International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN - 2395-1990
Online ISSN : 2394-4099
Available Online at :www.ijsrset.com
doi :https://doi.org/10.32628/IJSRSET24115108

LLM for Retail Business (Optimizing Clothing Sales with AI)

Deepali Narwade[1], Aditya Kanhere[2], Sahil Mulla[2], Atish Sanap[2], Abhay Patil[2]
[1]Assistant Professor, [2]BE Student
Department of Artificial Intelligence & Data Science Engineering DYPCOEI, Varale, Pune (SPPU), Maharashtra, India

ARTICLEINFO

Article History:
Accepted: 25 Sep 2024
Published: 08 Oct 2024

Publication Issue :
Volume 11, Issue 5
Sept-Oct-2024

Page Number :
176-179

ABSTRACT

This research paper presents an end-to- end implementation of a chatbot system tailored for the retail industry, utilizing a large language model (LLM). The chatbot is designed to assist employees of retail stores, such as clothing outlets, by providing real-time access to critical business data, including inventory levels, sales metrics, and profit margins. The solution aims to streamline decision- making processes, enhance operational efficiency, and improve information accessibility by reducing dependency on manual data retrieval. This approach leverages advanced natural language processing to simplify the interface between business systems and employees, ensuring accurate and timely responses to queries.

Keywords - Natural Language Processing, Large Language Model (LLM), Retail Industry, Google Palm, Streamlit.

I. INTRODUCTION

The retail industry is becoming increasingly data-driven, with real-time access to information playing a critical role in day-to-day operations. Retail employees are often tasked with accessing critical information, such as inventory levels, sales figures, product availability, and customer preferences, to ensure smooth store operations. Traditionally, accessing such data has involved manual processes or navigating complex business systems, leading to inefficiencies and potential errors in decision- making. Modern technologies like artificial intelligence (AI) and natural language processing (NLP) have emerged as powerful tools in transforming how employees interact with business systems. Specifically, chatbot systems powered by large language models (LLMs) offer an intuitive interface, enabling employees to access relevant business information quickly and accurately using simple, conversational queries. The chatbot integrates seamlessly with existing business systems and is capable of retrieving real-time data, including inventory, sales metrics, and profit margins, in response to user queries. By streamlining information retrieval, this chatbot helps reduce the reliance on manual data access and enhances decision-making, ultimately contributing to improved operational efficiency within retail stores. This research proposes an innovative solution by developing a question-answer system using a large language model (LLM) tailored for retail environments.

- **Motivation:**

To address the challenges faced by retail businesses in managing data, predicting sales, and understanding customer preferences, thereby enabling more effective, data-driven decision- making and boosting overall efficiency.

II. RELATED WORK

1. Niksa Alfirevic, Daniela Garbin and Pranicevic in their paper titled "Custom- Trained Large Language Models as Open Educational Resource" presented the contribution of custom-trained Large Language Models (LLMs) to developing Open Education Resources (OERs) in higher education. Feedback on the clarity of responses, their awareness of the information available from the chatbot, the immediacy of information delivery provided, and the conversational quality had high scores [1].
2. Stefano Filippi and Barbara Motyl in their paper titled "Large Language Models (LLMs) in Engineering Education: A Systematic Review and Suggestions for Practical Adoption" proposed a method based on Large Language Models (LLMs) is now spreading in several areas of research and development. This work is concerned with systematically reviewing LLMs' involvement in engineering education. LLM- based tools for enhancing educational activities and measuring impact, with certain tools demonstrating advantages such as enhanced student engagement, improved problem-solving [3].
3. M.F. Mridha & Talha Bin Sarwar in their paper titled "A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM" presented sentiment analysis classifies text into positive, negative, or neutral categories and is crucial for fields like marketing and politics. However, it becomes challenging with foreign languages and limited labeled training data. The metrics used—Accuracy, Precision, Recall, F1 Score, and Specificity evaluate the performance of sentiment analysis across two translation services (LibreTranslate and Google Translate) and four models (Twitter Roberta Base, Bertweet-Base, GPT-3, and a new Proposed Ensemble model) [2].
4. Rajvardhan Patil and Venkat Gudivada in their paper titled "A Review of Current Trends, Techniques, and Challenges in Large Language Models (LLMs)" focused on reasoning capabilities of large language models (LLMs) for tasks such as arithmetic reasoning, commonsense reasoning, and math word problems. PaLM-540B using CoT surpassed fine tuned GPT-3 on the GSM8K benchmark for math word problems [4].
5. Mohaimenul Azam Khan and Saddam Mukta in their paper titled "Exploring the Latest Applications of OpenAI and ChatGPT: An In-Depth Survey" proposed to the paper discusses the rapid growth and the challenges associated with understanding the overall impact of Large Language Models (LLMs) in various natural language processing (NLP) tasks. LLMs have demonstrated exceptional performance across a variety of NLP tasks, showcasing their effectiveness in real-world applications [10].

III. PROBLEM STATEMENT

The traditional retail approach struggles with processing large volumes of customer data, resulting in poor inventory management and ineffective sales strategies.

IV. PROPOSED METHODOLOGY

The chatbot system for retail, utilizing large language models (LLMs) and vector databases for efficient query handling and response generation.

connection between the natural language input and the structured data systems using tools like SQL Database Chain and Few Shot Prompt Template.

