Journal of Electronic Research and Application, 2026, Volume 10, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
Online ISSN: 2208-3510
Print ISSN: 2208-3502

Application Research on a Large Language Model-Based Auxiliary Design System for Nuclear Power Engineering Modification

Lihong Tang[1]*, Hengtao Pu[2], Yusong Li[2], Shanhong He[2]

[1]Suzhou Nuclear Power Research Institute Co., Ltd., Suzhou 215000, Jiangsu, China
[2]Daya Bay Nuclear Power Operations and Management Co., Ltd., Shenzhen 518000, Guangdong, China

***Corresponding author:** Lihong Tang, 834397404@qq.com

Abstract: Aiming at the core pain points in nuclear power engineering modification, such as low efficiency of knowledge retrieval, cumbersome document preparation processes, insufficient standardization, and difficulties in expert experience inheritance, this paper proposes and constructs an intelligent auxiliary design system based on Large Language Models (LLMs). The system adopts a hybrid storage strategy combining vector databases and knowledge graphs to build a high-quality vertical knowledge base for engineering modification, successfully transforming unstructured tacit knowledge scattered in historical documents into computable and retrievable structured intelligence. On this basis, the system establishes a dynamic multi-dimensional tag system and an intelligent evolution mechanism, and innovatively creates an intelligent auxiliary design engine based on dynamic questionnaire interaction. Integrating template guidance, questionnaire interaction, and Retrieval-Augmented Generation (RAG) technology, the engine realizes the automatic preparation and compliance verification of design documents. Application results indicate that the system effectively guides the sorting of design inputs, significantly improves design efficiency and quality, and reduces the risk of human error. Finally, the paper outlines the future integration with external drawing platforms, 3D design platforms, and computational analysis platforms, aiming to build an online collaborative design platform covering the full life cycle and achieve an intelligent closed-loop of the entire engineering modification process.

Keywords: Large language model; Nuclear power engineering modification; Intelligent auxiliary design; Knowledge graph; Retrieval-augmented generation

1. Introduction

China Nuclear Energy Industry Association, February 2, 2023: As of December 31, 2022, China has 55 operating nuclear power units (excluding Taiwan region), with a total installed capacity of 56,985.74 MWe (rated capacity) [1]. Under the dual carbon goals, nuclear power is increasingly important in China's power industry; national requirements for nuclear safety are also becoming stricter, demanding absolute assurance of nuclear safety [2]. Amid rising nuclear power share and growing importance of nuclear safety, the digitalization level of certain nuclear power business areas still fails to keep pace. Nuclear power engineering modifications belong to high-risk activities within nuclear configuration management. They refer to changes made to systems, equipment, and structures during a nuclear power plant's lifetime to meet objectives such as regulatory updates, technological upgrades, equipment replacement, performance enhancement, or life extension.

These modifications require designers to have a thorough understanding of the original design, including deep knowledge of the original design intent, design codes, and historical change records. Designers must also follow the latest design intent and codes. This process involves reviewing extensive historical documents and reference materials, placing extremely high demands on designers. Any issue in any stage of the design process could compromise the safety of the final design outcome.

The process spans the entire life cycle from requirement proposal, scheme design, safety analysis, construction and commissioning to document updates, characterized by multi-disciplinary coupling, strong safety constraints, high knowledge dependency, and long-cycle complexity. Traditional retrofitting models heavily rely on domain experts' tacit experience and vast, scattered procedural documents, facing inherent challenges such as knowledge transfer gaps, low collaboration efficiency, and high risk of human error [3].

In recent years, large language models (LLMs), represented by ChatGPT, have demonstrated outstanding performance in code generation, technical Q&A, document summarization, and logical reasoning tasks, with their 'intelligence' increasingly penetrating research and development processes in complex industrial systems such as aerospace and automotive manufacturing. Introducing LLMs into nuclear power engineering retrofitting holds core value in transforming unstructured knowledge scattered across drawings, reports, procedures, and expert minds into structured, computable, searchable, and reasoning-capable intelligence, thereby building an intelligent auxiliary decision-making system that spans the entire retrofitting process.

