Available Online: 18 June 2026 DOI: 10.54254/2977-5701/2026.34692

A study on improving the acquisition efficiency of supermarket promotion information based on agent workflow orchestration

Haoyu Zhang

Jinan University, Jinan, China

2507329521@qq.com

Abstract. Retail promotion stacking and price confusion have intensified search frictions, exposing the fragility of conventional rule-based approaches. This study proposes a supermarket discount decision support system that replaces multi-agent architectures prone to intention drift with a dual-model collaborative framework, decoupling intent routing from conversational generation. The system integrates three key mechanisms. First, a structured interaction scaffold is constructed to reduce ambiguity in user inputs. Second, a dynamic fallback and replanning loop is designed to evaluate the confidence of CNN-based Optical Character Recognition (OCR) in real time, enabling autonomous global rerouting whenever confidence falls below a predefined threshold to improve robustness. Third, a memory table is introduced for data consistency verification, establishing a risk-control foundation through a traceable closed loop that spans image feature extraction, cross-validation against publicly available data, and confidence assessment. Experimental results demonstrate that, compared with the 8–10 operational steps typically required in manual workflows, the fully orchestrated system (Agent_full) consistently compresses highly constrained tasks into two interaction steps. The success rate of the T2 price comparison task reaches 86.96%, significantly outperforming B2 (46.15%) and A2 (73.08%). Although task completion time increases to 159.09 seconds, the system achieves higher-quality outcomes and minimal user interaction by shifting the cognitive burden of decision-making from users to system-level computation.

Keywords: supermarket discount decision support system, dual-model collaboration, structured interaction, memory table, price comparison

1. Introduction

Between 2024 and 2026, artificial intelligence systems evolved from prompt-driven question-answering tools into intelligent systems capable of invoking external tools, orchestrating workflows, and autonomously executing tasks. At the same time, the increasing frequency of retail promotions and the growing complexity of pricing rules have made high-quality discount information an essential resource for consumer decision-making, store procurement, and supply chain coordination.

In practice, supermarket promotional information is fragmented, highly dynamic, and insufficiently transparent [1]. Threshold-based discounts, multi-item promotions, and dynamic pricing strategies further exacerbate price confusion. Meanwhile, client-side rendering, dynamic Document Object Model (DOM) obfuscation, and anti-automation mechanisms undermine the reliability of conventional web crawlers, often resulting in data collection failures and data contamination. Manual price comparison requires users to repeatedly switch between webpages, record prices, and integrate shopping routes, imposing substantial cognitive and operational costs. Although single Large Language Model (LLM)-based tool-calling frameworks offer greater flexibility, their limited control over execution flows can still lead to repetitive invocation loops, context loss, intent drift, and numerical hallucinations during cross-store price comparison and voucher verification, making it difficult to satisfy the requirements of reliability, interpretability, and traceability.

To address these challenges, this study develops a supermarket discount decision support system that replaces loosely coupled multi-agent invocation with a dual-model collaborative architecture separating planning and dialogue functions. The planning model is responsible for intent recognition, parameter extraction, and component routing, whereas the dialogue model manages clarification interactions, result integration, and response generation. A structured interaction scaffold is employed to impose upfront constraints on location, product, and voucher information. In addition, a dynamic fallback mechanism is triggered whenever critical information is missing, recognition confidence is insufficient, or inconsistencies arise between the memory table and publicly available data, enabling supplementary information collection, graceful degradation, or task replanning. The study primarily investigates the effects of dual-model collaboration and multi-level fallback strategies on mitigating numerical hallucinations and improving task completion rates. It further evaluates whether the traceability chain—comprising image recognition, public data verification, confidence assessment, and memory table status logging—can support crowdsourced verification of promotional vouchers and facilitate the transformation of unstructured user inputs into reusable digital assets.

2. Literature review

2.1. Search costs and price dispersion in supermarket pricing information

Information economics provides the fundamental theoretical framework for research on supermarket discount information acquisition. Stigler [2] argued that, in markets characterized by price dispersion, consumers incur search costs to obtain low-price information and terminate the search process when the marginal benefit falls below the marginal cost. Subsequent studies on the Diamond paradox and price confusion [3] further demonstrate that, even in technologically advanced market environments, retailers can maintain information opacity through complex pricing strategies and thereby extract excess profits. In modern supermarkets, layered promotional schemes, such as threshold discounts and multi-buy offers, further increase the computational complexity of price comparisons and the psychological burden on consumers, substantially reducing the practicality of traditional manual comparison methods. By leveraging parallel information retrieval and automated reasoning, multi-agent systems partially transfer the time and cognitive costs traditionally borne by consumers into system-level computational and token costs [4], thereby reshaping the mechanism through which search costs are incurred. Existing studies [5] suggest that the automated interpretation of promotional rules by intelligent agents can reduce the marginal cost of information search and alleviate information asymmetry. However, this shift also introduces new challenges, including platform defense mechanisms and the so-called "AI tax", which constitute an important practical backdrop for the system architecture and evaluation framework developed in this study.

