Research on the Application of Intelligent Agents in WeChat Mini Programs: An Analysis Based on System Architecture and Scenario Practice

Jiaqi Tang, Xiao Liu, Sile Xu
Guangzhou Railway Vocational and Technical College, Guangzhou, Guangdong, China

Abstract: The artificial intelligence industry was identified as a key development direction by the Central Government and local governments across the country in 2024.The integration of artificial intelligence (AI) and mobile internet technologies is deepening, and Intelligent Agent technology is becoming increasingly crucial, as it can enhance the intelligent service capabilities of WeChat Mini Programs. Our research has clarified all implementation methods of Intelligent Agents within WeChat Mini Programs. We integrated advanced technologies such as IoT devices, natural language understanding (NLU), and machine learning to develop an Intelligent Agent technology framework suitable for the WeChat ecosystem, and also created a corresponding system topology diagram. We selected representative application scenarios including life services, tool support, and health management. Through practical cases, we have proven that Intelligent Agents are effective in improving user interaction experiences and simplifying service processes. Additionally, we discussed existing challenges such as data privacy protection, system response efficiency, and user adaptability, and proposed improvement measures. Experimental data shows that Intelligent Agent technology can enhance the interactive experience and intelligent service capabilities of Mini Programs. In the future, with the application of new technologies like 5G communication and augmented reality/virtual reality (AR/VR), the WeChat Mini Program ecosystem will become more efficient and better able to meet users' personalized needs.

Keywords: Intelligent Agent; WeChat Mini Program; System Architecture; Artificial Intelligence; Internet of Things (IoT)

1.Introduction

Since its launch in 2017, WeChat Mini Programs have rapidly penetrated into various fields such as life services, medical and health care, education, and e-commerce, relying on advantages such as "use-and-go", lightweight, and cross-platform, with daily active users exceeding 500 million. However, as users' demands for personalized and real-time services increase, traditional Mini Programs have gradually shown limitations in interaction methods and service intelligence.

As a software entity with autonomy, reactivity, proactivity, and sociality, Intelligent Agent can simulate human behavior, independently perceive the environment, make decisions, and execute tasks[1]. Embedding Intelligent Agent technology into WeChat Mini Programs can realize functions such as voice interaction, intelligent recommendation, and automated services, significantly improving user experience and service efficiency.

Currently, multi-agent systems adopting agent technology have been widely applied in numerous fields such as traffic control, smart grids, manufacturing, drone control, and many others[2].This paper aims to systematically study the application mechanism of intelligent agents in WeChat Mini Programs, propose a reusable system architecture model, and verify its effectiveness through practical application scenarios, providing theoretical support and practical reference for the subsequent development of intelligent Mini Programs.

2.Technical Foundation for the Integration of Intelligent Agents and WeChat Mini Programs

2.1 Core Technologies of Intelligent Agents

WeChat Mini Programs are lightweight applications with prominent technical features and advantages, providing solid technical

support for the adaptation of Intelligent Agent applications.

From the perspective of system architecture, their lightweight design enables stable operation on mobile terminals with minimal resources. This not only reduces requirements for terminal hardware configurations but also improves application startup speed and interactive feedback performance. In terms of user experience, their "ready-to-use" interaction mode eliminates the need for downloading and installation, unlike traditional apps. This zero-threshold access method significantly enhances the efficiency of users accessing services. Moreover, leveraging WeChat's massive user base, Mini Programs can facilitate rapid service dissemination, connect various scenarios, and provide numerous practical venues and user traffic entry points for Intelligent Agent applications. In terms of technical implementation, they support the mixed use of multiple programming languages and development frameworks. This creates a highly compatible engineering environment for the integrated deployment of Intelligent Agent technology, enabling the smooth integration of intelligent functional modules into the development system.

