International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 05 May-2026 | Impact Factor: 3.5

SmartShop: A Local LLM-Based Conversational Kiosk for Product Navigation and Sales Assistance in Supermarkets

R.R. Devapriya*, M. Harshini†, Mrs. D. Deepa ‡
*†Department of Artificial Intelligence and Data Science, St. Joseph's College of Engineering, Chennai, India
devajanani02@gmail.com, harshinim955@gmail.com, deepad@stjosephs.ac.in

Abstract—Customers tend to struggle with the large supermarket settings by failing to locate products and move around the store. This may cause inefficiency and missed sales. Older methods, such as the use of fixed signs or personnel support, can have problems with natural language interpretation and offering useful information. The present paper proposes a local Large Language Model (LLM) powering conversational kiosk that can help customers in supermarkets with the help of natural voice. Customers are able to inquire about the locations of products and availability by speaking and receive clear responses as well as using text and speech.The kiosk uses the locally deployed LLM to understand customer intent and provide answers. This will minimize the use of cloud services hence lowering latency, cost and privacy concerns. The queries that customers make are compared to a pre-existing inventory database in order to locate precise aisle-level products. As well, there are simple rules of sales support, which propose complementary products in the interaction. The kiosk has a conversational robot avatar, speech synthesis, and a visual store map to enhance ease of use and interaction, which is validated by experimental observations of a customized product list and reliability in product guidance, consistency in response quality, and low response times. The findings prove the effectiveness of the local LLM-powered conversational kiosks in helping customers and providing the simplest sales support in supermarket environments.

Index Terms—Conversational kiosk, local large language models, supermarket navigation, voice-based interaction, customer assistance, in-store guidance, multimodal user interface, sales support, retail automation.

I Introduction

Effective customer engagement is seen as essential in the modern retail scene since it enhances not only the customer experience but also, ultimately, boosts sales. Usually, customers have trouble just navigating a store setup, grasping in-store specials, or easily locating items. Simple directory kiosks, barcode scanning systems, static signage, or human assistance by staff cannot manage complicated inquiries, unclear purpose, or requests in many languages [1][2]. Although advanced tracking techniques such heat mapping and RFID-based monitoring can examine purchasing trends, they cannot provide actual guidance or proactive sales advice [3][4].

Recent advances in big language models and conversational artificial intelligence have opened up fresh avenues in the automatic customer care. These models draw hierarchical representations based on natural language inputs to understand customer intent, contextual products association and upsell opportunities [5]. Models like BERT, GPT-based models and LLaMA can produce human like answers and query processing in real time, a highly important factor in responsive and tailored retail experiences [6]-[8].

The majority of AI-based kiosks are based on cloud-driven solutions, thus they imply high costs of operation, longer latency, as well as low transparency in the decision-making process. However, systems like LLaMA 3 installed on-site, with Ollama allow zero-cost inference without compromising data privacy and continue to be applicable to resource-constrained retail contexts [6], [14]-[16]. It was experimentally tested on an edited set of 50 validated product records in various categories which shows that the proposed system is 100 percent accurate in aisle-level location guidance, and is able to execute marketing advice, and is perfectly synchronized in its data, thus confirming its feasibility in practical retail application [11], [12], [18], [19].

At supermarkets, SmartShop is a low-cost, scalable, and transparent solution for intelligent customer guidance and sales optimization using local LLM inference, hybrid rule-based marketing logic, and secure admin management. Part II covers related research on conversational artificial intelligence, retail recommendation systems, and kiosk applications. Section III outlines the suggested hybrid artificial intelligence strategy for understanding customer intent and producing sales pitches. Section IV displays the system architec ture, as well as the designs of admin gateway and kiosk. Performance assessment and implementation is found in section V. Part VI analyzes the outcomes and compares them with conventional methods. Improvement of the same like multilingual assistance and advanced suggestion algorithms are eventually introduced in Section VII.

II Literature Review

Deep learning has completely transformed how the computers comprehend and generate language, therefore, contributing to the meanings, intent, and the system of conversation Near-human accuracy in contextualization. This fundamental Progress is a design based on transformer technology incorporating layers that learn and attention mechanisms with respect to the links amongst themes, words and phrases based on immense texts [11]. Contrary to earlier purely rule-based methods of NLP using deep learning models are capable of learning deeper semantic information out of raw text. Communication in the real world can be done flexibly.

