position, θ_0)
self.position = new_position
    Return self. position
Class AHSpanBERTModel:
def __init__(self, span_bert_model, archerfish_model):
self.span_bert = span_bert_model
self.ah_optimizer = archerfish_model
defoptimize_training(self, training_data, scenario_data):
fine_tuned_model = self.span_bert.fine_tune(training_data)
training_scenarios = generate_scenarios(scenario_data)
for scenario in training_scenarios:
updated_position = self.ah_optimizer.simulate()
predictions = self.span_bert.predict(scenario['input'])
scenario['actions'] = predictions
scenario['updated_position'] = updated_position
    Return training_scenarios
```

Through the simulation of domain-specific operational situations, such as task management, system monitoring, and emergency protocols, the AH-SpanBERT model is optimized to improve power grid dispatcher training. The proposed technique makes interactive, scenario-based learning, which enhances flexibility, knowledge retention, and decision-making. The approach facilitates context-aware, real-time responses, improving human-machine cooperation in power grid systems and dispatcher performance across a range of operational tasks.

3 Results and discussions

This section deliberates on the results produced by the implementation of the model, including parameter setup, evaluation criteria, and comparative phase.

3.1 Experimental setup

For fast processing and effective multitasking, the research makes use of an AMD Ryzen 7 5800X CPU and 32GB DDR4 memory. Graphics-intensive operations are managed by the NVIDIA GeForce RTX 3080 GPU. For GPU computing performance and scalability, the system runs on Windows 11 and uses Python 3.9, scikit-learn, and TensorFlow libraries

3.2 Parameters setup

AH-SpanBERT hyperparameters as described in Table 2.

Table 2: Parameter setup

Hyperparameter	Value
Learning Rate	0.0001
Epochs	50
Batch Size	32
Layers Frozen	The first few layers of SpanBERT model are frozen to maintain the pre-trained knowledge.
Loss Function	Combination of Masked Language Modeling (MLM) loss and Span Boundary Objective (SBO) loss
Loss Function Values	Regularization values of 0.5 for both MLM and SBO losses

3.3 Evaluation criteria

The evaluation criteria demonstrate the cost consumption of the proposed AH-SpanBERT technique's efficiency in the internal training of power grid dispatchers, and resulted in the 0 to 100 iterations & 0 to 500 iterations (Figure 3 a-b). The 100-iteration run is effective for quicker results because it quickly converges and maintains performance. Over time, better optimization could become possible due to the 500-iteration run, which allows for additional exploration and improvement. The AH-SpanBERT method shows significant outcomes in the cost function in that provides less cost consumption in performing the power grid dispatcher's internal training process.

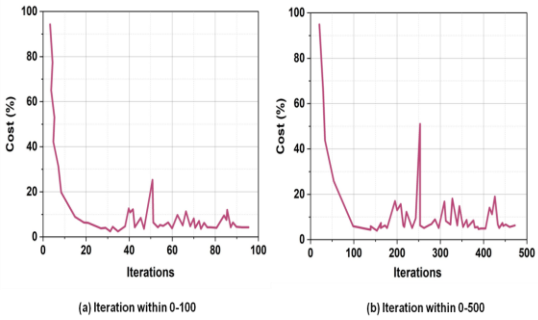

Figure 3: Outcomes of cost assessment with (a) iteration within 0-100 and (b) iteration within 0-500

3.4 Comparison phase

The proposed AH-SpanBERT and the existing models such as Generative Pre-trained Transformer 4 (GPT-4) [23] and Grid Artificial Intelligent Assistant (GAIA-70B) [23] are compared for providing an effective performance in internal training of power grid dispatchers along with various parameters like factuality, logicality, stability and security for various scenarios like general, dispatch, operation monitoring and black start.

- ***Factuality*** indicates that the information is valid and the findings accurately reflect the actual circumstances, which is crucial for valid decision-making in the operation of power systems.
- ***Logicality*** is concerned with the precision of logical inference and the dependability of the data utilized in the enhancement of power grid dispatcher learning assessment.
