Journal of Electronic Research and Application, 2025, Volume 9, Issue 6
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Intelligent Agent Analysis and Measurement Application Based on DeepSeek

Fengxi Gao*, Mingze Sun

State Grid Liaoning Electric Power Compant Limited Bconomic Research Institute, Shenyang 110015, Liaoning, China

**Author to whom correspondence should be addressed.*

Abstract: In the context of the integrated development of artificial intelligence and the power industry, the State Grid has carried out in-depth research on the research and development and application of the Guangming Power model. The model is based on the knowledge of the power industry, cognitive computing as the core, and knowledge service as the goal, and has the characteristics of large sample, large calculation, large parameters, large knowledge, and large tasks, providing important support for the intelligent transformation of the power industry. This study focuses on the efficient operation of Guangming Power large model, builds a trinity operation system of "model iterative optimization, sample full-process governance, and computing power resource collaboration", and drives the continuous improvement of model capabilities and compliance development through a two-level collaboration mechanism. With the goal of "building a strong and excellent Bright Power Large Model", the study clearly states that it is necessary to accelerate the construction of a two-level collaborative operation system, and promote model iterative optimization, capability evaluation, service monitoring and compliance review on a regular basis. At the same time, focusing on the whole process of R&D and application of large and small models, starting from four aspects: computing power planning and layout, allocation and scheduling, adaptation and optimization, monitoring and analysis, strengthening the application monitoring and analysis of two-level intelligent computing centers, and building an efficient computing power resource application and supply system to continuously improve its operation and service capabilities.

Keywords: Lightning power model; Two-level collaboration; Computational power optimization

Online publication: December 16, 2025

1. Introduction and significance of study

As a strategic and basic technology leading the future, artificial intelligence is becoming a key chess and winner in global scientific and technological competition and the strategic game of major powers. Generative large models represented by Deepseek, GPT, Sora, Doubao, and Qianwen have accelerated the evolution of AI from perception and understanding to generation and creation, and have continuously spawned new products, new formats and new models, and have increasingly become the core engine for creating new quality productivity.

At present, AIGC technologies such as DeepSeek and GPT are driving the explosive development of artificial intelligence. AIGC (Artificial Intelligence Generated Content) is a new type of content creation method after professional production content (PGC) and user-generated content (UGC), which includes tasks such as generating text, speech, code, images, videos, robot actions, and more. Power artificial intelligence is a "special artificial intelligence" formed by the integration and innovation of related theories, technologies and methods of artificial intelligence with the physical laws of the power system, technology and knowledge. The large model of the power industry will be based on industry knowledge, with cognitive computing as the core, and knowledge service as the goal, with the characteristics of large sample, large calculation, large parameters, large knowledge, and large tasks.

2. Guangming power large model capabilities and characteristics

The Guangming Power Model is a 100-billion-parameter, multimodal system distinguished by its comprehensive power-domain knowledge, large-scale architecture, and strong professional competencies. Characterized by "full-series, multi-modality, multi-scale, full autonomy, and full specialization," it demonstrates five core power-specific capabilities:

(1) Its power knowledge memory is highly robust, enabling deep comprehension of electrical power theories, guidelines, standards, and operational systems, along with flexible application of this knowledge;

(2) It excels in multimodal fusion analysis, effectively integrating power-related diagrams, textual materials, and time-series data to uncover intermodal correlations and support diagnostic assessments;

(3) The model shows strong power business logical reasoning, conducting inference based on industry workflows and operational rules to assist in system-state evaluation and decision support;

(4) It performs reliably in basic electric-power numerical computation, providing accurate analytical outputs to support system analysis and stability verification [1,2];

(5) The model offers advanced power-content assisted generation, producing design schemes, emergency plans, analytical reports, and other technical documents aligned with professional power-industry specifications.

In terms of development strategy, the model follows an R&D route defined by an "open-closed parallel, pre-training + fine-tuning" approach. Built upon a general foundational model, Guangming Power undergoes comprehensive parameter-enhanced training and domain-specific fine-tuning, integrating extensive expert knowledge and industry experience to elevate the base model into a power-industry expert. The construction concept adopts "incremental pre-training + fine-tuning" to develop a sector-oriented foundational model for the power domain, followed by "task-level fine-tuning" to produce specialized scenario models tailored to specific operational needs [3,4].

