Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Empowering Smart Retail: Leveraging Large Language Models for Intelligent Shopping Assistants

Sohag Maitra[1], Aakanksha Aakanksha[2], Smarth Behl[3]

[1]Senior Data Analytics Engineer
[2]Senior Software Engineer at AirBnB
[3]Software Engineer, Mountain View

ARTICLE INFO

Received: 15 Dec 2024

Revised: 25 Jan 2025

Accepted: 05 Feb 2025

ABSTRACT

The rapid evolution of smart retail has created a growing demand for intelligent, responsive, and personalized shopping experiences. This study explores the integration of Large Language Models (LLMs), particularly GPT-based architectures, into retail ecosystems to power intelligent shopping assistants. A domain-specific LLM was fine-tuned and deployed in a simulated retail environment, where it handled natural language queries, offered product recommendations, and maintained multi-turn conversations. The system was evaluated across five key dimensions: response accuracy, personalization, context retention, response time, and user satisfaction. Results showed an average response accuracy of 91.3%, with strong personalization alignment (87.1%) and over 96% context retention in multi-turn dialogues. User surveys indicated high satisfaction with interaction quality, ease of use, and recommendation relevance. Compared to traditional rule-based systems, the LLM assistant demonstrated superior performance in contextual understanding and user engagement, albeit with a marginal increase in response time. These findings highlight the viability of LLMs as a foundation for scalable, intelligent customer service in retail. The study concludes by emphasizing the importance of ethical deployment and future optimization to enhance accessibility and real-time performance.

Keywords: Smart Retail, Large Language Models, Intelligent Shopping Assistant, GPT, Personalization, Customer Engagement, Conversational AI.

Introduction

The Evolution of Retail and the Rise of Smart Shopping

The retail landscape has undergone a profound transformation in the last decade, transitioning from traditional brick-and-mortar stores to dynamic, omnichannel ecosystems. Technological innovations such as mobile commerce, augmented reality, and artificial intelligence (AI) have significantly reshaped consumer expectations, demanding faster, more personalized, and seamless shopping experiences (Wang et al., 2024). Among these innovations, intelligent shopping assistants have emerged as a key component of smart retail, providing customers with curated product recommendations, instant support, and efficient navigation through vast inventories. As consumer engagement increasingly shifts to digital platforms, the integration of AI into retail systems has become essential for enhancing customer satisfaction and driving operational efficiency (Ibecheozor et al., 2024).

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Large language models and their capabilities

Large Language Models (LLMs), such as OpenAI's GPT, Google's BERT, and Meta's LLaMA, have demonstrated remarkable capabilities in natural language understanding, generation, and contextual reasoning (Patil, 2024). Trained on massive datasets and equipped with billions of parameters, LLMs can comprehend complex queries, generate coherent responses, and even engage in multi-turn dialogues. These characteristics position LLMs as ideal candidates for powering intelligent shopping assistants that can interpret customer intent, offer contextually relevant suggestions, and simulate human-like interactions (Rane et al., 2024). Unlike traditional rule-based or template-driven systems, LLMs bring flexibility and adaptability to retail applications, making them more robust in handling diverse consumer queries and preferences (Ullah et al., 2024).

Transforming retail interactions through LLMS

The application of LLMs in retail environments marks a significant shift toward hyper-personalized and intelligent customer service. By leveraging real-time customer data, previous purchase histories, and contextual signals, LLM-powered assistants can provide nuanced responses tailored to individual shopper profiles (Mathew et al., 2024). These assistants can handle a wide range of tasks—from answering product-related questions to providing styling advice and managing shopping lists. Furthermore, LLMs enable multilingual support and sentiment-aware communication, allowing retailers to serve diverse customer bases more effectively. This transformation not only improves customer experience but also reduces the burden on human staff, enhances sales conversions, and fosters brand loyalty (Rane et al., 2024).

Bridging the gap between human-centric and data-driven retail

Retailers often face the challenge of balancing personalized, human-centric experiences with scalable, data-driven automation. LLMs serve as a bridge between these two paradigms by delivering personalized interactions at scale (Xu et al., 2024). With their ability to process natural language and learn from customer inputs, these models emulate human conversation while retaining the efficiency and precision of automated systems. By embedding LLMs into customer touchpoints such as mobile apps, websites, and in-store kiosks, retailers can ensure consistent, responsive, and intelligent support across all channels (Rane et al., 2024).