- ***Stability*** maintains operational continuity and reliability. The parameter also assesses LLMs' capacity to sustain similar outputs in dynamic situations.
- ***Security*** indicates the significance of ensuring that operational security provides model applications that can never damage the power system's security.

The evaluation metrics (factuality, logicality, stability, and security) scaled from 0 to 10, were adapted from the Elecbench benchmark (Zhou et al., 2024), which provides a standardized framework for assessing LLM performance in power dispatch tasks. The scoring system follows Elecbench's normalized evaluation procedure, where expert raters assess model responses based on alignment with ground truth, operational logic, and scenario safety. Each metric was averaged over multiple annotated responses across four scenarios to ensure consistency and comparability. This alignment ensures that our reported results are interpretable within the context of existing power grid LLM evaluation standards.

The purpose of the general scenario design is to assess the LLM's proficiency in handling data analysis, forecasting, and basic knowledge question-and-answer tasks related to day-to-day power system management. AH-SpanBERT achieves high scores in factuality (8.32), ensuring accurate information delivery. Logicality is very high at 9.74, supporting sound reasoning. Security (9.38) and stability (8.76) confirm the model's reliability and safety in everyday power system management and training tasks. Table 2 and Figure 4 represent the numerical outcomes of factuality in four scenarios.

Table 2: Numerical outcomes of General power systemwith four scenarios

Models	General power system			
	Factuality	**Logicality**	**Security**	**Stability**
GPT-4 [23]	7.05	9.71	9.28	7.52
GAIA-70B [23]	7.79	7.79	9.29	8.64
AH-SpanBERT [Proposed]	8.32	9.74	9.38	8.76

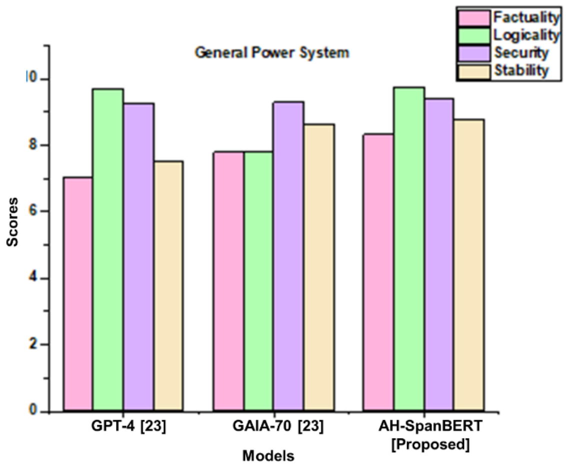

Figure 4: Graphical illustration of factuality results in general scenarios.

The dispatch sub-scenario is essential for assessing the efficiency of LLMs data analysis, forecasting, and optimal decision-making, all of which are vital for improving the stability and efficiency of the sector. In dispatch operations, AH-SpanBERT maintains solid factuality (7.59) and logicality (9.37), facilitating optimal decision-making. The security score of 9.40 indicates that the model safeguards operational safety, while stability at 8.15 reflects consistent output, essential for reliable performance in dynamic dispatch scenarios. Mathematical results of four scenarios in logicality are provided in Table 3 and Figure 5.

Table 3: Mathematical results of dispatch.

