

# Research on Application Analysis and Architecture Design of Artificial Intelligence in Power Big Data

Xin He, Yangang Qu, Senda Zhang, Xin Cui, Weigang Yin

Beijing China-Power Information Technology Co., Ltd., Beijing, China  
Email: zhangsenda93@163.com

**How to cite this paper:** He, X., Qu, Y.G., Zhang, S.D., Cui, X. and Yin, W.G. (2025) Research on Application Analysis and Architecture Design of Artificial Intelligence in Power Big Data. *Journal of Computer and Communications*, 13, 160-170.  
<https://doi.org/10.4236/jcc.2025.137008>

**Received:** June 9, 2025

**Accepted:** July 20, 2025

**Published:** July 23, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

As the construction of smart grids advances, the volume of data generated by power systems is experiencing explosive growth, covering all aspects of the business process, including power generation, transmission, distribution, and consumption. This study focuses on the integration of artificial intelligence with power big data, analyzing the characteristics and application needs of power big data. It explores the application scenarios of AI technology in areas such as power system operation monitoring, fault diagnosis, and load forecasting. By integrating practical power business needs, the study proposes an AI-driven architecture design for power big data, which includes data collection, processing, analysis, and application. The aim is to enhance the intelligence level of power systems through technological integration, providing theoretical support and practical guidance for their safe, efficient, and economical operation.

## Keywords

Artificial Intelligence, Power Big Data, Application Analysis, Architecture Design, Power System

## 1. Introduction

Multi-source heterogeneous data refers to a collection of data from various sources, such as sensors, databases, and web pages, with different structures, storage formats, and communication protocols. This type of data includes a mix of text, images, videos, and numerical values. Large-scale model technology involves complex AI models trained on vast amounts of data. These models use deep learning algorithms to learn the characteristics and patterns of the data, enabling them

to understand, generate, and reason effectively. They can be applied in various fields, including natural language processing and image recognition. The scale of power big data is vast, diverse, and highly real-time, making traditional data processing and analysis methods inadequate for the precise management and intelligent decision-making required in power systems. AI technology, with its robust data analysis, pattern recognition, and learning capabilities, offers a new approach to unlocking the value of power big data. Researching the application and architecture design of AI in power big data is crucial for improving the operational efficiency of power systems, ensuring energy security, and promoting energy transition.

## **2. Analysis of Characteristics and Application Requirements of Power Big Data**

### **2.1. Features**

In the realm of power big data, real-time performance is a core feature. Smart meters collect electricity usage data every 15 minutes, with some industrial users' metering devices capable of sampling at a frequency of seconds. The power monitoring system (SCADA) can monitor the power grid's operational status in milliseconds, allowing for real-time tracking of key parameters such as voltage, current, and power. This high-frequency data collection mechanism ensures immediate awareness of the power system's operational status, but it also imposes stringent requirements on data transmission bandwidth, storage performance, and processing efficiency [1]. The high-dimensional nature of the data is characterized by its diversity, including not only structured measurement data and equipment records but also unstructured inspection images, fault recording files, and semi-structured log data. The complexity of data dimensions increases the difficulty of analysis. Power big data is highly correlated, stemming from the tight coupling of various system components. Fluctuations in power generation directly impact the load on transmission lines, and faults in distribution equipment can trigger regional power supply anomalies. Changes in user electricity consumption behavior also have a reverse effect on power generation scheduling. This complex network of cause and effect requires data analysis to uncover potential connections between data from a systemic perspective. The spatiotemporal characteristics further complicate data processing, as electricity usage patterns vary significantly across different regions and times, and the operational characteristics of the power grid in different geographical areas are influenced by climate and industrial structure, exhibiting distinct regional features. For example, the monitoring data of ice accumulation on transmission lines during the plum rain season in southern China and the heating load data during winter in northern regions show entirely different patterns.

### **2.2. Application Requirements**

The application of big data in the power sector is essential throughout the entire

lifecycle of power system planning, construction, operation, and maintenance, covering various areas such as production operations, equipment management, and business decision-making. In grid operation and control, data analysis is crucial for optimizing power flow distribution, stabilizing voltage levels, and quickly locating faults [2]. To address the intermittent and fluctuating issues caused by large-scale integration of new energy sources, data modeling is used to predict wind and solar power generation, thereby enhancing the grid's capacity to absorb these renewable energies. For instance, by integrating meteorological data, historical power generation records, and numerical weather forecasts, the prediction error of wind power can be kept within acceptable limits, providing a solid scientific basis for dispatching decisions.