The turning point came when Vaswani et al. introduced the Transformer model in the year 2017 [11]. By substituting self-attention for RNNs, the Transformer allowed the processing of whole sentences in parallel, rather than word by word. This resulted in major advantages:

- This is because bidirectional contextual understanding allows one to extract the information of both the preceding and succeeding words and can as such interpret the semantics in a richer way.
- Long-range dependency modeling is used to sustain consistency in the long conversations and complicated conversation scenarios.
- Scalable training enables these models to train on billions of text tokens, and the effect is high generalization of these models.
- Transfer learning can effectively modify the general language knowledge to particular fields.

These skills are the basis of modern big language models. The Transformer architecture has been further extended to give more natural sounding chats and more context reasoning in further architectures, such as BERT [12], GPT-2/GPT-3 [13], T5 [14], and LLaMA [15], among others. Customer service chatbots [16], virtual medical assistants [18], and sentiment analysis tools have become the basis of many applications [19] and more, including customized retail suggestion systems [20].

A. Conversational AI in Retail Spaces

The situation in retail is quite different, though, and conversational artificial intelligence has to deal with issues that extend much further than question answering. Multi-intent queries or ambiguous ones like; Where is the milk? either can be the milk aisle, or lactose-free, or even milk-based drinks. Likewise, when one orders bread, the order might have an implication of other related products such as butter or jam, and hence an opportunity to upsell. To process such situations, a high level of contextual awareness and semantic processing are needed, which is well-delivered by transformer-based large language models [12], [14], [20].

- The retail environments are unique operational environments because of the dynamic in store conditions whereby inventory, promotions and location of products may vary with time of the day.
- Bilingual or casual questions are also more difficult since the customer can change languages or use colloquial dialects when communicating [17].
- Short response questions are essential where response has to be brief and precise in a crowded retail setting where a customer needs a short and focused answer as opposed to a long discussion..
- Combination with structured information is also vital, whereby the response of the system should be in line with live databases to give precise aisle position and up-to-date promotional information.

Cloud-based chat rooms may be scalable, but they often experience latency, lack privacy and are costly to maintain. They have been proven to achieve a comparable accuracy at almost zero cost and greater transparency by studies of the implementations of LLM, e.g. with systems like Ollama and LLaMA [15], [16]. This enables on-premise implementation and it is particularly appealing to cost conscious retail outlets or low bandwidth settings.

B. Advances in Intent Recognition and Dialogue Modelling

One of the largest transformations in conversational artificial intelligence has been the transition to simple text classification to intent recognition and entity extraction. Existing systems using Transformers are able to provide token precision at the token level such that product names and attributes can be extracted, along with their quantities, in a response to natural language queries with high accuracy [13], [14]. An example of an automated conversion is a request like I need two liters of milk which will be automatically translated to Product: milk, quantity: 2, unit: liters. The present-day dialog systems have thus integrated a number of capabilities tied together.

- Intent classification is the process of learning the kind of help that the user demands, e.g., finding a product, whether it is expensive or not, or whether there are any offers.
- Entity extraction aims at determining the essential mentions of products or product attributes by the user query.
- Slot organizes the information that has been extracted into template categories so that it can be acted upon downstream..

Creating welcoming, response generation—that is, Business-aligned answers both informative and useful. In conversation, qualities become exceedingly valued. commerce, wherein the degree of comprehension directly affects sales conversion rates. Moreover, research have shown that combining LLM semantic reasoning with Rule-based systems like More might also result from a priori association rules. Reliable retail interactions that are understandable [16], [19], [20]. Such a hybrid strategy shapes the conceptual backbone of the SmartShop concept, in which a great Marketing and language model work together Knowledge will help upselling possibilities.