Models	Dispatch			
	Factuality	**Logicality**	**Security**	**Stability**
GPT-4 [23]	7.42	9.30	9.35	8.07
GAIA-70B [23]	7.45	7.52	9.37	7.44
AH-SpanBERT [Proposed]	7.59	9.37	9.40	8.15

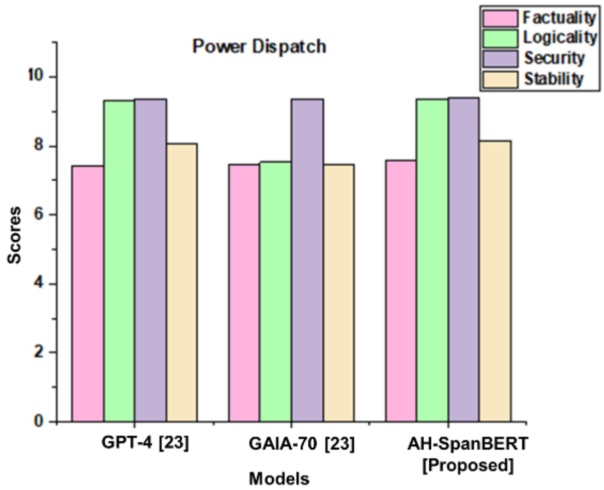

Figure 5: Mathematical outcomes of logicality with dispatch

Operation monitoring places significant demands on data speed and accuracy in the design of the assessment framework, requiring LLMs to have strong data gathering and processing skills along with the capacity to promptly recognize and address possible problems. For operation monitoring, AH-SpanBERT delivers high factuality (8.48) and logicality (8.97), demonstrating effective data interpretation and problem identification. The security score of 9.39 ensures safe application, while stability (7.14) remains adequate, supporting dependable performance despite real-time monitoring issues. Figure 6 and Table 4 show the outcomes of operation monitoring.

Table 4: Operation monitoring outcomes with several parameters

Models	Operation Monitoring			
	Factuality	Logicality	Security	Stability
GPT-4 [23]	8.21	8.92	9.10	6.16
GAIA-70B [23]	8.42	7.78	9.35	7.00
AH-SpanBERT [Proposed]	8.48	8.97	9.39	7.14

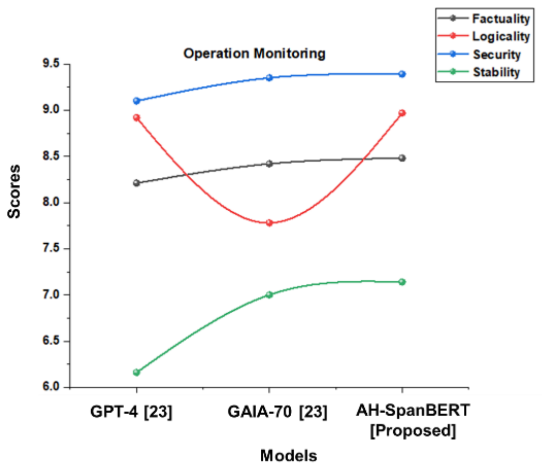

Figure 6: Visual depiction of operation monitoring

The accurate assessment of LLMs in this context is crucial for comprehending their practical utility and potential in identifying faults and recovery procedures, as Black Start is necessary to preserve power system continuity and avert large-scale grid failures. During black start procedures, AH-SpanBERT achieves high factuality (8.43) and locality (8.99), supporting accurate fault recovery decisions. This model scores highest in security (9.62), and stability (9.15), emphasizing the model's robustness and reliability in critical emergency power system restoration scenarios. Table 5 and Figure 7 determine the algorithmic outcomes of black start.

Table 5: Algorithmic outcomes of black start

Models	Black Start			
	Factuality	Logicality	Security	Stability
GPT-4 [23]	8.39	8.84	9.53	9.03
GAIA-70B [23]	7.94	7.85	9.50	7.28
AH-SpanBERT [Proposed]	8.43	8.89	9.62	9.15

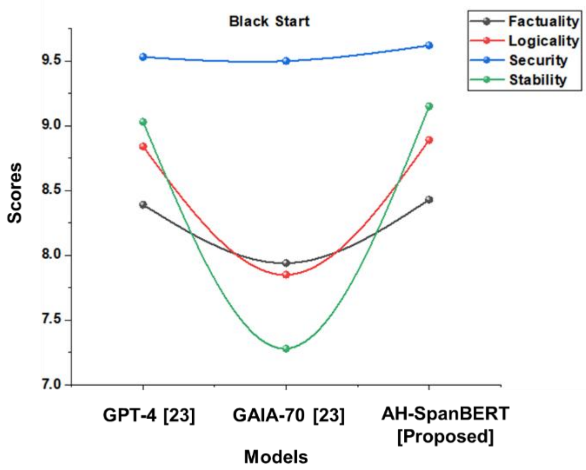

Figure 7: Results of the black start scenario with various performance matrices

Table 5 shows the standard deviation values for four metrics (Factuality, Logicality, Security, and Stability) applied to the various power system tasks using the AH-SpanBERT Model. The standard deviation for the four metrics is observed between a range of 0.05 to 0.11. This indicates slight differences in model performance across the power system tasks. Factuality displays little variation in its standard deviations (0.06 to 0.10); Logicality had slightly more variation, especially in the General Power System and Dispatch tasks (0.07 – 0.09). Security had a higher standard deviation observable during the Black Start task (0.10), delineating there were slight fluctuations in security predictions in this task. Stability shows the highest standard deviations in two tasks, General Power System and Dispatch (0.10 and 0.11, respectively) and indicate less stable predictions in these two tasks compared to the other tasks. Overall, the model performance shows little variation, although these particular tasks (Dispatch and General Power System) may vary more in some metrics.