- Hugging Face Embeddings and Vector Database: To ensure fast and relevant data retrieval, the system uses Hugging Face models to generate vector embeddings of data. These embeddings are stored and managed in a ChromaDB vector database. This allows the system to handle complex queries related to retail metrics such as inventory levels, sales figures, and discount calculations efficiently.
- SQL and Data Access: The system can translate natural language questions into SQL queries using the LLM. For example, the question, "How much is the total price of the inventory for all S-size T-shirts?" is converted into an SQL query:

System Architecture:

- Large Language Model (LLM): At the core of the system is the LLM, in this case, Google's PaLM 2 model, integrated through a platform like LangChain. The LLM interprets user queries, which might involve structured or semi-structured data requests (e.g., SQL queries), and provides meaningful results based on data stored in the retail databases.
- LangChain Framework: The system uses LangChain, which provides a framework to interact with LLMs and external systems like SQL databases. It facilitates the connection between the natural language input and the structured data systems using tools like SQL Database Chain and Few Shot Prompt Template.
- Hugging Face Embeddings and Vector Database: To ensure fast and relevant data retrieval, the system uses Hugging Face models to generate vector embeddings of data. These embeddings are stored and managed in a ChromaDB vector database. This allows the system to handle complex queries related to retail metrics such as inventory levels, sales figures, and discount calculations efficiently.
- SQL and Data Access: The system can translate natural language questions into SQL queries using the LLM. For example, the question, "How much is the total price of the inventory for all S-size T-shirts?" is converted into an SQL query:

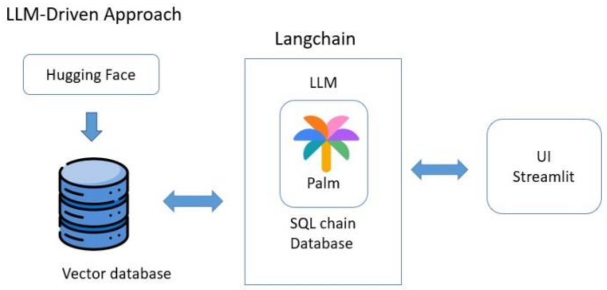

Fig. 1) System Architecture

Algorithm:

The algorithm for processing and responding to user queries consists of the following steps:

1. Input Handling:

The system receives the user's natural language question QQQ through the chatbot interface.

2. Query Parsing and Intent Recognition (LLM):

The LLM parses the natural language input and recognizes the intent behind the query.

It identifies relevant entities (e.g., product names, size, date ranges) and maps them to specific fields in the retail database schema.

3. SQL Query Generation:

Based on the parsed query, the LLM generates a corresponding SQL query using the SQL DatabaseChain.

4. Embedding Search (Vector Database):

For more complex or fuzzy queries, the system generates vector embeddings using Hugging Face models.

5. Data Retrieval and Aggregation:

The SQL query is executed on the relational database, while the embedding-based query retrieves approximate matches from the vector database.

6. Response Generation:

The system formats the retrieved data into a user-friendly response using the function g(·)g(\cdot)g(·). If the user requested a summary or aggregated data (e.g., total sales, inventory counts), the system computes the necessary metrics and presents them in a readable format.

7. Response Output:

The final response is sent back to the user via the chatbot interface, completing the query-response cycle.

V. CONCLUSION

The implementation of an LLM-powered system in the retail industry significantly enhances operational efficiency by providing real-time access to critical business data, simplifying decision- making processes, and improving information accessibility for employees, ultimately streamlining store operations and reducing manual data retrieval.

VI. REFERENCES

[1]. Niksa Alfirevic, Daniela Garbin and Pranicevic, “Custom - Trained Large Language Models as Open Educational Resource” published on MDPI, June 2024.

[2]. M.F. Mridha & Talha Bin Sarwar, “A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM” published on Scientific reports, April 2024.

[3]. B. C. Surve, B. Nemade, and V. Kaul, "Nano-electronic devices with machine learning capabilities," ICTACT Journal on Microelectronics, vol. 9, no. 3, pp. 1601-1606, Oct. 2023, doi: 10.21917/ijme.2023.0277.

[4]. G. Khandelwal, B. Nemade, N. Badhe, D. Mali, K. Gaikwad, and N. Ansari, "Designing and Developing novel methods for Enhancing the Accuracy of Water Quality Prediction for Aquaponic Farming," Advances in Nonlinear Variational Inequalities, vol. 27, no. 3, pp. 302-316, Aug. 2024, ISSN: 1092-910X.

[5]. B. Nemade, S. S. Alegavi, N. B. Badhe, and A. Desai, “Enhancing information security in multimedia streams through logic learning machine assisted moth-flame optimization,” ICTACT Journal of Communication Technology, vol. 14, no. 3, 2023.

[6]. Stefano Filippi and Barbara Motyl, “Large Language Models (LLMs) in Engineering Education: A Systematic Review and Suggestions for Practical Adoption” published on MDPI, June 2024.

[7]. Rajvardhan Patil and Venkat Gudivada, “A Review of Current Trends, Techniques, and Challenges in Large Language Models (LLMs)” published on MDPI journal, March 2024.

[8]. Olamilekan Shobayo, Swethika Sasikumar Sandhya Makkar, “Customer Sentiments in Product Reviews: A Comparative Study with GooglePaLM” published on MDPI journal June 2024.

[9]. Mohaimenul khan, Saddam Mukta, “Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges”, published on IEEE, Jun 2024.

[10]. Chang Victor, Hall Karl, “Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine Learning Models”, published on Scientific reports, June 2024.

[11]. Saadat Izadi, Mohmad Forouzanfar, “Error Correction and Adaptation in Conversational AI: A Review of Techniques and Applications in Chatbots”, published on published on MDPI journal, June 2024.

[12]. Alfirevic Niksa, Mabic Mirela, “Large Language Models for Business: Current State and Future Directions”, published on Scientific reports, April 2023.

[13]. Hong Zhang, Haijian Shao, “Exploring the Latest Applications of OpenAI and ChatGPT: An In-Depth Survey”, published on Tech Science Press, Dec 2023.