3. Application scenarios of Guangming power large model

In the marketing scenario, inefficiencies and suboptimal quality in power supply planning have prompted the development of an automated power-supply-plan generation agent. Utilizing single-grid map path-planning capabilities in combination with grid data, line-breaking capacity, and related planning information, the system employs retrieval-augmented generation (RAG) techniques to automatically infer and generate electricity-pricing strategies and billing schemes.

In the scheduling scenario, where load-transfer decisions in distribution networks have traditionally relied heavily on expert experience and are influenced by numerous complex factors, technologies such as RAG and

Tree-of-Thought reasoning are applied to comprehensively incorporate twelve categories of constraints, including power loss and equipment overload, to achieve intelligent generation and optimal evaluation of load-transfer plans for events such as fault handling and heavy-load mitigation.

In the materials scenario, challenges related to the labor-intensive preparation of bidding documents and the risk of human oversight during auditing are addressed by focusing on critical procurement processes. Through RAG-based knowledge retrieval enhancement and prompt-engineering techniques, a knowledge-slice index is constructed to ensure contextual consistency and traceability in the generation and review of bidding documents, thereby enabling intelligent assistance and advancing business-model innovation. These capabilities have been deployed in production environments. In the smart office scenario, multimodal inputs, interactive guidance, and related techniques have been used to meet the high-frequency daily office needs of personnel across organizational levels.

A suite of nine capabilities spanning intelligent writing, format conversion, and precise retrieval has been implemented and made accessible through multiple channels, including AI assistants. With nearly 30,000 daily active users and more than one million cumulative service calls, these tools have facilitated broad adoption and have made large-model capabilities perceptible and accessible to all employees [5–7].

4. Research on intelligent agent driven systems

For the current artificial intelligence project, the functional levels of the existing system should be further divided and the project measurement standards at each level should be clarified. To meet the differentiated needs of employees at all levels of the enterprise for legal and regulatory knowledge, the construction of a specialized intelligent question-answering engine has become a key measure to enhance the efficiency of compliance management. This engine focuses on “precise response, dynamic adaptation, and full-scenario coverage”, and achieves the entire process service from information retrieval to decision support through the collaboration of multiple modules [8–10].

DeepSeek drives an efficient and accurate intelligent evaluation system through its self-developed large model technology. This system can automatically parse complex text, code, and data, achieve multidimensional analysis and cross validation, and greatly improve the consistency and reliability of reviews. At the same time, DeepSeek supports flexible customization of review rules and knowledge bases, combined with continuous optimization based on feedback from human experts, to ensure that review results are both compliant with standards and have practical implementation value. Its human-machine collaboration mechanism effectively reduces labor costs, accelerates decision-making processes, and is widely used in high standard industries such as finance, scientific research, and engineering [11,12].

4.1. The fundamental principle

By integrating the RAG theory with the multi-agent system theory, a closed-loop architecture of “knowledge retrieval, role understanding, professional decision-making, collaborative arbitration” is realized.

$$\text{AgenticRAG}=\{A.KR,C)$$

It is a collection of role definitions, where represents the set of agents, represents the set of knowledge bases, represents the set of role definitions, and represents the collaborative mechanism. Research on the integration of knowledge enhancement and agent decision-making technology has broken through the limitations of the traditional “RAG” model.

4.2. The multi-agent intelligent agent construction technology based on the DeepSeek large model

Traditional large language models often lack clearly defined professional role positioning and therefore struggle to accurately reproduce the cognitive patterns and behavioral characteristics of specialized review experts, resulting in limited domain-specific review capability. To address this limitation, it is essential to analyze and delineate expert review roles in alignment with actual business requirements, thereby providing a structured foundation for professional role construction [3,4]. In this context, role-based prompt engineering algorithms offer a viable solution. The technical challenge lies in designing prompt mechanisms that enable large models to faithfully represent the attributes of specific expert roles. The core principle involves developing structured prompt templates that guide the model to emulate designated professional identities and execute specialized behaviors. The role-based prompt generation process can be expressed as:

$$P_{\text{role}} = f(\text{Role}_i, \text{Task}, \text{Context}, \text{Examples})$$

where P_{role} denotes the role-oriented prompt, comprising multiple layers, including role definition, knowledge activation, task guidance, and behavioral norms. This multi-level prompt structure supports precise role positioning and enhances the model's ability to simulate professional reasoning and behavior patterns effectively.