2.2. The evolution of information acquisition: from search engines to intelligent agents

The evolution of digital information acquisition has progressed through three stages: keyword-based retrieval, personalized recommendation, and intent-driven intelligent agent collaboration [6]. Conventional search engines rely primarily on document indexing and are ill-suited to supporting complex decision-making in environments characterized by fragmented commercial data. AI-powered search systems and zero-click commerce, driven by Large Language Models (LLMs), integrate natural language interaction with automated tool invocation, internalizing the processes of filtering, comparing, and selecting information that previously required explicit user actions, thereby fundamentally reshaping the information acquisition chain. Cognitive Load Theory provides a psychological perspective for understanding this transformation. Sweller [7] categorizes working memory load into intrinsic, extraneous, and germane cognitive load. Traditional price comparison significantly increases extraneous cognitive load, often resulting in decision fatigue. Highly automated intelligent agents can reduce this extraneous burden, enabling users to focus primarily on final purchasing decisions. However, previous research [8] has shown that insufficient system transparency or frequent hallucinations may induce automation bias and increase supervisory costs. Recent studies therefore emphasize the importance of maintaining a human-in-the-loop framework and introducing appropriate levels of friction, advocating the explicit presentation of decision pathways and uncertainty. These insights motivate the design of the observable tool invocation trajectories and fallback mechanisms adopted in this study.

2.3. Research on agent workflow orchestration and dual-model collaboration

Multi-agent systems and workflow orchestration represent a convergence of distributed artificial intelligence and AI systems engineering. Research in distributed AI [9] suggests that collaborative mechanisms can enhance system robustness, whereas industrial practice [10] has revealed that fully autonomous agents lacking explicit control flows are susceptible to repetitive execution loops, leading to uncontrolled costs and latency. Consequently, recent studies [11] have shifted toward constraining nondeterministic behaviors through explicit workflows and safety guardrails. Rather than adopting a rigid graph-based orchestration framework, this study implements a lightweight dual-model collaborative architecture. A planning model is responsible for intent recognition and component routing, while a dialogue model handles interactive clarification and response generation. Information extracted from multimodal user uploads is processed through a general Optical Character Recognition (OCR) module and recorded in a memory table. Specifically, a verification table stores verification status and outcomes, whereas a recognition log preserves recognition states and OCR confidence scores. By combining upfront completeness checks with verifiable memory records, the system minimizes ineffective reasoning paths and supports practical deployment in scenarios requiring high robustness and low latency.

2.4. Advances in the application of intelligent agents in retail

Extensive research has been conducted on supply chain management and pricing strategies in the retail sector [12]. However, end-to-end intelligent agent systems for promotional information acquisition and price comparison remain relatively underdeveloped. Existing approaches generally overlook the challenge of obtaining highly dynamic front-end retail information. Industry reports [13] indicate that most consumer-oriented price comparison assistants rely heavily on open Application Programming Interfaces (APIs) and fail to address the complexities of promotional rule interpretation and cross-platform information alignment. Research on agentic commerce [14] further suggests that consumer behavior is gradually shifting toward agent-assisted decision-making, with intelligent agents increasingly acting on behalf of users in commercial transactions. Although intelligent agents have the potential to deliver frictionless decision support, increasingly sophisticated anti-crawling measures employed by digital platforms have created a highly adversarial data acquisition environment. Academic research has therefore explored the integration of crowdsourced feedback with localized knowledge bases, encouraging users to upload physical purchase receipts and promotional vouchers for verification by multimodal intelligent agents. The dual-source data ecosystem proposed in this study, which combines a daily updated price database with crowdsourced promotional vouchers, represents a practical engineering implementation of this emerging research direction.

2.5. Evaluation dimensions and analytical framework for intelligent agent systems

Traditional information retrieval metrics are insufficient for evaluating multi-step reasoning and workflow-oriented intelligent agent systems. Research on agent workflow evaluation [15] emphasizes that relying solely on final task success rates may conceal critical failure modes, such as excessive token consumption or intermediate hallucinations. Emerging evaluation frameworks therefore incorporate trajectory quality and node-level assessment, characterizing system behavior through dimensions such as tool parameter accuracy and hallucination prevention mechanisms [16]. For production-grade systems, operational costs and the extent of cognitive load reduction should also be incorporated into the evaluation process. In accordance with the experimental design, this study establishes a comprehensive evaluation framework that integrates conventional performance indicators, including Task Completion Time (TCT_sec) and Task Success Rate (Success), with process-oriented metrics such as the total number of interaction Steps (Steps_total), Tool invocations and Null invocations (Tool_calls and Null_calls), Memory hits (Memory_hit), and Fallback occurrences (Fallback_count). This multidimensional evaluation framework provides a holistic assessment of the effectiveness of intelligent agents in improving the efficiency and reliability of supermarket discount information acquisition.

