Deep Reinforcement Learning-Based Real-Time Trading Decision Support for Virtual Power Plants with Intelligent Assistant Integration

Shaohua Zhao , Cong Zhang , Fuming Liu , Yuqian Tian , Jifan Ouyang, Zhenkai Hu*
Cgs Power Generation (Guangdong) Energy Storage Technology Co.,Ltd, Guangzhou,Guangdong, 510630, China
E-mail: 18933909033@163.com
*Corresponding author

Keywords: virtual power plant, intelligent assistant, hydropower, real-time trading, deep reinforcement learning, decision support system

Received: October 22, 2025

With the profound transformation of the energy structure and the advancement of the "dual carbon" goal, the virtual power plant (VPP), centered on distributed energy resources, has emerged as a key technology for enhancing the flexibility and integration capacity of modern power grids. However, the diversity and volatility of VPP internal resources, coupled with the complexity of the electricity market, impose significant challenges on the response speed and economic efficiency of real-time trading decisions.To address these challenges, this paper proposes and develops a real-time trading decision support system for VPPs driven by an Intelligent Assistant (IA). The system leverages hydropower plants—with their fast response and energy storage capabilities—as the core regulating resources and coordinates multiple distributed energy sources, including photovoltaics, wind power, and energy storage systems. At its core, the IA integrates deep learning–based forecasting models, reinforcement learning–based decision modules, and a natural language processing (NLP)–based interaction component. The IA assists operators in real time by analyzing multidimensional data such as market prices, grid loads, meteorological information, and hydropower inflows, accurately predicting generation and price trends, and dynamically optimizing bidding and regulation strategies through reinforcement learning algorithms to maximize overall benefits.This paper details the overall architecture and key technological components of the proposed system and conducts a simulation case study using a regional VPP containing multiple hydropower plants. Specifically, the core decision-making module employs the Soft Actor-Critic (SAC) deep reinforcement learning algorithm. The system was trained for 2 million steps over 72 hours on a V100 GPU, utilizing one year of real historical operational and market data from a VPP in Southwest China. The simulation results demonstrate that the proposed IA-DRL strategy outperforms a traditional rolling-horizon Mixed-Integer Linear Programming (MILP) method, achieving a 14.7% increase in net profit, a 47.8% reduction in deviation assessment costs, and a remarkable 99.7% acceleration in decision-making time. These results confirm the significant technical and economic advantages of the proposed framework, providing a new theoretical foundation and practical solution for intelligent VPP operation and business model innovation, while also offering enhanced interpretability through the intelligent assistant.

Povzetek:

1 Introduction

With the global energy transformation towards cleaner and lower-carbon forms, the penetration rate of intermittent renewable energy sources represented by wind energy and solar energy in the power system has been rapidly increasing. While bringing environmental benefits, this transformation also poses unprecedented challenges to the real-time balance and safe and stable operation of the power system[1]. As an advanced energy management technology, the virtual power plant (VPP) aggregates distributed energy resources (DERs) such as geographically dispersed distributed generation (DG), controllable loads (CL), and energy storage systems (ESS) into a unified entity that can interact with the power grid through information and communication technologies and intelligent aggregation algorithms [2], effectively enhancing the power grid's acceptance capacity for distributed energy and the overall operation efficiency of the system [3].

However, the economically efficient operation of VPP faces two core problems: First, the high heterogeneity and uncertainty of internal resources. The output of wind and solar power is highly random and volatile, and the load response is uncertain, which makes the overall controllability of VPP poor. Second, the complexity and high dynamics of the electricity market environment. The price of the electricity market

(especially the spot market) fluctuates frequently, and there are various trading varieties (such as electricity energy, ancillary services, etc.), requiring VPP to have millisecond-level response and precise decision-making capabilities. Traditional optimization methods based on mathematical programming often have huge computational amounts when dealing with high-dimensional, non-linear, and strongly uncertain problems, and it is difficult to meet the real-time requirements [5].

To address the above challenges, artificial intelligence (AI) technologies, especially machine learning and deep learning, provide new solutions for the intelligent decision-making of VPP [6]. In recent years, researchers have begun to attempt to use AI models for load forecasting, electricity price forecasting, and optimal scheduling. However, existing research mainly focuses on the application of single models, lacking an intelligent and integrated system that can integrate perception, prediction, decision-making, and interaction [7]. In particular, the concept of an "intelligent assistant" that can deeply understand the intentions of operators, assist in performing complex analyses, and interact in a natural language manner has not been fully exploited. This study believes that the core value of the intelligent assistant lies not only in simplifying operations, but more importantly, in acting as a bridge between human experts and "black box" models by transforming complex AI decision-making processes into interpretable and understandable suggestions, thereby building trust and achieving truly efficient and reliable human-machine collaborative decision-making. This model, which aims to empower operators rather than replace their ultimate decision-making power, is the core of the safe application of advanced AI technologies in critical infrastructure fields [8].

1.1 Related work

The management and optimization of Virtual Power Plants (VPPs) in dynamic electricity markets have been extensively researched. Early efforts often relied on traditional optimization techniques, primarily Mixed-Integer Linear Programming (MILP) or Quadratic Programming, for day-ahead scheduling and intraday re-dispatch [9][10]. While effective for well-defined problems, these methods struggle with real-time demands, high dimensionality, and inherent non-linearities, often leading to significant computational burdens and sensitivity to prediction errors [5]. With the advent of Artificial Intelligence (AI), Deep Reinforcement Learning (DRL) has emerged as a promising paradigm for sequential decision-making under uncertainty, showing potential for optimal bidding and scheduling in VPPs due to its ability to learn complex policies through interaction with dynamic environments [4][12][13]. However, many existing DRL applications in VPPs primarily focus on optimization objectives and often treat the DRL model as a "black box," lacking explicit mechanisms for human operators to understand and trust automated decisions [8]. Furthermore, the specific role of hydropower as a core regulating resource for its unique fast-response and energy storage capabilities has not been fully leveraged within integrated AI-driven trading systems. This study aims to address these identified gaps.

Table 1: Comparative analysis of related work in VPP decision-making

Feature/Study	Optimization Methods (e.g., MILP, QP) [5, 10]	DRL-based VPP Optimization (e.g., [4, 9, 12, 14])	This Work (IA-DRL with Hydropower Core)
Methodology Core	Mathematical programming (linear/quadratic solvers)	Model-free or model-based DRL algorithms (e.g., DDPG, PPO, SAC)	SAC-based DRL for real-time optimization, integrated with DL forecasting.
Uncertainty Handling	Relies on accurate forecasts; sensitive to prediction errors; often uses stochastic programming.	Learns robust policies through interaction; inherently handles dynamic uncertainties.	Strong endogenous ability to manage uncertainty, non-linearities, and forecasting errors through continuous environmental interaction and policy refinement.
Real-time Performance	Can be computationally intensive for large-scale/complex problems; slower for real-time.	Extremely fast inference post-training (millisecond-level).	Millisecond-level decision speed, critical for intraday/real-time markets.
Human-Machine Interaction	Command-line interfaces, basic dashboards; minimal decision support.	Typically a “black-box” model; limited direct human interaction or interpretability.	Intelligent Assistant (IA) with NLP for natural language interaction and explainable AI (XAI).
Resource Focus	General DERs; hydropower’s unique role	General DERs; specific resource advantages (like	Hydropower as core regulating resource; full exploitation of its fast response and storage.

Feature/Study	Optimization Methods (e.g., MILP, QP) [5, 10]	DRL-based VPP Optimization (e.g., [4, 9, 12, 14])	This Work (IA-DRL with Hydropower Core)
	often simplified or not central.	hydropower's regulation) not always highlighted.	
Scalability & Adaptability	Requires re-modeling for new market rules; limited online adaptation.	Strong adaptability through online learning; can generalize to changing conditions.	High adaptability to market changes; robust across different VPP resource ratios.
Interpretability/Trust	Transparent if model is simple, but complex models are hard to interpret.	Low transparency ("black box"); difficult for operators to trust or understand decisions.	Transforms DRL "black box" into "gray-box" through real-time explanations, building operator trust.
Validation/Benchmarking	Often uses deterministic or simplified stochastic scenarios.	Benchmarked against other DRL algorithms or simplified rule-based methods.	Benchmarked against a strong deterministic MILP baseline in a high-fidelity simulation.

In addition, among many distributed resources, hydropower (especially small and medium-sized hydropower stations) is an ideal "stabilizer" and "regulator" in VPP due to its excellent regulation performance, fast start-stop ability, and certain energy storage characteristics (through reservoir regulation). Taking hydropower as the core regulation resource of VPP can effectively suppress the fluctuations of wind and solar power output and improve the overall reliability and market competitiveness of VPP.