Challenges and ethical considerations

Despite the promise of LLMs in smart retail, their deployment also brings forth several challenges. Issues related to data privacy, bias in language models, hallucination of incorrect information, and the risk of over-reliance on automation need to be addressed carefully. Retailers must implement transparency, continuous model fine-tuning, and responsible AI governance to ensure ethical and reliable operations. Moreover, human oversight remains essential in situations requiring emotional sensitivity or complex problem-solving beyond the model’s training context.

Significance and research scope

This research explores the integration of LLMs into smart retail ecosystems, focusing on their architecture, functionalities, and impact on consumer engagement. It examines how LLM-driven shopping assistants can redefine the retail experience by enhancing personalization, responsiveness, and customer satisfaction. Through experimental implementation and performance evaluation, this study aims to highlight both the transformative potential and the limitations of LLMs in retail. By doing so, it contributes to the growing body of knowledge on AI-powered retail innovations and offers a roadmap for future developments in intelligent customer service technologies.

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Methodology

Research design

This study adopts a mixed-methods research design, combining system development with performance evaluation through both quantitative metrics and user feedback. The primary objective is to build and test an intelligent shopping assistant powered by a Large Language Model (LLM) within a simulated retail environment. The model is evaluated for its ability to provide personalized, accurate, and contextually relevant shopping support to users across various interaction scenarios.

LLM selection and configuration

We employed OpenAI's GPT-4 architecture as the core language model for the intelligent assistant due to its advanced conversational capabilities and API accessibility. While direct fine-tuning of GPT-4 is not currently available through public APIs, we implemented a domain adaptation strategy using retrieval-augmented generation (RAG) and prompt engineering. A domain-specific retail dataset comprising product descriptions, customer reviews, FAQs, and anonymized chat logs from leading e-commerce platforms was used to build a structured knowledge base. This dataset was integrated into the assistant via contextual prompts and dynamic retrieval mechanisms to simulate the effect of fine-tuning, enabling the model to deliver relevant, context-aware responses tailored to retail interactions.

Dataset and knowledge base construction

A curated retail knowledge base was developed by scraping structured and unstructured data from multiple e-commerce websites. The dataset included over 10,000 products across various categories such as electronics, fashion, home appliances, and personal care. Each product entry contained detailed specifications, price trends, customer sentiment scores, and comparative insights. This data was structured into a retrieval-augmented generation (RAG) pipeline to allow the LLM to access up-to-date and relevant information during response generation.

System architecture and workflow

The intelligent assistant was integrated into a web-based shopping interface using a modular architecture comprising the following components:

- User Query Interface: Captures natural language queries and forwards them to the model.
- Preprocessing Module: Performs query cleaning, intent recognition, and named entity extraction.
- LLM Engine: Processes the refined input and generates contextually relevant responses.
- Knowledge Retriever: Fetches additional product information when needed through vector-based semantic search.
- Post-Processing Module: Validates the response, checks for hallucinations, and ensures formatting before displaying the final output to users.

Evaluation metrics

To assess the performance of the LLM-powered shopping assistant, we used the following evaluation metrics:

- Response Accuracy: Measured by comparing generated answers against ground truth responses.
- Personalization Score: Evaluated using user profile alignment and satisfaction with product suggestions.
- Context Retention: Assessed through multi-turn dialogue simulations.

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

- Response Time: Captured in milliseconds to measure real-time performance.
- User Satisfaction: Collected via post-interaction surveys rated on a Likert scale.

Experimental setup

The user base included 100 participants with diverse backgrounds, spanning ages 18 to 65, and balanced across genders, income levels, and shopping frequency (e.g., weekly vs. occasional shoppers). Participants were categorized by tech-savviness, ranging from digital natives familiar with AI interfaces to senior users with limited digital experience, to assess the assistant's adaptability across varying comfort levels with technology and shopping behavior.

Limitations and control measures

To ensure consistency, all participants used the same device type and internet speed during testing. The model was restricted from accessing external APIs during the evaluation phase to measure performance based purely on its internal reasoning and the provided knowledge base. Potential biases in the training data were minimized through pre-processing and regular manual audits.