Table 5: Standard deviation values

Metric	Standard Deviation – AH-SpanBERT [Proposed]			
	General power system	Dispatch	Black Start	Operation Monitoring
Factuality	0.07	0.10	0.07	0.06
Logicality	0.09	0.07	0.06	0.08
Security	0.08	0.06	0.10	0.05
Stability	0.10	0.11	0.09	0.07

Table 6 indicates the performance metrics (Accuracy, Precision, Recall, F1-Score) of a model across 5 folds. The model's performance is consistently strong, with accuracy ranging from 0.90 to 0.99 (mean = 0.97), and high precision (range 0.90 - 0.99; mean = 0.97) indicates positive predictions are generally correct. Recall ranges from 0.90 - 0.99 (mean = 0.97), confirming the model's effectiveness in targeting positive instances. Finally, F1-Score also exhibits the same pattern of reliability (range 0.90 - 0.99; mean 0.97) as well as consistency across folds.

Table 6: Results of 5-Fold Cross-Validation

Fold	Accuracy	Precision	Recall	F1-Score
1	0.99	0.97	0.90	0.99
2	0.91	0.99	0.92	0.90
3	0.99	0.96	0.99	0.97
4	0.90	0.99	0.91	0.99
5	0.92	0.90	0.93	0.91
Average	0.97	0.98	0.97	0.97

4 Discussion

The internal training of power grid dispatches, enhanced by leveraging fine-tuning of general LLMs with domain-specific data, was the main objective of the research. Existing methods, including traditional training approaches and generic LLMs like GPT-4 [23] and GAIA-70B [23], have limitations such as insufficient domain adaptation, lack of scenario-specific understanding, and inconsistent performance in critical operational tasks. By applying specialist knowledge and optimizing a dataset of power grid operational data, the proposed AH-SpanBERT model overcomes the shortcomings of power system models. The domain-aware adaptation enhances stability, security, logical thinking, and factual correctness in a variety of situations. By filling up the gaps in previous models and providing a strong tool for dispatcher training and power system management, AH-SpanBERT dramatically increases operational reliability, knowledge retention, and decision-making accuracy.

The AH-SpanBERT model, leveraging Span masking and AH optimization, outperforms existing models like GPT-4 and GAIA-70B in key areas such as factuality, logicality, stability, and security. Span masking enables more accurate contextual understanding of domain-specific terminology, which is crucial for power grid dispatchers, while AH optimization simulates environment-specific problem-solving to enhance decision-making. In performance metrics, AH-SpanBERT excels, with higher factuality, logicality, and security scores, particularly in critical tasks like operation monitoring and black start scenarios. Unlike GPT-4 and GAIA-70B, which lack domain-specific focus, AH-SpanBERT's fine-tuning on power grid data ensures more reliable and accurate performance in real-time decision-making. However, its reliance on historical data may limit adaptability to unforeseen grid changes, and in highly uncertain scenarios, AH optimization may occasionally lead to suboptimal outcomes. Overall, AH-SpanBERT demonstrates substantial improvements in power grid dispatcher training, addressing the shortcomings of previous models, though it still has room for enhancement in handling dynamic, real-time disruptions.