In terms of application scenarios, WeChat Mini Programs have penetrated numerous fields such as life services, utility tools, education and training, and healthcare. These scenarios offer landing grounds for the application of Intelligent Agent technology. In local life services, Mini Programs for food delivery, navigation, and mobile payments integrate users' movement trajectory data and consumption preferences. This provides a data foundation for Intelligent Agents to deliver precise service recommendations. For example, in food-service Mini Programs, intelligent systems can build recommendation algorithm models based on users' consumption records and review data to provide personalized dish recommendations. In utility tool applications, functional modules for document processing, schedule management, and information query allow Intelligent Agents to exert capabilities in automated assistance and task processing. Just like intelligent customer service systems, which can quickly understand user instructions and guide operations through natural language interaction technology. In educational Mini Programs within scenarios such as online courses and knowledge assessments, Intelligent Agents can develop personalized learning plans and match learning resources based on students' learning behavior data[3].In the healthcare field, through scenarios like physical sign monitoring and disease screening, Intelligent Agents can assess risks and provide intervention suggestions based on users'health data. These segmented scenarios not only demonstrate the wide application scope of WeChat Mini Programs but also provide a systematic analytical framework for researching the application of Intelligent Agent technology across different scenarios.

2.2 Technical Characteristics of WeChat Mini Programs

WeChat Mini Programs have the following characteristics suitable for intelligent agent deployment:

Lightweight and Fast Startup: Suitable for deploying lightweight intelligent agent modules.

Rich API Interfaces: Supports calling hardware resources such as cameras, microphones, and geographic locations, facilitating intelligent agents to perceive the environment.

Cloud Development Support: WeChat CloudBase provides databases, storage, and cloud functions, facilitating intelligent agents to realize data storage and remote control.

Ecological Closed Loop: Seamlessly integrates with WeChat Pay, official accounts, WeChat Work, etc., facilitating the construction of a complete service chain.

3.System Architecture Design of Intelligent Agents in WeChat Mini Programs

3.1 Overall System Architecture Diagram

To clearly present the technical implementation logic of the intelligent agent in the WeChat Mini Program, this paper draws a system overall architecture diagram (Figure 1). This architecture is divided into four layers from bottom to top: the physical layer (optional), the service and data layer, the intelligent agent middle layer, and the user layer. Each layer cooperates to complete the intelligent interaction process.

3.2 Analysis of Architecture Characteristics

Modular Design

As the basic support of the architecture, it adopts the design principle of high cohesion and low coupling, decomposing the system into multiple

independent functional components such as user interaction, data processing, intelligent analysis, and result output. Each component communicates through standardized interfaces, focusing on core function development internally and avoiding redundant dependencies between components. This design not only allows the development team to carry out research and development of different modules in parallel, greatly improving development efficiency, but also facilitates subsequent independent upgrades and maintenance—when optimizing intelligent analysis algorithms, only the corresponding module needs to be iterated without modifying the overall architecture; when a module fails, it can be accurately located and repaired to avoid fault diffusion affecting the overall operation of the system.