This study aims to systematically explore the core application scenarios, technical implementation pathways, and key challenges of LLMs in nuclear power engineering retrofitting. First, this paper analyzes the business pain points in nuclear power engineering retrofitting and the alignment with LLM capabilities.

2. Current status and issues in nuclear power engineering retrofitting design

2.1. Business status and issues

Engineering retrofitting design, as a core business segment in nuclear configuration management, is a typical document-intensive task. This type of improvement involves changes to various design documents (including but not limited to installation drawings, as-built drawings, process flow diagrams, wiring diagrams, system design manuals, equipment operation and maintenance manuals, safety analysis reports, etc.), with documents serving as the primary medium for knowledge and information transfer. The main activity centers on processing and utilizing nuclear domain knowledge. The design process for retrofitting schemes requires searching for and comprehensively analyzing a large volume of documents, including regulatory standards, design documents and design bases from multiple disciplines, professional expertise, and internal and external experience feedback [4].

Nuclear power engineering retrofitting is a closed-loop management process strictly constrained by regulations (e.g., HAF series), typically following the standard workflow: retrofit application → scheme design and safety analysis → review and approval → implementation and commissioning → document update. Within this highly structured process, multiple efficiency bottlenecks and risk points exist. The overall design process specifically includes four steps: design survey → problem and root cause analysis → retrofit scheme preparation → design verification and review. Although these four steps complete the entire engineering retrofitting design process, the current document-driven design approach faces bottlenecks, and its quality directly impacts engineering safety and efficiency.

Currently, the preparation of design documents faces multiple challenges, including complex knowledge management, difficulty in ensuring information quality, and insufficient standardization. There is an urgent need to enhance design efficiency and quality through intelligent means. Design personnel currently encounter the following difficulties:

2.1.1. Low efficiency in knowledge retrieval

First, during the stage of retrofitting requirement identification and scheme conception, engineers must search through massive historical data to find similar cases, failure lessons, and best practices. This knowledge is typically scattered across retrofit reports from different periods, operational event reports (LER), equipment maintenance records, and individual experts' experience. Manual retrieval is inefficient and leads to 'reinventing the wheel' or repeating past mistakes due to incomplete information. Research indicates that up to 30% of engineering issues in complex industrial systems are related to ineffective use of historical knowledge.

2.1.2. Lack of information quality assurance mechanisms

Design input information lacks a unified validation mechanism. Inconsistent formats and standards across different sources easily lead to design deviations. Meanwhile, due to the absence of effective quality control measures, the accuracy and completeness of design documents are difficult to ensure.

2.1.3. Insufficient standardization

The preparation of design documents lacks a unified template, resulting in significant differences in format and content expression across disciplines and projects. This low level of standardization hinders the effective accumulation of tacit knowledge, leads to low reuse of experience, and increases the risk of human error, thereby affecting the overall quality of engineering retrofitting.

2.1.4. Complex and cumbersome document preparation process

During the document preparation and review phase, a single retrofit often requires simultaneous updates to dozens or even hundreds of related documents, such as P&ID diagrams, process flow diagrams, electrical wiring diagrams, operating procedures, maintenance procedures, and training materials. Ensuring consistency across all document changes is a tedious and error-prone task. Currently, this is mainly done manually, with traditional reviews relying on peer experts reviewing line by line, which is time-consuming and labor-intensive, prone to human error, and difficult to guarantee 100% coverage.

2.1.5. Difficulty in reusing experience

Historical renovation documents lack a classification management system, are scattered, and are difficult to retrieve. The persistent issue of knowledge loss and transmission is particularly prominent in the nuclear power industry. With the retirement of numerous experienced ‘first-generation’ nuclear power engineers, their tacit knowledge such as nuanced understanding of subtle interactions between specific systems is at risk of being lost. How to make this tacit knowledge explicit, structured, and passed on is a long-term challenge for the industry.