5 Conclusion

Power grid dispatchers play an essential role in maintaining the consistent, effective, and safe functioning of electrical networks. The use of domain-specific data to refine the generic LLMs to improve internal training procedures for power grid dispatchers was examined in the research. The AH-SpanBERT model was developed to facilitate black start processes, operation monitoring, and general dispatch, which were some of the many operating responsibilities and decision-making scenarios. An extensive dataset was assembled from several sources, including operational manuals, emergency procedures, internal communications, training materials, and historical dispatch logs. Utilizing tokenization and domain-specific word normalization, the data was preprocessed for contextual relevance and consistency. Rapid techniques were created to mimic authentic dispatch situations, allowing trainees to learn interactively through scenario-based learning. The model's power dispatch performance was assessed to evaluate LLMs on important parameters that were necessary for efficient power system management, including factuality (8.48 in operation monitoring), logicality (9.74 in general), stability (9.15 in black start), and security (9.62 in black start). Experimental findings showed significant improvement in factuality (8.48 in operation monitoring), logicality (9.74 in general), stability (9.15 in black start), and security (9.62 in black start) across dispatchers with different scenarios. In power dispatch operations, the research showed that the suggested General Large Model greatly improved human-machine interaction, operational efficiency, and decision-making capacities.

Limitations and future scopes: The use of a small, simulated Kaggle dataset (1,000 records) may limit the model's generalizability to real-world dispatcher logs, affecting performance under authentic, complex grid conditions. Reliance on historical information limits the investigation and could affect the model's capacity to adjust to changes in the power grid. Enhancing operational effectiveness and cooperation across various grid environments could expand the model to more power system management categories in the future. In future revisions, will explicitly outline risk mitigation strategies, such as human-in-the-loop validation, real-time alert prioritization, and fallback protocols. These ensure AI recommendations serve as augmentative tools, supporting but not overriding the dispatcher's expert decisions in high-stakes operational scenarios.

Acknowledgement: This work was supported by the Science and Technology Project of State Grid Sichuan Electric Power Company under Grant 52199723003A

References

[1] Wan W, Liu Y, Han X, Wang H (2021) Evaluation model of power operation and maintenance based on text emotion analysis. *Math Probl Eng* 2021:2824689. https://doi.org/10.1155/2021/2824689.

[2] Chen T, Zhang S, Wang Y, Chen Z, Jing W (2021) Construction methods of knowledge mapping for full-service power data semantic search system. *J Signal Process Syst* 93:275-284. https://doi.org/10.1007/s11265-020-01591-6

[3] Li X, Yang N, Li Z, Huang Y, Yuan Z, Song X, Zhang L (2022) Confidence estimation transformer for long-term renewable energy forecasting in reinforcement learning-based power grid dispatching. *CSEE J Power Energy Syst*. https://doi.org/10.17775/CSEEJPES.2022.02050

[4] Wang L, He Y, Li L, Liu X, Zhao Y (2022) A novel approach to ultra-short-term multi-step wind power predictions based on encoder–decoder architecture in natural language processing. *J Clean Prod* 354:131723. https://doi.org/10.1016/j.jclepro.2022.131723

[5] Ji Z, Wang X, Zhang J, Wu D (2023) Construction and application of knowledge graph for grid dispatch fault handling based on pre-trained model. *Glob Energy Intconnect* 6:493-504. https://doi.org/10.1016/j.gloei.2023.08.009

