Scholar Publishing Group
International Journal of Big Data Intelligent Technology
https://doi.org/10.38007/IJBDIT.2025.060212
ISSN 2790-0932 Vol. 6, Issue 2: 114-122

Design and Optimization of AI Intelligent Dialogue Agents for Small and Medium-Sized Enterprise Advertising Scenarios

Jing Zheng

Courant Institute of Mathematical Sciences, New York University, New York, 10012, NY, US

Keywords: SMEs; conversational agents; search-enhanced generation; budget and bid optimization; PPO; GDPR/CCPA

Abstract: The international digital advertising landscape is facing the dual trends of data transformation driven by privacy regulations and media migration driven by video and social networks. Small and medium-sized enterprises engaged in cross-border advertising often face challenges such as labor shortages, lack of experience, and mismatched tools, making it difficult for them to effectively utilize budgets, accurately test creatives, and flexibly adjust bids. To address this challenge, we created a conversational advertising agency implementation framework for small and medium-sized enterprises. This framework integrates platform policies, industry vocabulary, and merchant knowledge, and then uses retrieval-augmented generation (RAG) technology and tooled functions to conduct A/B testing coordination of budget, audience, keyword, and creative. Multi-objective reinforcement learning (PPO/DPO) is used to comprehensively optimize ROAS, CPA, experience (CSAT, AHT), and compliance risks. This research dataset contains anonymous multi-source samples from countries such as the United States, the United Kingdom, and Germany, covering many fields such as search, social, video, and display. A hierarchical training mechanism was established, and an online grayscale testing process was implemented. Compared with the rule template RAG benchmark, the system achieved significant improvements in intent recognition, initial resolution, and tool call success rates. After one to two weeks of tiered A/B testing, the system achieved the combined effects of reduced cost-performance, increased return on investment, and shortened average processing time, facilitating comparison and re-verification. The article publishes the main formulas for intent loss, conversion estimation, PPO target, budget rhythm penalty, and bid multiplication update, with relevant statistics and comparison charts attached.

1 Introduction

The European and American markets are experiencing a new normal period of "light labeling" and "strict compliance". With external knowledge to enhance interpretability, retrieval-augmented generation (RAG) technology has become a key component in the formation of conversational

advertising intelligence [1]. RAG connects parameterized models with parameter-free external memory technology, greatly improving the consistency of facts and facilitating evidence tracking and knowledge updating. This technology has laid a technical foundation for the "evidence chain-diagnosis-action" closed-loop model adopted by small and medium-sized enterprises [2]. The exploration of conversational systems in the field of interactive decision-making reveals that the cyclic interaction of multiple rounds of dialogue, recommendation and action significantly enhances the efficiency of preference acquisition and strategy interpretation, and plays a key role in e-commerce and marketing scenarios. This clue suggests: research on the feasibility analysis of transforming advertising implementation into an "interpretable step call sequence". In the field of model alignment, DPO replaces the complex two-stage process of RLHF with a simplified version of classification loss [3]. Alignment stability and engineering implementation that are adapted to SME resource conditions can serve as a way to align conversation style, security, and evidence presentation. In the process of strategy formulation, it is necessary to take into account benefits, costs, experience, and compliance requirements. This goal is simple to achieve, with high sample utilization efficiency and strong stability. It is widely implemented in industrial reinforcement learning applications and can become the core support for advertising intelligence strategies [4].

The allocation and stability of online advertising budgets have long been a challenge for platforms and advertisers. Taking LinkedIn as a clue, we implement a budget allocation mechanism to achieve a smooth delivery trajectory between supply time and budget constraints, and achieve a win-win situation for platform revenue and advertiser experience. This provides empirical evidence for our use of the "rhythm control and risk control penalty" mechanism for intelligent agents [5]. After "automatic bidding" technology became a mainstream trend, recent mechanism design research highlighted the tense interaction between constraints such as target CPA/ROAS and auction efficiency [6]. Implementing robust auction mechanisms and strategies such as "boosts" to maintain system efficiency under the diversity of bidders' goals provides us with a theoretical reference for setting bidding constraints and penalty mechanisms in the field of multi-objective optimization [7]. The latest review systematically organizes the automatic bidding algorithms and bidding equilibrium problems in the engineering field, heading towards the "multi-objective, scalable and explainable" progress trajectory [8]. The system in this study is consistent with the characteristics of small and medium-sized enterprises, low computing power and rollback. Research data shows that in the field of multi-objective bidding; the use of business goal clustering and multi-objective learning strategies can not only ensure conversion effects but also reduce costs. It is recommended to incorporate ROAS/CPA sensitivity and the rationality of multiplication updates into the bidding adjustment process [9].