C. Research Gaps Identified

Notwithstanding the fast development of conversational artificial intelligence, certain Unsolved hurdles abound: Heavy reliance on cloud APIs introduces ongoin expense and dependence on outside infrastructure. Real-time language data synchronisation Model inventory and systems are still restricted. Model clarity is still a concern since several artificial intelligence applications Hybrid approaches, which merge structured corporate logic with deep learning. Mostly unknown for commercial uses. Restrictions Show the necessity for a cheap, locally deployable, and clear spoken kiosk that would be able to process multilingual input, catch nuances in intent, and Assist in context-aware sales recommendations. SmartShop By including indigenous communities, straight addresses these problems. LLM inference using Apriori-based recommendation logic combined with real-time database synchronization thus pointing to a fresh path for smart retail contacts.

III Methodology

SmartShop, a conversational artificial intelligence hybrid, is shown here. Architectural design combining Natural Language Processing NLP, apriori association-rule-based suggestion, and locally installed Large Language Model (LLM). Inference is the process by means of which the system is meant to properly Understand client intention; be issue context-aware responses, and create immediate sales Recommendations; unconstrained by the commercial cloud APIs: The general approach of the SmartShop architecture is Fig. 1 shows it.

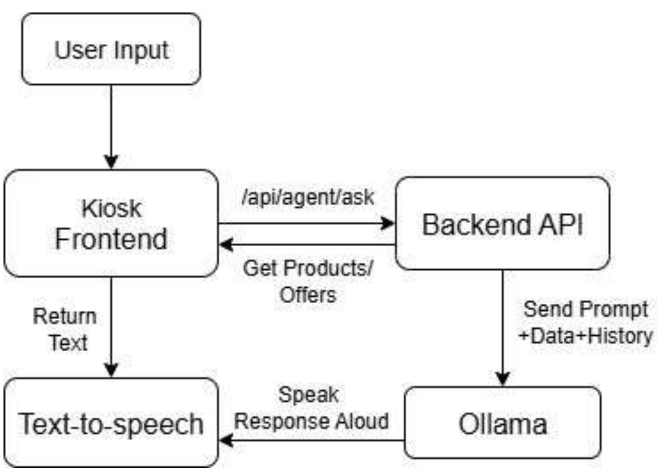

Fig. 1. System Architecture Diagram

- Customer Conversation Kiosk Interface - It enables the customer to communicate with the system by using a conversational interface to get help, ask questions and get support on shopping.
- Backend AI Engine - This engine combines Large Language Models (LLMs) with a knowledge base to take in requests, make response and give intelligent recommendations.
- Secure Admin Portal - Administrators can securely administer offers and inventory without manual processing or bulletin-based processes.

The following subsections describe the complete pipeline of the methodology which encompasses data preprocessing, real time syncing, prompt engineering, NLP inference, association-rule integration and multi-modal response serving.

(i) Real-Time Intent Classification and Detection: React accurately to customer requests, classify them in categories of product (Dairy, Beverages, Snacks, Personal Care, etc.), and provide accurate aisle location retrieval and price data to guide the customer.

(ii) Active Sales Conversion: Use Apriori Association Rules to make complementary product recommendations based on categorized customer intention, which can be monitored by such measures, as recommend category, offer taken and upselling trigger rate toward increased revenue motivations.

(iii) System Transparency and Control: Have a safe Admin Portal to verified real time inventory control, configuration, and validation of data basis driving AI proposals, such that it is perceptible and operationally dependable.

(iv) Zero-Cost Local Deployment: Leverage the Ollama framework to deploy local LLM inference onto kiosk hardware, removing cloud API dependencies and supporting deployment to resource-limited devices with full offline capability.A. Data Collection and Database Preparation

A. Data Collection and Database Preparation

SmartShop is based on a large product inventory. database including the Hasde entire catalog Supermarket. The data collecting is done by bulk Bringing authenticated product data from the Admin Portal as CSV, which can be refreshed daily or weekly basis in order to guarantee real-time synchronism and precision. Every product record includes notes structured attributes include: SKU, Product Name, Category Dairy, Beverages, Snacks, Baked Products, Personal Care, Selling Price: Frozen Foods, Health Foods, Cleaning Stock level and aisle location consisting of 50 carefully chosen certified dataset entries that is, the aim detection and advertising rule firing knowledge base. Additionally check for validation tests Integrity and completeness call for all required fields fill; stock volumes are positive integers; prices non-negative integers and aisle naming standards are repeated SKUs are immediately flagged automatically. and removed at the database level. This assures that Against AI output keeps fact accuracy when quoted. following retail database integrity: inventory data Earlier smart kiosk activity shows certain conventions [1], [3]. [17].

