The "Wenlu" Brain System for Multimodal Cognition and Embodied Decision-Making: A Novel Security-Enhanced Architecture for Deep Integration of General-Purpose Large Models and Industry Knowledge

Authors: Geng Liang

Date: 2025-05-30T00:00:00+00:00

Abstract

With the rapid penetration of artificial intelligence across multiple industries and scenarios, how to integrate the language understanding advantages of general large models with industry-specific knowledge bases in complex real-world applications has become a key challenge in constructing next-generation intelligent hubs. This paper proposes an embodied brain system named "Wenlu" based on multimodal cognitive decision-making, aiming to achieve secure fusion of private knowledge and public models, integrated processing of multimodal data such as images and speech, and closed-loop decision-making from cognition to automatic hardware code generation. Through a brain-inspired memory tagging and replay mechanism, the system organically integrates user private data, industry-specific knowledge, and general language models, providing precise and efficient multimodal services for enterprise decision support, medical analysis, autonomous driving, robot control, and other applications. Compared with existing technical solutions, "Wenlu" exhibits significant advantages including multimodal processing, privacy security, end-to-end hardware control code generation, self-learning, and sustainable updates, thereby laying a foundation for constructing a new generation of intelligent hubs.

Full Text

Preamble

The "Wenlu" Brain System for Multimodal Cognition and Embodied Decision-Making: A Secure Architecture for Deep Integration of

General Large Models and Industry Knowledge

Geng Liang[1],[2],[3]

[1]Shijiazhuang Key Laboratory of Agricultural Robot Intelligent Perception, Shijiazhuang 050035
Email: liang_{geng}@bupt.edu.cn (L. Geng)

As artificial intelligence rapidly penetrates diverse industries and scenarios, a critical challenge in building next-generation intelligent hubs lies in effectively integrating the linguistic understanding advantages of general large models with proprietary industry knowledge bases. This paper proposes an embodied brain system called "Wenlu," designed to achieve secure fusion of private knowledge and public models, unified processing of multimodal data such as images and speech, and closed-loop decision-making from cognition to automatic hardware code generation. Through a brain-inspired memory tagging and replay mechanism, the system organically combines user private data, industry-specific knowledge, and general language models to provide precise and efficient multimodal services for enterprise decision support, medical analysis, autonomous driving, and robot control. Compared to existing technical solutions, "Wenlu" offers significant advantages in multimodal processing, privacy security, end-to-end hardware control code generation, self-learning, and sustainable updates, laying the foundation for constructing a new generation of intelligent hubs.

Keywords: multimodal cognition; embodied brain; private data; general large model; automatic code generation; brain-inspired memory replay

In the evolving wave of artificial intelligence technology, deep learning-driven large language models (LLMs) have become a cornerstone for building intelligent systems. However, current mainstream language models predominantly focus on pure text input-output modes, with relatively weak capabilities for processing multimodal data. Simultaneously, how to efficiently integrate industry-specific knowledge with general models while protecting user privacy remains a persistent challenge for both academia and industry. To fill this gap and advance next-generation intelligent applications, this paper proposes an embodied brain system named "Wenlu," aiming to achieve a complete closed loop from multimodal cognition to automatic hardware code generation while demonstrating security and efficiency compatibility in private data and industry knowledge application scenarios.

Over the past few years, challenges in intelligent decision-making and multimodal information processing have primarily stemmed from three aspects. First, different data types (text, images, speech, sensor readings, etc.) exhibit vast differences in dimensionality, feature distribution, and temporal characteristics, making it difficult to efficiently integrate them using a single network structure or general model. Second, deep knowledge for specific industries is often embedded in private or proprietary datasets that require more stringent security and compliance measures. However, general large models are typically trained on massive public datasets and lack built-in confidentiality and access control mechanisms, making direct application in sensitive domains difficult. Third, many AI systems remain stuck at the "perception-cognition" stage, unable to automatically translate intelligent decisions into machine control or code generation, hindering efficient deployment in embodied scenarios such as robotics, wearable devices, and autonomous driving.

To effectively address these issues, the "Wenlu" system proposes a multi-level architecture that tightly integrates user private data decision modules, industry multimodal decision and service modules, hardware control and automatic code generation modules, and general model fusion units. The private knowledge unit provides secure isolation and controlled access for sensitive information through encrypted sandboxes and tag-based management. The multimodal decision module enables the system to process not only text but also comprehensively evaluate and reason about multi-source data including images, speech, and sensors. The hardware control module further breaks down barriers between language understanding and physical execution, automatically generating control commands based on high-level cognitive or task descriptions to enable real-time feedback and actions on robots or other devices. The underlying general model fusion unit handles overall language understanding and generation, providing powerful semantic support for all upper-level modules while deeply coupling with private knowledge bases and industry knowledge bases.

Furthermore, drawing inspiration from how biological brains tag and replay memories, the "Wenlu" system incorporates a brain-inspired memory reinforcement mechanism: when executing complex decision tasks, the system automatically tags key information or important reasoning paths and replays these memories during subsequent idle or offline periods for reinforcement learning, thereby continuously optimizing multimodal understanding and reasoning performance. This brain-inspired memory mechanism not only enhances the system's deep cognition and reuse capabilities for specific scenarios but also lays the foundation for self-learning and continuous iteration.

At the application level, the "Wenlu" embodied brain system aims to cover multi-scenario and multi-industry needs, including automatic inspection and control in industrial manufacturing, diagnostic assistance for medical imaging, environmental perception for autonomous driving, human-robot interaction for service robots, and intelligent monitoring for wearable devices. By combining industry experience with general semantic understanding, "Wenlu" can provide specialized decision support and automated execution solutions while ensuring data security and privacy compliance. Compared to traditional systems that rely on manual configuration or plugin-based invocation, it significantly reduces communication costs and error rates while achieving self-learning and adaptation through continuous iteration.