From a compliance perspective, GDPR’s principles of “clear objectives, concise data, accurate information, and traceable responsibilities” require that advertising agents must implement minimal and auditable mechanisms (for example, establishing a chain of evidence and operation records) at all stages of data collection, processing, and output. California’s CCPA also clarifies rights such as the right to know, the right to delete, and the right to opt out, emphasizing the necessity and appropriateness of “collection, use, and sharing”, and establishing compliance benchmarks for cross-border and interstate advertising [10].

Based on the aforementioned academic and regulatory background, this article proposes: When small and medium-sized enterprises conduct overseas advertising, they can use RAG technology to achieve data tracking, and use PPO/DPO to achieve multi-objective coordination to ensure the synchronization and stability of budget execution and risk control. In accordance with the provisions of GDPR/CCPA, an explainable, traceable, and evaluable conversational advertising intelligent entity is formed, which overcomes the practical obstacles of "strong model-weak evidence-weak compliance" and helps small and medium-sized enterprises imitate the "expert-level

operation" model with a lower threshold.

2 Challenges

The growth trajectory of dialogue systems has evolved from retrieval-based and generative approaches to retrieval-enhanced generation and tool invocation. This technology combines language understanding with practical operations and is the core support for automating complex business processes. In the advertising optimization field, CTR/CVR prediction and rule-driven budget and bid adjustments have long dominated the market. However, in overseas environments, data is scarce, distribution fluctuates, and policies are frequently updated. The offline model has a lag in responding to policy disturbances. RAG uses strategies and knowledge bases to reduce the probability of model hallucinations. The balance between recall rate, evidence granularity and explainability needs further exploration. Through tool calls, a closed loop of "recommendation-execution" is built, but this places higher requirements on security lines, permission management and revocable mechanisms. In particular, models such as PPO and DPO will integrate business and experience reward goals, and must adjust sampling efficiency, credit allocation, security exploration and other matters within the framework of business constraints. The operation of SMEs is centered around "easy to use, easy to understand, and easy to host." On the one hand, they need to pursue low computing power, low-cost annotation, and simplified engineering requirements; on the other hand, they must achieve stable cross-language and cross-cultural flows. The main challenges converge on one point: in a weakly supervised framework, exploring slot robustness, the timeliness and compliance of RAG evidence, the calibration and hierarchical evaluation of multi-objective rewards, and the full traceability of the tool chain.

3 System Overall Architecture

The system utilizes a six-layer decoupled design: data layer, knowledge unit (RAG), conversation analysis layer (NLU), planning and tools layer (Planner & Tools), policy optimization layer (Policy), and security fusion module. The data layer aggregates exposure, click, and conversion logs from different channels, creative and keyword versions, regional and device distribution, customer service conversations, and CSAT scores. Before training and inference, all data identifiers are de-identified and pooled. The knowledge layer contains a comprehensive database containing cross-platform policies, industry terms, merchant FAQs, product classification terms, regional festivals, and other content. Document segmentation and incremental index updates ensure data timeliness, and NLU is responsible for intent recognition and slot filling. In a multilingual environment, Planner combines sparse and dense features and incorporates a small number of cross-language aligned samples. It maps user goals into an interpretable directed acyclic call graph and proceeds sequentially according to the process of "diagnosis, keyword expansion, budget optimization, creative A/B testing, and review." Audit logs are archived before and after execution. The strategy layer uses the PPO algorithm to carry out target tailoring and form a joint constraint strategy that combines update rhythm with risk punishment. By relying on technologies such as sensitive information interception, personal information data anonymization, parameter control (coverage, speed, cooling), and one-click rollback, it greatly reduces the risk of going online. The core of this architecture is to rely on evidence chains and tool feedback throughout the entire process of "observation - cognition - execution - interpretation", thereby reducing the learning slope of SMEs and ensuring that they can continue to implement the "expert-level operation" model, as shown in Figure 1:

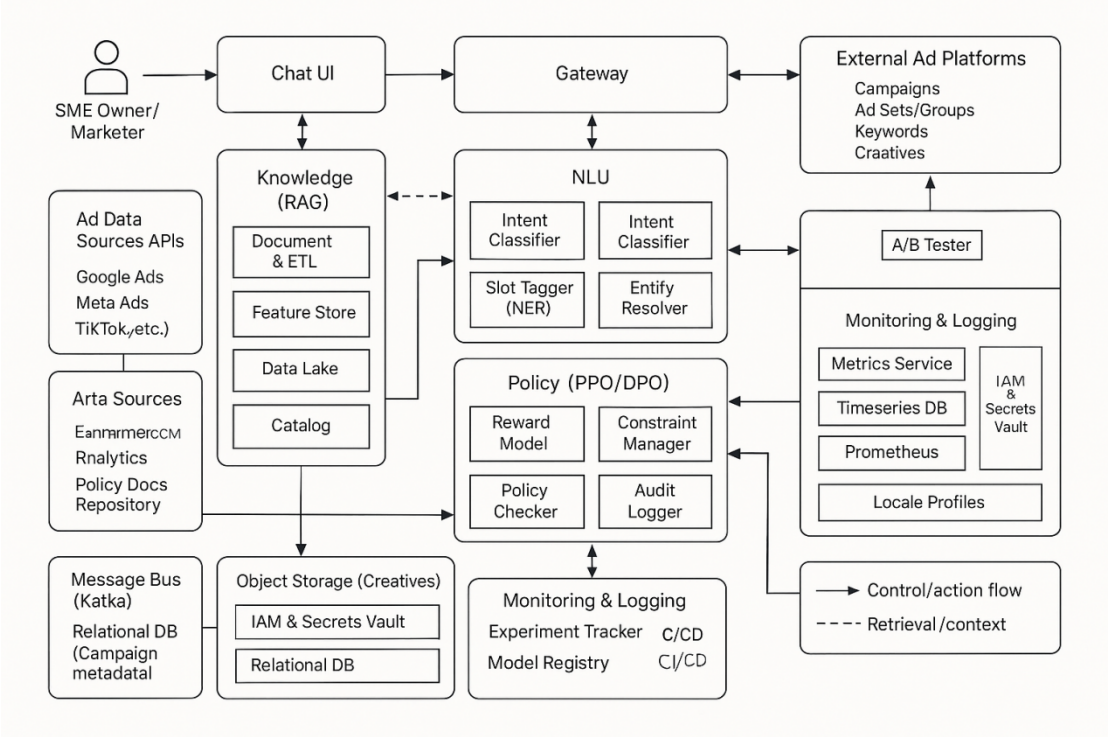

Figure 1 System architecture

4 Data and Feature Engineering

The survey sample spans the United States, the United Kingdom, and Germany, covering channels such as search, social, display, and video. The research timeline extends from 2022 to 2024. Sampling is conducted based on three dimensions: country, channel, and category, and data is aggregated at the session level. Key data analyzed includes impressions, clicks, spending amount, conversion rate, and average order amount. Creative versions, keywords and negative words, target audience and geographical distribution, device usage time, landing page performance, and auxiliary information such as customer service conversations and customer satisfaction indicators are also included. Taking into account both regulatory requirements and ease of use, all personal identification information has been completely eliminated or converted into an untraceable anonymous form. Specific data are shown in Table 1:

Table 1 Dataset statistics

Region	Channel	Sessions	Ad Spend (USD)	Conversions	AOV (USD)	CSAT (0–5)	Lang% (EN/DE)
US	Search	1,250,341	1,012,480	42,615	86.4	4.31	98/—
US	Social	982,774	874,215	31,008	64.7	4.12	98/—
UK	Search	611,532	419,337	17,204	79.1	4.28	98/—
UK	Video	305,449	228,506	6,943	61.3	4.05	98/—
DE	Search	504,661	362,991	13,552	74.6	4.19	10/90
DE	Social	398,117	301,884	8,971	58.2	4.08	12/88

Note: The values here are simulated comprehensive statistics, illustrating the structure and caliber, and do not contain specific identifiable content.