2.2. Adaptability analysis of large language models (LLMs) in nuclear power

Large language models are essentially probabilistic models trained on massive amounts of text, capable of understanding and generating human language. Their potential application in nuclear power engineering renovation stems from a deep alignment between several core capabilities and industry-specific requirements:

2.2.1. Deep semantic understanding and information extraction

LLMs can comprehend the contextual meaning of specialized terminology, for example, distinguishing the different emphases of ‘RCP’ (Reactor Coolant Pump) in thermal-hydraulic versus vibration monitoring contexts. This enables them to automatically extract key entities (equipment, defects, measures) and their relationships from unstructured engineering reports, laying the foundation for knowledge graph construction or intelligent question-answering systems [5].

2.2.2. Multi-step logical reasoning and task decomposition

Advanced LLMs possess certain logical reasoning abilities, allowing them to break down complex problems into sub-problems. For instance, when asked, ‘What aspects should be considered when adding a display terminal in the main control room?’, the model can systematically reason through dimensions such as power load and distribution systems, signal interfaces and network bandwidth, installation space and ergonomics, fire barriers, and updates to relevant procedures. This structured mind-map-like output effectively supports engineers in brainstorming and conceptualizing solutions.

2.2.3. Code generation and tool invocation potential

Many engineering analyses (e.g., simple thermal load calculations, cable voltage drop calculations) and document processing tasks (e.g., batch search-and-replace of specific text formats) can be automated. LLMs can generate corresponding script code (e.g., Python, VBA) based on natural language descriptions, or invoke professional software tools (e.g., CAD, simulation software) via predefined API interfaces, achieving a degree of automation [6].

3. Establishment plan for nuclear power engineering renovation knowledge system

Nuclear power expertise, due to its unique specialization and confidentiality, is not included in the training data of general models, making existing general models ill-suited for professional nuclear applications. Therefore, a combination of multiple technologies is required to enhance large models’ understanding of nuclear-specific knowledge.

Pre-trained neural language models have proven capable of learning extensive deep knowledge from data. They can achieve this without accessing external memory, serving as parameterized implicit knowledge bases. Although this advancement is promising, such models also have drawbacks: they are difficult to expand or modify, cannot intuitively reveal the basis for their predictions, and may generate 'hallucinations.' Hybrid models combining parameterized memory with non-parameterized (i.e., retrieval-based) memory can alleviate some of these issues, as knowledge can be directly revised and extended, and accessible knowledge can be inspected and explained [5].

3.1. Building a vertical knowledge base for engineering renovation fusing 'vector + graph'

Knowledge graphs are key to endowing LLMs with 'professional common sense' and precise reasoning capabilities [7]. A nuclear power renovation knowledge graph should include entities (e.g., equipment, systems, functions, failure modes, regulatory clauses), relationships ('belongs to', 'affects', 'causes', 'must comply with'), and attributes (design parameters, operational limits). The construction approach can be 'hybrid': for structured texts like design manuals, batch information extraction using LLMs; for unstructured texts such as experience feedback, a human-machine collaborative method combining 'expert annotation + model fine-tuning'. The constructed knowledge graph is not only used for RAG retrieval but also supports complex reasoning queries, such as 'identify all equipment connected to the emergency power bus and located within fire zone H-12'.

The knowledge base is the cornerstone of system intelligence. This solution adopts a hybrid storage approach, 'vector database + knowledge graph', to build a high-quality, traceable domain-specific knowledge base. First, a multi-source data ingestion pipeline is established, supporting automated parsing and cleaning of historical renovation documents, technical specifications, equipment inventories (Excel/PDF/Word), and scanned images. Leveraging the deep semantic understanding of large models (e.g., Qwen3), unstructured text undergoes secondary processing, transforming lengthy technical passages into high-density knowledge units. This includes generating 'question-answer' pairs (QA pairs) to enhance semantic matching, extracting related questions to cover user query intent, and writing technical summaries to preserve system design logic.