3. Research design and system implementation

This study focuses on the design and implementation of an intelligent agent architecture for the real-time acquisition and decision optimization of offline supermarket promotional information (see Figure 1). To address the highly fragmented, spatiotemporally heterogeneous nature of retail pricing information, the system establishes a dual-channel data acquisition mechanism at the data ingestion layer. On the one hand, distributed web crawler technology, combined with retailer-priority and periodic scheduling strategies, is employed to automatically collect structured promotional product metadata from the official platforms of major supermarket chains, such as Hualian, Carrefour, and RT-Mart. On the other hand, a multimodal crowdsourcing framework is introduced to complement long-tail retail scenarios. Leveraging computer vision techniques, including Convolutional Neural Networks (CNNs), the system extracts salient memory-point information from unstructured supermarket and product images uploaded by users and integrates these data with a dynamic memory table mechanism for multi-source cross-validation and confidence estimation.

At the level of underlying data flow and knowledge representation, massive heterogeneous product and pricing data are subjected to extensive preprocessing and cleaning. Through deep representation learning models, structured textual information and multimodal visual features are implicitly encoded into high-dimensional continuous vector representations and persistently integrated into the knowledge base [11]. To balance the semantic generalization capability required for information retrieval with the numerical precision necessary for price matching, the indexing and retrieval module adopts a hybrid architecture that combines inverted-index filtering with Approximate Nearest Neighbor (ANN) search, enabling millisecond-level, high-relevance knowledge retrieval under high-frequency agent invocation scenarios. During the prediction and inference stage, a deep feature-crossing network computes interactions among retrieved base features and dynamically generates higher-order interaction vectors, allowing the system to capture complex nonlinear pricing dynamics and spatiotemporal dependencies.

At the business decision and logical control layer, the system adopts a dual-model collaborative architecture that decouples perception from execution. The planning model is constrained to function primarily as an intent parser and parameter extraction engine. To mitigate the tendency of large language models to suffer from state-space explosion and hallucination propagation in global planning tasks formulated as Markov decision processes, the system fundamentally restricts the unconstrained heuristic search behavior of generative models. Through multi-turn interactions, the planning model grounds dispersed user intentions into structured parameters and relies on a knowledge graph to perform standardized clustering and semantic alignment of products across different retail platforms. Once the key variables have been extracted, the system immediately relinquishes the autoregressive control of the large language model and transfers execution to a deterministic finite-state automaton represented by a Directed Acyclic Graph (DAG). Expert-defined task sequence templates, implemented through constrained decoding and execution masking mechanisms, assign zero transition probability to unauthorized actions, thereby collapsing probabilistic exploration into deterministic state transitions. When searching for the lowest-priced product within a specified local radius and identifying the optimal retail platform, the search process exits the semantic generation space and is instead governed by structured memory tables and deterministic combinatorial optimization algorithms. Subsequently, the system invokes location service Application Programming Interfaces (APIs) and employs precise graph search algorithms, such as the A* and Dijkstra algorithms, to generate optimal multimodal travel routes supporting public transit, metro, and driving options. The resulting route plans provide unbiased spatial node recommendations and estimated travel times. Finally, the dialogue model unidirectionally receives the structured outputs within a secure sandbox environment and performs natural-language information integration and user-oriented response generation.