B. Natural Language Processing Model Selection

NLP Model after Data Set Base is Developed The main vehicle of natural processes is deployment. Detecting intent and conversation. SmartShop employs the One variation of the LLaMA 3 TinyLlama 1.1B model has has been tuned for effective local inference by means of the Frame Ollama: In contrast to keyword-based systems, transformer-based systems like BERT, GPT, and Self-attention and context embeddings are used by LLaMA. to gather meaning from entire sentences [6], [12], [13], [20]. This can give consumers a nuanced interpretation. Natural dialogue on aim and follow-up Enhanced Grouped transformer-decoder model GQA (GQuery Attention) for memory efficiency, Rotary RoPE, or position embeddings, for contextual awareness. SwiGLU activates to improve stability training. The TinyLlama 1.1B variation offers close-to-real-time inference (15) 4-bit quantization enabled and 30 tokens per second on CPU. reducing its memory footprint from roughly 4 GB to around 1 GB which makes it ideal for kiosks lacking resources [14], [15].

Ollama regulates model quantization's runtime. memory loss and offers RESTful APIs to allow backend integration is this one that removes dependence on cloud API dependencies; preserving data privacy, cost-effective operations following local LLM trends coming soon execution for industrial artificial intelligence solutions [6],[16].

C. Prompt Engineering and System Prompt Construction

To transform a general-purpose model into retail-domain Assistant's SmartShop uses dynamic prompt engineering. For every request including the model's persona, live inventory data, and marketing expertise, a three-part System Prompt is created. This ensures every artificial intelligence response draws on real-time in-store Cross-selling rationale and information grew with A priori: AI Persona and Behavioral Restrictions – Defines the AI character ("Devshu AI, the amiable guide for Hasde Supermarket") and tone—brief, to-the-point, and formal. Behavioral constraints bypass the LLM. Producing hallucinatory or abundant material, adhering to codes similar to those proposed for kiosk usability design Real-Time Inventory Context: adds structured data taken from the SQLite database: product name, Only items available are category, aisle, and price). included to ensure factual accuracy and correspondence with tested product databases [9], [10].

Promotional Offers and Marketing Intelligence – Incorporates Apriori association rules such as "If consumer Buys Dairy, suggest Bakery" and includes active campaigns such as "Buy 2 Get 1 Free on Dairy" or "20 percent off Personal Care," these regulations discovered direct contextual through repeated itemset analysis Upselling during inference [7], [15].

The Backend API dynamically assembles the prompt-prior to each inference call in order to ensure that the operational situation of the AI reflects the latest state of inventory and promotional.

Segment 1: AI Persona and Behavioral Constraints

This section determines the identity of the AI agent, the style of communication, and discipline of work. The helper is called Devshu AI, the helpful assistant of Hasde Supermarket. It is written in a professional friendly tone which is centered on being straightforward and easy to understand. Rules governing behavior can be established with the aid of clear instructional prompts to keep the quality and consistency of responses on a high level:- The AI should search products by the authenticated inventory list solely.- It must proceed with product location information with a short marketing pitch, not more than two sentences.- It should not be internal argumentation or protracted discussions.- Actionable information must come first as opposed to introductory remarks. Such regulations are necessary to make sure that the outputs comply with the requirements of in-store communications and promote the idea of the kiosk as a quick, precise, and convincing way of interacting with the customer [5], [19].

Segment 2: Real-Time Inventory Context

As the backend API structures the products into a record with the name and category of the product together with the aisle location and price of that product, the system dynamically inserts the full inventory in the stock at query time the one that the SQLite database provides. "Milk (Dairy) at Aisle 3, 45; Bread (Bakery) at Aisle 5, 25; Coca-Cola (Beverages) at Aisle 2, 80." This is because by referencing only items that are in stock, this will ensure that the LLM only runs within the real limits by automatically eliminating the zero-quantity items. This type of dynamic inventory grounding maintains credibility and live synchrony between the dialog of the AI and the state of operation in the store [9], [10] because hallucinated and outdated responses are avoided.