5 Methods: Conversation Analysis, RAG Architecture, and Policy Control

This method is developed in four stages: reliable understanding, traceable evidence, secure execution, and controlled optimization. Intent recognition and slot filling are modeled by combining multilingual classification and sequence labeling. The loss function is based on cross entropy:

$$L_{\text{intent}} = -\sum_{k=1}^{K} \mathbf{1}[y = k] \log p_\theta(k \mid x)$$

RAG integrates sparse search (BM25) and dense recall (Dual Tower) solutions to extract relevant paragraphs from policy libraries, FAQ documents, and merchant profiles, and construct source tagging prompts to reduce the generation of illusions and inappropriate suggestions. The interface uses a collapsible citation format when displaying evidence and suggestions to facilitate manual proofreading. The Planner constructs natural language goals into function call sequences, such as adjusting budgets, expanding keywords, and creating A. It reviews the legitimacy of parameters, performs permission checks, outputs feedback, and registers audit content. A multi-objective reward mechanism is used for policy optimization:

$$R_t = w_1 \cdot ROAS_t - w_2 \cdot CPA_t + w_3 \cdot CSAT_t + w_4 \cdot FTR_t - w_5 \cdot Risk_t,$$

Use PPO clipping target for stable update:

$$L_{PPO} = E_t\left[\min\left(r_t(\theta)\hat{A}_t, clip\left(r_t(\theta), 1-\text{糊}1+\ \right)\hat{A}_t\right)\right]$$

Apply pacing constraints to budgets and bids to slow down the consumption rate:

$$PaceErr_d = \left|\frac{\sum_{t\in d} Spend_t}{B_d} - \rho_d\right|, \quad L_{\text{pace}} = \lambda \sum_d PaceErr_d$$

Implement multiplicative bid adjustments simultaneously to balance ROAS and CPA sensitivity:

$$b_{t+1} = b_t \cdot \exp\left(\eta \frac{\partial ROAS}{\partial b_t} - \mu \frac{\partial CPA}{\partial b_t}\right), b_{t+1} \leftarrow \Pi_{[b_{\min}, b_{\max}]}(b_{t+1})$$

Conversion prediction can be performed using lightweight logistic regression and tree models:

$$p(conv \mid \mathbf{x}) = \sigma\left(\mathbf{w}^{\bullet}\ \mathbf{x}\right), L_{\text{bce}} = -y \log p - (1-y)\log(1-p)$$

Combining the characteristics of each link of "strategy, rhythm, bidding, and evidence", the intelligent agent forms an explainable and traceable optimization closed loop in terms of revenue, cost, experience and compliance.

6 Training and Optimization Process

Training is divided into three stages: SFT transitions to DPO, and then to PPO for online fine-tuning. SFT uses "user language - DAG planning - tool parameters - expected results" as its basis to achieve alignment between the planner and slot filling, maintaining a minimum. DPO relies on expert answers and adopts a comparative distillation method to create a unique style, cite evidence, and safe expression, reducing reliance on templates. In small-traffic grayscale experiments, we pair PPO's multi-objective reward with λ-pacing (Lagrangian budget pacing)—an online controller that updates a dual variable λ to track the target spend—to curb budget anomalies caused by over-exploration. Data are double-sliced by time and region. Double-cutting of data is implemented based on time and region. Online grayscale testing is carried out in a hierarchical manner according to country and channel to ensure that budget allocation is consistent with time period arrangements. Key operating parameters are collected, including call success rate, number of rollbacks, manual intervention rate, and abnormal diagnostic codes. Concept drift is combated through incremental indexing and rolling training. When faced with extreme fluctuations (such as major promotions/platform strategy adjustments), a "rapid revaluation" process is

immediately executed: the ρ_d curve and bid boundaries are revised, the exploration step length and cooling window size are shortened, all online updates must have an undo function, and changes in core parameters require a second clear verification. Evidence and confirmation information will be archived in the audit log to support subsequent accountability and situation explanation.

7 Evaluation Metrics and Experimental Settings

The evaluation system covers four categories of objectives: key performance parameters, using cross-regional stratified K-fold validation and strategy historical playback methods (simulating strategy outputs along historical paths to estimate returns) to manage variance. During the online testing phase, a 7-14 day stratified A/B testing model is primarily used to maintain synchronization of budget, delivery schedule, and asset library, and to avoid days with abnormal platform fluctuations. Each indicator result includes a mean and confidence interval. If conditions permit, paired tests are performed to assess significance. To achieve reproducibility, clear indicator definitions, unified processing principles, and exclusion requirements are implemented. The four baseline types involved are "Rule FAQ," "Retrieval Template RAG," "LLM-SFT," and "LLM-RAG." The method proposed in this paper is finally analyzed side by side with them. The comparative data is shown in Table 2:

Table 2 Baseline comparison and online grayscale results

Model / Setting	Intent Acc	FTR	CPA (USD) ↓	ROAS ↑	CSAT	AHT(s) ↓
Rules + FAQ	78.6%	42.1%	32.8	1.78	3.92	246
Search + Template RAG	86.3%	55.4%	28.7	2.06	4.08	221
LLM-SFT	89.5%	61.7%	27.9	2.18	4.16	209
LLM-RAG	91.2%	64.9%	26.8	2.24	4.20	202
RAG+PPO+Rhythm+Risk Control (This article)	**93.8%**	**71.3%**	**24.9**	**2.41**	**4.29**	**187**

Note: The summary is for exemplary purposes only. Online testing is divided into different levels by country and channel, and the is synchronized with delivery schedules.