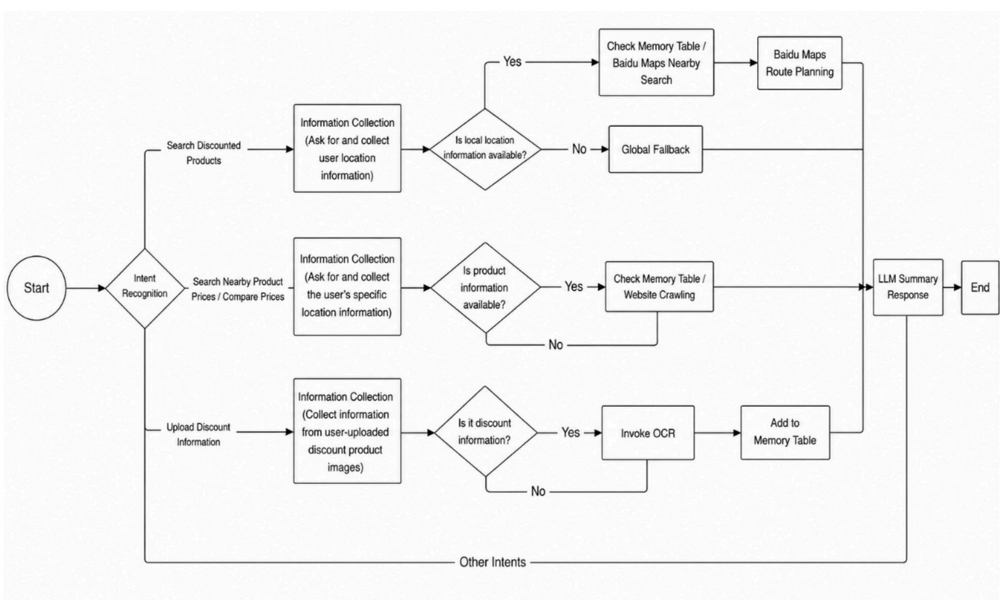

Figure 1. System analysis workflow

At the output layer, the system seeks to bridge the digital divide experienced by older adults by avoiding high-density tabular interfaces that conflict with established principles of cognitive aging. To prevent cognitive channel congestion and working memory overload associated with dense information presentation, the user interface is optimized to minimize visual distractions and adopts a low-density, high-contrast card-based interaction paradigm. Price comparison results and route recommendations are presented without requiring users to interpret complex matrices, while natural language text-to-speech synthesis and two-dimensional graphical geographic rendering are seamlessly integrated. By combining textual, auditory, and visual modalities, the system reduces the cognitive and operational burden on older users and delivers a streamlined, intelligent decision-support experience for complex and high-frequency offline shopping scenarios.

To ensure the integrity of the crowdsourced data repository and the traceability of decision-making processes, the system incorporates a memory-table-based state tracking mechanism that enables full life-cycle verification and persistent recording of user-generated promotional information. When a real-world image containing product price tags is detected, the system simultaneously invokes a CNN-based image understanding model and a general Optical Character Recognition (OCR) module to extract high-precision product names and numerical price values from noisy visual backgrounds. An automated verification workflow is then executed, cross-validating the extracted entity features against publicly available online databases, including map points of interest and official brand websites [17]. Information that satisfies predefined confidence thresholds—including product identifiers, price variations, and the spatial coordinates of physical retail outlets—is decomposed and persistently stored in dedicated product ontology, dynamic pricing, and geolocation memory tables. This closed-loop mechanism establishes a robust data infrastructure for large-scale historical price fluctuation analysis and provides a reliable foundation for efficient, personalized information retrieval based on user behavioral patterns.

At the final stage of the overall workflow, the dialogue model performs information orchestration and summary aggregation. It consolidates dispersed promotional product attributes, historical minimum prices across multiple time horizons, algorithmically recommended retail stores, and precise route plans containing specific transit nodes and estimated travel times into structured, hierarchically organized information cards for presentation to end users. At present, the proposed intelligent agent system remains in the stages of system integration and prototype validation. Nevertheless, its underlying dual-model architecture and multimodal data processing framework have already demonstrated strong engineering stability and considerable potential for intelligent digital transformation in the retail sector.

4. Experimental design and results analysis

4.1. Experimental groups and task scenario design

To quantitatively evaluate the practical effectiveness of the collaborative workflow proposed above, this study employs a controlled experimental design based on standardized task scripts and real-world experimental data. Different technical configurations are abstracted into four experimental settings for parallel comparison. B0_manual serves as the manual baseline. Participants complete the tasks without the assistance of generative AI or automated web crawlers, relying solely on standardized operating procedures to manually search, filter, and compare prices across multiple supermarket websites and publicly available information sources. This setting provides a benchmark for human task completion time and interaction complexity under predefined conditions. B2_no_orchestration represents the non-orchestrated dialogue model baseline. A single large language model equipped with basic tool-calling capabilities is employed, while the dual-model collaborative architecture and strict consistency verification mechanisms are removed. The system is instructed to complete tasks using as few interaction rounds as possible, thereby capturing the native performance characteristics and hallucination baseline of an unconstrained large language model. Agent_full constitutes the proposed framework. This configuration fully activates the collaborative division of labor between the planning model and the dialogue model while integrating the goods, location, and price memory tables, multimodal image recognition capabilities, and public information consistency verification mechanisms. A2_no_home_template serves as the interaction-input ablation setting. While preserving the dual-model collaborative architecture and data access capabilities, this configuration removes the predefined high-constraint input template and allows users to submit requests entirely in unrestricted natural language. The purpose of this design is to quantify the marginal contribution of structured input constraints to overall system efficiency and task success.

The four experimental configurations are evaluated under three standardized task scenarios, T1, T2, and T3, to ensure consistency in input conditions and evaluation criteria. T1 represents an exploratory promotional discovery task, assessing the system's ability to identify same-day promotional products across multiple categories within a specified geographic area and to generate an integrated shopping route under broad and diverse user demands. T2 focuses on target-product price comparison and decision support, evaluating the system's ability to compare prices for a specific product and generate an optimal shopping route under clearly defined geographic and product constraints. T3 is designed as a voucher-driven consistency verification task. This multimodal scenario requires the system to process image-based inputs, perform visual recognition, and integrate publicly available information with memory table records to classify promotional information as either pending verification or verified while updating the corresponding records.