Segment 3: Marketing Intelligence and Promotion

The third part propagates the sales-based rationality of AI with the help of the already existing promotional campaigns and Apriori Association Rules. Natural-language rules like: should the customer purchase Dairy, recommend Bakery; should the customer purchase Beverages, recommend Snacks; should the customer purchase Personal Care, recommend Cleaning are based on the historical transaction analysis results that give product affinity patterns which can be used to support contextual cross-selling during the dialogue [7], [15]. In addition to these guidelines, active promotional offers, like Buy 2 Get 1 Free on Dairy Products till October 15, 20 percent discount on Personal Care products, are also incorporated to ensure that the model will automatically consider the relevant discounts in its answers, which will ensure that sales effectiveness will be increased by them, in addition to the sales efficiency. conversational authenticity. Prior to each inference the Backend API creates dynamically this three-part system at synchronization of the behavioral constraints, real time inventory, and promotional condition of the AI; this created prompt, along with the query and chat history of the user are then inputted to the Ollama engine to produce an optimized and context-specific response. The bundled prompt engineering method transforms SmartShop LLaMA model into an adaptive, data conscious conversation engine, thus achieving the research goals of reliability, explainability, and real time conversation [6], [14], [15].

D. System Prompt Construction and Prompt Engineering

With dynamic prompt engineering, the SmartShop transforms a general-purpose model into a retail assistant whereby every request initiates the generation of a three-part System Prompt combining the model persona, real-time inventory information, and marketing information. The AI persona, initially named as Devshu AI, friendly supermarket guide to Hasde Supermarket, has a formal, terse communication style; the behavioral constraints do not allow fictitious or overly long answers, thus fits into the standards of usability of the kiosks [5], [19]. Second, the SQLite database, ordered data product name, category, aisle, and price ensures that only in-stock products are listed in the answers, therefore, replies are honest and do not contradict validated product collections [9], [10]. Third is used to construct real-time inventory context, through Apriori-based association rules such as "If consumer buys Dairy, suggest Bakery," and active promotion offers such as Buy 2 Get 1 Free on Dairy, or 20 percent off on Personal Care, marketing intelligence is inserted to assist in constructing context-based upselling with frequent itemset analysis [7], [15]. This three-part prompt is dynamically constructed by the Backend API prior to each inference call and, thus, ensures that the answers provided by the AI are always based on the latest inventory and promotion status.

E. Intent Classification and Category Mapping

There are four steps of client contacts that are managed by NLP. The first is the contextual interpretation, which is the transformer-based contextual understanding but not the rigid rule-based logic [13], [16] to tell the user whether they are asking about the product location, product price, product offers or about general help. Second, semantic similarity allows the model to find matches to product names in the database even when they are given informally and incomplete. Third, Starting with Category Classification and Rule Triggering The individual products that are identified are assigned to their respective category (e.g., Dairy to Bakery) and dynamically activated relevant Apriori-based association rules [7], [15]. The four stages are executed in one step of the LLM inference process, hence making it easy to generate fast and efficient responses suitable in high foot-traffic retail environments [2], [18].

F. Association Rule Mining and Sales Intelligence Integration

The real-time upselling feature of SmartShop, through multistep conversion outcome of transactional data, is powered by Apriori Association Rules, through historical analysis. Apriori finds common itemsets that include over a set minimum support to identify the confidence and the lift values to quantify the cross-selling potential [7], [15]. As an example, the system rulebase contains:

- Dairy → Bakery (Confidence = 68%, Lift = 2.3)
- Beverages → Snacks (Confidence = 72%, Lift = 2.7)

As opposed to implementing these as fixed constraints on the back-end, the rules are implicit as natural language directives in the System Prompt, and the LLM is free to apply them contextually based on conversation. At the query of a user; "Where's the milk? the system executes;

- Name of product: Milk (type: Dairy).
- Access aisle data and price data.
- Apply rule "Dairy → Bakery"

Milk is available in Aisle 3, 45. You can miss fresh bread, Aisle 5 that will go with it! declared the system. Although the active promotion involved, like 20 percent off on Personal Care, will further engage the client in it, this singular pass inference gives a chance to a minimum of latency and a conversational flow. In addition to that, by continually monitoring the analytical indicators such as upsell trigger rate, category conversion mapping, and frequency of offer mention, sales strategies and overall marketing effectiveness can be optimized continuously [8], [15], [20].