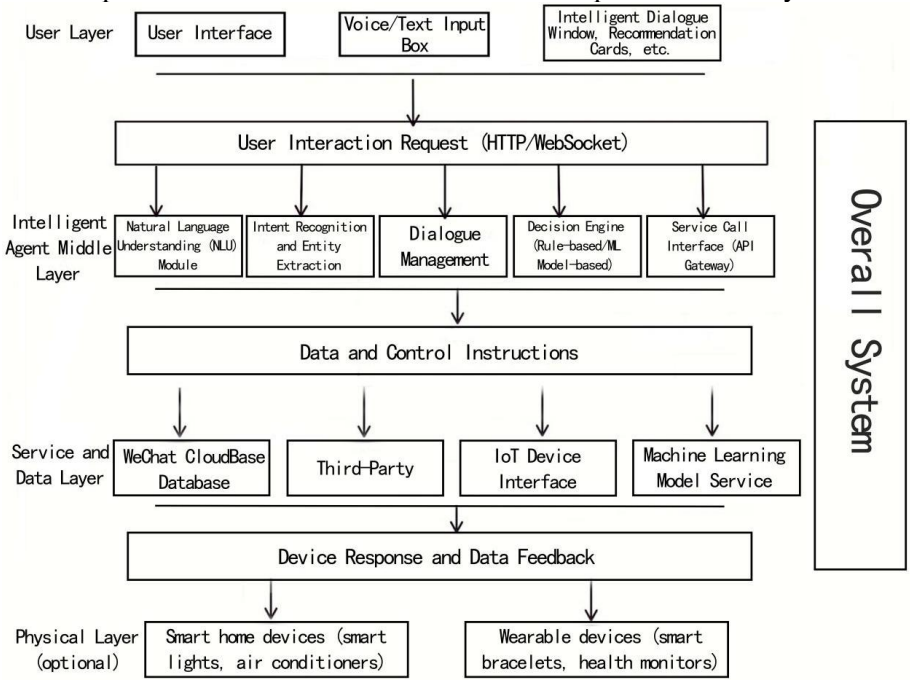

Figure 1. Overall System Architecture Diagram of Intelligent Agents in WeChat Mini Programs

Cloud-End Collaboration: The architecture achieves efficient resource allocation and experience optimization, adopting the mode of "heavy cloud-side logic + lightweight end-side interaction". The core logic of intelligent agents, including heavy-load tasks such as complex algorithm operations, large-scale data processing, and model reasoning, is deployed on the cloud server cluster, relying on the powerful computing power and storage capacity of the cloud to ensure processing efficiency; the Mini Program side only retains a lightweight interactive interface, responsible for simple operations such as user instruction reception and result display. This mode not only reduces the installation package size and running memory usage of the Mini Program, allowing users to use it quickly without waiting for long loading, but also realizes real-time updates of core logic through the cloud, ensuring users always use the latest version of the service.

Scalability: The design provides a flexible path for the scenario extension of the system, building a standardized API gateway as the service access hub, and establishing a unified interface specification and access process. When it is necessary to expand new application scenarios, such as adding intelligent customer service, data statistical analysis and other services, only the corresponding service modules need to be developed in accordance with the specifications and quickly connected to the existing system through the API gateway, without large-scale adjustments to the core architecture. This design enables the system to quickly respond to changes in market demand and realize rapid expansion from a single scenario to multiple scenarios.

Security Design: Constructs a full-link user privacy protection system, adopting a multi-layered protection strategy: relying on the login authentication mechanism of the WeChat ecosystem to achieve accurate verification of user identity and avoid unauthorized access; using SSL encryption technology during data transmission to ensure that user information, operation data, etc., are not stolen or tampered with during transmission; establishing a refined

permission classification control system, assigning different permissions according to user roles and operation scenarios, strictly limiting the scope of data access, and comprehensively protecting user privacy and security.

4. Application Scenarios and Case Analysis

4.1 Life Service Scenario: Voice Interaction-Based Intelligent Food Ordering Assistant System

In the field of life services, to improve users' food ordering efficiency and interaction experience, this paper designs and implements an intelligent food ordering assistant system based on voice input.Users can express their catering needs through natural language, such as "I need a spicy chicken set meal". The system performs semantic parsing and intent recognition on the input sentences using Natural Language Processing (NLP) technology, accurately extracting key entities from users' needs (e.g., dish names, taste preferences, set meal types, etc.).

The backend of the system integrates a menu database and realizes automatic matching of dishes and set meals in combination with an intent classification model. Matching results are presented through voice feedback or a visual interface, and users can further complete order confirmation and submission via voice commands. At the technical implementation level, the system adopts a deep learning-integrated natural language understanding model[4](e.g., BERT-based Intent Classification) and combines the cloud function architecture of the Mini Program platform to achieve efficient calls and data interaction with the order management system.

Application practice shows that the system reduces the average order processing time by approximately 40%, significantly improving the service response speed. User satisfaction survey results indicate that scores for interaction convenience and service accuracy have both increased significantly, verifying the feasibility and application value of voice-driven intelligent services in catering scenarios.

4.2 Tool Scenario: Intelligent Parsing Assistant for Document Understanding

To address the issues of high information density and low retrieval efficiency in traditional document reading, this paper constructs an intelligent document parsing assistant[5], aiming to improve users' efficiency in understanding and utilizing unstructured text content. After users upload documents, the system automatically performs operations such as text preprocessing, key information extraction, and summary generation. It also supports semantic-level content retrieval through natural language queries, such as "What are the specific clauses regarding liability for breach of contract in the contract?"