4.2. Evaluation metrics and data cleaning procedures

To ensure consistency in evaluation, all performance measures are organized into a three-dimensional framework comprising efficiency, quality, and cost [15]. The efficiency dimension [8] evaluates task completion speed and user interaction burden, including end-to-end Task Completion Time (TCT_sec) and the total number of user interaction Steps (Steps_total). The quality dimension is measured by the task Success rate (Success), indicating whether a task achieves its predefined objectives. The cost dimension characterizes the computational and operational overhead incurred during task execution. Key indicators include the number of Tool invocations (Tool_calls), the frequency of consecutive errors or repetitive request loops (Null_calls), and the Memory table hit rate (Memory_hit). Memory table hits reflect the extent to which the system successfully reuses existing records stored in the goods, location, and price memory tables. Before statistical analysis, two data cleaning procedures are uniformly applied to all experimental observations. First, tasks with Null_calls ≥ 3 are excluded to eliminate samples affected by persistent execution errors or repetitive request loops. Second, tasks with TCT_sec > 300 seconds are removed to reduce the influence of abnormal system stalls and excessively prolonged executions. All subsequent analyses are conducted using the cleaned dataset.

4.3. Results analysis

4.3.1. Interaction efficiency: step compression and cognitive offloading

By integrating multimodal interaction channels, including text, voice, and image inputs, Agent_full substantially reduces the physical and cognitive burden imposed on users. Experimental logs (Table 1) indicate that the B0_manual baseline requires an average of 9.64, 8.07, and 8.00 interaction steps to complete tasks T1, T2, and T3, respectively. Under this setting, users must repeatedly switch between different webpages, record prices, and manually compare shopping routes.

In contrast, B2_no_orchestration requires an average of 2.43, 2.15, and 2.00 interaction steps for T1, T2, and T3, respectively. Agent_full consistently completes all three task types in exactly 2.00 interaction steps, whereas A2_no_home_template requires 2.79, 3.00, and 3.15 steps, respectively.

Table 1. Comparison of the average total number of interaction steps (steps_total) across experimental groups

Experimental Group	Average Interaction Steps in T1	Average Interaction Steps in T2	Average Interaction Steps in T3
B0_manual	9.64	8.07	8.00
B2_no_orchestration	2.43	2.15	2.00
Agent_full	2.00	2.00	2.00
A2_no_home_template	2.79	3.00	3.15

These findings suggest that Agent_full effectively transfers the burden of navigating multiple information sources, performing manual visual comparisons, and integrating shopping routes from users to the system backend. Owing to its built-in workflow for automatic product categorization, cross-platform price comparison, and route optimization, users are only required to submit structured requests at the beginning of the interaction and receive the lowest available prices, recommended retail platforms, and optimized travel routes as the final output. This design directly addresses one of the major barriers faced by older adults and other user groups with limited digital literacy.

4.3.2. Task success rate: system robustness under complex constraints

Among the three experimental scenarios, T2, the target-product price comparison task, provides the most stringent evaluation of system robustness because it requires the simultaneous satisfaction of strict geographic and product constraints while generating both the lowest-price recommendation and an accurate travel route. Experimental logs (Table 2) show that the B0_manual baseline achieves a 100% success rate through sustained human attention and effort, albeit at the cost of substantial manual time investment.

The experimental results reveal that removing the dual-model collaborative mechanism reduces the T2 success rate from 76.92% to 53.85%. Furthermore, when dual-model collaboration is retained but the front-end structured input template is removed, requiring the system to extract all constraints from unrestricted natural language, the T2 success rate declines further to 46.15%. These findings indicate that, in open and dynamic data environments, both the dual-model collaborative execution framework and structured input constraints are indispensable. Together, they constitute the key mechanisms for suppressing hallucinations and maintaining operational robustness.

Table 2. Comparison of task success rates across experimental groups

Experimental Group	T1 Task Success Rate	T2 Task Success Rate	T3 Task Success Rate
B0_manual	100%	100%	100%
B2_no_orchestration	92.86%	53.85%	100%
Agent_full	92.31%	76.92%	83.33%
A2_no_home_template	78.57%	46.15%	76.92%

Compared with the ablation settings, Agent_full leverages the planning model for task identification, orderly invocation of external components, and rigorous verification against publicly available information to improve the T2 success rate to 76.92%. The progressive improvement from 46.15% to 53.85% and ultimately to 76.92% not only demonstrates the engineering effectiveness of separating planning and dialogue functions but also suggests that verification-state logging and duplicate-information filtering help prevent route planning failures caused by noisy or inconsistent data.

