Journal of Analysis and Computation (JAC)
(An International Peer Reviewed Journal), www.ijaconline.com, ISSN 0973-2861
Volume XVIII, Issue II, July-December 2024

RETAIL Q & A TOOL USING GENERATIVE AI

[1]Anshul Suryawanshi, [1]Anuj Sable1, [1]Mayur Ganvir, [1]Rohan Raut, [1]Shantanu Thakur, [2]Pinky Gangwani

[1]Student, Department of Artificial Intelligence, St.Vincent Pallotti College of Engineering and Technology Nagpur, India

[2]Assistant Professor, Department of Artificial Intelligence, St.Vincent Pallotti College of Engineering and Technology, Nagpur, India

ABSTRACT:

Retail shops often experience high customer traffic, which can lead to inadequate monitoring of stock levels and uncertainty about inventory. To address this challenge, we are introducing the Retail Q & A Tool Using Generative AI.[1] This innovative tool helps retail shops accurately track their inventory by converting natural language queries into SQL queries. These SQL queries are executed to provide precise and up-to-date information about stock levels.[2] In our project, when a user asks a question in natural human language, it is transformed into a SQL query using a Large Language Model (LLM).[3] We are utilising Google PaLM for this conversion. Google PaLM is integrated into the LangChain framework, which also includes the SQL Database Chain class to manage SQL queries effectively.[4] Additionally, we will use the Hugging Face library to generate word embeddings, which will be stored in a vector database to facilitate efficient and accurate query processing. The final component of the project involves developing a user-friendly interface using Streamlit. This interface will allow users to interact seamlessly with the tool, improving inventory management and ensuring that retail shops have timely and reliable information about their stock levels. This solution aims to streamline inventory tracking and enhance operational efficiency in busy retail environments.

Keywords- Generative Ai, Gemini, Real Time Update, User Friendly Accurately Track.

[1] INTRODUCTION

In the fast-paced and highly competitive world of retail, effective inventory management is vital for maintaining operational efficiency, maximizing sales[5]. Retail shops often encounter significant challenges due to fluctuating customer traffic, leading to inadequate stock level monitoring.[6] When inventory is not managed properly, it can result in both overstock and stockouts, creating uncertainty about product availability.[7] The traditional methods of inventory tracking often fall short in dynamic retail environments. Employees may struggle to retrieve accurate stock information quickly.[8] This lag in response can frustrate customers and lead to lost sales opportunities, highlighting the need for a more efficient solution that allows for real-time access to inventory data.

To tackle these pressing challenges, we propose the development of the Retail Q & A Tool Using Generative AI. This innovative solution aims to revolutionize how retail shops manage their inventory by seamlessly converting natural language queries into actionable SQL queries.[9] By enabling users to ask questions in everyday language, this tool provides a more intuitive approach to inventory management.

At the heart of our solution is using a Large Language Model (LLM), specifically Google Gemini, which has been chosen for its advanced capabilities in natural language understanding. By integrating Gemini into the LangChain framework, we can leverage its power to interpret user inquiries accurately and convert them into SQL queries that retrieve relevant data from the database[10]. This conversion process ensures that users receive precise and timely information regarding stock levels, empowering them to make informed decisions quickly.

In addition to the backend capabilities, the project emphasizes the importance of user experience. We are developing a user-friendly interface using Streamlit, which will facilitate seamless interaction with the tool. This interface will allow users to easily input their queries and receive real-time responses about their inventory. By simplifying the process of accessing inventory information, we aim to empower retail staff, enabling them to respond promptly to customer inquiries and improve overall operational efficiency.

Ultimately, the Retail Q & A Tool Using Generative AI represents a significant advancement in inventory management for retail shops.[11] By streamlining inventory tracking and enhancing access to reliable information, this solution aims to transform how retailers operate in busy environments.

[2] LITERATURE REVIEW

The evolution of artificial intelligence (AI) and natural language processing (NLP) has revolutionized how users interact with complex systems. In retail, where large datasets are generated daily, extracting actionable insights is often a challenge for non-technical users. Traditional database query methods require technical expertise, creating a gap between decision-makers and the valuable insights hidden in the data.[12] To address this, integrating AI-powered NLP tools like Gemini AI with SQL databases offers an intuitive way for users to query and visualize retail data using simple natural language.

This approach aims to bridge the technical gap by enabling retail professionals to retrieve meaningful insights without requiring programming knowledge. By combining robust data handling capabilities of SQL with the contextual understanding of AI models, this tool provides an accessible and efficient solution for retail data analysis[13]. This section reviews existing technologies, methodologies, and studies relevant to the development of the Retail Q & A Tool, emphasizing their application in addressing real-world challenges in the retail sector.

Text-to-SQL Systems

Text-to-SQL systems aim to simplify database querying by converting natural language inputs into SQL queries.[14] Early approaches relied on rule-based techniques and template matching, where predefined patterns were used to interpret user inputs and generate corresponding SQL commands. These methods were limited in handling complex or ambiguous queries and struggled with diverse language structures.