In summary, the goal of the "Wenlu" system is to become a universal platform for multimodal cognition and embodied decision-making. It inherits the advantages of large language models in text understanding and generation while strengthening capabilities to process images, audio, sensor data, and other information.

It can utilize private information in secure environments while deeply coupling proprietary knowledge for various industry scenarios. Simultaneously, it can automatically translate intelligent decision results into hardware execution code, achieving a true "perception-cognition-decision-action" integrated closed loop. This paper systematically elaborates on the technical approaches of the "Wenlu" embodied brain in multimodal fusion, private information management, automatic code generation, and brain-inspired memory reinforcement, and discusses its potential value across different industries and application scenarios. Through such system design, we hope to provide a new path for AI to transition from academic research to industrial application and lay a feasible technical foundation for building next-generation general intelligent hubs.

2 Theoretical Background and Existing Technical Limitations

Over the past decade, artificial intelligence has rapidly evolved from early statistical learning models to deep learning, large-scale language models, and multimodal fusion technologies. Particularly in natural language processing (NLP), Transformer-based general language models (LLMs) have emerged, sparking widespread discussion about the potential for "general artificial intelligence." However, current mainstream technical systems still face numerous challenges in multimodal information integration, private data security, automatic code generation, and deep coupling with industry knowledge. To thoroughly analyze these issues, this section combines several typical solutions to elaborate on the existing technical background and limitations.

2.1 Challenges in Multimodal Fusion and Cognitive Decision-Making

Early AI applications primarily focused on single modalities, such as image recognition, speech recognition, and text generation. However, information in the real world is typically multimodal: images, audio, text, and sensor data often appear simultaneously and are interrelated. In this context, multimodal fusion technology has emerged, attempting to establish unified feature spaces or semantic representations for different modal data to enhance perception and decision-making accuracy.

Although existing multimodal technologies have achieved preliminary success in tasks such as image-text matching and visual question answering, they remain inadequate for deep-level cognition and complex decision-making, primarily due to three factors. First, feature alignment and semantic projection are difficult because different modal data exhibit significant differences in dimensionality and semantic distribution, requiring unified network structures or feature projection mechanisms. Once the alignment approach between modalities is inappropriate, model reasoning effectiveness becomes limited. Second, there is a lack of cross-modal long-term memory and reasoning frameworks. Most multimodal fusion research focuses on short-term or one-time reasoning and cannot support long-term, multi-step cognitive processes. For instance, industrial inspection or medical diagnosis often requires continuous temporal information and iterative judgment, placing higher demands on multimodal systems. Third, model interpretability and causal reasoning are insufficient. Multimodal models often rely on large-scale training data for pattern matching and have not yet formed "common sense" or "causal" understanding of the world. In industry deployment, the lack of interpretable mechanisms also affects user trust and decision-making safety.

2.2 Limitations of General Large Models and Industry Knowledge Integration Bottlenecks

Since the advent of Transformer and its improved models, general language models trained on massive public corpora (such as the GPT and BERT families) have achieved breakthroughs in language understanding and generation. However, directly transplanting such large models to specific industries or private data scenarios encounters several difficulties. First, industry-specific knowledge is scarce. While the training data used by general large models covers a wide range, their deep mastery of each domain is limited. When facing rigorous professional scenarios (such as medical diagnosis or financial analysis), the lack of deep domain knowledge significantly reduces the accuracy and reliability of model outputs. Second, privacy and compliance challenges arise. General large models are often open-ended, with training data and inference paths lacking strict security isolation. Once user private information is mixed with public corpora, it can easily lead to data leakage or misuse. As industry regulations on data compliance and user privacy become increasingly stringent, how to safely link private data with large models has become an urgent problem. Third, there is a lack of continuous learning and memory mechanisms. Although current large models have massive parameters, their "memory" is mainly reflected in trained weight parameters and is not flexible enough. When new knowledge needs to be absorbed during operation, they often face the high cost of "secondary fine-tuning" or even "large-scale retraining," making rapid iteration and knowledge accumulation impossible. Fourth, there is a "breakpoint" from abstract understanding to concrete execution. Even if large models can accurately generate text descriptions, they cannot directly translate them into automatically executable hardware instructions or robot control scripts. This prevents "general models" from fully leveraging end-to-end intelligent potential in embodied scenarios.

2.3 Traditional Robot and Hardware Control Schemes and Their Drawbacks

In the field of robotics or hardware control, mainstream development models typically require human engineers to manually program action strategies and control logic, translating natural language requirements into low-level instructions or API calls. This process has multiple limitations. First, manual programming and interface adaptation incur high costs. Each robot platform or wearable device has its specific API or SDK, and transplantation and adaptation often require engineers to spend considerable time writing code and debugging, resulting in lengthy development cycles and high error rates. Second, there is insufficient adaptation to dynamic external environments. When the physical environment changes or device hardware is updated, programs require substantial modifications again, making it difficult to form a dynamically adaptive closed loop. Third, it is difficult to integrate with high-level semantic reasoning. Robot control mostly remains at the level of sensor data processing and behavior planning, lacking deep understanding of high-level information such as natural language and industry knowledge, and cannot be uniformly managed with multimodal decision-making systems.

2.4 Plugin/External API Integration Compromise Solutions and Problems

To bridge the gap between general large models and domain applications, some technical practices attempt to add plugins or call external APIs on top of large models. For example, robot APIs or wearable device interfaces are encapsulated as plugins, and expansion is achieved by providing corresponding calling examples to language models. While such methods provide certain hardware interaction capabilities for large models, they still suffer from deficiencies in practical applications. First, there is a lack of context management. Plugin-based solutions essentially rely on temporary API calls and lack unified memory and reasoning management mechanisms, limiting large models' ability to connect context across multiple conversations or multimodal scenarios. Second, cross-domain call chains are complex. If a task requires multiple plugins to collaborate simultaneously (for example, reading sensor data, controlling a robotic arm, and outputting diagnostic reports), the plugin call chain becomes cumbersome and highly coupled, with prohibitive maintenance costs. Third, decision results are unstable. Stability issues with plugin interfaces and call sequencing often lead to unexpected system failures or errors, and once anomalies occur, large models themselves lack global awareness of hardware or external APIs and cannot self-correct.