4.3. Multi-agent collaborative decision-making optimization algorithm

In the context of expert review tasks, a key technical issue lies in optimizing multi-agent collaborative decision-making so that multiple role-based agents can coordinate effectively while preserving the distinct professional perspectives of each role. This challenge highlights the research significance of developing algorithms capable of achieving balanced, cross-perspective coordination. To address this need, a collaborative decision-making optimization method based on the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) framework is proposed. The learning update for the role-specific agent is expressed as:

$$\nabla_{\theta_r} J(\mu_r) = \mathrm{E}_{s,a\sim \mathrm{D}}\left[\nabla_{\theta_r}\mu_r(o_r)\nabla_{a_r} Q_r^{\mu}(s, a_1, \ldots, a_R)|_{a_r=\mu_r(o_r)}\right]$$

By designing role-differentiated reward functions and observation spaces, the algorithm enables each agent to retain its professional independence while contributing to a coordinated and globally optimized decision-making process. This approach substantially enhances the system's ability to integrate multi-perspective expertise within complex review scenarios.

4.4. Role-based knowledge-enhanced retrieval

The traditional RAG technology adopts a unified search strategy, which is unable to meet the differentiated knowledge needs of different professional roles. The core principle of multi-agent collaboration lies in the communication, coordination, and cooperation among multiple autonomous agents to jointly complete complex tasks. Each agent has independent perception, decision-making and execution capabilities, and exchanges information, negotiates goals, and assigns roles through shared environments or communication protocols (such as blackboard systems, messaging, or agent communication language ACLs). Systems typically adopt centralized, distributed, or hybrid architectures, utilizing collaborative planning, contract network protocols, voting mechanisms, or reinforcement learning methods to achieve task decomposition and dynamic coordination.

The key challenges include avoiding conflicts, resolving divergent goals, ensuring consistency, and optimizing global efficiency. Multi agent systems, through division of labor, cooperation, and knowledge fusion, can surpass the limitations of individual intelligence and achieve more efficient, robust, and scalable solutions to complex problems. They are widely used in fields such as autonomous driving fleets, distributed energy grids, and swarm robots.

Another key technical challenge lies in optimizing knowledge retrieval to meet the differentiated needs of various professional roles. To address this, a role-based hybrid similarity search algorithm is proposed, enabling the search strategy to be tailored to the characteristics and knowledge demands of distinct expert roles. The algorithm integrates the strengths of dense vector retrieval and sparse lexical matching, and its core computation is defined as:

$$\mathrm{Sim}_r(q,d)=\alpha_r\cdot\cos\left(v_q^r,v_d\right)+(1-\alpha_r)\cdot\mathrm{BM25}_r(q,d)$$

where the weighting factor α_r adjusts the balance between semantic similarity and lexical relevance according to the requirements of role r.

This role-aware customization provides differentiated and precise retrieval capabilities that enhance knowledge support across varying professional review contexts.

To further strengthen knowledge acquisition, an active knowledge retrieval strategy is designed to shift intelligent agents from passive response patterns to proactive information seeking. Under this mechanism, each agent autonomously plans and executes retrieval actions based on the assigned review task. The strategy is formalized as:

$$Q_r=\mathrm{Plan}_r(T,K_r,E_r)=\{q_1,q_2,\ldots,q_n\}$$

$$D_r=\bigcup_{i=1}^{n}\mathrm{Retrieve}_r(q_i,K)$$

where Q_r denotes the set of queries generated by role r for task T, and D_r represents the resulting knowledge set obtained from iterative retrieval.