1.2 Research questions and contributions

Based on the identified gaps, this study proposes to construct a real-time trading decision-making system for virtual power plants driven by an intelligent assistant, with a focus on its application in VPPs with hydropower as the core. This research aims to answer the following questions:

RQ1: Can a DRL-based real-time trading system, driven by an intelligent assistant and leveraging hydropower as a core regulating resource, significantly outperform traditional optimization methods (e.g., MILP) in terms of economic benefits (e.g., net profit, deviation costs) for VPPs in a dynamic electricity market?

RQ2: How does the proposed IA-DRL system enhance the operational efficiency and reliability of VPPs, particularly in terms of renewable energy curtailment and real-time decision-making speed, compared to established benchmarks?

RQ3: Can an Intelligent Assistant, integrated with a DRL decision engine, effectively transform complex AI decisions into interpretable natural language explanations, thereby improving human-machine collaboration and operator trust in critical VPP operations?

RQ4: What are the key architectural and algorithmic components required to realize such an integrated and intelligent VPP real-time trading decision support system, and how do they interact to achieve optimal performance?

The main contributions of this paper are summarized as follows:

Proposed a novel, integrated IA-DRL framework for VPP real-time trading, with a specialized focus on hydropower as a core regulating resource, which is a unique contribution in the context of intelligent VPP operation.

Developed and validated a comprehensive system architecture that deeply integrates data perception, deep learning-based prediction, DRL-based decision-making, and NLP-based human-machine interaction.

Demonstrated the significant economic and operational superiority of the IA-DRL strategy over a traditional MILP method through detailed case studies, showcasing enhanced net profit, reduced deviation costs, and vastly improved decision-making speed.

Pioneered the integration of explainable AI (XAI) principles through the Intelligent Assistant's natural language explanation capabilities, fostering operator trust and facilitating effective human-machine collaborative decision-making in critical energy infrastructure.

Provided a robust solution for renewable energy integration, proving the system's ability to significantly reduce renewable energy curtailment by efficiently coordinating diverse DERs.

The structure of this paper is arranged as follows: Chapter 2 introduces the basic theories of virtual power plants and electricity market transactions; Chapter 3 elaborates on the overall design and architecture of the intelligent assistant-driven decision-making system in detail; Chapter 4 deeply explores the key technologies and core algorithm models in the system; Chapter 5 conducts the simulation analysis of system implementation and application cases;Chapter 6 presents a comprehensive discussion of the results and their implications; and finally, a summary and outlook are presented in Chapter 7.

2 Overview of virtual power plants and electricity market transactions

2.1 Concept and composition of virtual power plants

A virtual power plant is not a physical power plant but an energy aggregation management system. It aggregates different types of distributed energy resources (DERs) through advanced communication and control technologies and participates in the operation of the electricity market as a special market entity [10]. Its basic composition is shown in Figure 1, mainly including three types of resources: distributed generation (DG), energy storage system (ESS), and controllable load (CL). These resources are coordinated by the VPP control center and interact with the upstream electricity market and grid operators in terms of information and energy.

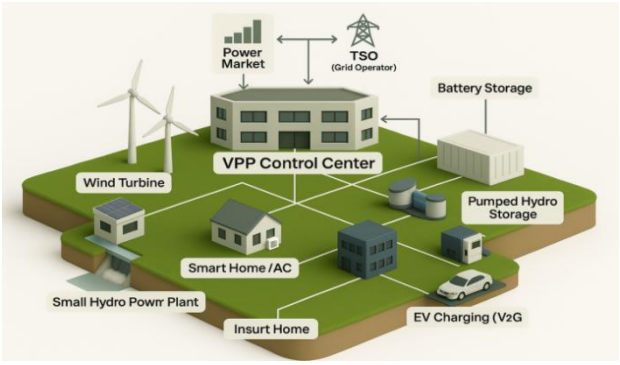

Figure 1:Schematic diagram of the basic composition of virtual power plants

The aggregated resources of VPP usually include:

Distributed generation (DG): Such as rooftop photovoltaic, decentralized wind power, small hydropower stations, gas turbines, etc.

Energy storage system (ESS): Such as electrochemical energy storage (lithium batteries), pumped storage, flywheel energy storage, etc.

Controllable load (CL): Such as intelligent air conditioners, industrial production lines, electric vehicle charging and discharging (V2G), etc. [11].

2.2 Core role of hydropower in virtual power plants

Among the numerous resources of VPP, hydropower stations (especially small and medium-sized hydropower stations with regulating reservoirs) play a crucial role:

Fast regulation ability: The start-stop and output adjustment speed of hydropower units is much faster than that of large thermal power units, which can effectively meet the rapid response requirements of the intraday market and ancillary service market.

Natural energy storage characteristics: The reservoir itself is a large-scale and long-term energy storage facility, which can realize the transfer of electric energy in the time domain and perfectly hedge the intermittency of wind and solar power generation.

Relatively high predictability: Compared with wind and solar, the medium- and long-term hydropower inflow runoff has certain regularity, and the short-term (hourly) output is basically controllable, providing a deterministic basis for the planning of VPP.

Basic output equation of hydropower station:

$$P_{\text{hydro}}=\eta \cdot g \cdot Q \cdot H \quad (1)$$

Among them, P_{hydro} is the output power of the hydropower station (W), η is the comprehensive efficiency coefficient (including the efficiency of the water turbine and generator), g is the acceleration due to gravity (m/s²), Q is the flow rate through the water turbine (m^3/s), and H is the effective head (m). This formula is the basis for VPP to dispatch hydropower resources

.

2.3 Power market trading mechanism

VPP mainly participates in the following types of power markets [12]:

Day-ahead market: Declare the power generation/consumption curve and price for the next day one day in advance, which is the main market for electricity trading.

Intraday/real-time market: Conducted within the operating day to correct the deviation between the day-ahead plan and the actual operation, usually with a trading cycle of 15 minutes or 5 minutes.

Auxiliary service market: VPP obtains revenue by providing services such as frequency regulation, reserve, and reactive power support, with extremely high requirements for response speed.

Table 2: Comparison of requirements for VPP in different power markets

Market type	Regulation cycle	Response time requirement	Decision-making complexity	Advantages of hydropower resources
Day-ahead market	24 hours	Hourly level	Medium	Reliable baseload/shoulder load provider
Intraday market	15 minutes/5 minutes	Minute level	High	Quickly adjust output and correct deviation
Frequency regulation market	4 seconds	Second level	Extremely high	Quick response and provide high-quality frequency regulation service
Reserve market	10 - 30 minutes	Minute level	Medium	Reliable reserve capacity and quick start-stop

2.4 Challenges faced by VPP real-time trading

VPP needs to solve a high-dimensional stochastic optimization problem in the real-time market: under the premise of meeting its own physical constraints and grid security constraints, how to coordinate internal resources, formulate optimal bidding and scheduling strategies to cope with the changing market prices and power generation/load forecasts. This requires the system to have strong sensing capabilities, forecasting capabilities, and decision-making assistance capabilities [13].

3 Overall design of the decision system driven by intelligent assistants

To address the above challenges, we designed a VPP real-time trading decision system with an intelligent assistant at its core.

3.1 Design concept

Data-driven: All decision-making suggestions of the system are based on the real-time analysis of multi-source heterogeneous data [14].

Model core: An advanced AI model is used as the engine for generating prediction and optimization strategies.

Human-machine collaboration: The intelligent assistant serves as a bridge between human operators and complex systems, enabling efficient human-machine collaboration, allowing operators to focus on strategic supervision and decision-making, rather than complex calculations.

Continuous evolution: The system has self-learning capabilities and can continuously optimize its strategy model in the continuous interaction with the market.

3.2 Overall system architecture and decision-making process

To address the complex challenges of virtual power plant (VPP) real-time trading, we designed a highly integrated decision system based on the concepts of data-driven, model core, and human-machine collaboration. The overall architecture and core decision-making process of the system are shown in Figure 2.

Figure 2 adopts a hierarchical and decoupled modular design, which is divided into a data layer (Data Layer), a model layer (Model Layer), a decision layer (Decision Layer), and an interaction layer (Interaction Layer) from top to bottom. This design ensures the independence of functions and the scalability of the system.