At the storage level, a hybrid strategy is employed: Milvus vector database stores vector features of knowledge chunks, enabling semantic similarity retrieval; Neo4j graph database constructs entity relationship networks among equipment, systems, failure causes, and renovation solutions, supporting logical reasoning; MySQL manages structured metadata and tag attributes. This architecture not only solves the 'findability' of knowledge but also achieves 'logical coherence' through graph connections, providing precise contextual support for LLMs to generate professional, coherent design proposals.

3.1.1. Knowledge validation and regulatory alignment protocol

To ensure the accuracy and compliance of the extracted tacit knowledge, a rigorous validation protocol is established. The process involves domain experts from multiple disciplines (e.g., mechanical, electrical, safety analysis) who participate in a multi-round annotation and review cycle.

For the validation workflow, initially, the LLM performs batch information extraction from historical documents. The extracted entities and relationships are then presented to expert annotators via a dedicated interface. Each knowledge unit is independently reviewed by at least two senior engineers. Discrepancies are resolved through discussion or adjudication by a third lead expert.

About the accuracy metrics, the quality of the extraction is quantified using Inter-Annotator Agreement (IAA) metrics, specifically Cohen's Kappa (κ) for binary decisions (e.g., entity existence) and Krippendorff's Alpha (α) for relational data. Our pilot study on a dataset of 500 historical modification reports yielded an average κ score of 0.82 and α score of 0.78, indicating substantial to strong agreement among experts. The paper mentions using Cohen's Kappa coefficient to measure inter-annotator agreement for knowledge extraction quality ,as shown in **Equation (1)**.

$$\kappa = \left(P\_o - P\_e\right) / \left(1 - P\_e\right) \tag{1}$$

P_o: Observed proportion of agreement.

P_e: Expected proportion of agreement by chance.

κ: Cohen's Kappa coefficient, measuring inter-rater reliability beyond chance. κ=0.82 indicates a high level of agreement.

Regarding the regulatory version control, the system integrates with an internal regulatory document management system that tracks versions of key standards (e.g., HAF series, ASME codes). Each piece of knowledge in the graph is tagged with its source document and version. When a new version of a regulation is released, the system automatically flags all dependent knowledge units for expert re-review. Similarly, retired procedures are marked as "deprecated" in the knowledge base, and the RAG engine is configured to prioritize active regulations during retrieval, ensuring design outputs remain compliant.

3.1.2. Hybrid query resolution mechanism

The system employs a unified query resolution layer that seamlessly integrates semantic search from Milvus and logical reasoning from Neo4j.The workflow is as follows:

(1) Initial Semantic Retrieval: A user query (or a system-generated query from the questionnaire engine) is first embedded and used to retrieve the top-K (K=20) most semantically similar knowledge chunks from Milvus;

(2) Graph-Based Re-ranking: The entities mentioned in these K chunks are then used as anchor points to query the Neo4j graph. The system traverses the graph to find direct and indirect relationships (e.g., Equipment-AFFECTS-System-MUST_COMPLY_WITH-Regulation). A graph-based relevance score is computed for each of the K initial results based on the strength and depth of these connections;

(3) Fusion and Final Ranking: The final ranking of retrieved documents is determined by a weighted fusion of the vector similarity score and the graph-based relevance score. This hybrid approach ensures that results are not only semantically relevant but also contextually coherent and logically sound within the nuclear domain's complex relational structure. To ensure the retrieved results are not only semantically relevant but also contextually coherent, the system employs a weighted fusion mechanism (**Equation (2)**) to calculate the final relevance score (S_final) for ranking:

$$S\_final = \left(1 - \alpha\right) \times S\_graph + S\_vector \times \alpha \tag{2}$$

where S_vector denotes the semantic similarity score from the vector database, S_graph represents the graph-based relevance score derived from the knowledge graph, and α is a tunable weight coefficient (empirically set to 0.6) balancing the two scores.

This hybrid ranking approach ensures that the most contextually appropriate knowledge is prioritized for the subsequent RAG process.

This joint resolution mechanism is exposed through a single, unified API endpoint, abstracting the underlying complexity from the RAG generation engine.

