Advances in Computer, Signals and Systems (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2025.090410
ISSN 2371-8838 Vol. 9 Num. 4

Research on Multimodal Reasoning and Self-Verifying Agents Based on the Brightness Large Model for Report Materials

Jing Xie[1,a,*], Shilong Li[1] , Chuan Huang[1] , Xiangjun Kong[1] , Yongjie Zhu[1]

[1]State Grid Shanghai Electric Power Company Shinan District Power Supply Company, Shanghai, China
[a]jingxie1115@163.com
**Corresponding author*

Keywords: Multimodal Reasoning, Intelligent Agent, Smart Grid Management

Abstract: Manual review of complex State Grid documentation suffers from inefficiency and oversight limitations. To address these issues, this research proposes an intelligent agent framework based on the Brightness Large Model for automated verification. The methodology integrates three core components. First, a shared semantic space fuses text, table, and diagram data to enable deep understanding and structured summarization. Second, a Retrieval-Augmented Generation system maintains a dynamic knowledge base to ensure strict alignment with evolving regulations. Third, a multi-agent pipeline facilitates collaborative rule matching, inconsistency detection, and automated revision. This system provides robust risk warnings and decision support, optimizing resource allocation while advancing smart grid development and national energy security.

1. Introduction

As a cornerstone of national energy security and socioeconomic development, the efficient management of project lifecycles in the power industry directly influences smart grid construction, carbon neutrality goals, and the cultivation of new quality productive forces. Within this framework, project documentation verification serves as a critical safeguard for quality assurance. This process fulfills essential functions including validating the authenticity of technical achievements, quantifying resource-benefit conversions, and standardizing scientific research management.

In 2024, the State Grid Corporation of China released the Brightness Large Model. This multi-trillion-parameter multimodal foundation model provides core technological support for the intelligent upgrade of the power industry chain [1]. Despite this advancement, current report verification processes rely primarily on labor-intensive manual reviews and multi-departmental collaboration. This traditional approach suffers from low processing efficiency, oversight limitations, and susceptibility to human error. These deficiencies create a bottleneck for corporate digital transformation and fail to meet the demands of large-scale projects. Consequently, leveraging the technological capabilities of the Brightness Large Model to automate multimodal data fusion and ensure precise alignment with specialized rules has become an urgent priority.

To address these challenges, this paper proposes an intelligent agent framework for the multimodal reasoning and autonomous verification of project documentation. The methodology first constructs a multimodal reasoning module to achieve the deep integration and structured extraction of information from texts, tables, and diagrams. Furthermore, the study establishes a local knowledge base driven by Retrieval-Augmented Generation to ensure the authority and dynamic adaptability of verification rules [2]. Finally, a collaborative multi-agent pipeline is implemented to execute precise automated verification [3]. This system aims to drastically reduce review cycles from weeks to hours while minimizing oversight risks. The outcomes provide enterprises with comprehensive management tools that deliver data-driven insights for financial decision-making.

The contributions of this research are summarized as follows.

- First, the proposed framework transcends the limitations of traditional single-modal validation by achieving unified semantic understanding of multimodal data.
- Second, the system addresses the challenge of evolving industry standards through a dynamic rule repository with incremental update capabilities.
- Third, the multi-agent architecture enhances reasoning flexibility and decision-making precision in complex scenarios.

The remainder of this paper is organized as follows. Section 2 reviews the current state of relevant research. Section 3 details the proposed solution and system architecture. Section 4 validates the system performance through experiments. Section 5 summarizes the findings and outlines future research directions.

2. Relevant Work

This section reviews related work across three specific dimensions: Retrieval-Augmented Generation and industry knowledge base construction, multimodal data processing and reasoning techniques, and the application of Large Language Model agents in automated verification.

2.1. RAG and Industry Knowledge Base Construction

Retrieval-Augmented Generation technology addresses inherent limitations of large language models such as knowledge obsolescence and factual inaccuracies by integrating external knowledge bases [4]. As a core component, vector databases enable efficient indexing and precise matching of unstructured data through high-dimensional vector embeddings and semantic similarity search. Techniques including text chunking, dynamic data ingestion, and hybrid retrieval facilitate rapid access to industry regulations and technical standards [5].

In the power domain, the Meta-RAG framework proposes a metadata-driven architecture tailored for knowledge-intensive tasks. This approach utilizes modules for document conversion and hybrid encoding retrieval to construct datasets based on power industry specifications [3]. Regarding dynamic updates, existing research employs incremental update pipelines and re-ranking strategies to integrate new regulations without the high cost of model retraining [6].