The operation process of the system follows two tightly coupled information flows:

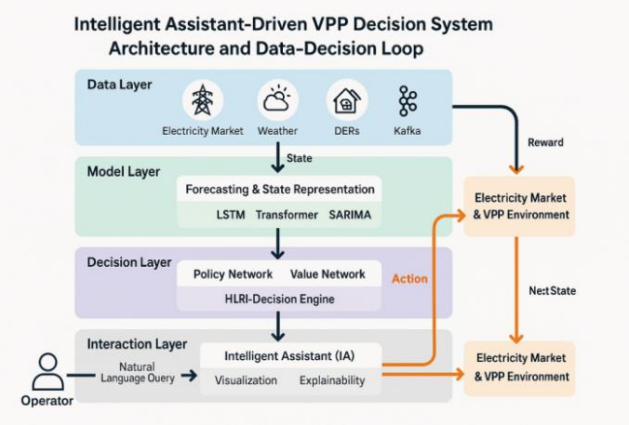

Figure 2: System architecture and closed-loop process of the intelligent assistant-driven VPP

Main data processing flow (black vertical downward arrow in the figure): The information flow starts from the data layer, which is responsible for aggregating multi-source heterogeneous data from the power market, meteorological system, and each unit (DERs) within the VPP, and storing and transmitting it using technologies such as InfluxDB and Kafka. The data then flows into the model layer, where it undergoes in-depth analysis and processing through prediction models such as LSTM and Transformer and physical resource models, and is finally encapsulated as a state vector (State) describing the global information of the current system **Error! Reference source not found.**.

Reinforcement Learning Decision Closed-Loop (the orange highlighted circular arrow in the figure): This is the core innovation of the system. After receiving the state, the deep reinforcement learning (DRL) engine in the decision-making layer generates the optimal action, that is, the bidding and scheduling strategies of the VPP. After this action acts on the external power market and the VPP environment, the environment will feedback an immediate reward (such as trading revenue) and the next state after the environment evolves. This feedback information of "state-action-reward" constitutes the closed-loop of reinforcement learning, enabling the agent to conduct autonomous learning and policy iterative optimization through continuous interaction with the environment [15].

At the same time, the interaction layer reflects the design concept of human-machine collaboration in the system. Operators can initiate "queries" or instructions to the Intelligent Assistant through natural language. The Intelligent Assistant can not only present the complex strategies in the decision-making layer to the operators in the form of visual charts and "decision suggestions", but also "explain" the logic behind them, thus realizing efficient and trustworthy intelligent assistance and empowering operators to strategically supervise the system.

In summary, this architecture integrates data collection, model prediction, intelligent decision-making, and human-machine interaction, forming a complete and dynamic closed-loop system from raw data to optimal strategies and then to learning feedback. Next, the key functions of each layer will be elaborated in detail.

3.3 Detailed explanation of the functions of each layer

In the hierarchical architecture, each layer forms an organic whole through close information flow transmission, jointly supporting the efficient operation of real-time trading decisions.

Data Layer: As the basic data entry of the system, this layer is responsible for collecting, cleaning, integrating, and storing multi-source heterogeneous data from external information sources (such as weather forecasting systems, power market platforms, grid dispatching centers) and telemetry terminals of each distributed energy resource (DERs) within the VPP. This layer is not only a data aggregation point but also the starting point of data governance. We use a time series database (such as InfluxDB) for efficient storage and query, and perform preliminary cleaning and outlier removal on telemetry data through a data collection gateway deployed on the edge side. External data is obtained by calling standardized RESTful APIs. The data sources are aligned through a unified data model to ensure data consistency and availability. In addition, this layer uses technologies such as encrypted channels to ensure the security and integrity of data during transmission and storage [17].

Model Layer: As the analysis and cognitive core of the system, this layer consists of a series of precise mathematical and artificial intelligence models, mainly including: Prediction Model Portfolio: Integrate the wind and solar power prediction model based on long short-term memory network (LSTM), the hydropower reservoir inflow prediction model based on seasonal autoregressive integrated moving average model (SARIMA), and a time series model using attention mechanism (such as Transformer) for high-precision electricity price prediction [18].

Physical Resource Model: It accurately mathematically represents various types of energy units within the VPP. In particular, for the core regulation resource, the hydropower station, it conducts detailed modeling including reservoir capacity, head-output relationship, flow rate limits, and operating boundary conditions.This physical model is internally validated against a simplified traditional hydropower dispatch model to ensure accurate representation of hydropower dynamics.

Market Rule Model: It establishes a digital description of the trading rules, bidding structures, and clearing mechanisms in power markets at all levels (such as day-ahead, intra-day, and ancillary service markets), thus ensuring the compliance and effectiveness of all decisions.

The models do not operate independently but are uniformly managed through a model management and scheduling engine. This engine is responsible for triggering the corresponding prediction models at regular intervals according to the task type (such as hourly prediction, minute-by-minute prediction), and publishing the results to a message queue (such as Kafka) for the decision-making layer to subscribe and use. This loosely coupled architecture ensures the scalability of the system, and future model modules can be easily added, deleted, or replaced.

The interaction mechanism between the decision-making layer and the interaction layer: The DRL engine of the decision-making layer is deployed in the form of a service. After the intelligent assistant in the interaction layer receives the user's natural language query, its built-in intent recognition module will parse it into a structured API call request and send it to the decision-making layer. After the decision-making layer executes the corresponding calculations or simulations, it returns the results in JSON format, which are then encapsulated by the natural language generation module in the interaction layer into human-readable text or charts and presented to the user. This asynchronous call mechanism ensures the smoothness of the front-end interaction.

4 Key technologies and model construction

This chapter will introduce in detail the mathematical models and algorithms of the core modules of the system. The core of a successful decision-making system lies in choosing the right tools for the right problems. Therefore, when constructing the prediction and decision-making models, we strictly followed the principle of "problem-driven, model adaptation", and carefully selected and optimized the corresponding algorithms according to the characteristics of different subtasks.

4.1 Hydropower inflow and output prediction model

In the selection of the prediction model combination, we followed the principle of "adapting to local conditions". For wind and solar power prediction, its physical process is highly correlated with the continuous changes in short-term meteorological conditions (such as wind speed, light intensity) and has obvious time series dependence. The long short-term memory network (LSTM) can effectively capture and remember this short-term to medium-term time series dependence through its unique gating mechanism, so it becomes the first choice for this task. For hydropower reservoir inflow, it not only has daily, weekly, and annual periodicities but also shows strong seasonal trends. The seasonal autoregressive integrated moving average model (SARIMA) is a classic statistical model designed to handle such time series with both trends and seasons and can provide a robust baseline for runoff prediction. Since hydropower is a core regulation resource, accurate prediction of its available water volume is crucial. We adopt a hybrid model that combines physical mechanisms and data-driven methods.

Reservoir Inflow Prediction Based on Time Series:

$$Q_{\text{in},t+1}=\text{SARIMA}(P,D,Q)(p,d,q)_s+f\left(\text{Rainfall}_t,\text{Temp}_t\right)+\varepsilon_t \quad (2)$$

Among them, $Q_{\text{in},t+1}$ is the predicted inflow for the next moment. The SARIMA model captures the seasonal and cyclical trends of the flow itself. f is a non-linear function (such as a neural network) used to model the impact of external meteorological factors such as rainfall and temperature on runoff. ε_t is the error term. Consider the dynamic power output constraint of hydropower considering the head change:

$$P_{\text{hydro, min}}(V_t) \leq P_{\text{hydro},t} \leq P_{\text{hydro, max}}(V_t) \quad (3)$$

Among them, the maximum/minimum power output $P_{\text{hydro, min}}$ of hydropower is a function of the current reservoir water storage V_t because the reservoir water level directly affects the head height H . This reflects the dynamic constraint of hydropower output.

4.2 Real-time market electricity price prediction model

The electricity spot market price has high volatility and complex non-linear characteristics. We adopt a Transformer model based on the attention mechanism.

Attention mechanism weight calculation:

$$\text{Attention(Q,K,V)}=\text{softmax}\left(\frac{\text{QK}^{\text{T}}}{\sqrt{\text{d}_{\text{k}}}}\right)\text{V} \quad (4)$$

This formula is the core of the Transformer model. Through the self-attention mechanism, the model can capture the complex dependencies between different time points in the electricity price time series, such as identifying similar "morning peak" and "evening peak" patterns in a day, thus improving the prediction accuracy.

4.3 Real-time trading decision-making model based on deep reinforcement learning

We model the real-time trading decision-making problem of the VPP as a Markov decision process (MDP), and its core elements are defined as follows:

State space (State, S): A vector containing all relevant information of the VPP at time t .