3.2. Building a dynamic multi-dimensional tagging system and intelligent evolution mechanism

To enable precise knowledge delivery and scenario-based matching, the system establishes a multi-dimensional dynamic tagging system. Tag dimensions include basic tags (e.g., discipline, equipment type, scope of renovation) and semantic tags (e.g., technical terms, failure causes, performance metrics). During construction, a collaborative mechanism of 'model-based automatic tagging + followed by manual review and optimization' is adopted, using large models to automatically classify and extract tags for incoming knowledge, with human verification via an interface to ensure tag quality.

In addition, by introducing a dynamic evolution mechanism, the system automatically discovers new tags and adjusts the weights of existing tags based on user search logs and incremental document data, using algorithms such as Word2Vec to calculate semantic distances.

The system periodically cleans up redundant tags to ensure the tag system remains highly aligned with business scenarios, thereby establishing a strong associative mapping between 'scenarios-template-knowledge' to provide accurate indexing references for document generation.

3.3. Building an intelligent auxiliary design engine based on dynamic questionnaire interaction

This is the core innovation of the system in its vertical domain application. The system abandons the traditional single-instruction generation model and establishes a closed-loop intelligent auxiliary design process of 'template guidance–questionnaire interaction–content generation'.

First, a customizable design document template library is built, supporting users in configuring standardized document structures and chapter strategies. During document preparation, the system introduces a dynamic questionnaire interaction mechanism. After the user selects a template and uploads foundational materials (such as an ESR assessment form), the system automatically parses key information and, for ambiguous parameters or subjective design intentions, generates dynamic questionnaires via a reasoning engine to guide users in supplementing critical design inputs through single-choice or multiple-choice formats.

In the content generation phase, workflow orchestration technology is adopted to combine AI dialogue nodes, knowledge base retrieval nodes, and document parsing nodes. Based on questionnaire responses and tag information, the system invokes RAG (Retrieval-Augmented Generation) technology to precisely retrieve relevant historical cases and regulatory standards from the knowledge base, then drives the large model to generate design content chapter by chapter according to predefined prompts (Prompts) and workflow strategies. Meanwhile, an automated document quality verification function is integrated to perform real-time checks on logical conflicts, formatting standards, and citation sources in the generated content, ensuring compliance and accuracy of the design outcomes.

Regarding the retrieval performance and safety mechanisms, the reliability of the RAG component is critical for safety-critical design. We evaluated the hybrid Milvus/Neo4j retrieval pipeline on a test set of 200 anonymized historical modification cases. For each query derived from a design questionnaire, we measured standard IR metrics. The system achieved a Recall@5 of 92.3% and a Mean Reciprocal Rank (MRR) of 0.87, demonstrating high precision in surfacing relevant precedents and standards.

To mitigate the risk of LLM hallucinations due to poor retrieval, a confidence threshold mechanism is implemented. The system calculates a combined relevance score based on vector similarity and graph path strength. If the top-k retrieved documents fall below a predefined confidence threshold (empirically set to 0.65), the generation module is blocked, and the user is prompted with a warning: "Insufficient authoritative sources found. Please consult a domain expert before proceeding." This ensures that non-compliant or unsupported design content is never generated autonomously.

4. Summary

4.1. Summary of application outcomes

To objectively assess the system's impact, a controlled pilot study was conducted with two teams of 5 designers each, working on comparable modification tasks over a 3-month period. The control group used traditional methods, while the experimental group used our LLM-based auxiliary system. To objectively assess the system's impact, a controlled pilot study was conducted with two teams of 5 designers each, working on comparable modification tasks over a 3-month period. The detailed quantitative results are summarized in **Table 1** .

The results demonstrate statistically significant improvements:

(1) Design Time: The average time to complete a design task was reduced by 35.7% (from 14.2 days to 9.1 days, $p < 0.01$, t-test);

(2) Revision Cycles: The number of revision cycles required during the review phase decreased by 42.1% (from an average of 3.8 to 2.2 cycles per task);

(3) Error Rate: Post-implementation audits revealed a reduction in minor non-conformities from 4.8% to 1.2% of total design elements.