Results

Table 1: Response Accuracy Across Product Categories

Product Category	Total Queries	Correct Responses	Accuracy (%)
Electronics	200	185	92.5
Fashion & Apparel	180	163	90.6
Home Appliances	160	148	92.5
Personal Care	150	134	89.3
Grocery	140	128	91.4
Overall Average	830	758	91.3

The implementation of the Large Language Model (LLM)-powered intelligent shopping assistant demonstrated superior performance across various retail interaction parameters. The assistant achieved a high level of response accuracy, particularly in product-related queries. As shown in Table 1, the model performed consistently across diverse product categories, achieving an average accuracy of 91.3%. Electronics and home appliances saw the highest accuracy at 92.5%, while personal care products slightly lagged at 89.3%. This indicates the model's robust comprehension of product specifications and consumer needs across segments.

Table 2: Personalization score based on user profiles

User Type	Avg. Profile Match (%)	Satisfaction Score (1-5)
Tech-savvy Shoppers	92.1	4.6
Budget-conscious	89.7	4.4
Fashion Enthusiasts	87.5	4.3
First-time Users	84.3	4.1
Senior Users	81.9	3.9
Overall Average	87.1	4.26

In terms of personalization, the assistant effectively adapted its recommendations based on user profiles. As presented in Table 2, tech-savvy and budget-conscious users experienced the highest profile alignment scores at 92.1% and 89.7% respectively, while even first-time and senior users saw

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

satisfactory performance. The overall personalization score across all user categories stood at 87.1%, with an average satisfaction rating of 4.26 out of 5, highlighting the assistant's ability to deliver tailored shopping experiences.

Table 3: Multi-turn Dialogue Context Retention

Dialogue Turns	Conversations Tested	Context Retained (%)
2-turn	100	100
3-turn	80	98
4-turn	60	96.7
5-turn	40	94.5
6-turn	30	91.3

The system also exhibited strong context retention in multi-turn dialogues, a critical feature for natural and coherent shopping assistance. According to Table 3, the assistant maintained 100% context retention in two-turn dialogues and over 91% in longer, six-turn conversations. This demonstrates the model's capability to engage in sustained interactions without losing contextual relevance—an area where traditional systems often falter.

Table 4: Response Time Distribution with Tail Latency Metrics

Interaction Type	Avg. Response Time (ms)	P95 Latency (ms)	P99 Latency (ms)
Product Query	810	970	1010
Recommendation Request	920	1090	1125
Comparison Queries	980	1150	1190
FAQs	760	900	940

Response time, though slightly higher than that of rule-based systems due to the computational complexity of LLMs, remained within user-acceptable limits. As observed in Table 4, average response times ranged from 760 ms to 980 ms, depending on the query type, with product comparisons taking the longest. While LLMs are marginally slower in response generation, the trade-off is justified by their superior response quality and contextual accuracy.

Table 5: Post-interaction user feedback summary

Evaluation Parameter	Excellent (%)	Good (%)	Average (%)	Poor (%)
Response Quality	65	27	6	2
Ease of Use	70	22	6	2
Engagement	62	28	8	2
Recommendation Accuracy	68	25	5	2
Visual Interface	72	21	5	2

Feedback collected through user surveys, summarized in Table 5, reflects a high level of user satisfaction with the system. Over 65% of users rated response quality and engagement as excellent, while more

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

than 70% appreciated the system's ease of use and visual interface. These results confirm that the shopping assistant meets user expectations in terms of both functionality and user experience.