State space vector:

$$S_t=\left[P^{\text{pred}}_{\text{wind},t},P^{\text{pred}}_{\text{solar},t},Q_{\text{in},t},V_{\text{hydro},t},\text{SOC}_{\text{ess},t},\lambda^{\text{pred}}_{\text{market},t},D_{\text{grid},t},T_t\right] \quad (5)$$

It includes: predicted wind and solar power output, hydropower inflow, reservoir water storage, energy storage SOC, predicted market electricity price, grid demand, and time stamp, etc. Action space (Action, A): The operations that the VPP can execute at each decision-making moment, and these operations strictly comply with physical and market rule constraints. Action space vector:

$$A_t=\left[P^{\text{bid}}_{\text{hydro},t},P^{\text{ch/dis}}_{\text{ess},t},P^{\text{buy/sell}}_{\text{grid},t}\right] \quad (6)$$

It includes: the quoted/scheduled power of hydropower, the charge/discharge power of energy storage, and the power of purchasing/selling electricity to the market.

Reward function (Reward, R): The immediate reward feedback by the environment (electricity market) after the agent executes an action, and the optimization goal is to maximize the cumulative reward. Reward function design:

$$R_t=\text{Revenue}_{\text{market},t}-\text{Cost}_{\text{op},t}-\text{Penalty}_{\text{dev},t} \quad (7)$$

$\text{Revenue}_{\text{market},t}$ is the market trading revenue. $\text{Cost}_{\text{op},t}$ is the operating cost (such as energy storage loss, unit start-stop cost). $\text{Penalty}_{\text{dev},t}$ is the deviation assessment fine caused by prediction error.The weights for each component of the reward function (e.g., penalties for deviation or costs of energy storage degradation) are determined through a combination of historical market data analysis, expert domain knowledge, and iterative hyperparameter tuning during the training phase. This ensures a balanced optimization towards both economic gains and system reliability.

Policy (Policy, π): The core of the agent, which is a mapping function from state to action, $\pi(A|S)$, usually represented by a deep neural network. Bellman optimal equation (Q-learning idea):

$$Q^*(s,a)=\text{E}\left[R_{t+1}+\gamma\max_{a'}Q^*(s',a')|S_t=s,A_t=a\right] \quad (8)$$

This is the theoretical basis of reinforcement learning, indicating that the optimal value of executing action a in state s is equal to the expectation of the immediate reward and the discounted sum of the future optimal value. Our DRL algorithm (such as SAC) is to learn this optimal Q function.

4.3.1 Algorithm selection and model structure

In this study, we selected the Soft Actor-Critic (SAC) algorithm as the core of the DRL decision engine. Compared with deterministic policy algorithms such as DDPG, SAC is a stochastic policy algorithm based on the maximum entropy framework. Its core advantage is that by introducing an entropy term into the objective function, it encourages the agent to explore more fully, effectively avoiding premature convergence of the policy to a local optimum, which is particularly important for dealing with the highly volatile electricity market environment. At the same time, the stability and sample efficiency of SAC are also better than other commonly used algorithms such as PPO.

Both the policy network (Actor) and the value network (Critic) of the agent adopt a three-layer fully connected neural network. The input layer receives the state vector S , both hidden layers contain 256 neurons, and ReLU is used as the activation function. The output layer outputs the action or state value according to the definition of the action space. To ensure the stability and efficiency of the training process, we carefully tuned the

key hyperparameters, and the specific settings are shown in Table 3.

Table3: Key hyperparameter settings of the DRL model

Hyperparameter	Value	Rationale
Learning Rate	3e-4	Balance the convergence speed and stability
Discount Factor, γ	0.99	Encourage the agent to focus on long-term cumulative rewards
Experience Replay Buffer Size	1,000,000	Store diverse enough experiences for learning
Batch Size	256	Balance computational efficiency and gradient estimation accuracy
Entropy Regularization Coefficient (α)	Auto-tuning	The core mechanism of SAC, dynamically balance exploration and exploitation
Target Network Update Coefficient (τ)	0.005	Adopt soft update to ensure stable training

4.3.2 Model training

We conducted offline training on the agent for up to 2-million-time steps using historical data from the past year. To ensure the generalization ability of the model, the selected training dataset comprehensively covered different seasons and market conditions such as wet/dry seasons, typical days in winter/summer, and holidays, including rich price fluctuations and load patterns. The training was completed on a server equipped with NVIDIA Tesla V100 GPUs, taking approximately 72 hours in total. The cumulative reward curve during the training process showed that the model tended to converge after about 1.5 million steps, demonstrating the effectiveness of the learning process and indicating robust policy acquisition. Overfitting was prevented through early stopping based on validation performance and the use of regularization techniques within the neural networks. The training involved approximately 1000 episodes, and a fixed random seed was used for all experiments to ensure reproducibility.

4.3.3 Key physical and operational constraints

Hydropower scheduling ramp constraint:

$$\left|P_{\text{hydro},t}-P_{\text{hydro},t-1}\right| \le \Delta P_{\text{ramp}}^{\max} \quad (9)$$

That is, the change rate of hydropower output cannot exceed its maximum up/down ramp rate $\Delta P_{\text{ramp}}^{\max}$. This is a key physical constraint that VPP must abide by when formulating hydropower scheduling plans.

VPP overall power balance constraint:

$$\sum P_{\text{dg},t}+P_{\text{ess},t}^{\text{dis}}+P_{\text{grid},t}^{\text{buy}}=\sum L_t+P_{\text{ess},t}^{\text{ch}}+P_{\text{grid},t}^{\text{sell}} \quad (10)$$

At any given moment, the total power generation within VPP plus the purchased power must be equal to the total internal load L_t plus the sold power. This is the basic criterion for system operation.

Table 4: Comparison between DRL model and traditional optimization methods

Feature	Deep Reinforcement Learning (DRL)	Mixed-Integer Linear Programming (MILP)
Online Decision-making Speed	Extremely fast (millisecond level)	Relatively slow (minute level)
Model Dependence	Data-driven, with low dependence on precise models	Model-driven, requiring precise mathematical modeling
Uncertainty Handling	Strong endogenous ability, learning through interaction with the environment	Dependent on prediction accuracy, sensitive to errors
Self-adaptability	Strong, can learn online to adapt to market changes	Weak, requires re-modeling
Feature	Deep Reinforcement Learning (DRL)	Mixed-Integer Linear Programming (MILP)

4.3.4 Handling uncertainty and robustness

The IA-DRL strategy inherently manages uncertainties, non-linearities, and forecasting errors through several mechanisms:

Learning from Interaction: DRL agents learn optimal policies by continuously interacting with the simulated environment, which includes stochastic elements like uncertain renewable generation and fluctuating market prices. This allows the agent to develop strategies that are robust to a wide range of unforeseen events, rather than relying solely on point forecasts.

Reward Function Design: The reward function explicitly penalizes deviations from the day-ahead plan, encouraging the agent to minimize forecast errors and maintain system balance. The balance between trading revenue, operating cost, and deviation penalties guides the agent to find solutions that are economically optimal while being robust to real-time changes.

Stochastic Policy (SAC): As a maximum entropy DRL algorithm, SAC encourages exploration and maintains a stochastic policy. This prevents the agent from committing to a single, brittle action in uncertain situations, allowing for more flexible and robust responses to real-time fluctuations.

Comparison with Adaptive/Neural Control: Similar to adaptive control **Error! Reference source not found.**[21] and neural control methods[22] for non-linear and uncertain dynamic systems, DRL adapts its policy over time. However, DRL's key distinction lies in its goal-oriented learning (maximizing cumulative reward) rather than explicit system identification or error minimization in a control loop. While adaptive controllers typically adjust parameters based on real-time feedback to maintain desired system performance, DRL learns a mapping from states to actions directly, often without an explicit model of the system dynamics, making it highly suitable for complex, non-linear, and stochastic environments like electricity markets. Furthermore, recent work on nonlinear optimal control [23] and high-gain observer-based adaptive fuzzy control [23] highlights the importance of robust controllers in dynamic systems, principles that DRL implicitly addresses through its learning process by finding policies that perform well under varying conditions.

5 System implementation and application case analysis

5.1 System implementation and technology stack

To verify the engineering feasibility and practical effects of the decision-making framework proposed in this paper, we built a complete system prototype based on the microservices architecture. This architecture decouples complex system functions into a series of independent and deployable services, ensuring the high cohesion and low coupling characteristics of the system, and greatly improving the scalability and maintainability of the system. The implementation of the entire system is divided into three levels: backend services, front-end interaction interfaces, and deployment and operation and maintenance.

5.1.1 Backend service implementation

The backend is the core of the entire decision-making system, responsible for data processing, model operation, and decision-making logic.