Table 1. Comparative analysis of the auxiliary design system's application impact

Evaluation Dimension	**Specific Metric**	**Baseline Group (Traditional Method)**	**Experimental Group (LLM-Assisted System)**	**Improvement/ Change**
Efficiency Improvement	Avg. Task Completion Time (days)	14.2 ± 2.3	9.1 ± 1.5	-35.7%
	Time Spent on Initial Data Retrieval & Organization (%)	~40% (5.7 days)	~15% (1.4 days)	-75.4%
	Time Spent on Drafting Initial Document (%)	~35% (5.0 days)	~30% (2.7 days)	-46.0%
Quality Improvement	Avg. Number of Review Revision Cycles	3.8 ± 0.9	2.2 ± 0.6	-42.1%
	Conformity Check Pass Rate for Design Elements (%)	95.2%	98.8%	+3.6% (Error Rate -75%)
	Avg. Missing/Incorrect Regulatory Citations per Document	4.2	1.1	-73.8%
Knowledge Utilization	Avg. Number of Historical Cases/ Standards Cited per Document	5.7	12.5	+119.3%
	Designer Satisfaction Score with System Assistance (1-5)	N/A	4.3	N/A

These quantitative findings validate that the system not only improves efficiency but also enhances the quality and consistency of engineering outputs, effectively reducing the cognitive load and error-proneness associated with manual document handling.

4.2. Future planning

Although the system has achieved results in document auxiliary generation, to realize deep digitalization across the entire engineering retrofit process, future work can focus on breaking down data silos, advancing deep integration with external engineering software ecosystems, and building an online collaborative design platform covering the full life cycle.

First, integration with the nuclear power digital drawing platform will be achieved. The system will gain the capability to parse 2D design drawings and automatically associate or trigger drawing change processes based on design document content, ensuring consistency and synchronization between document information and drawing designs.

Second, deeper integration with 3D design platforms will be pursued. By establishing mapping relationships between document data and 3D models, design parameters will be automatically driven and updated within the 3D model, supporting visual validation based on 3D scenarios and resolving challenges related to spatial interference and configuration management among complex systems.

Finally, integration with specialized computational analysis platforms will be implemented. Leveraging the large model’s semantic understanding capabilities, key physical parameters will be automatically extracted from design specifications, driving simulation software to perform specialized analyses such as thermal-hydraulic and stress analysis, with results automatically fed back into the design report, achieving an automated closed-loop process from ‘design input - model computation - result output’.

Through these integrations, the ultimate goal is to achieve full online and intelligent workflow, from requirement proposal, scheme design, drawing creation, simulation computation, to final document archiving, fully breaking through the engineering retrofit design chain and significantly enhancing both quality and efficiency in nuclear power engineering retrofit projects.

Disclosure statement

The author declares no conflict of interest.

References

[1] Lihong Tang, et al.Application research of full chain topology inference technology in the field of Nuclear Power Drawings[J].Engineering technology research,2024(4).

[2] Lihong Tang.Application research of full chain topology inference technology in the field of Nuclear Power Drawings[J].

[3] International Atomic Energy Agency. Plant Life Management for Long Term Operation of Nuclear Power Plants: Implementation and Experience [R]. IAEA, Vienna, 2023.

[4] Zhang ming,Jianguo Wang. Research on Aging Management and Technological Transformation Strategies of Nuclear Power Plants[J]. Engineering of nuclear power, 2022, 43(S1): 76-81.

[5] Lewis P, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks[J]. Advances in Neural

Information Processing Systems, 2020, 33: 9459-9474.

[6] Frieder S, et al. Mathematical Capabilities of ChatGPT[J]. arXiv preprint arXiv:2301.13867, 2023.

[7] Ji S, et al. A Survey on Knowledge Graphs: Representation, Construction and Application[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 33(2): 1619-1637.