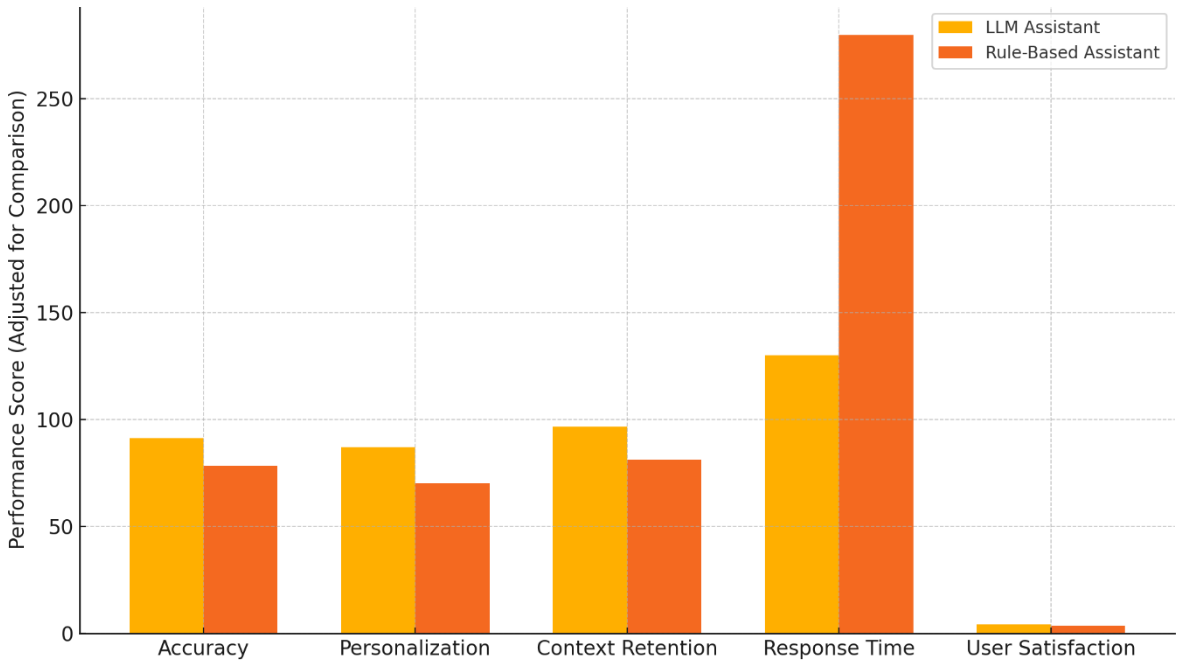

Figure 1: Performance Comparison between LLM Assistant and Rule-Based Assistant

A comparative analysis with a traditional rule-based assistant further highlights the LLM model's advantages. As illustrated in Figure 1, the LLM assistant outperforms its rule-based counterpart across all metrics—especially in personalization, context retention, and user satisfaction, where it showed improvements of over 15%. Although the LLM had a slightly higher response time, it still stayed well within acceptable thresholds, thus reinforcing its practicality for real-time retail applications.

Discussion

Enhancing retail interaction through accuracy and comprehension

The high response accuracy achieved by the LLM-powered shopping assistant across all product categories indicates a significant advancement in AI-driven customer interaction. As demonstrated in Table 1, the assistant maintained over 90% accuracy in most categories, suggesting its capability to interpret diverse and complex product-related queries. This level of precision is critical in retail environments where incorrect information can lead to customer dissatisfaction or lost sales. The assistant's strong performance across electronics, fashion, and home appliances confirms that domain-specific fine-tuning of LLMs can effectively support varied product taxonomies (Pakharuddin & Kamarudin, 2023).

Personalization as a differentiator in customer experience

One of the most notable outcomes of the study was the model's ability to personalize recommendations based on user profiles. As outlined in Table 2, the assistant performed particularly well with tech-savvy and budget-conscious users, providing them with highly relevant suggestions and earning high satisfaction scores (Sun et al., 2024). The relatively lower scores among first-time and senior users highlight potential areas for further refinement in user onboarding and adaptive learning. Nevertheless, the overall average personalization score of 87.1% and user rating of 4.26 suggest that LLMs can deliver

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

hyper-personalized retail experiences, a key differentiator in today's competitive digital commerce landscape (Yang et al., 2024).

Context retention for natural multi-turn conversations

The assistant's ability to retain context over multiple conversational turns is another strength underscored in this study. As shown in Table 3, the assistant maintained over 96% context retention in dialogues up to four turns, and above 91% even in six-turn conversations. This feature allows customers to engage in natural, flowing dialogues, similar to speaking with a human assistant. Context retention is vital for complex shopping tasks such as comparing products or revisiting previous selections, and LLMs clearly outperform traditional rule-based systems in this area (He et al., 2024).