In the field of equipment management, data applications primarily focus on two areas: fault diagnosis and life prediction. Over time, power equipment generates vibrations, temperature changes, and oil chromatography data that reflect the health status of the equipment. Data analysis technology is essential for early warning of potential faults to prevent sudden incidents. For instance, analyzing the trends in gas composition dissolved in oil can help identify internal overheating and partial discharge faults in transformers. It is essential for management to analyze power consumption data to understand load characteristics and predict market demand, which supports the formulation of electricity prices and the optimization of marketing strategies. For example, LSTM is suitable for power load time series prediction due to its ability to capture long-term dependencies, although it has low computational efficiency and is difficult to parallelize. GNN is used for power grid topology analysis, capable of handling node relationship information, but it performs poorly in complex dynamic scenarios. Transformer excels in power text analysis, relying on self-attention mechanisms to enhance global perception, but it is prone to overfitting when data is insufficient and lacks the ability to model temporal causality.

### **3. The Application Scenarios of Artificial Intelligence Technology in Power Big Data**

Currently, the industry predominantly employs a traditional hierarchical architecture of “data collection, storage, and analysis,” which heavily relies on manual feature engineering. This architecture integrates edge computing and federated learning to create a collaborative system between endpoints, edges, and the cloud. It uses AI models to automatically extract power data features, supporting real-time processing of multi-source heterogeneous data and ensuring privacy protection. Compared to traditional methods, this approach significantly enhances timeliness, intelligence, and security. Specifically:

#### **3.1. Intelligent Monitoring of the Operation Status of the Power System**

Artificial intelligence (AI) technology has provided an intelligent method for

monitoring the operational status of power systems, transforming manual inspections into intelligent perception. By utilizing deep learning-based image recognition, automated analysis of inspection images of transmission lines can be performed. Using convolutional neural networks (CNNs), key features such as insulator damage, conductor breaks, and hardware corrosion can be extracted. Drones equipped with high-definition cameras can periodically patrol transmission lines, capturing tens of thousands of images in a single day. Combined with edge computing technology, this process completes preliminary on-site analysis, significantly enhancing inspection efficiency. When monitoring the operational parameters of the power grid, long-term short-term memory (LSTM) networks are used to analyze key data collected by SCADA systems, such as voltage and current, in time series, enabling real-time prediction of system operation and early detection of abnormal conditions like voltage limits and power oscillations [3]. By constructing multi-variable time series models and considering load changes, new energy output, and weather conditions, dynamic evaluations of the power grid's operational status are achieved. The intelligent alarm system uses natural language processing technology to automatically classify and generate disposal suggestions from a large volume of alarm information, reducing fault response time from minutes to seconds.

By applying digital twin technology to the monitoring of power grid operations, the physical grid is mapped in real-time to a virtual grid. A 3D model of the power grid is created to map real-time operational data and equipment status data into a virtual environment, providing a visual representation and immersive monitoring of the grid's operational status. To enhance the efficiency and safety of power grid maintenance, maintenance personnel can use virtual reality (VR) tools for virtual inspections, directly observing the operation of equipment and simulating fault scenarios for emergency drills.

### 3.2. Knowledge Questions and Answers

In large-scale, complex power systems, knowledge question and answer (KQA) systems serve as a key application of AI technology, providing efficient service and decision support. By leveraging natural language processing (NLP) and deep learning, AI can swiftly understand complex questions about power big data from users and extract precise answers from a vast amount of industry knowledge and data. For example, in the context of power equipment maintenance, when a staff member asks, "What measures should be taken when the oil temperature of a certain type of transformer abnormally rises?" The AI system can quickly provide a comprehensive answer by analyzing operational data, historical fault cases, and relevant technical standards, including fault analysis, emergency procedures, and follow-up inspection recommendations. This system not only processes structured data but also deeply understands unstructured power technical documents and expert experience reports, converting the knowledge into information that users can directly utilize. In the field of power dispatching, when dispatchers face

complex issues like “how to optimally allocate grid load under extreme weather conditions,” the AI KQA system can simulate the operation of the power grid under different dispatching strategies using real-time meteorological data, grid topology, historical load curves, and other multi-source information, providing data support and solution recommendations for dispatch decisions. Additionally, this system features continuous learning capabilities, updating its knowledge base by learning about new power technologies, industry standards, and solutions to practical problems, thereby enhancing the accuracy and timeliness of its answers [4]. Meanwhile, for service scenarios involving power users, when a user inquires about the potential causes of household power outages, the system can provide step-by-step guidance to identify and address issues based on the limited information provided by the user, along with regional power grid fault data and details about the user’s electrical equipment. This approach offers targeted solutions, supports power services with intelligence, and significantly enhances the efficiency and quality of power services.