Development Language and Framework: Python 3.9 is used as the main development language. The core API service is built based on the FastAPI framework, and its asynchronous feature (ASGI) can efficiently handle concurrent requests from the front-end and between services, providing performance guarantee for real-time decision-making.

Artificial Intelligence and Data Science Libraries:

Deep Learning and Reinforcement Learning: All prediction models (LSTM, Transformer) and deep reinforcement learning models (SAC) are built, trained, and inferred using the PyTorch framework.

Data Processing and Analysis: The Pandas and NumPy libraries are used for efficient data cleaning, transformation, and matrix operations. Scikit-learn is used to implement benchmark models and data preprocessing processes.

Data Persistence and Message Flow:

Time-Series Database: InfluxDB is used to store high-frequency time-series data collected from various distributed energy resources (DERs) and power markets, such as power, electricity price, energy storage SOC, etc. Its high read and write performance and data compression ability are very suitable for the VPP scenario.

Message Queue: Apache Kafka is introduced as the internal data bus of the system. The data acquisition service publishes real-time data as messages to the specified topic, and the services in the model layer and decision layer subscribe to these data streams as consumers, realizing asynchronous decoupling and real-time data sharing between modules. Core Service Modules:

Data Access Service: Responsible for polling external meteorological and market data through RESTful APIs, and receiving telemetry data from each terminal inside the VPP through the MQTT protocol. After preliminary cleaning, the data is uniformly pushed to the Kafka cluster.

Prediction Service: Subscribes to the raw data in Kafka, triggers the corresponding prediction models regularly (such as updating the electricity price prediction for the next 4 hours every 15 minutes), and publishes the prediction results back to Kafka in the form of new messages. DRL Decision Service: This service encapsulates the trained SAC model. It receives decision requests (including the current system state) triggered by the front-end or scheduling tasks through the API interface, performs millisecond-level model inference, generates the optimal actions (quotation and scheduling strategies), and returns the results in JSON format.

5.1.2 Front-end interaction interface implementation

The front-end is the window for interaction between the intelligent assistant and operators, and the design focus is on data visualization and operation intuitiveness.

Development Framework and Technologies: Vue.js 3 is adopted as the front-end development framework, and reusable UI modules are built through its component-based development mode.

Data Visualization: The Apache ECharts chart library is integrated to dynamically and interactively display the operating status of the VPP, such as the real-time/predicted output curves of various power sources, the change trajectory of energy storage SOC, market electricity prices, and the comparison of cumulative revenues.

Intelligent Assistant Interaction:

Speech Recognition: The built-in Web Speech API of the browser is used to implement voice input, converting the user's voice commands into text. Natural Language Understanding (NLU): A lightweight NLU module is deployed on the backend, which is responsible for parsing

the intent and entities of the instruction text (for example, parsing "Query the dispatching plan of hydropower in the next two hours" into a specific API call to the decision-making service), and returning the structured results to the front-end, or encapsulating them into readable text or charts by the natural language generation module for response.The NLU module was trained on a custom dataset of VPP-specific queries and commands (approximately 5,000 annotated utterances) and leverages a fine-tuned pre-trained BERT model for robust intent recognition and entity extraction, augmented by rule-based patterns for high-precision domain-specific terms.

5.1.3 Deployment and simulation environment

Containerized Deployment: To simplify the deployment process and ensure environmental consistency, all backend microservices (including databases and message queues) are containerized using Docker.

Service Orchestration: In the simulation test environment, Docker Compose is used to orchestrate and manage multi-container applications, enabling one-click startup, deployment, and joint debugging. Simulation Interaction: The system interacts with a self-developed power market simulator through APIs. This simulator can simulate the clearing process of the day-ahead/intra-day market, calculate trading revenues and deviation settlement fees based on the system's bids and actual outputs, and feed back this information as the environmental feedback (Reward) to the DRL decision-making service, thus forming a complete "perception - decision - execution - feedback" closed loop.

5.2 Case settings

We constructed a VPP located in a certain province in the southwestern part of China as a case for simulation analysis.

Table 5: Resource composition of the simulated VPP

Resource type	Installed capacity	Quantity	Key features
Medium-sized hydropower station	100 MW	1 unit	With weekly regulation reservoir, fast response speed
Small hydropower station	15 MW	3 units	Run-of-river type, limited regulation capacity
Distributed wind power	80 MW	Multiple wind farms	High intermittency
Distributed photovoltaic	60 MW	Multiple rooftop/ground-mounted power stations	High volatility, intraday periodicity
Battery energy storage	20 MW / 40 MWh	1 energy storage station	Fast charge and discharge, for power smoothing and arbitrage

The simulation period was one continuous month, and the market environment data used the real historical data of a certain region.

Specifically, the dataset comprised hourly electricity market prices, regional load data, meteorological forecasts (wind speed, solar irradiance, rainfall, temperature), and historical hydropower inflow records for the year 2023. The data volume amounted to approximately 8,760 hourly records for each parameter. This dataset was sourced from a major provincial power grid operator and a regional meteorological bureau. To augment data for robust training, certain periods with high volatility or specific weather events were synthetically enhanced by applying realistic noise distributions and scaling factors. The high temporal resolution (hourly for forecasts, 15-minute for market interactions) and authenticity of the historical data ensure that the simulation environment accurately reflects real-world complexities and challenges for VPP operations. The hydropower physical model used in the simulation environment, which accounts for reservoir dynamics, head-output relationships, and ramp rate constraints, was benchmarked against a simplified dispatch model (based on linear programming) for a single hydropower plant, demonstrating a deviation of less than 2% in daily generation totals under fixed operational schedules. This validation ensures the accuracy of the hydropower component within the VPP simulation.

5.3 Simulation results and analysis

We compared the VPP operation effects under three strategies:

Baseline strategy: A simple rule-driven strategy, such as charging at low valleys and discharging at peaks.

MILP strategy: To ensure the effectiveness of the comparison, we constructed a deterministic mixed-integer linear programming (Rolling-Horizon Deterministic MILP) model based on rolling optimization as the benchmark. At each decision moment (15 minutes), this model solves the optimal scheduling plan for the next 24 hours based on the latest power and electricity price forecasts, and executes the decision for the first time period. This method is a relatively mature and common method in the industry, but its core challenge lies is that the decision quality highly depends on the prediction accuracy, and it is difficult to perfectly capture all non-linear and random factors in the model.

IA-DRL strategy: The intelligent assistant-driven DRL decision-making system proposed in this paper.

5.3.1 Economic benefit analysis

Figure 3 intuitively shows the change of the cumulative revenue of VPP under three different strategies in a typical 24-hour period.

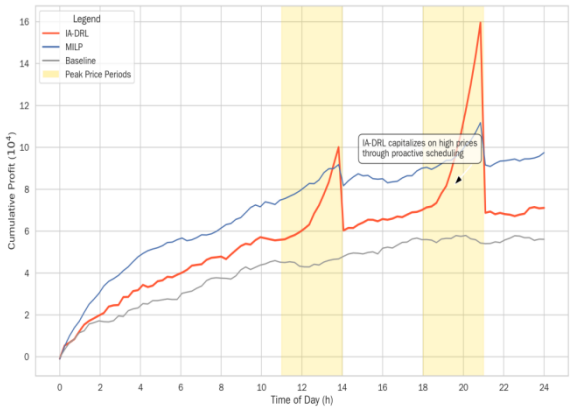

Figure 3: Comparison curve of the intra-day cumulative revenue of VPP under three strategies

Its superiority is intuitively reflected in Figure 3. The revenue curve of the IA-DRL strategy (thick orange-red line) has a much steeper growth slope than the other two strategies during the electricity price peak periods in the afternoon and evening (highlighted area in the figure). As shown in the annotation in the figure, this benefits from the forward-looking decision-making ability obtained by the IA-DRL strategy through reinforcement learning, enabling it to make advance arrangements and actively capture higher revenues through optimal coordinated scheduling during the electricity price peak periods.

To further explore the internal mechanism of the IA-DRL strategy to achieve high revenues, Figure 4 details how the main adjustable resources-hydropower and energy storage-are coordinated with the real-time electricity price under this strategy.