2.5 Knowledge Graph-Based Q&A and Decision Systems

In specific industries such as healthcare, finance, and manufacturing, there are also Q&A or reasoning systems centered on knowledge graphs. They encapsulate structured expert knowledge into retrievable graph nodes and edge relationships to assist reasoning and decision-making. However, such solutions also face several constraints. First, construction and update costs are high. Knowledge graphs require dedicated teams for data annotation and maintenance, and multimodal features (images, audio, sensors, etc.) are difficult to express efficiently in graphs. Second, there is a lack of strong support for language generation. Compared with general large models, traditional knowledge graphs mainly handle retrieval-based Q&A or logical reasoning, lacking natural language understanding and generation capabilities, and insufficient ability to present complex domain data in readable form. Third, they cannot achieve automated hardware output. Knowledge graph-based systems focus more on "information retrieval" and "logical reasoning," with a large gap from automatically generating robot or device control instructions, making it difficult to form an embodied closed loop.

2.6 Summary and Limitation Analysis

Combining the above typical solutions, existing technologies exhibit significant limitations in four aspects. First, multimodal fusion is insufficient. Whether relying on general large models or knowledge graphs, most existing systems lack the ability to simultaneously process multi-source information such as images, audio, text, and sensors for deep decision-making, or can only perform limited functional splicing through plugin-based approaches. Second, there is a lack of secure and controllable private data management. Existing large models mostly ignore privacy protection needs, while traditional industry systems often lack support for open language understanding; efficient and secure integration between the two is difficult. Third, there is a "break" from cognition to execution. While general large models can perform text understanding and generation, physical hardware control still requires manual "translation," leading to a disconnect between actual decision-making and execution, hindering self-learning and rapid iteration. Fourth, there is a lack of continuous self-learning and memory reinforcement mechanisms. Whether large models or knowledge graphs, updates and fine-tuning often require high costs. Most solutions lack brain-inspired memory tagging and replay stages, preventing the accumulation of new experiences and knowledge through continuous interaction and application.

Thus, to break through the comprehensive barriers of multimodal cognition, private information protection, and embodied decision-making, a new architecture is needed that combines general language understanding, efficient domain knowledge integration, private data security protection, and automatic hardware code generation capabilities. It is against this background and driven by these needs that the "Wenlu" embodied brain system emerges, committed to building an intelligent hub for multimodal cognitive decision-making that emphasizes private information security and can directly output executable instructions, forming a complete closed loop from "perception-cognition-decision-action."

3 Wenlu System Architecture

The "Wenlu" embodied brain system aims to bridge the gaps in AI across three key areas—multimodal cognition, private information protection, and hardware control execution—thereby providing more complete, flexible, and secure intelligent services for real-world scenarios. To achieve this goal, "Wenlu" adopts a multi-level, modular architecture that includes both front-end perception and interaction components and a core back-end where large models are deeply integrated with private knowledge bases. It also reserves modules for hardware control generation and industry knowledge docking to meet end-to-end application requirements.

The first layer is the user private knowledge understanding and Q&A decision unit, which primarily addresses how sensitive user information and industry-specific data can be safely used within large models. This module employs technical means such as secure sandboxes, encrypted storage, and tag-based management to ensure that read/write permissions for private data are strictly controlled after entering the system. Once a user' s query touches sensitive content, the system automatically invokes specific access strategies to parse and desensitize the required private data. This unit not only effectively isolates public training data from user private data but also provides highly customized support for subsequent decision-making.

The second layer is the industry multimodal decision and service support unit, whose core function is to fuse industry-specific multimodal inputs such as images, speech, text, and sensors, and perform deep semantic alignment with general large models. By converting underlying processing such as image recognition, audio analysis, and text understanding into unified feature vectors or embedding representations, this module can comprehensively evaluate information correlations across multi-source signals and perform deep decision-making. In high-demand fields such as industrial inspection, medical imaging analysis, or autonomous driving monitoring, this multimodal decision unit can significantly improve result accuracy and reliability while also outputting richer, more explainable analysis reports.

The third layer is the robot control and hardware device code generation unit, which serves as the bridge "from cognition to action" throughout the "Wenlu" system. When the multimodal decision unit and private knowledge decision unit collaboratively determine an execution plan, the robot control unit can automatically generate low-level code scripts from high-level semantic instructions, adapting to various hardware platforms such as ROS2, wearable device APIs, and intelligent robotic arm interfaces. Through this end-to-end automation process, the system forms an integrated closed loop from natural language description and image/sensor fusion analysis to specific device execution. Unlike traditional manual programming approaches, "Wenlu" can achieve rapid migration by simply updating the adaptation layer when hardware or environmental changes occur, significantly reducing engineering costs.

The underlying foundation supporting the entire system is the deep language model base and the "general-specialty integrated" knowledge fusion unit. This unit incorporates both the general language understanding and generation capabilities of large models (such as DeepSeek) and integrates industry expert knowledge bases to improve Q&A and reasoning accuracy in vertical domains. Simultaneously, this unit embeds a brain-inspired memory tagging and replay mechanism: during each interaction or reasoning process, the system automatically annotates key decisions or important scenarios, and replays and reinforces this information during idle or offline periods, enabling the model to establish stronger "long-term memory" for high-value knowledge and frequent tasks. This brain-inspired memory consolidation approach can significantly enhance system reliability and customization capabilities after multiple iterative interactions, providing theoretical and technical guarantees for integrating large models with private information and industry knowledge.