Despite these advancements, current applications in the power sector exhibit distinct limitations. First, knowledge bases predominantly focus on textual documents and lack multimodal retrieval capabilities for tabular parameters or graphical engineering indicators. Second, rule matching often relies on keywords or coarse semantic similarity rather than the fine-grained validation required for numerical range verification or format specification adherence. Third, incremental update mechanisms lack the flexibility to accommodate the phased rule changes characteristic of power engineering projects.

2.2. Multimodal Data Processing and Reasoning Techniques

Multimodal data fusion and reasoning constitute core technologies for processing complex documents by achieving deep semantic alignment of heterogeneous information from texts, tables, and diagrams. Recent advances in Multimodal Large Language Models have demonstrated significant progress in document understanding. The DocRefine framework employs a multi-agent approach for scientific documents utilizing specialized agents for layout analysis and consistency checking [1]. This closed-loop architecture offers valuable references for document validation. To enhance reasoning capabilities, LMM-R1 utilizes a two-stage training strategy with rule-based reinforcement learning [2]. This method improves logical reasoning in tasks involving geometry and visual question answering and demonstrates the feasibility of transferring textual reasoning skills to multimodal scenarios. Furthermore, the RH-BrainFS strategy proposes a regional heterogeneity fusion mechanism using brain-inspired networks to extract local features while preserving modality-specific characteristics [7].

However, the application of multimodal technology within the power industry remains nascent. Frameworks such as Meta-RAG focus primarily on text and metadata Q&A while neglecting the deep analysis of tables and charts [3]. Existing research generally lacks coordinated validation solutions for technical reports and fails to address the necessity for consistency checks across textual descriptions, tabular data, and graphical trends.

2.3. Application of LLM Agents in Automated Verification

Large Language Model agents deliver flexible automation solutions through modular collaboration and workflow orchestration. The AutoAgent platform demonstrates robust capabilities in enterprise scenarios by automating tasks including invoice review and risk alerts [8]. This validates the utility of agents in compliance verification contexts. To ensure security and reliability, the AgentSpec framework introduces a runtime constraint mechanism. Through structured rule definitions involving triggers and predicates, it enables real-time monitoring and violation blocking of agent operations [9]. This modular design supports the definition of compliance boundaries in report verification and prevents the generation of non-compliant revision suggestions.

Regarding multi-agent collaboration, existing research predominantly adopts specialized module division and unified messaging protocols. Pipeline-based operations involving input preprocessing and rule querying enhance processing efficiency for complex tasks [10]. Nevertheless, applications in the power sector remain confined to generic workflows and lack adaptation to the specific verification requirements of complex engineering documentation [11].

3. Methodology

To address multimodal heterogeneity in power reports, dynamic validation rules, and process automation needs, this paper constructs a tripartite technical framework integrating multimodal reasoning, RAG knowledge base, and LLM agents to realize autonomous validation of report materials. As shown in Figure 1, the overall technical workflow is as follows: the RAG local knowledge base provides accurate support from industry rules and standards; the multimodal reasoning module parses heterogeneous data and extracts structured features; the LLM agent pipeline connects all modules to automate the entire process from input processing and rule matching to final validation decisions. Implementation methods of each core module are detailed below.

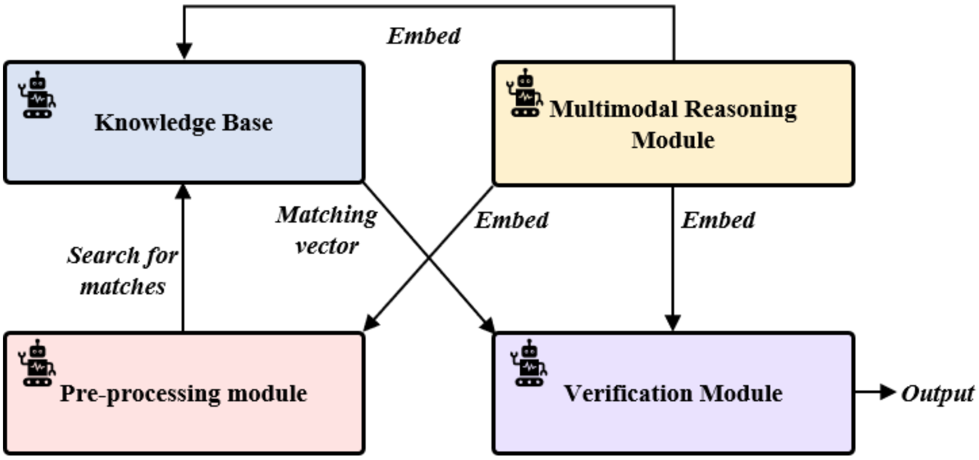

Figure 1: Technical Workflow of Autonomous Validation Framework.