### 3.3. Data Retrieval

In the big data environment of the power industry, data retrieval is crucial for effectively obtaining information and uncovering the value of data. Artificial intelligence technology, with its powerful data analysis and processing capabilities, has revolutionized traditional data retrieval methods, transforming them from simple keyword matching to intelligent semantic retrieval. The power system stores a vast amount of structured and unstructured data, including equipment operation parameters, grid monitoring data, maintenance records, and design drawings. The AI data retrieval system can use deep learning algorithms to extract the features and semantics of this data, building multi-dimensional data indexes. During the user’s search process, the system is not limited to accurately matching keywords but can understand the semantic intent behind the user’s query. For example, when a user enters “find substations that have experienced 3 or more failures in the past 6 months,” the system will automatically link relevant data fields and store them in the power equipment database. It will also enable quick searches in the fault record system and display data lists and visual analysis results based on user requirements [5]. In the context of power planning and design, engineers need to find “substation construction plans with similar terrain conditions.” The AI retrieval system can use image recognition and natural language processing technologies to analyze historical project documents and design drawings, identify instances that meet the criteria, and use similarity algorithms to rank these instances, providing engineers with valuable reference information. Additionally, the data retrieval system has an intelligent recommendation feature that actively pushes relevant data resources and analysis reports based on the user’s search history and behavior habits. In diagnosing power system faults, after technicians search for a specific type of fault data, the system can automatically recommend similar fault cases, relevant handling experiences, and the latest research

findings. This helps technicians quickly identify the cause of the fault and develop solutions. The intelligent data retrieval using artificial intelligence significantly enhances the efficiency of power big data utilization, providing strong support for the safe, stable operation, and innovative development of power systems.

### **3.4. Power User Behavior Analysis and Demand Response**

Artificial intelligence technology assists power users in conducting behavior analysis and demand response management to achieve efficient allocation of power resources and balance supply and demand. The user classification model uses cluster analysis to categorize users based on their electricity load curves, peak-to-valley difference rates, and usage time distribution, dividing them into three categories: industrial, commercial, and residential. It further classifies high-energy-consuming users and flexible load users into subgroups, mining and analyzing the patterns of user electricity consumption behaviors. This process identifies the differences in load between commercial users on weekdays and weekends, as well as the seasonal electricity usage characteristics of residential users, among other usage patterns of different user groups. Demand response management uses incentive mechanisms to guide users in adjusting their electricity consumption behaviors, formulating differentiated demand response strategies based on the results of user behavior analysis. For industrial users, real-time electricity prices and interruptible load compensation are implemented to guide users to avoid peak electricity usage times. For residential users, smart meters provide energy-saving suggestions to encourage participation in load aggregation. The intelligent demand response platform uses reinforcement learning algorithms to dynamically optimize incentive strategies and effectively utilizes demand-side resources based on the real-time operation status of the power grid and user feedback information.

By analyzing user behavior, the effectiveness of demand response can be evaluated, providing data support for policy formulation. Machine learning models predict user responses to various incentives, optimizing incentive schemes. Additionally, the analysis of electricity usage data and the performance of energy-saving devices has been conducted, evaluating the effectiveness of energy-saving renovations and laying the groundwork for the promotion and application of green electricity. In the power market environment, user behavior analysis can also predict electricity purchase decisions, aiding in the formulation of marketing strategies by power companies and enhancing their market competitiveness.

## **4. Principles of Power Big Data Architecture Design Driven by Artificial Intelligence**

### **4.1. Architecture Design Objectives and Positions**

The architecture design of power big data driven by artificial intelligence needs to serve the goal of achieving intelligent transformation of the power system, with the core positioning being the construction of an efficient data processing, analysis and application system. The architecture design is oriented towards supporting

the safe and stable operation of the power system, improving operational efficiency, promoting the consumption of new energy, and maximizing the value of data. By integrating multi-source heterogeneous data, a unified data resource pool has been constructed to provide high-quality data input for the implementation of artificial intelligence algorithms. The architecture needs to have strong computing and storage capabilities to meet the requirements of real-time processing and massive storage of power big data. Functionally, this architecture should achieve full-process integration of data collection, transmission, storage, analysis and application (see **Table 1**). The data acquisition module ensures the completeness and accuracy of data by adapting to multi-device interfaces, integrating Kafka message queues and Flink real-time computing, and supporting protocols such as IEC 61850 and Modbus. The transmission network is based on 5G and optical fiber, and is combined with SDN technology to achieve dynamic bandwidth scheduling, ensuring real-time and reliable data transmission. The storage system adopts the Hadoop distributed architecture and ClickHouse columnar storage to build a hierarchical data warehouse system, supporting the mixed storage of structured and unstructured data. The analysis platform integrates AI algorithms such as Spark MLlib and TensorFlow, and is combined with the Dolphinscheduler task scheduling tool to achieve flexible modeling. The application layer, in combination with the business scenarios of the power grid, utilizes visualization tools such as Table to promote the realization of data value.