[6] Tan Z, Chen Y, Liang Z, Meng Q, Lin D (2023) Named Entity Recognition for Few-Shot Power Dispatch Based on Multi-Task. *Electronics* 12:3476. https://doi.org/10.3390/electronics12163476

[7] Dong Z, Zhang Z, Huang M, Yang S, Zhu J, Zhang M, Chen D (2024) Research on day-ahead optimal dispatching of virtual power plants considering the coordinated operation of diverse flexible loads and new energy. *Energy* 297:131235. https://doi.org/10.1016/j.energy.2024.131235

[8] Li B, Liao Y, Liu S, Liu C, Wu Z (2025) Research on Short-Term Load Forecasting of LSTM Regional Power Grid Based on Multi-Source Parameter Coupling. *Energies* 18:516. https://doi.org/10.3390/en18030516

[9] Rancilio G, Cortazzi A, Viganò G, Bovera F (2024) Assessing the nationwide benefits of vehicle–grid integration during distribution network planning and power system dispatching. *World Electric Vehicle* J 15:134. https://doi.org/10.3390/wevj15040134

[10] Zheng Q, Chen H, Wu H, Zheng H, Pan P, Bian J, Wang X (2024) Performance analysis of a medical waste gasification-based power generation system integrated with a coal-fired power unit considering dispatching optimization. *Energy* 310:133200. https://doi.org/10.1016/j.energy.2024.133200

[11] Bokopane L, Kusakana K, Vermaak H, Hohne A (2024) Optimal power dispatching for a grid-connected electric vehicle charging station microgrid with renewable energy, battery storage and peer-to-peer energy sharing. *J Energy Storage* 96:112435. https://doi.org/10.1016/j.est.2024.112435

[12] Yu G, Zhang Z, Cui G, Dong Q, Wang S, Li X, Shen L, Yan H (2024) Low-carbon economic dispatching strategy based on feasible region of cooperative interaction between wind-storage system and carbon capture power plant. *Renew Energy* 228:120706. https://doi.org/10.1016/j.renene.2024.120706

[13] Chen Y, Lin D, Meng Q, Liang Z, Tan Z (2023) Named entity identification in the power dispatch domain based on RoBERTa-attention-FL model. *Energies* 16:4654. https://doi.org/10.3390/en16124654

[14] Zhou H, Wang J, Ouyang F, Cui C, Li X (2023) A two-stage method for ultra-short-term PV power forecasting based on data-driven. *IEEE Access* 11:41175-41189. https://doi.org/10.1109/ACCESS.2023.3267515

[15] Zhang T, Zhao W, He Q, Xu J (2025) Optimization of microgrid dispatching by integrating photovoltaic power generation forecast. *Sustainability* 17:648. https://doi.org/10.3390/su17020648

[16] Ioannou II, Javaid S, Christophorou C, Vassiliou V, Pitsillides A, Tan Y (2024) A distributed AI framework for Nano-Grid power management and control. *IEEE Access* 12:43350-43377. https://doi.org/10.1109/ACCESS.2024.3377926

[17] Yang S, Fu Y, Li S (2024) Fault traceability of power grid dispatching system based on DPHS-MDS and LambdaMART algorithm. *Energy Informatics* 7:85. https://doi.org/10.1186/s42162-024-00391-7

[18] Xu Y, Wang T, Yuan Y, Huang Z, Chen X, Zhang B, Wang Z (2025) LLM-Enhanced Framework for Building Domain-Specific Lexicon for Urban Power Grid Design. *Appl Sci* 15:4134. https://doi.org/10.3390/app15084134

[19] Zhang Y, Gao Y, Zhao Z (2025) Research on operation and anomaly detection of smart power grid based on information technology using CNN+ Bidirectional LSTM. *Informatica* 49:7037. https://doi.org/10.31449/inf.v49i7.7037

[20] Liu J, Zhao W, Gu F (2025) Panoramic Intelligent Monitoring of Power Equipment Using Dynamic Black Hole-Driven DCGAN Under New Power Systems. *Informatica* 49:8037. https://doi.org/10.31449/inf.v49i8.

[21] Hong Q, Chen X, Zhai X, Sun X, Zong Y (2025) Hybrid Deep Learning-Based Renewable Energy Classification for Smart Grid Optimization. *Informatica* 49:8366. https://doi.org/10.31449/inf.v49i25.8366

[22] https://www.kaggle.com/datasets/ziya07/power-grid-dispatcher-operations-dataset/data

[23] Zhou X, Zhao H, Cheng Y, Cao Y, Liang G, Liu G, Zhao J (2024) Elecbench: A power dispatch evaluation benchmark for large language models. *arXiv:*2407.05365. https://doi.org/10.48550/arXiv.2407.05365