4.3.3. Time cost analysis: backend computation and the reconstruction of search costs

Contrary to the intuitive expectation that automation should inevitably reduce execution time, experimental logs (Table 3) indicate that the fully orchestrated intelligent agent system exhibits a noticeable increase in end-to-end Task Completion Time (TCT_sec). For the T2 task, the experienced human baseline (B0_manual) completes the task in an average of 65.36 seconds, whereas Agent_full requires 157.77 seconds. Similarly, for T1, the manual baseline averages 78.00 seconds compared with 173.15 seconds for Agent_full.

Table 3. Comparison of end-to-end Task Completion Time (TCT_sec) across experimental conditions

Experimental Group	Average T1 End-to-End Completion Time (TCT_sec, s)	Average T2 End-to-End Completion Time (TCT_sec, s)	Average T3 End-to-End Completion Time (TCT_sec, s)
B0_manual	78.00	65.36	58.14
B2_no_orchestration	100.48	140.38	81.92
Agent_full	173.15	157.77	129.85
A2_no_home_template	225.57	136.08	221.15

The longer execution time of the system relative to manual operation should not be interpreted as an engineering failure. Instead, it reflects a fundamental restructuring of the underlying cost architecture. Within the Agent_full execution pipeline, the planning model must first invoke web crawlers to retrieve official retail data, then query the price and goods memory tables, compare the retrieved information, and eliminate duplicate records before generating a verifiable recommendation for the lowest price and the optimal shopping route.

In effect, the system substitutes the user's manual effort with backend computational processes involving API invocations and network communication delays that may last from several tens to over one hundred seconds. These machine-side operations replace the time and cognitive resources that users would otherwise expend on navigating multiple webpages, comparing prices, and performing geographical reasoning. For user populations that generally lack the time, motivation, or technical proficiency required for manual price comparison, this strategy of trading longer backend computation for a low-friction, two-step interaction process offers substantial marginal utility and practical value. From the perspective of information economics, the system reconstructs search costs by transferring them from human cognitive labor to computational resources, thereby enhancing the accessibility and usability of complex retail decision-making tasks.

4.3.4. Process-level analysis: memory hits and tool invocation characteristics

At the process level, this study distinguishes between Memory table retrieval (Memory_hit) and external data acquisition. The former refers to the system's ability to match historical records or previously identified product attributes stored in the goods, location, and price memory tables, whereas the latter involves obtaining incremental information through online searches or official website crawlers. Experimental logs reveal that Agent_full exhibits pronounced asymmetries in its invocation patterns across different task types. In T3, where the system is required to process uploaded images, perform visual recognition, estimate recognition confidence, and query the local memory tables for verification status, the memory hit rate remains consistently high. By contrast, T1 involves more generalized and open-ended user requests with relatively weak product-specific constraints. When users submit broad queries, such as "recommend several daily necessities", the granularity of these requests may not align with the existing entries stored in the memory tables, resulting in ineffective memory retrieval. The experimental log for T1-04 indicates that, under such circumstances, the planning model is compelled to invoke external web resources and map-based services more frequently to obtain environmental feedback, thereby increasing the Tool_calls metric. At the same time, B2_no_orchestration, which lacks both the dual-model collaborative architecture and consistency verification mechanisms, is more susceptible to repetitive request loops and abnormal retry behavior when encountering unstable external interfaces or incomplete crawler data. In comparison, Agent_full benefits from the stability isolation provided by the dual-model framework and the explicit state-machine-based execution design, demonstrating superior performance in exception handling and execution stability. These findings suggest that memory tables should not be regarded merely as passive storage components but as active control mechanisms that regulate external tool invocation frequency and enhance workflow robustness.

4.4. Ablation study and system limitations

4.4.1. Ablation analysis of input constraints

The performance of A2_no_home_template provides direct evidence for the importance of predefined interaction templates. After the removal of front-end structured input fields, the system retains both the dual-model collaborative architecture and the memory table mechanisms. Nevertheless, substantial ambiguity and uncertainty remain during task parameter extraction. These results indicate that, although natural language interaction offers considerable flexibility, tasks such as cross-supermarket price comparison and route planning depend heavily on the accurate identification of critical constraints. Moderate front-end input guidance, including explicit specifications of geographic locations and product categories, can substantially reduce the search space available to the planning model and improve overall workflow efficiency. Consequently, structured input constraints constitute an important safeguard for intelligent agent workflows rather than a limitation on user interaction.