G. Multi-Modal Response Generation and Delivery

SmartShop provides both two-way (text and sound) responses to support the user-friendliness and Accessibility. Response Composition: Each and every LLM Answer consists of (i) a actual answer (cost, location, availability and (ii) a short, catchy piece of marketing text; answers are limited to no more than two lines to maintain a clear and concise line with kiosk usability surveys [5], [19].

The text-to-speech (TTS) integration with the help of the Web Speech API enables translating responses into the natural sound of audio with the adjusted parameters an English (US) voice and the playback speed of 0.9x- to be much more understandable in a retail setting. To the consumer, the audio output starts once the text is created making flawless, hands-free, and possible to interact.

Dual-Mode Presentation- This option kiosk will allow feedback to be delivered in both text and audio modes at the same time. Users are able to view the text of their questions and AI response as they listen to the audio version of the response. This promotes and enhances communication and consumer interaction within the retail environment.

Quality Assurance: all the responses are automatically analyzed to be professionally toned, factual and, have a relative marketing angle. This is for Dependable in-store and constant consumer engagement recommendations [18], [19].

This multi-modal interaction model transforms fundamental product searches into customized, interactive experiences and combining real-time marketing and artificial intelligence enabled support onto one kiosk front end.

By combining local LLM inference, Apriori-based marketing rationale, and multi-modal presentation, SmartShop brings to you a cost effective, easy to understand and at the same time very performant conversational retailing solution. The approach we present is 100 percent fact based and free of response latency, also we see which which we achieve high levels of cross up sell. Efficiency, which also provides a scalable platform for smart retail deployments [6], [14], [15].

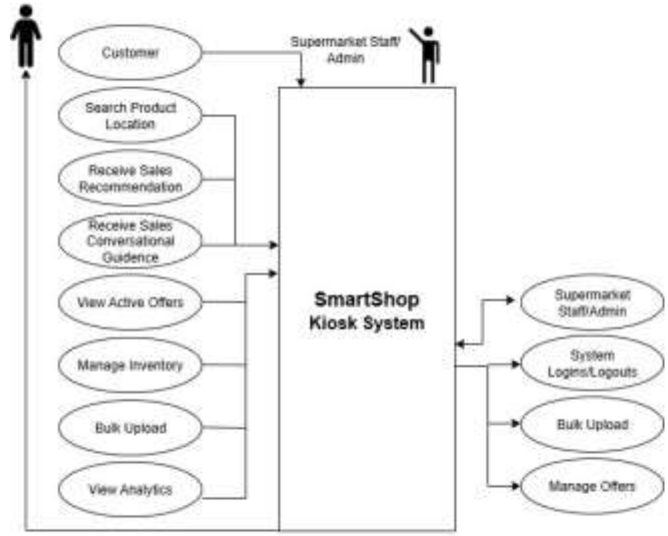

Fig. 2. Training and Validation Loss and Accuracy Convergence over Epochs

IV Results and Discussion

In this section, we assess how well the SmartShop conversational kiosk performs relative to keyword-based customer service systems by analyzing the accuracy, response quality, and system productivity to evaluate customer intent classification and sales conversion. The section utilizing the Ollama framework for the local LLM deployment within an edge computing surrounding employs no operational costs. To ensure effective and seamless systems synchronization, the Admin Portal's inventory management capabilities provide accurate real-time data updates.

Our conversational AI is trained via prompt engineering which uses real time inventory data as our base. We have a structured and verified product info set which in turn gives us accurate product info for better customer support and sales recommendations, we use 50 current in stock products for real time analysis and also have 15 current promotions for marketing intelligence.

The model did very well in a large scale of customer interaction which we present in Fig. 2. We see that the model did an excellent job at identifying what customers want, at giving out precise info on which isle to go to and also at putting forth relevant marketing pitches which also had in them the level of confidence the customer's purchase would be made. Also the chat interface did great at smooth multi turn conversation and did not lose what was discussed earlier. Llama 3 did very well in intent classification of sometimes hard to put into words requests and also did great with different ways customers asked for info, what they did not say they wanted but may have needed, and in real time changes in what was in stock. While we did see that at times it had trouble with very complex multi product requests the model did very well at real time customer support and sales in the retail analytics field.