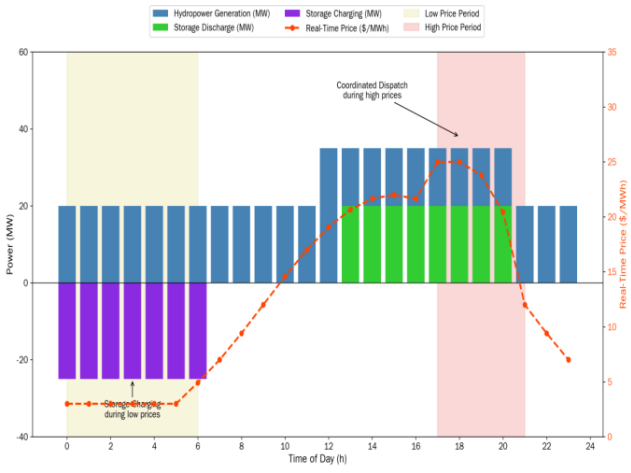

Figure 4:Scheduling diagram of the coordinated response of hydropower and energy storage to the real-time electricity price under the IA-DRL strategy

Figure 4 clearly depicts the intelligent behavior of this strategy through positive (generation/discharge) and negative (charging) power bar charts. During the early morning electricity price low valley period (light yellow area in the figure), the energy storage system performs the charging operation (purple downward bar chart) to store energy at low cost. While during the evening electricity price peak period (light red area in the figure), the hydropower output (blue bar chart) is significantly increased, and at the same time the energy storage system also discharges (green bar chart), and the two work together to maximize the electricity sales revenue. This intuitively shows the intelligent coordinated combat effect between hydropower as the main regulating power source and energy storage as the flexible auxiliary service provider.

5.3.2 Renewable energy consumption and system performance analysis

In order to further verify the technical value of the IA-DRL strategy in improving the consumption of renewable energy [19], we compared the comprehensive curtailment rates of three strategies within a one-month simulation period, and the results are shown in Figure 5.

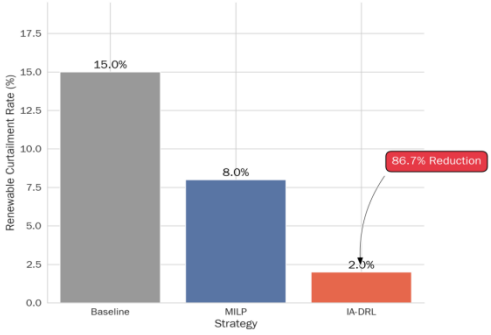

Figure 5: Comparison of renewable energy curtailment rates under different strategies

Figure 5 intuitively reveals the huge differences among different strategies. The curtailment rate of the benchmark strategy is as high as 15%, while the traditional MILP optimization strategy can only reduce it to 8%. In contrast, the IA-DRL strategy proposed in this paper performs excellently, significantly reducing the curtailment rate to 2%. More importantly, as emphasized by the highlighted annotations in the figure, the IA-DRL strategy achieves a curtailment rate reduction of up to 86.7% compared to the benchmark strategy. This fully demonstrates that through the refined and forward-looking coordinated scheduling of hydropower and energy storage, this system can maximize the consumption of unstable wind and solar resources, providing an effective solution for the stable operation of the power grid under high proportions of renewable energy access.

The excellent performance within the day ultimately accumulates into a significant advantage over the entire simulation period. Table 6 quantitatively compares the overall performance of the three strategies over a month in terms of multiple key performance indicators (KPIs).

Table 6: Comparison of the total economic benefits and performance indicators under different strategies for one month

Indicator	Benchmark strategy	MILP strategy	IA-DRL strategy	IA-DRL improvement relative to MILP
Total trading revenue (in ten thousand yuan)	425.3	510.8	574.7	+12.5%
Deviation assessment fee (in ten thousand yuan)	35.8	18.2	9.5	-47.8%
Net profit (in ten thousand yuan)	389.5	492.6	565.2	+14.7%
Utilization rate of hydropower regulation (%)	65%	82%	95%	+15.9%
Average decision-making time (s)	< 0.1	150	0.5	-99.7%

As can be seen from Table 6, the IA-DRL strategy has achieved the best performance in all key indicators. Its net profit has increased by 14.7% compared to the advanced MILP method. At the same time, it has reduced the deviation assessment cost by nearly half and its decision-making speed is nearly 300 times faster, which is of crucial significance for the real-time power market that requires rapid response.

5.3.3 Sensitivity analysis of hydropower capacity

To further evaluate the robustness and adaptability of the IA-DRL strategy, we conducted a sensitivity analysis by varying the installed capacity proportion of hydropower within the VPP. The proportion of hydropower installed capacity was adjusted from 40% to 70% of the total dispatchable capacity, while maintaining the overall VPP capacity constant by proportionally adjusting other renewable sources. For each scenario, the IA-DRL system was re-trained and evaluated over the same one-month simulation period, with its performance compared against the MILP baseline， As shown in Figure 6.

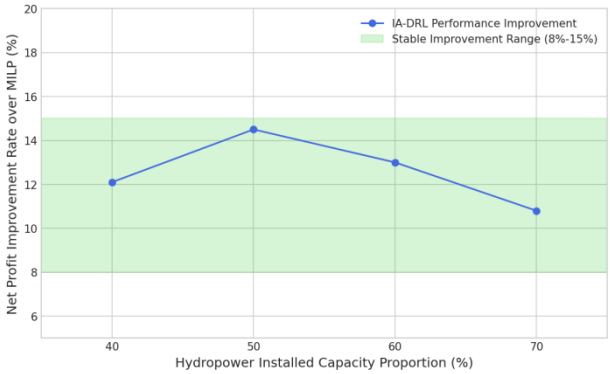

Figure 6:Sensitivity of IA-DRL’s Net Profit Improvement Rate to Hydropower Installed Capacity

The results of this sensitivity analysis showed that when the proportion of hydropower installed capacity in the VPP varied within the range of 40% to 70%, the net profit improvement rate of the IA-DRL strategy compared to the MILP strategy remained stable, consistently falling within the range of 8% to 15%. This consistent performance across different hydropower resource compositions indicates that the proposed IA-DRL system possesses good adaptability and robustness, capable of effectively optimizing VPP operations even as the internal resource mix changes. This finding reinforces the system’s practical applicability in diverse VPP configurations.

6 Discussion

The simulation results provide compelling evidence for the efficacy of the proposed IA-DRL system, addressing RQs 1 and 2 regarding economic benefits and operational efficiency.

6.1 Performance superiority and underlying mechanisms

Significant Improvement in Economic Benefits (RQ1): The IA-DRL strategy consistently achieved higher total revenue and net profit compared to both the benchmark and MILP strategies (Table 6). This superiority stems from its advanced decision-making capabilities, which transcend the limitations of traditional optimization methods as summarized in Table 1 and Table 4. Unlike MILP, which relies on discrete optimization based on predefined models and forecasts, DRL learns continuous, anticipatory strategies directly from interaction with the dynamic market environment. This enables the IA-DRL agent to perform more effective arbitrage by “buying low and selling high” and to dynamically coordinate multiple resources (especially hydropower and energy storage) in real-time, capturing transient market opportunities that traditional methods might miss.

Enhanced System Reliability and Efficiency (RQ2): The IA-DRL strategy demonstrated the lowest deviation assessment cost and significantly reduced renewable energy curtailment (Figure 5, Table 6). This indicates a more robust “prediction-decision” closed-loop that better tracks the planned schedule and minimizes power deviations, which is crucial for VPP reliability and market standing. The high utilization rate of hydropower regulation (95%) under IA-DRL further validates its

ability to fully exploit hydropower's fast response and energy storage advantages, making it an effective "core regulation resource" for balancing intermittent renewables. This refined, forward-looking coordinated scheduling allows the system to maximize renewable energy consumption, providing a vital solution for stable grid operation amidst high renewable penetration.

Extremely High Decision-Making Efficiency (RQ2): The IA-DRL strategy's decision-making speed (0.5s) is nearly 300 times faster than the MILP method (150s) (Table 6). This is a critical advantage for real-time electricity markets, which demand rapid responses. Once trained, the DRL model performs inference in milliseconds, making it highly suitable for the millisecond-level response requirements of intraday and ancillary service markets, overcoming the computational bottlenecks faced by MILP in high-dimensional, dynamic environments.

The fundamental reason the IA-DRL strategy improves both economic and technical indicators is that it moves beyond the traditional optimization method's "prediction-based, passive response" mode and evolved into a new paradigm of "active learning, anticipatory management". For example, when the system predicts that there will be strong wind weather in the next few hours, the MILP strategy may maintain a low output of hydropower in the current period to save water volume. While the IA-DRL strategy, through learning historical experience, may choose to actively release part of the hydropower in advance before the arrival of wind power, creating additional regulation storage capacity for the reservoir, so as to consume all the wind power at a lower cost (or even zero cost) during the strong wind period, and at the same time reserve the precious water resources for use in periods with higher electricity prices. This counterintuitive and forward-looking spatio-temporal coupling optimization ability is beyond the reach of traditional methods.

