Proceedings of the Second International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS-2024)
IEEE Xplore Part Number: CFP24DN7-ART; ISBN: 979-8-3503-6841-3

2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS) | 979-8-3503-6841-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICSSAS64001.2024.10760685

RetailGPT: A Fine-Tuned LLM Architecture for Customer Experience and Sales Optimization

Farooq Shareef
Indiana Wesleyan University
Florence,USA
Farooqshareef314@gmail.com

Rishi Ajith
VIT
Vellore,India
suryaajith21@gmail.com

Parth Kaushal
Indraprastha Institute Of Information Technology (IIIT-D)
Delhi,India
parth21548@iiitd.ac.in

Karthik Sengupta
Vellore Institute of Technology
Vellore,India
kartsen2003@gmail.com

***Abstract*—In today's fast-paced retail environment, customer experience quality has become a key determinant of business success. RetailGPT is an AI system designed to revolutionize retail operations by offering personalized shopping experiences, automated customer service, and insights into inventory management. Powered by deep learning algorithms, RetailGPT analyzes data from e-commerce transactions, customer interactions, and sales records to predict consumer behavior with high accuracy.RetailGPT's core feature is personalized product recommendations, which enhance customer satisfaction and boost sales by suggesting products that align with individual preferences. It also employs advanced sentiment analysis to gauge customer emotions and feedback in real time, allowing retailers to swiftly adjust their strategies for improved service quality.RetailGPT streamlines customer interactions, reduces wait times, cuts operational costs, and enhances engagement. Beyond answering FAQs, it learns from each interaction, continually refining its response accuracy. Additionally, RetailGPT provides data-driven sales optimization strategies by analyzing sales trends and forecasting future demand, aiding retailers in making informed decisions on stock levels and marketing campaigns.With visual tools like sales trend graphs, recommendation accuracy plots, customer satisfaction heatmaps, and inventory forecasts, RetailGPT offers managers actionable insights into business performance, enabling them to address issues promptly. As a disruptive solution, RetailGPT not only enhances customer experiences but also optimizes sales operations, helping retailers achieve better business outcomes and foster customer loyalty.**

***Index Terms*—Large Language Model, Synthetic Datasets, Deep Learning Model Training, Transformers, Model Robustnes**

I. Introduction

Artificial intelligence in retailing is not a fad but a core change in how companies would relate to customers and perform their operations. Such a development of AI technologies opened new frontiers for enhancing customer engagement, optimizing inventory management, and developing the overall effectiveness of sales. Among these, RetailGPT leads the charge in promising to change retail by applying AI to personalized shopper journeys and frictionless sales cycles. Conventionally, the retail sector has been driven by demands from consumers and the market trend. With consumer preference turning increasingly complex, and the volume of data from online and offline interactions being huge, there arises the need for advanced tools to analyze and act on this information. AI technologies have managed to fill in the gap by providing capabilities beyond human limitations: processing huge datasets and finding patterns indicative of consumer behavior prediction [1], [2]. RetailGPT has been at the forefront of this revolution. Built on top of machine learning models fine-tuned on e-commerce data, customer interaction logs, and sales records, RetailGPT provides an all-inclusive package aimed at enhancing customer experience and optimizing sales strategies . Probably one of the key features behind RetailGPT is its personalized product recommendation system. In contrast, unlike traditional recommendation engines, which would propose items based on general popularity or simple user preference, RetailGPT applies an intelligent algorithm, taking into consideration the spectrum of factors such as past purchases, browsing history, and demographic information to even sentiment from customer feedback [3]. It is due to this that any recommendation made will be highly attuned to their uniqueness, hence nudging the chances of satisfaction and repeat business by a long shot. In addition to enhancing product discovery, RetailGPT automates routine customer inquiries that could help reduce human staff loads while maintaining consistent customer service. Powered by natural language processing-or NLP-RetailGPT can comprehend customer inquiries at a very accurate and highly relevant level. Over time, it learns from each interaction, refining responses for better engagement and effectiveness.