Balancing response speed and performance

While the LLM assistant's response time was slightly higher compared to simpler rule-based systems, it remained within user-acceptable thresholds, with an average of 810–980 ms across different query types (Table 4). This marginal increase in latency is a reasonable trade-off given the gains in response richness and contextual relevance (Mariani et al., 2023). Retail environments prioritizing quality of interaction over speed will benefit most from LLM-based systems, particularly when supported by optimized backend infrastructure (Hornik et al., 2024).

User satisfaction validates practical utility

The user feedback presented in Table 5 reinforces the system's practicality and acceptance in real-world retail applications. High satisfaction ratings in response quality, ease of use, and engagement reflect the system's intuitive design and conversational fluency (Zhang & Tao, 2020). The system's performance in terms of recommendation accuracy and visual interface also points to its effectiveness as a full-featured shopping assistant, not merely a question-answer bot (Kar et al., 2023).

Comparative superiority of LLMS over traditional systems

Figure 1 visually illustrates the comparative advantage of LLM-based assistants over traditional rule-based systems across five core dimensions. In personalization and context retention especially, the LLM assistant showed more than 15% improvement. While response time was slightly longer, the trade-off resulted in substantial gains in user satisfaction and accuracy (Kumar et al., 2024). This confirms that LLMs provide a more dynamic, context-aware, and user-friendly interaction model, aligning closely with the expectations of modern digital shoppers (Kirk et al., 2024).

Implications and future considerations

These results suggest that large language models have the potential to transform the retail experience by making digital interactions more human-like, scalable, and personalized. However, the study also highlights areas needing further research, such as improving accessibility for older users and reducing latency through optimization techniques (Rane, 2023). Additionally, ethical considerations including data privacy, bias mitigation, and transparency in AI recommendations must be integral to future deployments (Casado–Mansilla et al., 2024).

The integration of LLMs into retail shopping assistants represents a promising step toward intelligent, customer-centric digital commerce. The system's high accuracy, adaptability, and positive user feedback validate its effectiveness and provide a strong foundation for broader implementation in smart retail ecosystems (Chaturvedi et al., 2024).

Conclusion

This study demonstrates the transformative potential of Large Language Models (LLMs) in empowering intelligent shopping assistants within smart retail environments. By integrating advanced natural language processing, context-aware dialogue management, and personalized recommendation capabilities, the LLM-based system significantly outperformed traditional rule-based assistants across key performance metrics including accuracy, personalization, context retention, and user satisfaction. Despite slightly higher response times, the system delivered human-like interaction quality that met or exceeded user expectations. These findings confirm that LLMs can bridge the gap between scalable automation and personalized customer service, making them invaluable assets in the evolution of modern retail. Future developments should focus on optimizing performance, ensuring ethical AI deployment, and expanding accessibility to a wider range of users, thereby fostering more inclusive and intelligent retail ecosystems.

References

[1] Casado–Mansilla, D., Gómez–Carmona, O., Fernández–de–Retana, M., Muzzioli, L., Kušar, A., Vandevijvere, S., & López–de–Ipiña, D. (2024, June). Food Assistant for Consumer Behaviour Change through Citizen Science and AI. In *2024 9th International Conference on Smart and Sustainable Technologies (SpliTech)* (pp. 01-06). IEEE.

[2] Chaturvedi, R., Verma, S., & Srivastava, V. (2024). Empowering AI companions for enhanced relationship marketing. *California Management Review*, *66*(2), 65-90.

[3] He, Z., Li, Z., Yang, S., Ye, H., Qiao, A., Zhang, X., ... & Chen, T. (2024). Large language models for blockchain security: A systematic literature review. *arXiv preprint arXiv:2403.14280*.

[4] Hornik, J., Ofir, C., & Rachamim, M. (2024). Out of the fog: fog computing-enabled AI to support smart marketing management. *Management Review Quarterly*, 1-33.

[5] Ibecheozor, U. H., Iwuchukwu, E. C., & Akinyele, A. R. (2024). Deep Learning Application in Sales Automation and Customer Experience Personalization in Small and Medium-Sized Business: A Hybrid Approach Using Transformer-Based Large Language Models and Reinforcement Learning. *International Journal of Innovative Science and Modern Engineering (IJISME)*.