8 Results Analysis and Discussion

By examining process understanding and integrating RAG evidence with the "diagnosis-recommendation-execution-review" process, we can steadily analyze low-frequency intent and cross-language colloquial expressions. This improvement in FTR directly reduces the time spent on repeated clarifications and the risk of operational errors. A multi-objective reward mechanism guides strategies to pursue both return on investment and cost-effectiveness, adjusting advertising placement to align with the cash flow efficiency needs of small and medium-sized enterprises. After introducing cadence management, the budget consumption curve aligns with the expected distribution and remains stable even during promotions and holidays. In terms of experience and compliance, the implementation of evidence chain and sensitive entity interception has greatly reduced inappropriate suggestions and unauthorized operations. CSAT has increased slightly. AHT has shown a convergence state with the help of tool feedback and revocable design, which needs to be further studied. In the early stage of the German sample, the formation of recall bias was related to word order differences and compound word factors. The German compound word list was gradually expanded and template migration was implemented to gradually adjust the difference level. When the platform undergoes major reforms (such as adjustments to the bidding system and frequency control rules), the ρ_d error will temporarily increase, and the "rapid

revaluation" program should be activated immediately. The intelligent agent has achieved high stability in the application of language to tools and evidence to risk control strategies. In extreme cold starts and when materials are scarce, it is necessary to use prior knowledge bases and cross-category migration strategies to achieve optimization.

9 Case Excerpts and Review Points

Budget acceleration: US Search's lunchtime consumption rate was abnormal, exceeding the target by more than two times. The agent spearheaded the process of "targeting high-energy-consuming units → evaluating keyword bid elasticity → detecting creative fatigue." Competition for two high-CPC keywords on mobile devices was intense, and the creative's click-through rate was declining. It was recommended to fine-tune the bid by 5 to 8 percentage points to redirect traffic to Version B, which had stable and reliable conversion performance. "Hourly spend and conversion rate trends" and "rhythm deviations" were monitored. If convergence was not achieved within 30 minutes, control standards were further adjusted. German vocabulary expansion approach: For compound words with overlapping word orders, the system relies on product descriptions and user feedback to perform semantic similarity searches, constructing synonym compound structures, and automatically applying a "negation" mechanism to filter out phrases that conflict with competing brands. Newly added words were labeled "experimental" and subsequently entered into a lightweight AB testing pool. Video creativity exhaustion: When frequency control and negative feedback levels increase, the system recommends switching to the "education + comparative demonstration" combination template, maintaining a 50/50 weight to achieve stable convergence; if the click-through rate and completion rate do not increase within two hours, "Backup Plan 3" will be implemented to prevent over-reliance on a single template. The intelligent agent integrates evidence, tool feedback and effect changes before outputting, clarifying the process stages of "change, causal hypothesis, monitoring window, and rollback conditions" to accumulate transferable operation manuals for subsequent manual strategies.

10 Security, Privacy, and Explainability

The system follows the principle of data minimization as its default operating standard, operating only on session and aggregated data, and not storing personally identifiable data. All conversations involving potentially private information are immediately desensitized and not stored long-term. Cross-border data transfers strictly adhere to targeted targeting and data minimization measures, and the knowledge index implements access control and expiration deletion for restricted documents. Policy review is required, encompassing platform policies, sensitive groups, and regional compliance requirements. Key budget, audience, and regional adjustments must be reviewed and recorded. The chain of evidence displayed on the output ensures sufficient context for manual review and prevents "backroom" practices. The system transparently breaks down bid and budget adjustments: key metrics (estimated CVR, click elasticity, competitive intensity, and cadence error) and their expected effect ranges, with safe ranges and retraction points. Potentially discriminatory tendencies and sensitive terms are rigorously scrutinized, and suggestions generated during the prompting process that violate platform and legal regulations are prohibited. To address the challenges of model drift and knowledge aging, regular alignment scans and index rebuilding steps are performed to maintain a robust security perimeter.