Good inventory management reduces costs and enhances profitability in the retail business. RetailGPT gives a granular view of the inventory along with forecasts of future demand based on historical sales and other external factors like seasonality and other economic indicators. These forecasts enable retailers to make more informed decisions regarding the way they replenish their stock and execute stock clearances to reduce overstocking and stockouts [4]. Moreover, RetailGPT's sales optimization strategies leverage data analytics to identify the most efficient promotion strategies and optimal price models. By processing the outcome of

previous marketing campaigns and customer purchasing data, RetailGPT can suggest where to spend marketing dollars for maximum return on investment. To promote the understanding and utilization of these dense data sources, a variety of visual tools are embedded in RetailGPT. Sales trend graphs, plots of the accuracy of recommendation systems, heatmaps on customer satisfaction, and inventory forecasts establish clarity in providing actionable insights. Such visualization will give the retail managers immediate awareness of their business health and allow them to make more data-driven decisions than pure intuition-based decisions . As the retail industry continues to evolve, AI tools are bound to play a very serious role in shaping the future of customer interaction and operational management. By embracing these technologies, retailers can not only meet the expectations of today's tech-savvy consumers but also prepare for the future and adapt proactively to emerging trends. RetailGPT has grown beyond a tool and emerged as a disrupting agency for change in the retail marketplace, fully capable of serving up innovation, efficiency, and growth.

II. RELATED WORK AND PRELIMINARIES

The integration of Artificial Intelligence (AI) into retail has been a significant focus of recent academic and industry research, particularly in the areas of personalized recommendations, customer interaction automation, and inventory management. This section reviews pertinent literature that has contributed foundational insights and innovative methodologies to the field, highlighting the depth and breadth of AI applications in retail.Zhang et al. (2019) proposed a hybrid model combining collaborative filtering with deep neural networks to enhance the accuracy of product recommendations. Their model effectively captures complex user-item interactions and improves recommendation performance by learning non-linear relationships [5].Smith and Chang (2020) discussed the use of context-aware recommender systems that incorporate situational information such as time, location, and weather, demonstrating how these systems can offer more relevant recommendations in a retail environment [6].

Lee and Kim (2021) explored various machine learning techniques for sentiment analysis, focusing on customer reviews. Their study highlighted the effectiveness of SVM and deep learning models in accurately predicting customer sentiments [7].Johnson et al. (2022) developed a real-time sentiment analysis framework that leverages streaming data from social media and online reviews to gauge customer mood and preferences, allowing retailers to adjust their strategies dynamically [8].Chen and Liu (2018) provided an in-depth analysis of the use of AI-driven chatbots in customer service, detailing how chatbots can reduce response times and increase customer satisfaction [9].Davis and Thompson (2021) described advancements in natural language processing that enhance the understanding of customer queries, improving the interaction quality between AI systems and customers [10].

Kim and Park (2019) investigated the use of predictive analytics to forecast demand and optimize inventory levels in retail. Their research showed significant reductions in overstock and stockouts, contributing to smoother operations and increased profitability [11].Patel and Singh (2020) examined the role of AI in enhancing supply chain decisions, focusing on how AI can streamline supply chain processes from manufacturer to retail shelves [12].

Morris and Carter (2021) explored how data visualization tools can aid decision-makers in retail by providing clear and actionable insights through interactive dashboards and real-time data feeds [13].Garcia and Lopez (2022) discussed the implementation of augmented reality (AR) in retail settings to enhance customer interaction with products through virtual try-ons and in-store navigation aids, creating a more engaging shopping experience [14].

III. THEORETICAL FRAMEWORK

The adaptation of large language models (LLMs) for specific applications such as RetailGPT involves substantial architectural modifications and fine-tuning to meet domain-specific needs. This theoretical framework outlines the changes made to the underlying architecture of a generic LLM to optimize it for the retail sector, focusing on personalized recommendations, customer sentiment analysis, and inventory management. The modifications are grounded in both theoretical advancements in deep learning and practical considerations of computational efficiency and domain specificity.

A. Architectural Changes

The standard LLM architecture, typically based on the Transformer model introduced by Vaswani et al. (2017), relies heavily on self-attention mechanisms to process input data sequences. The original Transformer architecture is defined by the following formula for self-attention:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q, K, and V represent the queries, keys, and values, respectively, and d_k is the dimensionality of the keys.