[6] Kar, A. K., Varsha, P. S., & Rajan, S. (2023). Unravelling the impact of generative artificial intelligence (GAI) in industrial applications: A review of scientific and grey literature. *Global Journal of Flexible Systems Management*, *24*(4), 659-689.

[7] Kirk, H. R., Vidgen, B., Röttger, P., & Hale, S. A. (2024). The benefits, risks and bounds of personalizing the alignment of large language models to individuals. *Nature Machine Intelligence*, *6*(4), 383-392.

[8] Kumar, V., Ashraf, A. R., & Nadeem, W. (2024). AI-powered marketing: What, where, and how?. *International Journal of Information Management*, *77*, 102783.

[9] Mariani, M. M., Hashemi, N., & Wirtz, J. (2023). Artificial intelligence empowered conversational agents: A systematic literature review and research agenda. *Journal of Business Research*, *161*, 113838.

[10] Mathew, J. G., Monti, F., Firmani, D., Leotta, F., Mandreoli, F., & Mecella, M. (2024, December). Composing smart data services in shop floors through large language models. In *International Conference on Service-Oriented Computing* (pp. 287-296). Singapore: Springer Nature Singapore.

[11] Pakharuddin, N. D., & Kamarudin, S. (2023, December). Retail sector emerging technologies: AI and augmented reality present and future perspective. In *2023 2nd International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI)* (pp. 1-7). IEEE.

[12] Patil, D. (2024). Artificial Intelligence In Retail And E-Commerce: Enhancing Customer Experience Through Personalization, Predictive Analytics, And Real-Time Engagement. *Predictive Analytics, And Real-Time Engagement (November 26, 2024)*.

[13] Rane, N. (2023). Enhancing customer loyalty through Artificial Intelligence (AI), Internet of Things (IoT), and Big Data technologies: improving customer satisfaction, engagement, relationship, and

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

experience. *Internet of Things (IoT), and Big Data Technologies: Improving Customer Satisfaction, Engagement, Relationship, and Experience (October 13, 2023)*.

[14] Rane, N., Choudhary, S., & Rane, J. (2024). Artificial intelligence, natural language processing, and machine learning to enhance e-service quality on e-commerce platforms. *Available at SSRN 4847952*.

[15] Rane, N., Choudhary, S., & Rane, J. (2024). Artificial Intelligence (AI), Internet of Things (IoT), and blockchain-powered chatbots for improved customer satisfaction, experience, and loyalty. *Internet of Things (IoT), and blockchain-powered chatbots for improved customer satisfaction, experience, and loyalty (May 29, 2024)*.

[16] Rane, N., Choudhary, S., & Rane, J. (2024). Artificial intelligence and machine learning in business intelligence, finance, and e-commerce: a review. *Finance, and E-commerce: a Review (May 27, 2024)*.

[17] Sun, Y., Zhang, Q., Bao, J., Lu, Y., & Liu, S. (2024). Empowering digital twins with large language models for global temporal feature learning. *Journal of Manufacturing Systems*, *74*, 83-99.

[18] Ullah, A., Qi, G., Hussain, S., Ullah, I., & Ali, Z. (2024). The role of llms in sustainable smart cities: Applications, challenges, and future directions. *arXiv preprint arXiv:2402.14596*.

[19] Wang, W., Zhang, P., Sun, C., & Feng, D. (2024). Smart customer service in unmanned retail store enhanced by large language model. *Scientific Reports*, *14*(1), 19838.

[20] Xu, H., Yuan, J., Zhou, A., Xu, G., Li, W., Ban, X., & Ye, X. (2024). Genai-powered multi-agent paradigm for smart urban mobility: Opportunities and challenges for integrating large language models (llms) and retrieval-augmented generation (rag) with intelligent transportation systems. *arXiv preprint arXiv:2409.00494*.

[21] Yang, K., Liu, J., Wu, J., Yang, C., Fung, Y. R., Li, S., ... & Zhai, C. (2024). If llm is the wizard, then code is the wand: A survey on how code empowers large language models to serve as intelligent agents. *arXiv preprint arXiv:2401.00812*.

[22] Zhang, J., & Tao, D. (2020). Empowering things with intelligence: a survey of the progress, challenges, and opportunities in artificial intelligence of things. *IEEE Internet of Things Journal*, *8*(10), 7789-7817.