**Table 1.** Architecture design objectives.

Architecture design objectives	Architecture design orientation	Architecture features
Realize the goal of intelligent transformation of power system	Support the safe and stable operation of power system, improve operational efficiency and promote the consumption of new energy	The whole process of data acquisition, transmission, storage, analysis and application is connected
The core is to build an efficient data processing, analysis and application system	Maximize the value of data	-
data acquisition module	data transmission network	inventory system
Supports a variety of device interfaces and IEC 61850, Modbus and DNP3 communication protocols	Use 5G, optical fiber and other communication technologies to establish high-speed transmission channels	Use a distributed architecture to support the hybrid storage of structured and unstructured data

In the architecture design, it is essential to reserve expansion interfaces to meet the future development needs of the power system. As new power systems are constructed, more data types and business requirements will emerge, such as monitoring data from distributed power sources and charging data from electric vehicles. The architecture must have excellent scalability to support the rapid ac-

cess to new data sources, the deployment of new algorithms, and the development of new applications, ensuring the system operates stably over the long term. Additionally, the design must address issues of data security and privacy protection, adhering to the cybersecurity level protection standards set by the power industry, and establish a multi-layered security protection system.

## 4.2. System Architecture Design

The system architecture design is guided by the principles of hierarchical decoupling, modularity, scalability, and high availability, ensuring the stability, flexibility, and maintainability of the architecture. The storage layer constructs a data lake using the Hadoop distributed file system, employing ClickHouse columnar storage and HBase distributed database to support the mixed storage of structured, semi-structured, and unstructured data. Data governance is achieved through a layered data warehouse architecture. The analysis layer integrates Flink real-time computing, Spark MLlib machine learning libraries, and TensorFlow deep learning frameworks, along with the Dolphinscheduler task scheduling system, to provide flexible modeling and intelligent analysis capabilities. The application layer leverages the Tableau visualization tool and self-developed algorithm models to optimize power grid operations, enhance equipment intelligent maintenance, and support business decision-making scenarios, promoting the deep integration of power big data and AI technology to realize the value of data. The storage layer uses distributed file systems (such as Ceph) and columnar databases (such as ClickHouse) to ensure efficient data storage and rapid retrieval. The analysis layer utilizes big data processing frameworks (such as Hadoop and Spark) and artificial intelligence platforms (such as TensorFlow) to provide data processing and model training capabilities. The application layer develops various business applications to realize the value of data, with a modular design that decomposes the system into independent functional modules, each implementing specific functions, making development, maintenance, and upgrades easier. For example, the data cleaning module is responsible for removing noisy data and filling in missing values; the feature engineering module handles data transformation and feature extraction; the model training module supports the training and optimization of various algorithms; the model evaluation module is used to validate and refine the training results. Each module communicates with others through standardized interfaces, reducing dependencies and enhancing system maintainability.

The scalability principle imposes stringent horizontal scaling requirements on the architecture, enabling it to adapt to data volume growth by adding computing and storage nodes. The system supports dynamic node addition and removal through a distributed computing framework and storage system, thereby linearly expanding performance. Additionally, the architecture must support the rapid deployment of new algorithms and applications to meet the evolving needs of the power industry. To ensure business continuity, the high availability design employs redundancy and failover mechanisms to ensure that the system continues to

function even when some components fail.

### **4.3. Sample Processing of Artificial Intelligence**

Under the power big data architecture, artificial intelligence sample processing operates efficiently with massive data resources and advanced technology systems. First, a data storage platform built using Hadoop distributed file systems and ClickHouse columnar databases can integrate multi-source heterogeneous data, including grid operation, equipment status, and user electricity consumption, laying a solid foundation for sample processing. ETL tools are used to clean, transform, and standardize raw data, removing noise and outliers, unifying data formats, and improving data quality. In the sample feature engineering phase, Spark MLlib machine learning libraries and Python data analysis tools are used to extract, filter, and optimize features from cleaned data. For example, in fault prediction for power equipment, core features are extracted from operational data and past maintenance records, and techniques like Principal Component Analysis (PCA) are used to reduce data dimensions and redundancy, thereby enhancing model training efficiency. Additionally, by combining knowledge from the power industry and business scenarios, features are merged and derived to enhance data representation. During the sample annotation stage, a combination of manual annotation and semi-supervised learning algorithms is used to effectively annotate large datasets. For data that is difficult to annotate manually, active learning algorithms are used to filter out key samples, improving annotation accuracy and efficiency. The high-quality sample dataset, constructed through data augmentation techniques, further expands the diversity of samples. This provides high-quality data input for training machine learning and deep learning models, ensuring that the model has strong generalization capabilities and high prediction accuracy in applications such as power load forecasting, fault diagnosis, and equipment life assessment. This promotes the intelligent upgrade of power systems.