The core technical architecture of the system is based on the pre-trained language model BERT, achieving deep semantic understanding of document content through fine-tuning. Meanwhile, knowledge graph technology is introduced to model the entity relationships within documents, constructing a structured semantic network to support cross-paragraph and multi-level information association and reasoning. Finally, the parsing results are visually presented in a highly readable manner through the Mini Program front-end, facilitating users to quickly locate key information.

The system effectively reduces the cognitive load and time cost for users when processing long texts. Experimental evaluation shows that the information retrieval accuracy rate has increased to over 92%, and the average reading time has been shortened by approximately 50%, significantly improving the efficiency and accuracy of knowledge acquisition. It has broad potential for application in document-intensive scenarios such as law, administration, and education.

4.3 Medical and Health Scenario: Health Monitoring and Management Platform Based on Multi-Source Data Fusion

To promote the transformation of health management from "treatment-centered" to "prevention-oriented", this paper proposes and implements an integrated health monitoring and management platform. Through communication protocols such as Bluetooth or Wi-Fi, the platform real-time accesses wearable devices like smart bracelets and body fat scales, continuously collecting multi-dimensional physiological time-series data of users, including heart rate, blood oxygen saturation, sleep quality, and step count.

The system conducts dynamic modeling of the collected data based on time-series analysis models (e.g., LSTM, Prophet, etc.), identifies

abnormal fluctuation patterns, and performs risk assessment combined with individual historical health records. When potential health risks are detected, the system automatically generates personalized intervention suggestions, such as "It is recommended to appropriately reduce exercise intensity today" or "Sleep quality has declined recently, please pay attention to a regular schedule".

In terms of technical implementation, the platform adopts a three-layer architecture: the IoT [6]device access layer, the data cleaning and feature extraction middle layer, and the machine learning-based risk early warning model layer, ensuring the stability of data flow and the reliability of analysis results. In addition, the system has a built-in privacy protection mechanism to safeguard the security of users' sensitive health data.

Practical applications show that the platform can effectively support early warning of chronic diseases and daily health interventions, enhance users' self-health management capabilities, and provide a feasible path for building an intelligent and personalized digital health service system.

4.4 Challenges and Optimization Strategies

To fully present the core challenges and corresponding solutions in the development of intelligent health assistants, this paper, through Table 1, combs their specific manifestations and optimization strategies from four dimensions: data privacy and security, performance bottlenecks, differences in user acceptance, and cross-platform compatibility.

5. Future Development Trends

5.1 In-depth Integration with Emerging Technologies

An AI agent refers to an AI system that can autonomously plan and execute actions[9].The deep integration of intelligent agents with emerging technologies such as 5G, AR/VR, and digital twins will revolutionize the interaction form and service boundaries of WeChat Mini Programs. 5G ensures the real-time response of intelligent agents with high bandwidth and low latency, realizing lag-free experiences such as high-definition voice interaction; AR/VR makes intelligent agents "tangible"—in AR navigation Mini Programs, they can act as virtual tour guides to explain scenic spot information, and in VR shopping Mini Programs, they can recommend outfits and support virtual fitting; digital twins endow intelligent agents with "environmental perception + predictive decision-making" capabilities, such as in smart home control Mini Programs, where they can replicate the home environment, monitor appliance status, predict faults, and adjust parameters according to user habits[10].

Table 1. Challenges and Optimization Strategies for the Development of Intelligent Health Assistants

Challenges	Specific Performances	Optimization Strategies
Data Privacy and Security	Risk of leakage of user health and behavior data	Strengthen end-to-end encryption, the principle of least privilege, and transparency of user authorization[7]
Performance Bottlenecks	Intelligent agent response delay affecting experience	Adopt lightweight models (such as TinyML), edge computing, and caching mechanisms[8]
Differences in User Acceptance	Elderly users are not adapted to voice interaction	Provide multi-modal interaction (voice + graphics and text), guided teaching, and personalized settings
Cross-Platform Compatibility	Adaptation issues with different devices	Adopt WeChat standard component libraries and adapt to mainstream models

5.2 In-depth Penetration into Vertical Fields

With technological maturity and refined scenario demands, intelligent agents will deeply penetrate vertical fields such as education, government affairs, and finance: in education, they can serve as personalized teaching assistants to customize learning plans and answer questions based on student data, supporting fragmented learning through the lightweight nature of Mini Programs; in government affairs, they can act as "government service consultants" to provide consulting and process guidance by integrating policy databases, linking systems to realize online pre-acceptance of businesses; in finance, they can serve as intelligent financial assistants to recommend products and push market dynamics according to users' risk-bearing capacity, with functions such as financial

knowledge popularization, and get close to users' lives through the convenience of Mini Programs.