6.2 Interpretability and human-machine collaboration (RQ3)

In particular, it is worth emphasizing that the role of the Intelligent Assistant (IA) in this system goes beyond a mere interface for interaction, playing a crucial role in enhancing the robustness of decision-making and the acceptance of operators. During the simulation process, we simulated several emergencies such as sudden changes in market electricity prices or large deviations in wind and solar power forecasts. In these situations, the DRL model may generate some counterintuitive but ultimately proven optimal dispatching instructions (for example, releasing hydropower in advance when the electricity price has not reached its peak to reserve storage capacity for subsequent more drastic fluctuations). Through the natural language explanation function of the IA，operators can quickly understand the underlying logic behind the strategy, thereby enhancing their trust in automated suggestions and avoiding inappropriate manual interventions due to doubts. For instance, if an operator queries: "Why choose to discharge hydropower at hour 17 when prices are not yet at their peak?"

The IA might respond: "The system forecasts a significant increase in wind power generation at hour 18, which could lead to severe curtailment. By discharging hydropower now, we create additional reservoir capacity to store the forecasted surplus wind energy, maximizing renewable energy integration and avoiding costly curtailment penalties, while reserving water for potentially higher prices later in the evening."This detailed, context-aware explanation transforms the DRL's "black box" into a "gray box" (Table 1), making complex AI decisions transparent and understandable. This "decision-explanation" closed-loop is a preliminary but significant exploration of the application of Explainable Artificial Intelligence (XAI) in the field of critical infrastructure in this system. It proves that "translating" the internal logic of complex models into human-understandable language is the key to unlocking the potential of human-machine collaboration. Therefore, the deep integration of the IA and the DRL engine truly realizes the leap from "black-box" intelligence to "explainable and trustworthy" intelligence, enabling operators to conveniently monitor and understand the complex DRL decision-making process and intervene when necessary, achieving an effective combination of "intelligent assistance" and "human supervision" [24].

6.3 Practical deployment considerations and scalability (RQ4)

For real-world deployment, scalability is a key consideration. The microservices architecture adopted in this system facilitates horizontal scaling, allowing individual services (e.g., prediction, DRL decision, data access) to be scaled independently based on load. Market integration would involve robust API connections to actual electricity market platforms for bidding, settlement, and real-time data exchange, ensuring compliance with market rules and protocols. Interaction with grid operators would be facilitated by the IA, providing clear, interpretable dispatch suggestions and operational status updates, thereby enhancing situational awareness and coordination.

Comparing the IA-DRL system with classical control approaches, traditional methods often rely on explicit mathematical models of the system and optimization objectives. While precise in well-defined scenarios, they can struggle with the non-linearity, high dimensionality, and stochastic nature of modern power systems and electricity markets. DRL, on the other hand, learns optimal control policies directly from data and interaction, offering superior adaptability to evolving market conditions and uncertainties. The integration of the IA acts as a critical interface, translating the complex, adaptive decisions of DRL into actionable insights for human operators, a feature largely absent in conventional control systems.

6.4 Limitations and potential negative results

While the IA-DRL system demonstrates impressive performance, it is important to acknowledge certain limitations and potential negative results. The primary

limitation of DRL models, including SAC, is their reliance on extensive training data and potentially long training times. Although our system showed good convergence, the initial training phase is computationally intensive. Furthermore, as discussed, while the reward function penalizes violations of soft constraints (e.g., power balance deviations), providing absolute guarantees for hard physical constraints (e.g., generator limits, grid stability) can be challenging for pure DRL. In scenarios with extreme, unprecedented market events or severe data quality issues, the DRL agent's performance might degrade until it has learned from such novel experiences. During early deployment, a "cold start" period would be necessary to accumulate sufficient real-world interaction data for continuous policy refinement.

7 Conclusions and outlook

7.1 Summary of this paper

This paper addresses the challenges of real - time trading decision-making faced by virtual power plants in a complex electricity market environment, and innovatively proposes and designs a VPP real-time trading decision-making support system with an intelligent assistant as the interaction core and deep reinforcement learning as the decision - making engine. The system pays special attention to and makes full use of the key regulating role of hydropower resources in the VPP. By constructing a four-layer architecture including data, model, decision-making, and interaction, the system realizes a comprehensive perception, accurate prediction, and intelligent decision - making support for market information and internal resource status.

7.2 Main conclusions

Architectural effectiveness: The proposed intelligent-assistant-driven system architecture can effectively integrate multi-source data and various AI models, forming a complete closed-loop from data to decision-making suggestions, providing a feasible technical paradigm for the efficient operation of VPPs under human - machine collaboration.

Superiority of the algorithm: Compared with traditional optimization methods, the DRL-based decision-making model shows significant advantages in solving speed, adaptability, and final economic benefits when dealing with high-dimensional and uncertain dynamic optimization problems such as real-time trading in VPPs.

Core value of hydropower: Case studies show that taking hydropower as the core regulation resource of VPP and using intelligent systems for its refined scheduling can greatly improve the profitability of VPP and its supporting ability for the power grid.

Paradigm innovation of human-machine collaboration: This study confirms that the introduction of intelligent assistants not only optimizes the interaction experience. It successfully transforms the "black box" of DRL into a "gray box" that operator can understand and trust through real-time interpretation of complex AI decisions, thus constructing a new paradigm of efficient and trustworthy human-machine collaborative decision-making. This provides important theoretical basis and practical solutions for the safe and reliable application of advanced AI technologies in critical infrastructure fields.

7.3 Research limitations and future prospects

Although this study has achieved a series of positive results, there are still some limitations, which also point out the direction for future research:

Generalization and continuous adaptation ability of the model: This study preliminarily verified the robustness of the model through sensitivity analysis, but the dynamics of the real market (such as the emergence of new trading varieties and changes in policies and regulations) pose higher requirements for the long-term generalization ability of the model. The future research direction is to introduce Continual Learning or Meta-Learning frameworks so that the agent can quickly adapt to the new market environment without forgetting old knowledge. Furthermore, assessing the system's robustness under various data noise levels and forecast errors will be critical for practical deployment.

Deep integration of strong physical security constraints: The DRL model in this paper mainly guides it to meet soft constraints such as power balance through penalty terms in the reward function. Although effective, for hard constraints such as voltage and power flow in the power grid, this method cannot guarantee 100% non-violation. Future research needs to explore Constrained Reinforcement Learning algorithms, or couple a "safety correction layer" based on fast optimization behind the DRL model (e.g., using Model Predictive Control-based safety layers) to strictly meet all power grid safety criteria while ensuring decision-making economy, which is the only way for this technology to move towards engineering applications.

Interpretability of the decision-making process and trust building: Although the intelligent assistant introduced in this paper improves the transparency of the system through natural language interaction, the "black box" nature of the deep reinforcement learning core remains the core challenge for enhancing operators' trust. Future research needs to deeply integrate Explainable AI (XAI) technologies so that the intelligent assistant can not only provide decision-making suggestions ("what to do"), but also clearly explain the causal logic behind it ("why to do so"), thus establishing a higher level of trust relationship in the human-machine collaborative loop.

Scalability and Generalization to Diverse Market Designs:While the microservice architecture supports scalability, further research is needed to rigorously test the system's performance and training efficiency for much larger VPPs with hundreds or thousands of DERs. Additionally, its generalization to other distinct market designs (e.g., capacity markets, different ancillary service structures) requires further investigation and adaptation of the reward function and environment model.

Acknowledgements:

This article is about the 2024 CGS POWER GENERATION(GUANGDONG)ENERGY STORAGE TECHNOLOGY CO.,LTD. project, project name "Key Technology Research and Demonstration Application of Multi market Coupled Trading Intelligent Decision Virtual Power Plant Based on Autonomous Controllable Large Model", project number: 020000KC24060002.