Together, these mechanisms enhance the forward-looking nature, completeness, and role alignment of knowledge acquisition in complex review scenarios.

5. Conclusion

This study centers on enhancing the operational efficiency of models, datasets, and computing resources within the Guangming Power large-model framework. By constructing an integrated "three-in-one" operational system, comprising iterative model optimization, full-lifecycle sample governance, and coordinated management of computing power, the study establishes a two-level collaborative mechanism that drives the continuous advancement and compliant development of Guangming Power's large-model capabilities. Guided by the objective of "strengthening and refining the Guangming Power Model," the two-level collaborative operation system will be further accelerated through routine activities such as model iteration and optimization, capability evaluation, service monitoring, and compliance review. From an end-to-end perspective on the R&D and deployment of models of varying scales, the work prioritizes four key areas: computing power planning and layout, resource allocation and scheduling, adaptation and optimization, and monitoring and analytics. By reinforcing application-level monitoring across both tiers of intelligent computing centers and establishing an efficient system for computing power application and resource provisioning, the operational and service capabilities of the two-level intelligent computing infrastructure will continue to be comprehensively enhanced.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Liu W, Gan Z, Xi T, et al., 2022, A Semantic and Intelligent Focused Crawler based on Semantic Vector Space Model and Membrane Computing Optimization Algorithm. Applied Intelligence, 53(7): 7390–7407.

[2] Yeh P, Puri C, Kass A, 2010, A Knowledge based Approach for Capturing Rich Semantic Representations from Text for Intelligent Systems. Int. J. of Advanced Intelligence Paradigms, 2(1): 33–48.

[3] Andon P, Rogushina J, Grishanova I, et al., 2020, Experience of the Semantic Technologies Use for Intelligent Web Encyclopedia Creation (On Example of the Great Ukrainian Encyclopedia Portal). CEUR Workshop Proceedings, 2866(2020): 246–259.

[4] Bai X, Wei G, Ye T, 2024, Analysis and Optimization Scheme of the Effectiveness of Word2vec in Intelligent Recommendation of Traditional Chinese Medicine Folk Prescriptions. Frontiers in Computing and Intelligent Systems, 10(3): 95–100.

[5] Hadi E, Kishoro B, Rachita M, et al., 2023, Semantics Aware Intelligent Framework for Content-Based E-Learning Recommendation. Natural Language Processing Journal, 2023(3).

[6] Kamaleddin Y, Ali F, 2022, An Approach for Semantic Interoperability in Autonomic Distributed Intelligent Systems. Journal of Software: Evolution and Process, 34(10).

[7] Zhou H, Shen Y, Liu X, et al., 2020, Survey of Knowledge Graph Approaches and Applications. Journal on Artificial Intelligence, 2(2): 89–101.

[8] Gong Y, Ma F, Wang H, et al., 2025, The Evolution of Research at the Intersection of Industrial Ecology and Artificial Intelligence. Journal of Industrial Ecology, 29(2): 440–457.

[9] Haghighi M, Joseph A, Kapetanios G, et al., 2025, Machine Learning for Economic Policy. Journal of Econometrics, 2025(249): 105970.

[10] Flores-Becerra G, Tlelo-Cuautle E, Polanco-Martagon S, 2009, Applying Fuzzy Sets Intersection in the Sizing of Voltage Followers. CEUR Workshop Proceedings, 2009(533): 209–216.

[11] Schulze L, Douglas B, Raffo G, 2022, Fast Computation of Binary Search Tree for PWA Functions Representation using Intersection Classification. Automatica, 141(5): 110217.

[12] Kishore D, George A, 2023, Empowering Novel Scholarship at the Intersection of Machine Learning/Deep Learning and Ecology. Ecological Informatics, 2023(77).

[13] Mohammad H, Reza A, Hoda M, Hyperideal-Based Intersection Graphs. Indian Journal of Pure and Applied Mathematics, 54(1): 120–132.

[14] Schwartmann R, Keber T, Zenner K, et al., 2024, Data Protection Aspects of the Use of Artificial Intelligence: Initial Overview of the Intersection between GDPR and AI Act. Computer Law Review International, 25(5): 145–150.