It is worth emphasizing that the various units are not isolated components but collaborate efficiently through a unified bus and communication protocol. The private knowledge decision unit can share the capabilities of the core large model with the multimodal decision unit and call encrypted data based on permission policies. Multimodal analysis results can be directly passed to the hardware control generation unit, ultimately presented as executable scripts or instructions. The underlying general model and industry knowledge base continuously iterate the essence of multiple interactions into the model core through the memory tagging mechanism, providing increasingly accurate support for subsequent Q&A and decision-making.

Through this layered, modular architecture, "Wenlu" constructs a closed system loop between multimodal perception, semantic decision-making, private data processing, and robot control, providing powerful and flexible solutions for intelligent applications across industries. Under this framework, the system not only possesses the ability to deeply reason about multi-source data in complex environments but can also output automated hardware execution scripts while fully protecting user-sensitive information. Ultimately, "Wenlu" is poised to become a new generation of general intelligent hubs, combining cross-industry scalability with personalized professional knowledge accumulation, providing reliable technical support for AI' s transition from laboratory to real-world production and life scenarios.

4 Core Modules and Implementation Mechanisms

Within the "Wenlu"embodied brain architecture, each module is designed around multimodal fusion, privacy protection, and end-to-end decision-making closed loops. To more clearly demonstrate its technical approach and internal logic, this section provides an academic and systematic exposition of the structure, mechanisms, and inter-module relationships of four core modules according to system functions and implementation processes.

4.1 Private Knowledge Decision Unit

4.1.1 Module Functions This module focuses on solving the problem of how to securely fuse sensitive user information with general large models, ensuring both strict access control and processing of confidential data within the system and meeting the needs for efficient reasoning and customized Q&A. Its core functions include: (1) sandbox-style encrypted storage and tag-based management of private data; (2) permission verification and desensitized output; and (3) collaboration between private knowledge and general large models.

4.1.2 Implementation Mechanisms

Secure Sandbox and Encryption Strategy: The system employs symmetric or asymmetric encryption algorithms for private data storage, supplemented by role-based access control (RBAC). A "private knowledge base index" table is constructed, assigning encryption keys and security tags to each piece of confidential information.

Permission Verification and Desensitized Output: When a user query touches data indicated by security tags, the permission verification module is first triggered to determine whether access is allowed or if desensitization procedures must be executed. After authorization, the private knowledge unit decrypts the original data and performs semantic vectorization; after merging with inference results from the general model, secondary review and sensitive information filtering are conducted.

Collaboration Between Private Knowledge and General Large Models: Using encrypted indexes, private corpora are shared with the general model in an embedded manner within an implicit feature space, enabling the large model to dynamically retrieve relevant confidential information during inference. Before the final Q&A or decision result is released, it undergoes security policy detection to ensure no unauthorized sensitive information is leaked.

4.2 Multimodal Decision Unit

4.2.1 Module Functions

The multimodal decision unit of "Wenlu" aims to organically fuse different types of industry data (images, speech, text, sensor signals, etc.) and output more refined and explainable decision results under the support of industry-specific knowledge bases and general language models. It performs three key functions: (1) feature fusion and semantic projection; (2) interaction between industry knowledge bases and general models; and (3) explainability and service interfaces.

4.2.2 Implementation Mechanisms

Feature Fusion and Semantic Projection: For image, audio, and sensor data, specialized deep neural networks extract their feature vectors, which are then projected and fused with text inputs in a unified semantic space. Contextual information across different modalities is interconnected through multi-head attention or cross-modal Transformer structures to generate comprehensive multimodal representations.

Interaction Between Industry Knowledge Bases and General Models: Based on this fused representation, the system calls upon both general models and domain-specific knowledge bases for reasoning. If private data is involved, tag matching with the private knowledge unit is also required. The final output forms explainable industry diagnoses, fault detection reports, predictive analysis results, or service solutions.

Explainability and Service Interfaces: To support high-reliability scenarios such as industrial and medical applications, the system includes an explainability component that can provide core evidence (such as annotated defect areas or semantic keypoints) alongside decision conclusions. Final results can be returned via API or front-end interface for user reference or further action.

4.3 Robot Control and Hardware Code Generation Unit

4.3.1 Module Functions This unit is the key link for "Wenlu" to achieve embodied decision-making, providing direct execution solutions for physical platforms such as robots and wearable devices through automatic translation from high-level semantic understanding to specific hardware control instructions. Its functions include: (1) mapping high-level semantics to instructions; (2) adaptation layers and interface management; and (3) real-time feedback and closed-loop iteration.

4.3.2 Implementation Mechanisms Mapping High-Level Semantics to Instructions: Based on the sequence-to-sequence generation capability of language models, "natural language requirements" or "multimodal reasoning results"are converted into hardware control languages such as ROS2 node scripts, embedded C++, or Python execution scripts. Compilation or interpretation can be performed at the output stage to adapt to different operating environment specifications.

Adaptation Layer and Interface Management: A unified adaptation layer is designed for mainstream robot frameworks and hardware platforms, translating intermediate instructions generated by the general model into function calls or configuration files that comply with specific APIs, facilitating rapid hardware device switching. When hardware or environment configuration changes, engineers only need to modify the adaptation layer without large-scale adjustments to high-level logic.

Real-Time Feedback and Closed-Loop Iteration: Status or sensor data generated by hardware devices during task execution can be fed back into the multimodal decision unit for subsequent decision updates, enabling adaptation and self-learning. Through continuous iteration, the system can better adapt to complex or dynamic external environments.

4.4 General Model Fusion Unit

4.4.1 Module Functions This unit is the foundational support for the entire "Wenlu"system, integrating open-source or commercial general large models (such as DeepSeek) with industry knowledge bases through a "general-specialty integrated" fusion approach. It is responsible for both cross-domain language understanding and generation and provides a unified underlying semantic foundation for multimodal and private data reasoning. Its functions include: (1) pre-trained model loading and industry knowledge base docking; (2) brain-inspired memory tagging and replay; and (3) multi-module collaboration and adaptive updates.