5.3 Building a "Mini Program Intelligent Agent Ecosystem"

In the future, the "Mini Program Intelligent Agent Ecosystem" will gradually take shape, forming an open, shared, and collaborative development pattern: a dedicated "Intelligent Agent Market" platform will emerge, where developers can release intelligent agent modules with standardized interfaces, including customer service and data analysis. Mini Program developers can select and configure modules for quick integration, reducing development costs; the ecosystem will also establish unified technical standards including interface protocols and security authentication to realize cross-Mini Program reuse of modules, such as e-commerce intelligent recommendation modules adapted for educational course recommendation; at the same time, a collaborative innovation mechanism will be formed to support secondary development and combination of modules, such as integrating voice interaction and IoT control modules to build new smart home Mini Programs. Through these, the ecosystem will break the functional boundaries of individual mini-programs and promote a more efficient, rich, and intelligent WeChat Mini Program ecosystem[11].

6. Conclusion

This paper systematically analyzes the application mechanism of intelligent agents in WeChat Mini Programs, proposes a four-layer system architecture including the user layer, intelligent agent middle layer, service layer, and physical layer, and verifies its feasibility through multiple application scenarios. The research finds that intelligent agents can significantly improve the intelligence level and user experience of Mini Programs, but continuous optimization is still needed in terms of data security, performance optimization, and user acceptance. In the future, with technological progress and ecological improvement, intelligent agents will become an indispensable core capability of WeChat Mini Programs, promoting mobile services to evolve towards "proactive, personalized, and scenario-based".

Acknowledgments

This research is supported by the "Climbing Program" Guangdong University Students' Science and Technology Innovation Training Program—"Study Partner Aladdin" (pdjh2025bk369).

References

[1] Russell S, Norvig P. Artificial Intelligence: A Modern Approach. Tsinghua University Press, 2018.

[2] Li Y, Xu F, Xie G Q, et al. Review on the Development and Application of Multi-Agent Technology. Computer Engineering and Applications, 2018, 54 (9):13-21.

[3] Sun L, Ning Y N, Wang Z, et al. The Role and Construction of WeChat Mini Programs in the New Era of "Internet + Education". Journal of Heihe University, 2023, 14 (09).

[4] Zhang X, He S Z, Zhang Y Z, et al. Review of Semantic Representation Methods for Natural Language Understanding. Journal of Chinese Information Processing, 2024, 38 (12).

[5] Zhao Y H, Zhang Y S. Review of the Application of Artificial Intelligence Technology in Document Parsing. Journal of Beijing Information Science and Technology University (Natural Science Edition), 2025, 40 (04).

[6] Zhao X W, Teng W, Chu H W. Large-Scale IoT Data Integrity Verification Based on Blockchain-Enabled Cloud-Edge Collaboration. Cybersecurity and Data Governance, 2025, 44 (10).

[7] Chen J. Design and Implementation of Intelligent Health Management Platform Based on Internet of Things. China Digital Medicine, 2021, 16(8): 88-91.

[8] Zhang Q, Liu F. Research on Lightweight Machine Learning Model Application on Mobile Terminal. Software Engineering, 2023, 26(2): 45-50.

[9] Chen Y W. The Rise of Agent Economy: AI Agents Reshape the Business World. Comparison, 2025, (7):15-17.

[10] Li Y, Xu F, Xie G Q, et al. Review on the Development and Application of Multi-Agent Technology. Computer Engineering and Applications, 2018, 54 (9):13-21.

[11] Zhao L. Analysis of Development Trends of Agent Technology in Mobile Applications. Information Technology and Informatization, 2022(12): 33-36.