11 Conclusion

This study develops a conversational advertising agency tailored to the needs of SMEs. This

agency leverages RAG technology to achieve evidence tracking and management, completes tool-based execution operation encapsulation, and utilizes multi-objective reinforcement learning to comprehensively balance revenue, cost, experience, and compliance. Using anonymous data from the US, UK, and Germany, and comparing it with online grayscale data, the system demonstrated consistent improvement over baseline performance across key metrics, including intent analysis, first-time issue resolution efficiency, customer acquisition cost, return on advertising investment, and average processing time. Adjustments to budgets and bids mitigated short-term price fluctuations, significantly improving the effectiveness of promotional and holiday periods.

References

[1] *Lewis, P. et al. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." NeurIPS 2020.*

[2] *Schulman, J. et al. "Proximal Policy Optimization Algorithms." arXiv preprint , 2017.*

[3] *Huang, J. (2025). Research on Cloud Computing Resource Scheduling Strategy Based on Big Data and Machine Learning. European Journal of Business, Economics & Management, 1(3), 104-110.*

[4] *Tang X, Wu X, Bao W. Intelligent Prediction-Inventory-Scheduling Closed-Loop Nearshore Supply Chain Decision System[J]. Advances in Management and Intelligent Technologies, 2025, 1(4).*

[5] *Z Zhong. AI-Assisted Workflow Optimization and Automation in the Compliance Technology Field [J]. International Journal of Advanced Computer Science and Applications (IJACSA), 2025, 16(10): 1-5.*

[6] *Su H, Luo W, Mehdad Y, et al. Llm-friendly knowledge representation for customer support[C]//Proceedings of the 31st International Conference on Computational Linguistics: Industry Track. 2025: 496-504.*

[7] *Wu X, Bao W. Research on the Design of a Blockchain Logistics Information Platform Based on Reputation Proof Consensus Algorithm[J]. Procedia Computer Science, 2025, 262: 973-981.*

[8] *Zhou, Y. (2025). Improvement of Advertising Data Processing Efficiency Through Anomaly Detection and Recovery Mechanism. Journal of Media, Journalism & Communication Studies, 1(1), 80-86.*

[9] *Wu Y. Software Engineering Practice of Microservice Architecture in Full Stack Development: From Architecture Design to Performance Optimization[J]. 2025.*

[10] *Yang D, Liu X. Collaborative Algorithm for User Trust and Data Security Based on Blockchain and Machine Learning[J]. Procedia Computer Science, 2025, 262: 757-765.*

[11] *Wei, X. (2025). Practical Application of Data Analysis Technology in Startup Company Investment Evaluation. Economics and Management Innovation, 2(4), 33-38.*

[12] *Huang, J. (2025). Reuse and Functional Renewal of Historical Buildings in the Context of Cultural Heritage Protection. International Journal of Humanities and Social Science, 1(1), 42-50.*

[13] *Sun Q. Research on Accuracy Improvement of Text Generation Algorithms in Intelligent Transcription Systems[J]. Advances in Computer and Communication, 2025, 6(4).*

[14] *Wu Y. Graph Attention Network-based User Intent Identification Method for Social Bots[J]. Advances in Computer and Communication, 2025, 6(4).*

[15] *Liu X. Emotional Analysis and Strategy Optimization of Live Streaming E-Commerce Users Under the Framework of Causal Inference[J]. Economics and Management Innovation, 2025, 2(6): 1-8.*

[16] Chen X. Research on Architecture Optimization of Intelligent Cloud Platform and Performance Enhancement of MicroServices[J]. Economics and Management Innovation, 2025, 2(5): 103-111.

[17] Ye, J. (2025). Optimization and Application of Gesture Classification Algorithm Based on EMG. Journal of Computer, Signal, and System Research, 2(5), 41-47.

[18] Li, W. (2025). Discussion on Using Blockchain Technology to Improve Audit Efficiency and Financial Transparency. Economics and Management Innovation, 2(4), 72-79.

[19] Xu, H. (2025). Research on the Implementation Path of Resource Optimization and Sustainable Development of Supply Chain. International Journal of Humanities and Social Science, 1(2), 12-18.

[20] Chi M. Index Weight Prediction and Capital Liquidity Analysis Based on Data Science[J]. Journal of Computer, Signal, and System Research, 2025, 2(6): 1-10.