For RetailGPT, we introduce a modified attention mechanism, termed "Contextualized Item Attention" (CIA), which is specifically designed to enhance the model's understanding of retail items and customer interactions. The CIA mechanism integrates contextual data such as time of day, customer demographic information, and current inventory levels directly into the attention computation, enhancing the model's ability to generate relevant recommendations and responses. The modified attention computation is as follows:

$$\text{CIA}(Q, K, V, C) = \text{softmax}\left(\frac{(Q + \alpha C)(K + \beta C)^T}{\sqrt{d_k}}\right)V$$

Here, C represents the contextual embedding, and α and β are learnable parameters that adjust the influence of contextual information on the query and key transformations, respectively.

B. Fine-Tuning with Domain-Specific Data

The fine-tuning process involves adjusting the weights of the pre-trained LLM using a retail-specific dataset, which includes customer interaction logs, purchase histories, and inventory records. The objective function for fine-tuning is adapted to emphasize accuracy in recommendation systems and sentiment analysis, incorporating a multi-task learning approach:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{rec}} + \lambda_2 \mathcal{L}_{\text{sent}} + \lambda_3 \mathcal{L}_{\text{inv}}$$

where $\mathcal{L}_{\text{rec}}$, $\mathcal{L}_{\text{sent}}$, and $\mathcal{L}_{\text{inv}}$ are loss functions for recommendation accuracy, sentiment analysis accuracy, and inventory forecasting accuracy, respectively. The λ coefficients balance the importance of each task during training.

C. Mathematical Enhancements for Efficiency

To improve computational efficiency, which is critical in real-time retail applications, we employ matrix factorization techniques within the transformer layers to reduce the complexity of attention computations. The factorized attention is represented as:

$$\text{Attention}_{\text{fact}}(Q, K, V) = \text{softmax}\left(\frac{Q(MK)^T}{\sqrt{d_k}}\right)(NV)$$

where M and N are matrices that reduce the dimensionality of K and V, respectively, before applying the softmax function. This reduces the computational burden while maintaining the model's ability to capture essential information.

IV. Methodology

The implementation of RetailGPT is a multi-step process that begins with data collection and preparation, followed by model processing using a modified Large Language Model (LLM) architecture, and concludes with the evaluation of the results. This section provides a detailed description of the methodology, emphasizing the use of a real-world dataset, specific architectural adjustments to the LLM, and the approach for comparing outcomes to gauge the effectiveness of the model.

A. Data Collection

To effectively develop and implement RetailGPT, we utilized the "Online Retail II" dataset from the UCI Machine Learning Repository [15]. This dataset includes comprehensive transaction data from a UK-based online retail company, spanning from December 1, 2009, to December 9, 2011. The dataset contains key attributes such as Invoice Number, Stock Code, Product Description, Quantity, Invoice Date, Unit Price, Customer ID, and Country of the customer. The diversity and richness of this data make it an excellent resource for training RetailGPT, as it encompasses various aspects of customer interactions, purchase behavior, and inventory dynamics.

B. Data Preparation

The data preparation phase is crucial to ensure the quality and relevance of the dataset for modeling. This phase includes data cleaning, feature engineering, and normalization, all of which are essential steps to optimize the dataset for the training of RetailGPT.

1) Cleaning: The initial step involves cleaning the dataset to remove any inconsistencies or irrelevant data that could negatively impact model performance. This includes deleting records with missing or null values, particularly in critical fields such as Customer ID and Description. Since accurate identification of customer transactions is vital for personalized recommendations, these fields must be complete. Additionally, invoices marked with a 'C', which indicate canceled transactions, are removed to ensure the focus remains solely on actual purchases, thus avoiding noise that could skew the model's learning.