3.1. Building Local Knowledge Bases Based on RAG

This module focuses on efficient storage, precise retrieval, and dynamic updating of power industry verification rules, providing authoritative knowledge support for intelligent verification and addressing knowledge lag and domain adaptability issues of large models.

3.1.1. Knowledge Base Data Preparation and Preprocessing

During the knowledge base construction phase, multi-source data is integrated, including power industry technical specifications, financial audit standards, project contract terms, and corporate validation requirements. This collection covers textual manuals, tabular parameter standards, and structured regulatory documents, ensuring comprehensive rule coverage across technical, financial, and formatting dimensions. The operational workflow is illustrated in Figure 2.

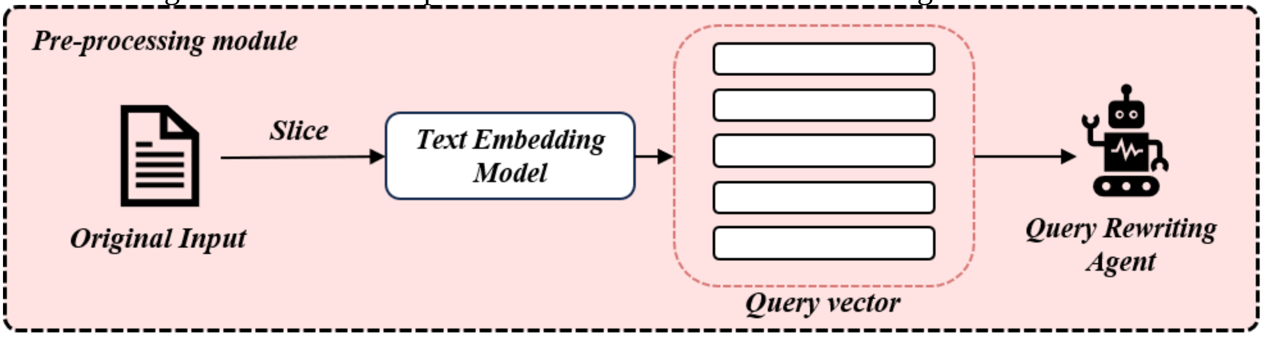

Figure 2: Preprocessing Workflow for Knowledge Base Retrieval.

For data cleansing and filtering, the Brightness Large Model's semantic comprehension capabilities eliminate verification-irrelevant redundant information while preserving core rule content. During text segmentation and annotation, the system employs a paragraph-semantic integrity prioritized strategy. This approach combines fixed-length segmentation with sentence boundary detection algorithms to decompose lengthy documents into independent knowledge fragments, effectively preventing semantic fragmentation. Concurrently, comprehensive labeling is performed across three validation categories: technical, financial, and formatting. This classification system enables precise subsequent retrieval while minimizing interference from irrelevant rules during the validation process.

3.1.2. RAG Architecture Integration and Retrieval Strategy

During RAG architecture integration, the retrieval module employs a hybrid strategy combining semantic similarity retrieval with keyword-based retrieval. As shown in Figure 3, structured query text from the multimodal reasoning module is first converted into vectors via an embedding model. Cosine similarity is then computed between these vectors and those in the knowledge base to obtain the top-5 most similar knowledge fragments. Simultaneously, keyword matching filters fragments containing core query terms. After merging and deduping both result sets, they are prioritized based on predefined rules to generate a candidate rule set. In the context fusion phase, the candidate rule set and original query text are concatenated into an enhanced prompt following the structure “Query Intent - Rule Basis - Related Cases,” ensuring the Brightness Large Model fully leverages rule information to generate verification conclusions. To ensure knowledge base timeliness, a rule update pipeline is designed. When new industry standards or corporate specifications are released, it automatically triggers text slicing, embedding, and vector storage processes. This enables rule synchronization through incremental updates without retraining the model, reducing maintenance costs.