4.4.2. Risk identification and practical limitations

The broader deployment of the proposed system requires careful consideration of data security and model transparency. User location information and consumer-uploaded images may contain sensitive personal data and should therefore be protected through encrypted storage, secure data transmission, access control mechanisms, and comprehensive audit trails to mitigate the risk of information leakage. In addition, price comparison recommendations and route planning results should retain supporting evidence, confidence scores, and detailed invocation histories to improve both the traceability of recommendation logic and the interpretability of system outputs. Such transparency is particularly important in applications involving automated decision support and crowdsourced information verification. It should also be noted that the current implementation primarily serves as a proof-of-concept prototype. The interface demonstrations and experimental results are intended to validate the proposed system architecture and interaction workflow rather than to imply comprehensive supermarket coverage or production-level deployment capabilities. Future work should focus on expanding data coverage, improving real-time information acquisition, enhancing system scalability, and conducting large-scale field evaluations under practical retail environments.

5. Discussion

The findings of this study suggest that, for complex retail decision-making tasks, the principal advantage of the dual-model collaborative architecture lies in its ability to achieve effective workflow governance through functional specialization. The planning model is responsible for intent recognition, task decomposition, and component routing, whereas the dialogue model manages interaction clarification, result integration, and response generation. By separating these responsibilities, processes such as price retrieval, cross-supermarket price comparison, and route recommendation are incorporated into an execution chain that is observable, reversible, and verifiable. This architecture reduces the risks of intent drift and numerical hallucinations associated with unconstrained one-shot generation while improving system interpretability and process manageability.

Input constraints constitute another essential component of workflow governance. Cross-supermarket price comparison relies on highly precise parameters, including location, product type, time, and promotional vouchers. Although unrestricted natural language interaction offers considerable flexibility, it also amplifies the uncertainty associated with parameter extraction. The structured interaction scaffold addresses this issue by requiring critical information to be specified in advance, transforming user intentions into executable constraints. This design not only improves the stability of component invocation but also reduces the operational burden associated with page navigation, price recording, and route integration, particularly for older adults and other users with limited digital literacy.

Whether multimodal crowdsourced data can evolve into reusable digital assets depends critically on the system's ability to distinguish between information that has merely been recognized and information that can be considered trustworthy. After the general Optical Character Recognition (OCR) module extracts key attributes, such as product names, prices, and store information, the information verification table and recognition log simultaneously preserve verification status, verification outcomes, and recognition confidence scores. This mechanism prevents unverified recognition results from being directly transformed into deterministic recommendations. Consequently, fragmented user-generated images and promotional information can be progressively converted into structured data through the verification pipeline, providing a reliable foundation for repeated queries, historical price tracking, and personalized information retrieval.

System reliability should therefore be understood as a controllable capability that extends throughout the entire lifecycle of information recognition, verification, storage, and retrieval. The planning model determines task routing, the dialogue model facilitates information exchange, and the memory tables preserve and update verification states, enabling the system to maintain essential contextual information across multiple interaction rounds. When confidence levels are insufficient or anomalies are detected in external data sources, supplementary information collection, graceful degradation, or fallback procedures are automatically triggered. The preservation of confidence scores, verification outcomes, and invocation histories helps mitigate the propagation of erroneous decisions and provides a transparent basis for subsequent auditing and accountability.

From the perspective of practical deployment, risks are embedded throughout the complete workflow, including data acquisition, information extraction, verification, storage, and recommendation generation. Future system development should therefore continue to strengthen privacy protection, improve resilience to fluctuations in external interfaces, and enhance the interpretability of recommendation outputs. Equally important is the need to prevent insufficiently verified information from being directly incorporated into recommendation results. These considerations further underscore that trustworthy intelligent agent systems require governance mechanisms that are inherently traceable, auditable, and capable of controlled rollback at the architectural level.

6. Conclusion and future research

This study addresses the challenges posed by fragmented supermarket promotional information, high search costs, and cumbersome price comparison procedures through the development and validation of a supermarket discount decision support system. The proposed framework employs a dual-model collaborative architecture that decouples planning and execution from dialogue generation. Combined with a structured interaction scaffold, completeness verification, exception handling and fallback mechanisms, and a dual-source data strategy integrating a daily updated price database with crowdsourced promotional vouchers, the system establishes a closed-loop workflow spanning information acquisition, price comparison and verification, and result presentation.

Experimental results demonstrate that Agent_full consistently reduces the approximately 8–10 interaction steps required under the manual baseline to a stable two-step workflow. In the highly constrained T2 target-product price comparison task, Agent_full achieves a success rate of 76.92%, outperforming both B2_no_orchestration (53.85%) and A2_no_home_template (46.15%). These findings indicate that the separation of responsibilities between the planning and dialogue models, together with front-end structured input constraints and memory-table-based verification mechanisms, jointly improves the quality of complex price comparison tasks while mitigating intent drift and numerical hallucinations associated with unconstrained generative processes.

Although the end-to-end execution time of the proposed system exceeds that of experienced human operators, the additional backend computational cost effectively substitutes for users' efforts in information retrieval, price comparison, and route integration. For individuals who lack the time, motivation, or technical expertise required for manual price comparison, this strategy of exchanging computational resources for low-friction interaction provides significant practical value. From the perspective of information economics, the system demonstrates how search costs can be restructured by transferring cognitive labor from users to computational infrastructure, thereby offering empirical support for the intelligent transformation of supermarket promotional information services.