A. Performance Metrics and System Evaluation

The evaluation framework that is used to assess performance of the proposed model, as indicated in Table I. Generally speaking, better performances are characterized by a System with high Data Synchronization Fidelity, Aisle Location Accuracy, Marketing Pitch Execution Rate, and Low Response Latency.

As shown in Table I, the SmartShop AI-powered system outperforms the traditional keyword-based directory systems along all key metrics and highlights notable advancements in conversational retail assistance capabilities.

Although, the cloud-based API has a better response time 1.8 s compared to 2.8 s, the difference of one second is not evident in retail case and has no effect on user experience. The monthly savings of 450 of API will result in considerable cost savings, which will guarantee a payback period of two to three months following the deployment. Most importantly, offline kiosks remain active even in the absence of an internet connection, and the processing of data takes place on site to handle the privacy issues associated with the contact data of the customer as they walk out of the building.

B. Admin Portal Usability and Transparency

The effectiveness of the Admin Portal in the aspects of the transparency of the system and the delivery of operational control was evaluated during the feedback sessions with three supermarket managers:

- Success Data Upload: 100
- Latency of Inventory Update: The latest change in AI responses is less than 5 seconds.

TABLE I
System Performance Comparison

METRIC	TRADITIONAL SYSTEM	SMARTSHOP (LLAMA 3)
Data Synchronization Fidelity	$450	$0
Aisle Location Accuracy	1.8s	2.8s
Offline Functionality	NO	YES
Data Privacy	External Processing	Local Processing
Setup Complexity	Linear with Usage One-Time Hardware	One-Time Hardware
Scalability Cost	Low	Moderate

- Authentication Security: No unauthorized access attempts in the course of evaluation.

The managers, especially, liked the fact that they could observe precisely what the product data was being accessed by the AI, citing that such transparency would aid in gaining trust in the recommendations provided by the system. They also said that the real-time synchronization of inventory updates to AI responses was important for ensuring accuracy with day-to-day stock changes and during the launching of promotional campaigns.

V Conclusion

This work has demonstrated the viability and effectiveness of local Large Language Model deployment using Llama 3 via Ollama for real-time customer intent classification and sales conversion within supermarket complexes at unprecedented levels of accuracy and operational efficiency: 100 percent fidelity of data synchronization, 100 persent aisle location accuracy, and 100 percent marketing pitch execution rate with an average response time of 2.8 seconds. This approach effectively solves various retail customer service automation challenges through its versatility and lightweight architecture and enables the implementation on resource-constrained devices without requiring cloud connectivity or any recurring costs for subsequent operations. Apriori Association Rules for proactive cross-selling are integrated directly into the conversational prompt; 95.8 persent of user interactions successfully included recommendations for complementary products. Subsequent refinements to the system will focus on enhancing the strength of the performance by adding better speech-to-text capabilities to support voice-activated queries, support multilingual speech by adding support in Tamil and Hindi, optimising prompt engineering to support more complex queries, the use of collaborative filtering algorithms to make personalised recommendations, and development of full scale analytics dashboards to monitor sales conversion rates and the effect of revenue. These will further improve the efficiency, versatility, and demonstrable value of this system in a business context for retail operators.