4.4.2 Implementation Mechanisms

Pre-trained Model Loading and Industry Knowledge Base Docking: The system deploys pre-trained general large models in server or cloud environments and incorporates industry expert knowledge bases into the semantic representation space through fine-tuning or incremental training. When facing different domain requirements, corresponding knowledge sub-libraries can be dynamically loaded to achieve targeted industry reasoning capabilities.

Brain-Inspired Memory Tagging and Replay: Drawing inspiration from biological brain mechanisms for memory tagging and replay, the system tags important scenarios or reasoning paths after completing complex decisions or services. During idle or offline periods, the tagged information is replayed and reinforced, achieving secondary consolidation of potentially high-value knowledge and gradually improving prediction accuracy and execution efficiency for similar tasks.

Multi-Module Collaboration and Adaptive Updates: The private knowledge module, multimodal decision module, and hardware control module can all access the general language model or update internal knowledge indexes through bidirectional communication with the model base. When the system receives new industry-specific data or private information, this knowledge is gradually embedded into the model structure through encrypted indexes and incremental fusion strategies, enhancing its deep reasoning and customized service capabilities.

4.5 Inter-Module Collaboration and Overall Process

In actual operation, "Wenlu" forms a closed loop through a unified interface bus and communication protocol that integrates the four core modules described above. The workflow proceeds as follows: (1) external multimodal data and user request reception, where the multimodal decision unit completes feature extraction or semantic matching and calls the private knowledge unit if private data is involved; (2) core language model inference and memory tagging, where the general model fusion unit retrieves the base model and industry knowledge base for deep analysis of multimodal information and performs memory tagging; (3) private information security processing, where if the inference process or results involve sensitive data, the private knowledge unit is triggered for sandbox decryption, permission judgment, and necessary desensitized output; (4) decision result generation and hardware control, where after determining the high-level decision, the robot control unit converts it into scripts or API instructions and issues them to target hardware; and (5) execution feedback and self-learning, where feedback from devices or the external environment is processed again by the multimodal unit and core model, triggering memory replay and reinforcement learning to continuously optimize the overall performance of the "Wenlu" system.

Through this modular, extensible system design, "Wenlu" can flexibly handle the entire process from software-level Q&A decision-making to hardware execution, achieving true multimodal cognition and embodied artificial intelligence. The modules complement each other, ensuring both private data security and providing deep decision support and automatic execution capabilities for a wide range of industry scenarios.

5 System Innovations and Technical Advantages

Building upon the deep coupling of multimodal fusion, large model semantic understanding, and security privacy protection, the "Wenlu" embodied brain system forms a relatively complete, closed-loop AI innovation architecture. The following elaborates on the system' s main innovations and technical advantages from three levels: technology, application, and ecosystem.

5.1 Technical-Level Innovations

Multimodal Brain-Inspired Memory Tagging and Replay Mechanism: Rather than relying solely on the powerful reasoning and generation capabilities of general large models in text semantics, "Wenlu" incorporates a "memory reinforcement" principle similar to biological brains. It tags core decision paths for multimodal data (text, images, audio, sensor information) and completes replay and consolidation during system idle or offline periods. Through this mechanism, the system accumulates industry-specific experience through continuous interaction, gradually improving understanding and reasoning accuracy for complex scenarios.

Deep Integration and Security Management of Private Data: Addressing common privacy and compliance needs in industry applications, "Wenlu" specifically designs a private knowledge decision unit and encrypted sandbox mechanism. User confidential information and public general corpora are isolated and managed through tag-based approaches, eliminating the risk of mixing private data with public training data. During inference, if sensitive information is called, permission verification and desensitization processing are performed first, strictly ensuring controllable information leakage risks.

Embodied Closed Loop from Cognition to Automatic Code Generation: Traditional AI often remains at the "cognition-decision" level, requiring manual writing of hardware control logic for actual execution. In contrast, the "Wenlu" system includes a robot control and hardware device code generation unit that automatically maps high-level decisions to specific hardware instructions or executable scripts. Without manual translation, it significantly reduces labor costs and development cycles and enables rapid adaptation to environmental changes.

Deep Coupling of General Large Models and Industry Knowledge: To bridge the gap between industry-specific data and open large models, "Wenlu" implements a "general-specialty integrated" strategy in the general model fusion unit. Through small-scale fine-tuning or incremental training, professional content from industry knowledge bases is organically integrated into the representation space of general language models. When facing vertical domain scenarios, the model can leverage both general and professional knowledge advantages to complete more refined Q&A and reasoning.

5.2 Application-Level Advantages

Integrated Multimodal Processing: The system can simultaneously handle text, images, audio, and various sensor data, with multimodal inputs fused in a unified semantic space. In scenarios requiring multi-source information collaboration such as industrial inspection, medical diagnosis, and autonomous driving, decision results achieve higher precision and stronger explainability.

Secure and Controllable Private Data Management: Through an independent private knowledge decision module and tag-based encryption strategy, "Wenlu" constructs rigorous security barriers for user-sensitive information. While balancing data utilization value and compliance risks, it provides a feasible path for enterprise-level private deployment of large models.

End-to-End Embodied Code Generation: The system' s unique hardware control generation unit can directly translate natural language descriptions or multimodal analysis results into executable instructions, significantly shortening the path from requirement to action. This capability extends not only to robotics but also to various wearable devices, intelligent manufacturing robotic arms, and other hardware forms, achieving a true "perception-cognition-decision-action" closed loop.

Self-Learning and Sustainable Evolution: With the brain-inspired memory replay mechanism, "Wenlu" can reinforce high-value information after multiple decision-making and task execution cycles and relatively inexpensively incorporate new industry data or private information. Without frequent full retraining of large models, the system can continue to grow through iteration.

Balancing Generality and Specialization: Through the "general-specialty integrated" model fusion approach, "Wenlu" not only retains large models' broad coverage of general knowledge but also demonstrates professional standards in vertical domains. This dual capability is particularly valuable for enterprises, healthcare, autonomous driving, and other environments with high precision requirements for intelligent decisions and diverse application scenarios.