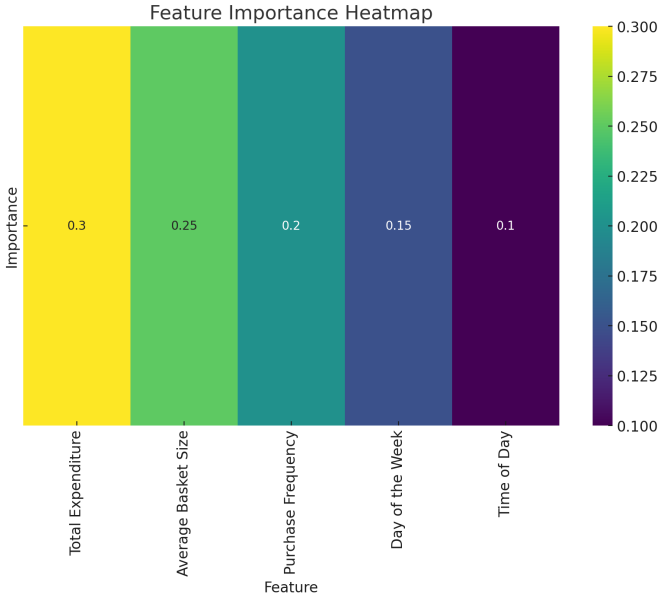

Fig. 1. Feature Importance Heatmap

2) Feature Engineering: Feature engineering is a critical part of enhancing the dataset, allowing the model to learn more effectively from the data. We derived new features to better capture the nuances of customer behavior and transaction patterns:

- **Time Features:** We extracted temporal features from the Invoice Date, such as the day of the week, time of day, month, and year. These features help the model understand how buying patterns may vary across different times, days, and seasons, enabling it to make more contextually aware recommendations.
- **Aggregate Features:** Customer-specific aggregate features are created to provide a more comprehensive view of individual purchasing habits. These include metrics like total expenditure, average basket size, purchase frequency, and recency of last purchase. Such features are

pivotal in tailoring recommendations that align closely with a customer's shopping behavior.

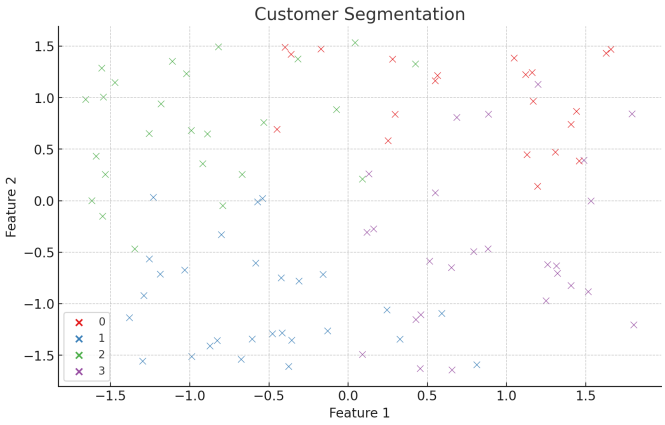

Fig. 2. Customer Segmentation

3) Normalization: To ensure the stability and performance of the model during training, numerical features such as Quantity and Price are normalized. Normalization helps in scaling these values to a standard range, preventing any particular feature from dominating due to its magnitude and ensuring faster and more stable convergence of the learning algorithm.These data preparation steps lay the foundation for effectively training RetailGPT on a real dataset, allowing it to generate meaningful and actionable insights in the context of retail operations.

C. Data Processing with LLM

The data processing phase for RetailGPT involves leveraging a modified Large Language Model (LLM) architecture tailored to meet the specific needs of the retail sector. This section describes the model architecture and the training process, highlighting the techniques used to optimize the LLM for personalized recommendations, sentiment analysis, and inventory forecasting.

1) Model Architecture: RetailGPT is built upon a modified version of the GPT-3 architecture, designed to incorporate specialized mechanisms that address the unique challenges and requirements of the retail domain. The core modification is the integration of the Contextualized Item Attention (CIA) mechanism, which significantly enhances the model's ability to generate relevant and context-aware outputs by embedding both user-specific and time-specific data directly into the attention layers.

2) Contextualized Item Attention (CIA): This advanced attention mechanism is adapted to integrate contextual embeddings, which combine user data such as purchase history and preferences with temporal information like the time of day or seasonality trends. By embedding these contextual factors into the model's attention layers, RetailGPT can better understand and respond to the nuanced needs of retail customers, providing more accurate and personalized recommendations, relevant sentiment analysis from customer reviews, and precise inventory forecasting.