Future research may proceed along three principal directions. First, the robustness of image recognition and general OCR technologies should be further improved to enhance the identification of supermarket storefronts, product images, price tags, and promotional rules, thereby increasing the usability and verification efficiency of user-contributed data. Second, closer integration among memory tables, knowledge storage, and retrieval mechanisms should be developed to improve the reuse efficiency of frequently accessed product, location, and pricing information, supporting both personalized recommendations and rapid information retrieval. Finally, greater attention should be devoted to the interaction characteristics of key user groups, particularly older adults, through the optimization of multimodal interfaces integrating voice, image, and text, as well as multimodal transportation route planning. At the same time, continued efforts should be directed toward strengthening data security, privacy protection, and the interpretability of recommendation logic to facilitate the safe and trustworthy deployment of intelligent retail decision support systems.

References

[1] Elberg, A., Gardete, P., Macera, R., & Noton, C. (2019). Dynamic effects of price promotions: field evidence, consumer search, and supply-side implications. *Quantitative Marketing and Economics, 17*(1), 1–58. https://doi.org/10.1007/s11129-018-9204-4

[2] Stigler, G. J. (1961). The economics of information. *Journal of Political Economy, 69*(3), 213–225. https://doi.org/10.1086/258464

[3] Ellison, S. F. (2016). Price search and obfuscation: an overview of the theory and empirics. In É. Basker (Ed.), *Handbook on the economics of retailing and distribution* (pp. 287–305). Edward Elgar Publishing.

[4] Sun, J., & Hu, D. D. (2024, November 15). How AI search engines are reshaping information acquisition [in Chinese]. *Economic Information Daily,* 005.

[5] Li, B. L., Zhang, H. X., Cao, Y. J., Sun, R. J., & Liu, Y. T. (in press). A*daptive power flow adjustment for power systems based on large language models* [in Chinese]. Automation of Electric Power Systems. Advance

online publication. https: //link.cnki.net/urlid/32.1180.TP.20250728.1548.004

[6] Wang, Y. B., & Zhang, H. Z. (2025). From search engines to AI search: an analysis of the transformation of information acquisition empowered by large language models [in Chinese]. *Editors Monthly, 3*, 74–80.

[7] Sweller, J. (2011). Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Elsevier. https: //doi.org/10.1016/B978-0-12-387691-1.00002-8

[8] Pan, F. (2025). *Research on QoS multi-objective optimization for distributed machine learning systems* (Master's thesis, Guizhou University) [in Chinese].

[9] Qiao, J., Li, Y., & Su, X. W. (2025). *Manus applications and AI agent design guide: from beginner to expert* [in Chinese]. Publishing House of Electronics Industry.

[10] Xu, Y. T. (2025). *Research on automatic construction technologies for machine learning tasks based on large language models* (Master's thesis, Guizhou University) [in Chinese].

[11] Xu, C. F., Zhao, W. J., & Zhu, S. L. (2026). Knowledge reorganization and narrative revitalization of local gazetteer records on traditional products through workflow-agent integration [in Chinese]. *Journal of Nantong University (Social Sciences Edition), 42*(1), 127–141.

[12] Qi, T. Y., Liu, Y., Dou, L. X., Wang, S. W., & Zhang, L. (2025). Intelligent applications of radio management based on large language models [in Chinese]. *China Radio, 11*, 55–58.

[13] Liu, J. (2025). *Research and application of large language model-enhanced enterprise end-to-end data governance* (Master's thesis, University of Electronic Science and Technology of China) [in Chinese].

[14] Qin, J. J., Zhang, J. T., Zhou, X. L., Zhao, M., & Wang, Q. (2025). Innovative applications of AI agents in electric power safety production management [in Chinese]. In China Electronics Enterprise Association (Ed.), *Typical cases and proceedings of technological innovation in electric power safety and emergency management* (2025) (pp. 25–27).

[15] Mu, M. (2025). *Research on the application of intelligent workflows based on large language models in Chinese-English translation of traditional Chinese medicine texts* (Master's thesis, Beijing Foreign Studies University) [in Chinese].

[16] Yao, L. W., Ren, F., Wang, S. Y., Zhang, C., & Li, H. (2025). A generative approach to the development of earthquake emergency response plans [in Chinese]. *Geomatics and Information Science of Wuhan University, 50*(6), 1159–1174.

[17] Chen, L. Q., Hu, H. F., & Zhang, S. F. (2026). *Workflow generation based on Monte Carlo tree search and self-feedback optimization* [in Chinese]. Computer Engineering. Advance online publication. https: //doi.org/10.19678/j.issn.1000-3428.0252940