5.3 Ecosystem-Level Advantages

Cross-Industry, Multi-Domain Deployment: Relying on its multimodal fusion and private data management architecture, "Wenlu" can provide adapted solutions for different domains such as healthcare, finance, manufacturing, and transportation. The professional knowledge and experience accumulated across industries will also feed back into the system, gradually forming a virtuous cycle of data and model co-evolution.

Compatibility with Existing Large Model Systems: "Wenlu" is compatible with mainstream open-source general models (such as DeepSeek, GPT family, etc.), enabling both migration deployment based on public platforms and customized development. This allows it to meet more application needs while absorbing cutting-edge research results from the industry.

Continuously Evolving Learning Capability: The system' s brain-inspired memory replay mechanism provides diverse experimental and improvement space for subsequent research, such as integrating causal reasoning models or reinforcement learning frameworks to further improve adaptability to complex scenarios and long-term dynamic environments.

In summary, the "Wenlu" embodied brain system technically breaks inherent barriers between multimodal fusion, privacy protection, and hardware decision-making, while demonstrating high scalability and portability potential at the application level. Through forward-looking architecture design and brain-inspired memory management, the system provides innovative value and technical pathways for AI' s deep industry deployment and continuous self-learning evolution.

6 Typical Applications and Extensions

Under the multiple advantages of multimodal cognition, private data protection, and hardware code generation, the "Wenlu" embodied brain system brings considerable application value and development potential to numerous industries and scenarios. The following elaborates on its typical application methods and extensible directions from different perspectives and levels.

6.1 Enterprise Management and Decision Support

Enterprise Management and Business Strategy: Using the "Wenlu" system, massive text reports, market data, and industry expert documents can be semantically analyzed and combined with real-time financial indicators and environmental perception data to assist enterprise executives in making scientific decisions. The private knowledge unit ensures strict protection of internal confidential documents and financial data during use, effectively reducing data leakage risks. Multimodal decision capabilities allow unified input of text, tables, images, and other data for comprehensive judgment, achieving more comprehensive and precise business strategy formulation.

Complex Prediction and Early Warning Analysis: For scenarios requiring cross-domain, multi-stage data analysis such as financial markets and supply chain management, "Wenlu" can deeply integrate textual news, historical transaction data, and sensor monitoring indicators to generate risk early warning or trend prediction reports. The brain-inspired memory replay mechanism can tag and reinforce recent abnormal fluctuations or high-risk events, continuously improving the model’ s responsiveness.

6.2 Robotics and Smart Manufacturing

Service Robots and Smart Home: Combining multimodal decision-making with automatic hardware code generation provides service robots with complete solutions from speech understanding to action execution. For example, home robots can directly perform cleaning, 搬运, or remote monitoring tasks after receiving voice commands. The private knowledge management module can protect family users’ privacy data while allowing robots to access restricted information when necessary (such as home security deployment plans) and execute corresponding actions.

Industrial Manufacturing and Unmanned Production Lines: In intelligent manufacturing, “Wenlu” can judge production line health status in real-time based on sensor data, surveillance images, and production logs, and automatically generate operation scripts for industrial robots to complete flexible manufacturing tasks. When abnormalities occur, the multimodal decision unit combines private knowledge (such as proprietary manufacturer technology) for fault localization, while the hardware generation unit issues temporary control commands to avoid risks, thereby constructing a highly reliable, low-latency production system.

6.3 Healthcare and Autonomous Driving

Medical Imaging Diagnosis: In medical scenarios, the system can integrate patient text records, imaging examination results (X-ray, CT, MRI, etc.), and physiological signals to output more accurate diagnostic suggestions or treatment plan recommendations. The private data unit ensures that sensitive patient case information and hospital internal materials are not leaked, while the general model fusion unit can deeply couple public medical knowledge with proprietary medical institution databases to form a personalized diagnostic assistant.

Autonomous Driving and Intelligent Transportation: For autonomous vehicles, the system can comprehensively analyze visual images, radar/lidar sensor data, voice navigation instructions, and traffic flow information to generate optimal driving strategies in real-time. If confidential maps or security strategies are involved, the private knowledge unit can manage them through encryption and retrieve encrypted data in critical decisions such as route planning and traffic prediction. Automatically generated execution scripts are directly applied to vehicle controllers, achieving highly autonomous vehicle operation.

6.4 Human-Computer Interaction

Smart Customer Service and Expert Systems: Enterprises or public institutions can deploy “Wenlu” in customer service domains to provide users with consultation and services integrating multimodal inputs such as voice, text, and images, significantly improving interaction experience. Through the "general-specialty integrated" knowledge base fusion approach, customer service systems can handle not only general Q&A but also call upon deep industry knowledge to provide authoritative answers for professional scenarios (such as legal consultation or insurance claims). If user privacy data is involved (such as ID cards or policy information), the private decision unit can strictly control it and output customized answers after security review.

Virtual Assistants and Wearable Devices: The system can interface with AR/VR devices, smartwatches, smart glasses, etc., analyze real-time camera footage or sensor data, generate personalized suggestions through language models, and feedback to users in natural language. In healthcare scenarios, virtual assistants can obtain sensor data such as heart rate, blood pressure, and exercise steps, combine them with the general model' s health knowledge base to output exercise and dietary recommendations, while strictly adhering to user personal privacy protection requirements.

6.5 Ecosystem Development

Open Source Models and Industry Data Fusion: "Wenlu" can be deeply integrated with various open-source large models. Through directed fine-tuning or incremental learning of general models, industry data, private materials, and public corpora can be efficiently coupled. Relying on the basic language understanding and generation capabilities provided by general models, supplemented by industry-specific knowledge bases and private data management, a next-generation intelligent hub with greater targeting and scalability can be achieved.