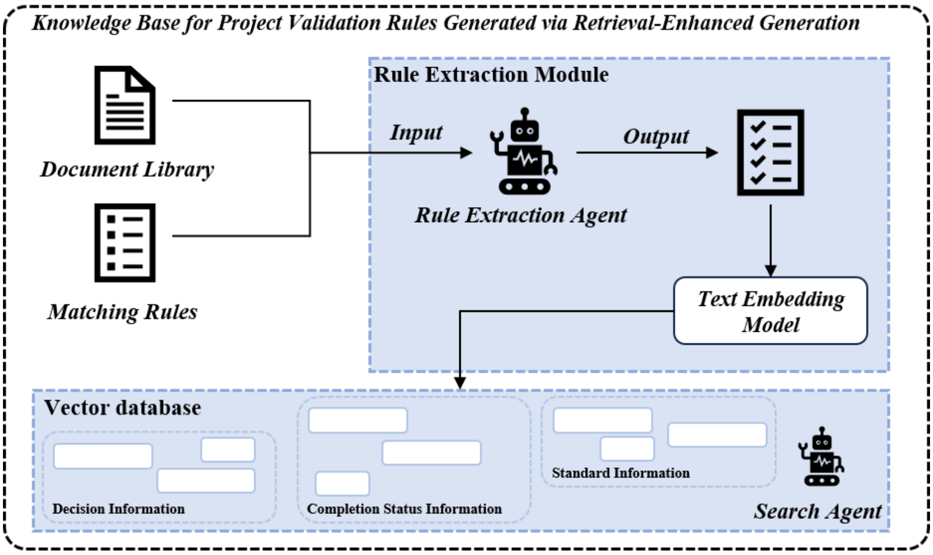

Figure 3: RAG Architecture for Project Validation Rules.

3.2. Multimodal Reasoning Module Based on Brightness Large Model

The core objective of this module is to overcome the limitations of single-modal analysis, enabling deep semantic understanding of multimodal data within power reports and cross-modal consistency verification. This provides highly reliable structured data support for subsequent agent decision-making.

3.2.1. Multimodal Data Preprocessing and Feature Extraction

This module adopts a multi-channel parallel processing architecture, as shown in Figure 4. It designs dedicated preprocessing procedures based on the characteristics of different modal data in report materials. For text data, first, a PDF and Word format standardization conversion tool unifies document formats, removing format noise like redundant spaces, special characters, and invalid HTML tags. Then, relying on the semantic segmentation technology of the Brightness Large Model, long texts are split into semantically complete sub-paragraphs, with core verification information such as technical terms, numerical indicators, and constraints retained. For table data, the Table Transformer table recognition algorithm is used to extract row-column structures and cell numerical information. Regular expressions verify the standardization of power-specific units such as “kW” and “kV”. Meanwhile, unstructured text in table notes is converted into “parameter-description” key-value pairs to ensure the association between numbers and semantic descriptions.

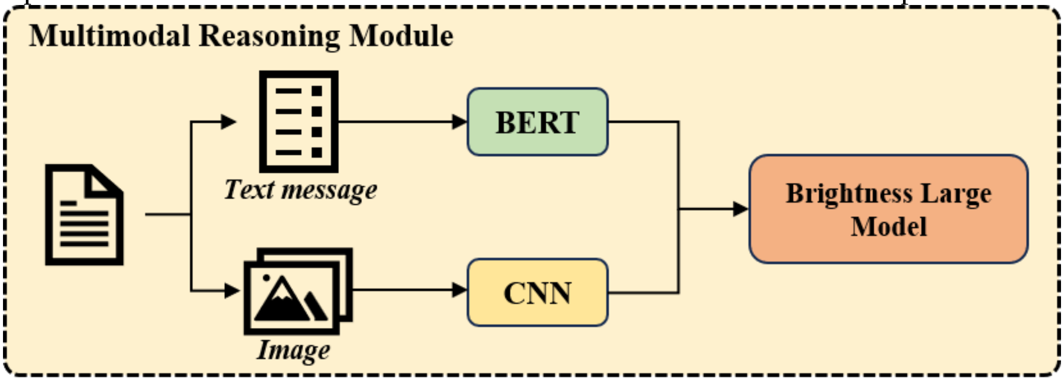

Figure 4: Workflow of Multimodal Reasoning Module.