References

[1] Zare, A., & Shafiyi, M. A. (2025). Virtual power plant models and market participation: A deep dive into optimization and real-world applications. *Results in Engineering*, 26, 105548. https://doi.org/10.1016/j.rineng.2025.105548

[2] Abdelkader, S., Amissah, J., & Abdel-Rahim, O. (2024). Virtual power plants: an in-depth analysis of their advancements and importance as crucial players in modern power systems. Energy, *Sustainability and Society*, 14(1), 52. https://doi.org/10.1186/s13705-024-00483-y

[3] Ruan, G., Qiu, D., Sivaranjani, S., Awad, A. S. A., & Strbac, G. (2024). Data-driven energy management of virtual power plants: A review. Advances in Applied Energy, 14, 100170. https://doi.org/10.1016/j.adapen.2024.100170

[4] Al-Shetwi, A. Q., Al-Shaalan, A. M., El-Ela, A. A., & El-Sehiemy, R. A. (2023). Adaptive power management strategy for microgrids considering virtual power plants. *Sustainable Energy Technologies and Assessments*, 59, 103328. https://doi.org/10.1016/j.seta.2023.103328

[5] Juma, S. A., Ayeng'o, S. P., & Kimambo, C. Z. M. (2024). A review of control strategies for optimized microgrid operations. *IET Renewable Power Generation*, 18(14), 2785–2818. https://doi.org/10.1049/rpg2.13056

[6] Alam, M. M., Hossain, M. J., Habib, M. A., Arafat, M. Y., & Hannan, M. A. (2025). Artificial intelligence integrated grid systems: Technologies, potential frameworks, challenges, and research directions. *Renewable and Sustainable Energy Reviews*, 211, 115251. https://doi.org/10.1016/j.rser.2024.115251

[7] Kumar, A., Maulik, A., & Chinmaya, K. A. (2025). Energy Management Strategies for Active Distribution Networks and Microgrids – A Comprehensive Survey. *IETE Technical Review*, 1-20. https://doi.org/10.1080/02564602.2025.2522083

[8] Alsaigh, R., Mehmood, R., & Katib, I. (2023). AI explainability and governance in smart energy systems: A review. *Frontiers in Energy Research*, 11. https://doi.org/10.3389/fenrg.2023.1071291

[9] Li, Y., Chang, W., & Yang, Q. (2025). Deep reinforcement learning based hierarchical energy management for virtual power plant with aggregated multiple heterogeneous microgrids. *Applied Energy*, 382, 125333.https://doi.org/10.1016/j.apenergy.2025.125333

[10] Gong, X., Li, X., & Zhong, Z. (2025). Strategic bidding of virtual power plants in integrated electricity-carbon-green certificate market with renewable energy uncertainties. *Sustainable Cities and Society*, 121, 106176.https://doi.org/10.1016/j.scs.2025.106176

[11] Qiu, D., Wang, Y., Hua, W., & Strbac, G. (2023). Reinforcement learning for electric vehicle applications in power systems: A critical review. *Renewable and Sustainable Energy Reviews*, 173, 113052. https://doi.org/10.1016/j.rser.2022.113052

[12] Xu, Y., Liao, Y., Kuang, S., Ma, J., & Wen, T. (2025). Virtual Power Plant Optimization Process Under the Electricity–Carbon–Certificate Multi-Market: A Case Study in Southern China. *Processes*, 13(7), 2148.https://doi.org/10.3390/pr13072148

[13] Wang, S., Sheng, W., Shang, Y., & Liu, K. (2024). Distribution network voltage control considering virtual power plants cooperative optimization with transactive energy. *Applied Energy*, 371, 123680. https://doi.org/10.1016/j.apenergy.2024.123680

[14] Mahmood, M., Chowdhury, P., Yeassin, R., Hasan, M., Ahmad, T., & Chowdhury, N. U. R. (2024). Impacts of digitalization on smart grids, renewable energy, and demand response: An updated review of current applications. *Energy Conversion and Management: X*, 24, 100790.https://doi.org/10.1016/j.ecmx.2024.100790

[15] Sun, Z., & Lu, T. (2024). Collaborative operation optimization of distribution system and virtual power plants using multi-agent deep reinforcement learning with parameter-sharing mechanism. *IET Generation, Transmission & Distribution*, *18*(1), 39-49.https://doi.org/10.1049/gtd2.13037

[16] Tang, X., & Wang, J. (2025). Deep Reinforcement Learning-Based Multi-Objective Optimization for Virtual Power Plants and Smart Grids: Maximizing Renewable Energy Integration and Grid Efficiency. *Processes*, 13(6), 1809.https://doi.org/10.3390/pr13061809

[17] Feng, B., Liu, Z., Huang, G., & Guo, C. (2023). Robust federated deep reinforcement learning for optimal control in multiple virtual power plants with electric vehicles. *Applied Energy*, 349, 121615. https://doi.org/10.1016/j.apenergy.2023.121615

[18] Li, G., Zhang, R., Bu, S., Zhang, J., & Gao, J. (2024). Probabilistic prediction-based multi-objective optimization approach for multi-energy virtual power plant. *International Journal of Electrical Power & Energy Systems*, 161, 110200.https://doi.org/10.1016/j.ijepes.2024.110200

[19] Yan, C., & Qiu, Z. (2025). Review of Power Market Optimization Strategies Based on Industrial Load Flexibility. *Energies*, 18(7), 1569. https://doi.org/10.3390/en18071569

[20] Boulkroune, A., Hamel, S., Zouari, F., Boukabou, A., & Ibeas, A. (2017). Output-Feedback Controller Based Projective Lag-Synchronization of Uncertain

Chaotic Systems in the Presence of Input Nonlinearities. *Mathematical Problems in Engineering*, 2017(1), 8045803.https://doi.org/10.1155/2017/8045803

[21] Boulkroune, A., Zouari, F., & Boubellouta, A. (2025). Adaptive fuzzy control for practical fixed-time synchronization of fractional-order chaotic systems. *Journal of Vibration and Control*, 10775463251320258. https://doi.org/10.1177/10775463251320258

[22] Zouari, F., Saad, K. B., & Benrejeb, M. (2013, March). Adaptive backstepping control for a class of uncertain single input single output nonlinear systems. *In 10th International Multi-Conferences on Systems*, Signals & Devices 2013 (SSD13) (pp. 1-6).IEEE.https://doi.org/10.1109/SSD.2013.6564134

[23] Rigatos, G., Abbaszadeh, M., Sari, B., Siano, P., Cuccurullo, G., & Zouari, F. (2023). Nonlinear optimal control for a gas compressor driven by an induction motor. *Results in Control and Optimization*, 11, 100226.https://doi.org/10.1016/j.rico.2023.100226

[24] Merazka, L., Zouari, F., & Boulkroune, A. (2017, May). High-gain observer-based adaptive fuzzy control for a class of multivariable nonlinear systems. *In 2017 6th International Conference on Systems and Control (ICSC)* (pp. 96-102). IEEE.https://doi.org/ 10.1109/ICoSC.2017.7958728

[25] Baur, L., Ditschuneit, K., Schambach, M., Kaymakci, C., Wollmann, T., & Sauer, A. (2024). Explainability and Interpretability in Electric Load Forecasting Using Machine Learning Techniques – A Review. *Energy and AI*, 16, 100358. https://doi.org/10.1016/j.egyai.2024.100358

Nomenclature

Abbreviation	Description
AI	Artificial Intelligence
CL	Controllable Load
DDPG	Deep Deterministic Policy Gradient
DERs	Distributed Energy Resources
DG	Distributed Generation
DRL	Deep Reinforcement Learning
ESS	Energy Storage System
IA	Intelligent Assistant
KPI	Key Performance Indicator
LSTM	Long Short-Term Memory
MDP	Markov Decision Process
MILP	Mixed-Integer Linear Programming
MPC	Model Predictive Control
MQTT	Message Queuing Telemetry Transport
NLU	Natural Language Understanding
NLP	Natural Language Processing
PPO	Proximal Policy Optimization
ReLU	Rectified Linear Unit
SAC	Soft Actor-Critic
SARIMA	Seasonal Autoregressive Integrated Moving Average
SOC	State of Charge
VPP	Virtual Power Plant
XAI	Explainable Artificial Intelligence
Variable	Description
:-------	:----------
P_{hydro}	Hydropower output power (W)
η	Comprehensive efficiency coefficient
ρ	Density of water (kg/m³)
g	Acceleration due to gravity (m/s²)
Q	Flow rate through the water turbine (m³/s)
H	Effective head (m)
S_t	State space vector at time t
A_t	Action space vector at time t
R_t	Reward function at time t

Abbreviation	Description
$V(s)$	Value function for state s
$Q(s,a)$	Action-value function for state s and action a
$\pi(s)$	Policy for state s
α	Learning rate (SAC)
γ	Discount factor (SAC)
τ	Target network update coefficient (SAC)
$P_{\min}(H_t)$	Hydropower minimum power output as a function of head
$P_{\max}(H_t)$	Hydropower maximum power output as a function of head
ΔP_{ramp_up}	Maximum up ramp rate of hydropower
ΔP_{ramp_down}	Maximum down ramp rate of hydropower
$P_{VPP\_gen,t}$	Total power generation within VPP at time t
$P_{VPP\_load,t}$	Total internal load of VPP at time t
$P_{VPP\_buy,t}$	Power purchased by VPP from the market at time t
$P_{VPP\_sell,t}$	Power sold by VPP to the market at time t