Self-Learning Ecosystem: The system accumulates massive multimodal task data through multiple interactions and continuous use, forming a self-learning cycle combined with the brain-inspired memory replay mechanism. In the long term, "Wenlu"will continuously enhance its mastery of key tasks and algorithmic strategies across industries, driving the evolution from a general large model platform to a broader general artificial intelligence ecosystem.

6.6 Future Expansion Directions

Deeper Causal Reasoning and Explainable Machine Learning: By integrating causal reasoning frameworks and reinforcement learning concepts, further improve system interpretability and robustness in complex decision-making and uncertain scenarios.

Cross-Modal Generation and Reverse Engineering: Beyond code generation, explore multimodal generation (such as text-to-image, text-to-3D action simulation) to assist in design, scientific research, art, and other fields.

Multi-Level Privacy and Compliance Strategies: For stricter or more detailed data compliance scenarios (such as medical privacy laws or financial regulation), more granular access control and auditing functions can be added within the private knowledge decision unit.

In summary, the "Wenlu"embodied brain system demonstrates significant advantages in multimodal fusion, efficient decision-making, and private data protection across multiple industry applications. In fields such as healthcare, finance, industrial manufacturing, and new human-computer interaction, the system can provide end-to-end intelligent services. With the continuous strengthening of its self-learning mechanism, it brings users AI experiences that combine both breadth and depth. As large models achieve further breakthroughs in multimodal understanding, "Wenlu" will continue to have vast space for innovation and expansion, laying a solid foundation for the development of next-generation general intelligent platforms.

7 Implementation Methods and Workflow

To maximize the effectiveness of the "Wenlu" embodied brain system in practical applications, systematic process design and implementation are required from initialization and deployment to daily operation. This section elaborates on its typical workflow and explains key implementation details around the collaboration and information flow of major system modules.

7.1 System Initialization and Deployment

General Model Deployment: In cloud or local server environments, pre-deploy open-source or commercial large-scale language models (such as DeepSeek) and configure corresponding computing resources (GPU/TPU). According to target industry requirements, perform necessary fine-tuning or incremental training on the model to achieve higher Q&A and reasoning accuracy in domain scenarios.

Industry Knowledge Base and Private Data Management Module Preparation: Import or create industry knowledge bases (such as healthcare, manufacturing, finance) to support multimodal decision-making and reasoning. Configure the private knowledge decision unit to encrypt and index materials that may contain sensitive information (text, tables, images, confidential documents, etc.). Establish access control strategies, including role permissions and desensitization output mechanisms, to ensure private data is securely stored independently from public training corpora.

Multimodal Input Channel Setup: Interface with image capture, voice collection, sensor data reading, and other interfaces to achieve real-time connection or scheduled data fetching with external devices/databases. For scenarios requiring physical interaction (such as robots or wearable devices), initialize communication protocols or adaptation layers to ensure smooth command transmission and reception after deployment.

7.2 Private Data Management and Encryption

Private Data Upload and Identification: Users or system administrators upload private files (such as internal corporate documents, medical records, confidential technical solutions) through the backend, and the system automatically performs encryption processing and generates unique identifiers. Using text analysis or predefined metadata, the system conducts preliminary semantic parsing of file content, dividing it into tags for sensitivity level, subject matter, and access permissions.

Permission Management and Index Construction: Based on role-based access control (RBAC) strategies, private data is assigned access levels (such as administrator-only or specific business department-only). A private index table is established within the system, pairing document content embedding vectors or keywords with the general model for subsequent Q&A or reasoning phase invocation.

Encrypted Storage and Sandbox Isolation: A secure sandbox mechanism separates private data from public corpora for storage, combined with encryption algorithms (symmetric or asymmetric) to protect files. When external applications initiate requests involving private information, they must first pass authorization verification before reading or writing sensitive content.

7.3 Multimodal Data Processing Pipeline

Feature Extraction Module: For image data, convolutional neural networks or other vision models are used for object detection, classification, or feature embedding; for audio data, acoustic feature extraction or speech recognition models can be utilized; sensor data is processed through filtering, normalization, and other methods. Raw information from different modalities is converted into unified or alignable vector representations and stored in an intermediate buffer.

Semantic Fusion and Context Encoding: Relying on the multimodal decision unit in the "Wenlu"system, feature vectors for images, speech, text, etc., are combined into the same context encoding model, using attention mechanisms to achieve inter-modal semantic association learning. When necessary, industry knowledge bases can be dynamically called to further identify professional information points contained in features (such as lesion locations in medical images or fault points in industrial inspection).

Privacy Detection and Filtering: When multimodal data contains content that may map to private indexes (such as internal identifiers or confidential parameters), privacy tag checks are performed first. If triggered, coordination with the private knowledge unit verifies permissions before deciding subsequent steps. For parts without sensitive tags, they are directly handed over to the general model for deep semantic parsing and reasoning.

7.4 Decision Generation and Output

User Request/Task Input: Users can initiate queries or decision requests through natural language, mixed text-image, voice, and other methods, or be triggered by external events (such as industrial equipment failure alerts or urgent medical diagnosis needs). If the task involves references to private data or requires high-sensitivity decision-making, the system immediately calls the private knowledge unit to verify access permissions.

General Model Fusion Unit Inference: User requests, multimodal inputs, and industry-specific knowledge (and private information if permissions allow) are input together into the general language model for reasoning. A brain-inspired memory tagging mechanism is introduced to preliminarily tag important knowledge points or reasoning paths involved in this inference process.

Output Decision or Q&A Result: The general model fusion unit outputs preliminary decision conclusions or response content, and the multimodal decision unit annotates and explains visual objects (such as diagnostic images). If private data or sensitive information is involved, desensitization checks are performed before result release to ensure that output text or images do not exceed user permission scopes.