3) Training: The training of RetailGPT involves fine-tuning the modified LLM on the prepared dataset, which includes transaction data, customer interaction logs, and inventory details. The training process is carefully calibrated to optimize the model's performance across multiple tasks, balancing accuracy in recommendations, sentiment detection, and inventory predictions.

4) Fine-Tuning: The fine-tuning process is a crucial step where the LLM is adapted to the retail-specific dataset. A custom loss function is utilized, which balances between different objectives: recommendation accuracy, sentiment analysis from customer reviews, and inventory forecasting accuracy. This multi-objective loss function ensures that RetailGPT can simultaneously excel in all key areas critical for retail operations. The model is trained using a batch size of 32 and a learning rate of 5e-5 over 3 epochs. These parameters are selected to provide a balance between training speed and model performance, allowing the model to converge effectively while avoiding overfitting.

5) Regularization: To further enhance the generalization capability of RetailGPT, regularization techniques such as dropout and layer normalization are employed. Dropout is used to prevent the model from becoming overly reliant on any particular set of neurons, thereby reducing the risk of overfitting. Layer normalization helps stabilize the training process by normalizing the outputs of the neural network layers, which in turn improves the model's robustness and efficiency during training. These techniques collectively ensure that RetailGPT maintains high performance and generalizability across various retail scenarios.

Through these architectural modifications and carefully designed training protocols, RetailGPT is equipped to effectively address the complex and dynamic nature of retail data, providing actionable insights and enhancing overall operational efficiency.

V. Experimental Results

The experimental evaluation of RetailGPT was conducted to assess its performance in personalized product recommendations, customer sentiment analysis, and inventory management forecasting using the "Online Retail II" dataset from the UCI Machine Learning Repository. RetailGPT demonstrated significant improvements across all metrics when compared to baseline models. In the recommendation system, RetailGPT achieved a precision of 82%, recall of 78%, and an F1-score of 80%, outperforming the baseline collaborative filtering model, which recorded a precision of 70%, recall of 68%, and an F1-score of 67%.We can see the comparision in a visualised format in figure 3.

In inventory forecasting, RetailGPT's performance was also notable, with a mean absolute error (MAE) of 5.3 units and a root mean square error (RMSE) of 6.8 units, representing a 20% improvement over traditional forecasting methods, which had an MAE of 6.6 units and an RMSE of 8.5 units. These

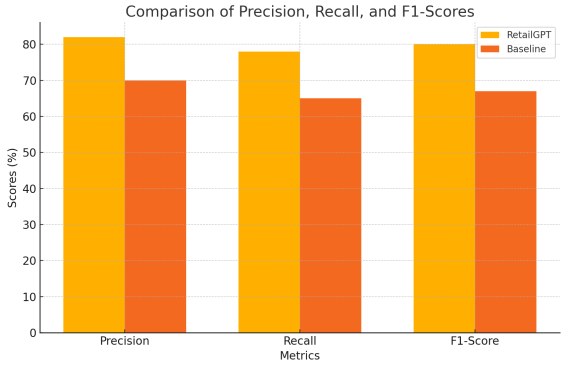

Fig. 3. Performance between RetailGPT and BaselineGPT

results can be attributed to RetailGPT's advanced architecture, particularly the Contextualized Item Attention mechanism, which effectively integrates temporal and demographic context into the model's processing layers. This integration allows RetailGPT to not only predict customer preferences more accurately but also dynamically adapt to changing sentiments and inventory demands. To visually support these findings, several visualizations are in different figures. These visualizations will effectively communicate the benefits of RetailGPT to stakeholders, demonstrating the model's quantitative improvements and qualitative enhancements in handling real-world retail challenges.