## **5. The Construction of Power Big Data Artificial Intelligence Architecture System**

### **5.1. Design of Data Acquisition and Preprocessing Layer**

The data collection and preprocessing layer is a fundamental component of the artificial intelligence architecture in power big data, responsible for collecting, cleaning, and standardizing multi-source heterogeneous data. The performance of this layer directly impacts the accuracy and efficiency of subsequent analysis. From the perspective of data collection, it is essential to establish a comprehensive perception network that covers the entire business chain, from power generation to distribution and consumption. On the generation side, sensors are deployed to collect operational parameters of units and environmental data from new energy stations. On the transmission side, intelligent inspection robots and drones are used to obtain line images and infrared temperature data. On the distribution side, intelligent switches and transformer monitoring devices provide status infor-

mation. On the consumption side, smart meters and load monitoring terminals assist in collecting user consumption data. For different equipment interfaces and communication protocols, edge computing gateways are used to convert protocols and aggregate data, ensuring the completeness and real-time nature of data acquisition. In the preprocessing stage, the principle of improving data quality is followed, using methods such as noise reduction, missing value filling, and outlier detection to provide high-quality data for upper-layer applications. Statistical methods and machine learning algorithms are used to identify noisy data, and data is smoothed using median filtering and Gaussian filtering. For missing data, multiple imputation and random forest interpolation are used based on the distribution characteristics of the data. Isolation forests and One-Class SVM algorithms are employed to detect outliers, ensuring the reliability of the data. In addition, a data standardization process has been established to unify the conversion and encoding of different source data formats, eliminating dimensional differences and laying the groundwork for subsequent data analysis. A data quality monitoring mechanism has also been introduced to monitor the data collection and processing processes in real time, generating quality assessment reports to promptly identify and address data quality issues.

## **5.2. Construction of the Support Layer of Artificial Intelligence Algorithm**

As a critical component of deep analysis in power big data, the artificial intelligence algorithm support layer constructs a comprehensive algorithm system using large model technology. By integrating the Transformer architecture with the TensorFlow framework, it performs deep feature extraction and pattern recognition on vast amounts of power data. The system uses the self-attention mechanism to uncover complex relationships between power equipment operation and user consumption behavior, achieving high-precision power load forecasting through a temporal Transformer model. By integrating graph neural networks (GNN) with large language models (LLM), it constructs a device fault diagnosis model. This model uses knowledge graph technology to integrate multi-source information, enabling intelligent analysis of fault causes and the generation of handling suggestions, thereby significantly enhancing the intelligence level of power business scenarios.

## **6. Conclusion**

In summary, the deep integration of artificial intelligence (AI) and big data in the power sector is an inevitable trend for the intelligent upgrade and development of power systems. The application analysis and architecture design proposed in this paper provide a practical approach to achieving efficient data processing and precise decision-making in power systems. By constructing an intelligent architecture system, the safety, reliability, and economic efficiency of power system operations can be significantly enhanced. Moving forward, it is essential to continue focusing

on the innovation and development of AI technology, further optimize architectural design, and strengthen data security measures. This will continuously deepen the application practices in various fields, promoting the power industry towards higher levels of intelligence and digitalization, thereby providing strong support for the high-quality development of China's energy sector.

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] Fan, Z.Y. and Gao, Z.X. (2023) Key Technologies of Power Big Data for Smart Grid. *Electrical Technology and Economy*, No. 5, 107-110.
- [2] Dong, L.X. (2023) Application of Key Technologies of Smart Grid Power Big Data. *Light Source and Lighting*, No. 3, 230-232.
- [3] Lu, S. and Guo, L.G. (2022) Analysis of the Application of Power Big Data Technology for Smart Grid. *Henan Science and Technology*, **41**, 23-26.
- [4] Deng, M. (2022) Analysis of Key Technologies of Power Big Data for Smart Grid Applications. *Science and Technology Innovation*, No. 8, 7-9+12.
- [5] Xie, Q.Y., Zhang, Y.K. and Li, J.W. (2021) Key Technologies of Power Big Data for Smart Grid. *Power Grid and Clean Energy*, **37**, 39-46.