7.5 Hardware Control and Code Generation

Hardware Control Request Trigger: Once the system decision result includes robot actions, wearable device scheduling, or other hardware control requirements, the hardware control generation unit is triggered. For example, a user inputs "Please control the robotic arm to grasp the red object and place it in area X," or multimodal detection results show "Need to close specific valve to prevent fault escalation."

Automatic Script/Instruction Generation: Based on ROS2 or target hardware API calling rules, the hardware control generation unit converts natural language or multimodal decision results into control scripts (such as Python or C++). For different models or brands of robot systems, the underlying adaptation layer translates unified intermediate instructions into commands for specific drivers.

Execution and Status Feedback: The system issues generated scripts or commands to target hardware in real-time. If devices have feedback capabilities (sensors, status logs, etc.), data is automatically incorporated into the multimodal analysis module for secondary evaluation. If obstacles or environmental mutations are encountered during execution, the system can quickly output new control strategies through re-reasoning, achieving dynamic adaptive closed loops.

7.6 Memory Marking and Reinforcement Learning

Key Decision Point Tagging: After completing a full interaction, reasoning, or hardware operation, the system adds memory annotations to key information and reasoning steps in the process, recording their weight distribution in multimodal fusion and knowledge base invocation. If private data was called during the task, its "invocation path" is also securely recorded for subsequent auditing and strategy optimization.

Offline Replay and Reinforcement: During system idle periods or offline batch processing stages, important decisions from past sessions are replayed according to memory tags. Reinforcement learning or parameter optimization methods are used to improve execution efficiency and accuracy for similar tasks. This process does not require retraining the entire large model; only policy parameters or fusion units within the system need fine-tuning to achieve continuous improvement.

Index and Strategy Updates: After replay completion, reinforcement results are written into knowledge index tables or private data management strategies, allowing subsequent tasks to directly benefit. This forms a positive cycle of "multiple interactions-memory tagging-offline reinforcement-performance improvement," driving system self-evolution.

7.7 System Maintenance and Updates

New Industry Knowledge Base Integration: When expanding to new industries (such as agriculture, construction, energy), relevant knowledge bases can be imported through incremental learning, with corresponding model plugins added to the multimodal feature extraction stage. The general language model only requires lightweight parameter updates, reducing the overhead of repeated training.

Private Policy and Security Compliance Updates: In response to evolving laws, regulations, or internal compliance requirements, the private knowledge unit's permission rules, tag management schemes, and desensitization algorithms can be flexibly configured. Encryption algorithms and security protocols are regularly checked to maintain the system's latest adaptation for privacy and compliance.

Hardware Interface and Adaptation Layer Maintenance: As hardware devices are upgraded or new platforms are connected, only the adaptation layer scripts and API mappings in the hardware control generation unit need updating, without affecting high-level decision logic. During long-term operation, data on different hardware failures or abnormal scenarios can be collected to further optimize code generation stability.

7.8 Example Workflow

Scenario Description: Industrial scenario: A production line triggers a failure alert, and the system receives data streams from sensors (temperature, pressure), surveillance camera images, and text descriptions input by operators.

Multimodal Fusion and Private Determination: The multimodal decision unit performs feature extraction to identify fault locations; if manufacturer proprietary component information is involved, the private knowledge unit performs secure invocation determination.

General Model Inference and Output: Combining industry knowledge base and general model to determine the most likely fault cause, recommending replacement of a specific part or parameter adjustment.

Automatic Maintenance Robot Instruction Generation: The hardware control unit translates the fault handling solution into robot movement, grasping, detection, and other action scripts; the robot feeds back data for secondary verification after execution.

Offline Replay and Reinforcement Learning: During system idle periods, the decision path of this session is tagged and replayed, recording features causing the fault, repair methods, and execution efficiency to form gains for next decision.

Through the above complete workflow, the "Wenlu" embodied brain system can continuously absorb multimodal information, industry experience, and user private data during daily operation, achieving self-learning and functional iteration with minimal additional training costs. Meanwhile, the end-to-end hardware execution capability gives the system closed-loop characteristics from cognition to action, significantly shortening the cycle from intelligent decision-making to actual deployment while providing higher reliability and security in critical scenarios.

Geng Liang received his master' s degree from Hebei University of Technology (Tianjin) in 2012. He is currently a doctoral student at the School of Artificial Intelligence, Beijing University of Posts and Telecommunications (Beijing), and concurrently serves as an assistant researcher at the School of Mechanical and Electrical Engineering, Shijiazhuang University, and an assistant researcher at the Shijiazhuang Key Laboratory of Agricultural Robot Intelligent Perception.

Figures

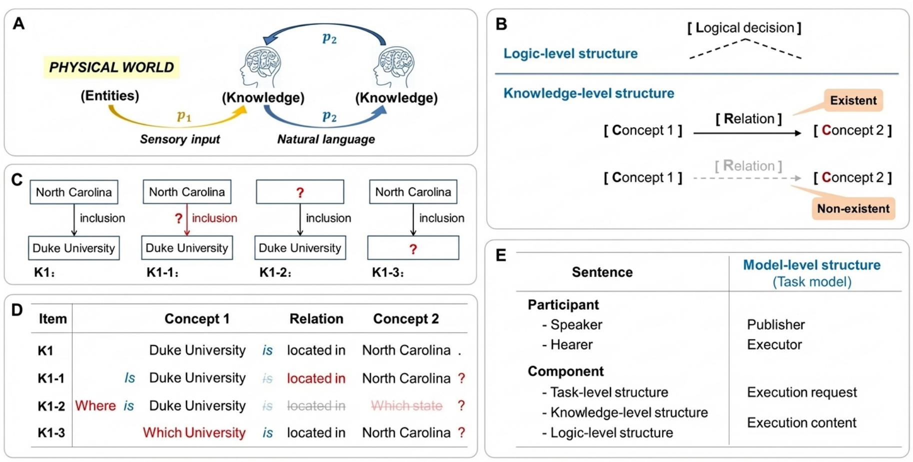

Figure 1: Figure 1