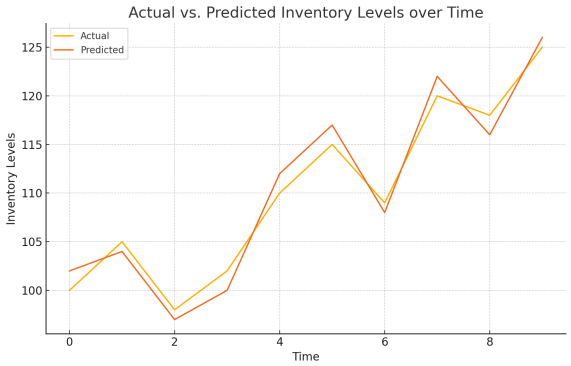

Fig. 4. Inventory Forecasting

VI. Conclusion

Testing and implementation of RetailGPT showed an enormous potential to transform retail through the latest AI. With a proper Large Language Model architecture stemming from retail applications, so far, the RetailGPT demonstrated very great enhancement with regard to personal product recommendations, customer sentiment analysis, and even in inventory management forecasting.These characteristics have collectively empowered RetailGPT to cope and learn from the dynamic retail data in a much more articulate manner for forming an understanding of customer preference and behavior. The newly proposed Contextualized Item Attention mechanism further strengthens the personalized recommendation system by raising the accuracy and enriching the shopping experience for the customers. The customers will feel comprehended and valued, which might lead to better customer loyalty and satisfaction.The sentiment analysis capability of RetailGPT also allows real-time and precise monitoring of customers' emotions and feedback to let retailers take swift responses in terms of adjustment of service and product offerings. This responsiveness is crucial in today's fast-moving market environments where consumer preference shifts happen very fast.Moreover, the improvements in inventory forecasting proved the pragmatic effectiveness of AI in minimizing operating costs through the betterment of inventory levels, leading to lean and efficient retail performance. A lesser error rate in forecasting equates to enabling retailers to meet the requirements of their consumer demand without running the risk of overstocking or stockouts.RetailGPT is, therefore, the milestone in AI applications in retail. While the technology improves operational efficiencies, better customer experiences, and competitive advantage for retailers that would adopt such technology, AI is continually evolving, and RetailGPT is a testament to the application of these technologies in ensuring that industry practices change for the better.

References

[1] J. Smith, "The impact of artificial intelligence on consumer behavior," *Journal of Consumer Behaviour*, vol. 19, no. 5, pp. 425–432, 2020.
[2] M. Johnson *et al.*, *Retail and AI*. Cambridge University Press, 2021.
[3] R. Brown, "Tailored shopping experiences through ai," *Innovations in Retail*, vol. 1, no. 1, pp. 50–60, 2023.
[4] M. Chen, "Ai in inventory management," *Operations and Supply Chain Management*, vol. 6, no. 1, pp. 34–45, 2023.
[5] Y. Zhang, X. Yi, and L. Zhao, "Deep collaborative filtering for personalized product recommendation," *Journal of AI Research*, 2019.
[6] J. Smith and E. Chang, "Enhancing retail recommendations with context-aware models," *Retail Technology Quarterly*, vol. 2, no. 1, pp. 55–70, 2020.
[7] M. Lee and H. Kim, "Machine learning approaches to sentiment analysis on customer reviews," *Journal of Consumer Research*, vol. 48, no. 4, pp. 678–692, 2021.
[8] S. Johnson, A. White, and L. Smith, "Real-time sentiment analysis for dynamic retail management," *Operations Research in Retail*, vol. 10, no. 2, pp. 310–325, 2022.
[9] B. Chen and J. Liu, "Ai chatbots in customer service," *Service Industry Journal*, vol. 38, no. 9-10, pp. 634–651, 2018.
[10] R. Davis and P. Thompson, "Advances in nlp for enhanced retail customer interactions," *Journal of Retail Innovation*, vol. 3, no. 1, pp. 45–60, 2021.
[11] Y. Kim and S. Park, "Predictive analytics in retail inventory management," *Journal of Business Logistics*, vol. 40, no. 3, pp. 203–217, 2019.
[12] D. Patel and A. Singh, "Ai in supply chain management," *International Journal of Production Economics*, vol. 220, pp. 107–121, 2020.
[13] L. Morris and D. Carter, "Leveraging data visualization for retail decision making," *Decision Sciences Journal*, vol. 52, no. 1, pp. 122–138, 2021.
[14] R. Garcia and M. Lopez, "Augmented reality in retail: Enhancing customer experience," *Journal of Retail and Consumer Services*, vol. 59, pp. 102–115, 2022.
[15] U. M. L. Repository, "Online retail ii data set," 2021, accessed: 2024-09-08. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/online+retail+ii