For chart data, a CNN visual model extracts image features, and OCR technology recognizes chart titles, axis labels, and data annotation texts. Visual information is converted into structured numerical sequences and descriptive texts, laying the foundation for subsequent cross-modal association. In the feature extraction stage, text features generate semantic vectors via BERT. For table features, a fusion strategy of adding structural features and numerical features is adopted—structural features are generated based on column name semantic embedding, and numerical features are normalized to build feature vectors. For chart features, visual embedding and OCR text embedding are combined to form multimodal feature vectors with dimensions unified with those of text and table.

3.2.2. Cross-Modal Information Fusion Mechanism

To achieve effective integration of multimodal data, this module adopts an intermediate fusion strategy, constructing a fusion framework that preserves modality specificity while ensuring global semantic consistency. First, through contrastive learning training, feature vectors from text, tables, and charts are mapped to a shared semantic space. By minimizing intra-modal similarity loss and maximizing inter-modal association loss, comparability across different modalities is ensured. Second, a multi-head cross-modal attention layer is introduced to compute semantic similarity weights between different modal features. This layer prioritizes strengthening the associative mapping between textual descriptions, tabular values, and graphical trends. It achieves semantic alignment with corresponding field values in tables and power trend curves in graphs, capturing consistency between numerical data and textual descriptions. For scenarios where reports lack charts or tables, the module leverages the generative capabilities of the Brightness Large Model to supplement missing modal features through textual descriptions. This includes generating illustrative tables based on parameter ranges in the text or creating trend diagrams from numerical sequences, ensuring the continuity and integrity of the verification process.

3.2.3. Rule Validation

To achieve traceable and precise validation, the module establishes a neural symbolic reasoning mechanism that integrates explicit rules with implicit knowledge. As shown in Figure 5, the process begins with the Brightness Large Model parsing foundational documents including power industry regulations, technical standards, and project contracts. This parsing automatically extracts explicit validation rules covering numerical boundaries, formatting specifications, and unit requirements. These rules are then formalized as logical “IF-THEN” expressions to create a structured rule repository.

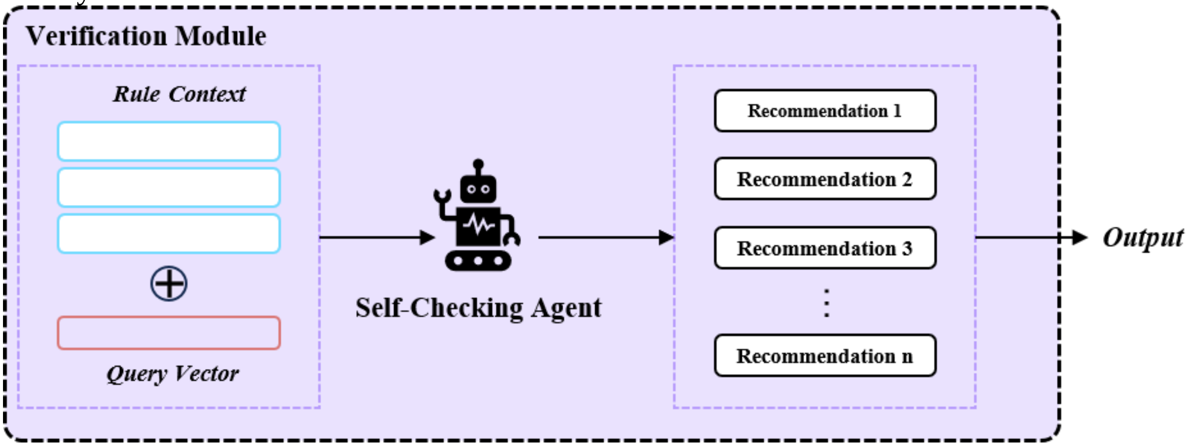

Figure 5: Workflow of Verification Module.

During the validation phase, the system compares multimodal feature vectors against these symbolic rules within a unified semantic space. When compliance conflicts are identified, the mechanism traces back through the symbolic reasoning chain to locate the precise source of each contradiction. The system subsequently generates comprehensive verification traces that document the specific conflicting content, relevant validation rules, and complete reasoning pathways. This structured approach provides auditable logical evidence for revision recommendations while effectively addressing the traceability limitations of conventional verification methods.

4. Experiments and Results Analysis

To evaluate the effectiveness of this system in real-world scenarios, we conducted a three-month user study. We recruited 20 professionals from power companies as volunteers, whose roles spanned core business functions including project management, technical review, and financial auditing. All participants possessed over one year of report review experience and received operational training on the system prior to the study. We comprehensively evaluated the system's performance by comparing key business metrics before and after deployment, supplemented by a post-study user satisfaction survey.