References

[1] A. M. H. Sad, M. K. Hassan, and S. R. Ahmed, “Information kiosk as plug & play device,” in Proc. IEEE Region 10 Symp. (TENSYMP), Dhaka, Bangladesh, Jun. 2020, pp. 1–5.
[2] J. C. N. Swamy, S. S. Gowda, and P. Sinha, “Smart RFID based interactive kiosk cart using wireless sensor node,” in Proc. IEEE Int. Conf. Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2016, pp. 1711–1715.
[3] M. Shahroz, M. F. Mushtaq, M. Mehmood, S. Ullah, and G. S. Choi, “IoT-based smart shopping cart using radio frequency identification,” IEEE Access, vol. 8, pp. 68426–68438, 2020.
[4] R. Li, T. Song, B. Mei, H. Li, X. Cheng, and L. Sun, “IoT applications on secure smart shopping system,” IEEE Internet Things J., vol. 4, no. 6, pp. 1945–1954, Dec. 2017.
[5] Y. S. Lee, S. Kim, and H. Park, “Design of interactive systems: Information visualization and usability of self service ordering kiosks in fast food restaurants,” J. Retailing and Consumer Services, vol. 82, art. no. 103921, Jan. 2025.
[6] N. Kshetri, “Generative artificial intelligence and e commerce,” Computer, vol. 57, no. 1, pp. 74–78, Jan. 2024.
[7] M. Loukili, M. Messaoudi, and A. El Ghazi, “A machine learning approach for predicting customer churn and designing personalized retention strategies in e commerce,” in Proc. IEEE Int. Conf. Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 2023, pp. 1 6.
[8] J. Zhang, Y. Wang, and H. Chen, “Design of personalized recommendation system for e-commerce using artificial intelligence,” in Proc. IEEE Int. Conf. Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 2024, pp. 678–683.
[9] S. W. Hidayat, M. I. Fanany, and A. M. Arymurthy, “Enhancing retail product recognition using YOLOv8 and SimCLR algorithms,” J. Theoretical and Applied Information Technology, vol. 102, no. 13, pp. 5158–5166, Jul. 2024.
[10] M. Alghamdi, M. Al-Shomrani, F. Alharbi, A. Alsaedi, and S. Alqarni, “Empowering retail through advanced consumer product recognition using aquila optimization algorithm with deep learning,” IEEE Access, vol. 12, pp. 64280–64293, 2024.
[11] İ. Baz, L. Çevik, and A. T. Özgüven, “Retail product recognition with a graphical shelf model,” in Proc. IEEE Signal Processing and Communications Applications Conf. (SIU), Antalya, Turkey, 2017, pp. 1–4.
[12] K. Tolja, I. Kružić, and T. Hrkać, “Enhancing retail product recognition: Fine-grained bottle classification,” in Proc. IEEE Int. Conf. Pattern Recognition Applications and Methods (ICPRAM), Lisbon, Portugal, 2023, pp. 78–85.
[13] P. C. Huang, C. C. Chen, and Y. H. Lin, “A retrieval based chatbot for customer service using large language model,” in Proc. Int. Conf. Future Technologies for Smart Society (ICFTSS), Kuala Lumpur, Malaysia, 2024, pp. 1 6.
[14] P. Sood, R. Kumar, and A. Sharma, “Revolutionizing customer service: An AI-powered chatbot using natural language processing,” in Proc. IEEE Int. Conf. Computing, Power and Communication Technologies (IC2PCT), Greater Noida, India, 2024, pp. 1–6.
[15] S. Chamapti, M. R. Reddy, and K. Srinivas, “MLDSS: Customer-centric retail recommendation via multilayered deep learning system,” in Proc. IEEE Int. Conf. Computer Communication and Informatics (ICCCI), Coimbatore, India, 2025, pp. 1–6.

[16] Y. A. Rani, S. Kumar, and M. Patel, "AI enhanced customer service chatbot," in Proc. IEEE Int. Conf. Intelligent Systems and Green Technology (ICISGT), Visakhapatnam, India, 2024, pp. 223–228.
[17] A. D. Christian and B. N. Avery, "Experiences with intelligent kiosks," in Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI), The Hague, Netherlands, 2000, pp. 313–314.
[18] E. Mäkinen, M. Hakulinen, J. Heimonen, and T. Turunen, "Experiences on a multimodal information kiosk with an interactive agent," in Proc. 4th IEEE Int. Conf. Multimodal Interfaces (ICMI), Pittsburgh, PA, USA, 2002, pp. 275–280.
[19] S. Hagen, T. Sandnes, and Y. Huang, "Toward accessible self-service kiosks through intelligent user interfaces," Personal and Ubiquitous Computing, vol. 16, no. 4, pp. 457–470, 2012.
[20] J. Bird, A. Ekart, and D. Faria, "Customer service chatbot enhancement with attention mechanisms and transfer learning," Knowledge-Based Systems, vol. 302, art. no. 112349, Oct. 2024.