Table 1: Comparison of Key Performance Indicators Before and After System Deployment.

Evaluation Metric	Before Deployment	After Deployment
Average Report Review Cycle	5 days	2 hours
Manual Review Workload	Baseline	Reduced > 80%
Error Omission Rate	15%	>3%
User Satisfaction	-	95%

After three months of practical application, the system demonstrated significant improvements across multiple key performance indicators. As shown in the Table 1, the report review cycle was drastically reduced from an average of 5 working days to approximately 2 hours, achieving a leap in efficiency from days to hours. Simultaneously, the system's automation capabilities reduced manual review workload by over 80%, enabling professionals to focus on higher-value analytical and decision-making tasks. Regarding validation quality, the error and omission detection rate has been successfully reduced from the original 15% to below 3%, substantially enhancing risk prevention capabilities. Final user satisfaction survey results further validate the system's value within actual workflows, with 95% of participants rating the system's overall performance as “satisfied” or “very satisfied."

The experimental results comprehensively validate the practicality and superiority of the proposed autonomous validation framework from four dimensions: efficiency, quality, cost, and user experience. The system not only successfully automated the validation process but also, through its accurate multimodal understanding and rule-matching capabilities, became a reliable and efficient intelligent assistant for professionals, providing robust tool support for the digital transformation of power enterprises.

5. Conclusion

This study addresses power report verification challenges via a framework integrating multimodal reasoning, RAG knowledge base, and LLM agents. Centered on the Brightness Large Model, it enables deep multimodal understanding, dynamic rule adaptation, and full verification automation.

Future work will improve low-resolution chart parsing, expand rare technical indicator rule coverage, and integrate the framework into more new power system scenarios.

Acknowledgements

This research was supported by the Scientific and Technological Project of State Grid Shanghai Municipal Electric Power Company (Project No.: 520933250005; Contract No.: SGSHSN00ZSJS2503900).

References

[1] K. Qian, W. Li, T. Sun, W. Wang, W. Luo, Docrefine: An intelligent framework for scientific document understanding and content optimization based on multimodal large model agents, arXiv preprint arXiv:2508.07021(2025).

[2] Y. Peng, G. Zhang, M. Zhang, Z. You, J. Liu, Q. Zhu, K. Yang, X. Xu, X. Geng, X. Yang, Lmm-r1: Empowering 3b lmms with strong reasoning abilities through two-stage rule-based rl, arXiv preprint arXiv:2503.07536(2025).

[3] M. Dadopoulos, A. Ladas, S. Moschidis, I. Negkakis, Metadata-driven retrieval-augmented generation for financial question answering, arXiv preprint arXiv:2510.24402(2025).

[4] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. K ̈uttler, M. Lewis, W.-t. Yih, T. Rockt ̈aschel, et al., Retrieval-augmented generation for knowledge-intensive nlp tasks, Advances in neural information processing systems, 33(2020) 9459–9474.

[5] T. Taipalus, Vector database management systems: Fundamental concepts, use-cases, and current challenges, Cognitive Systems Research, 85 (2024) 101216.

[6] A. Caione, A. L. Guido, A. Martella, R. Paiano, A. Pandurino, Knowledge base support for dynamic information system management, Information Systems and e-Business Management, 14 (3) (2016) 533–576.

[7] H. Ye, Y. Zheng, Y. Li, K. Zhang, Y. Kong, Y. Yuan, Rh-brainfs: regional heterogeneous multimodal brain networks fusion strategy, Advances in Neural Information Processing Systems, 36 (2023) 59286–59303.

[8] J. Tang, T. Fan, C. Huang, Autoagent: A fully-automated and zero-code framework for llm agents, arXiv preprint arXiv:2502.05957(2025).

[9] H. Wang, C. M. Poskitt, J. Sun, Agentspec: Customizable runtime enforcement for safe and reliable llm agents, arXiv preprint arXiv:2503.18666(2025).

[10] J. Juziuk, D. Weyns, T. Holvoet, Design patterns for multi-agent systems: A systematic literature review, in: Agent-Oriented Software Engineering, Springer, 2014, pp. 79–99.

[11] A. Gonz ́alez-Briones, F. De La Prieta, M. S. Mohamad, S. Omatu, J. M. Corchado, Multi-agent systems applications in energy optimization problems: A state-of-the-art review, Energies, 11 (8) (2018) 1928.