Recent advancements in NLP and deep learning have led to the development of more sophisticated models for text-to-SQL tasks. These models leverage contextual understanding to accurately interpret and generate SQL queries, even for complex inputs[15]. However, a major limitation of current systems is their reliance on English inputs, restricting their applicability in multilingual environments. This presents a significant challenge in linguistically diverse regions, where users might interact in regional languages.

Multilingual NLP Models

The growing adoption of AI technologies across the globe has emphasized the need for tools that support diverse linguistic communities. Multilingual NLP models have made significant advancements in understanding and processing text across various languages, including regional languages like Hindi and Marathi. These models are trained on extensive multilingual datasets, enabling them to identify patterns and relationships across languages with varying grammatical structures.[16]

In the retail domain, multilingual NLP models have demonstrated potential in applications such as customer support and chatbots, where users interact using their native languages. However, their application in structured query generation tasks, like text-to-SQL, is still emerging. In retail analytics, users may pose queries about sales or inventory in regional languages such as Hindi or Marathi. The system must accurately interpret these inputs, translate them if necessary, and preserve the original intent for SQL query generation.

Machine translation tools have been utilized to address this challenge by converting inputs in regional languages into English for further processing. However, achieving accurate translations is challenging when dealing with colloquialisms, idiomatic expressions, or domain-specific terms often used in retail contexts. For instance, the literal translation of phrases in Marathi or Hindi may fail to capture the specific context or meaning intended by the user.

The integration of multilingual capabilities with text-to-SQL systems represents a critical area of research for improving accessibility. By enabling natural language inputs in regional languages like Hindi and Marathi, such systems can cater to a broader user base, particularly in linguistically diverse regions like India. The challenge lies in creating models that can handle diverse grammatical structures and semantic nuances, ensuring seamless translation and query generation.

Voice Recognition Systems

Voice recognition technologies have gained widespread adoption due to advancements in speech processing and the popularity of virtual assistants. In retail analytics, voice input offers an intuitive and hands-free interface for querying databases, allowing stakeholders to verbally inquire about sales performance, inventory, and customer trends.[17]

While voice-enabled systems enhance user experience, their effectiveness is limited in multilingual settings. Most existing speech recognition systems are optimized for English, resulting in lower accuracy when processing inputs in regional languages. This is a critical issue in linguistically diverse regions, where users may communicate in local languages or dialects.

Recent research has focused on training voice recognition models on datasets representing various languages and accents to improve transcription accuracy. These efforts aim to address challenges such as regional accents and colloquialisms, which often deviate from standardized language models.[18] Enhancing the multilingual capabilities of voice recognition systems is essential for their effective application in retail analytics.

Integration of Multilingual NLP, Text-to-SQL, and Voice Recognition

The integration of multilingual NLP, text-to-SQL systems, and voice recognition technologies offers a promising solution for enhancing data accessibility in retail analytics. Studies have demonstrated the potential of combining voice input with natural language querying systems to improve user experience.[19] However, the integration of voice-enabled input, multilingual support, and SQL query generation remains underexplored, particularly in the retail domain.

Research on conversational AI platforms has highlighted the benefits of generating structured queries from natural language inputs. Integrating voice input with natural language processing and database querying frameworks could democratize access to data analytics tools[20]. This approach requires accurate language detection, effective translation, and the ability to preserve contextual nuances in queries.

The proposed Retail Q & A Tool seeks to address these challenges by combining advancements in NLP, machine translation, and speech recognition. By leveraging a unified framework for SQL generation, multilingual translation, and voice input, the tool aims to provide a comprehensive solution for querying retail databases. This approach is expected to enhance the accessibility and usability of data analytics tools for non-technical users, regardless of their language proficiency.

[3] SUMMARY AND RESEARCH GAPS

This literature review highlights significant advancements in text-to-SQL systems, multilingual NLP models, and voice recognition technologies.[21] While these developments have improved user accessibility in data querying tasks, existing solutions often address these areas in isolation. The lack of comprehensive tools integrating multilingual and voice-enabled natural language querying with accurate SQL generation represents a major gap in the current research.

The proposed project addresses this gap by developing a multilingual, voice-enabled tool for querying retail databases.[22] By integrating these technologies into a single framework, the tool aims to make data analytics more accessible to a broader range of users in the retail industry.

[4] METHODOLOGY

The development of the Retail Q & A Tool Using Generative AI involves a systematic approach, integrating multiple technologies to ensure effective inventory management in retail environments. The methodology comprises several key phases, each focused on a specific aspect of the tool's design and implementation.

1. Tools and Technologies

- Hardware: Laptop with decent processing (e.g., Intel core i5 processor, 8 GB RAM)
- Software: Anaconda Jupyter
- Programming Languages: Python, SQL
- AI Framework: LangChain

- Database Management System: SQLite
- Front-End Technology: Streamlit

2. System Design

- **Architecture Planning**: Design the system architecture, outlining how different components will interact.
- **Data Schema Design**: Develop a database schema to organize inventory data effectively, ensuring it supports the SQL queries that will be generated.

3. Integration of Generative AI

- **Model Selection**: Choose Google Gemini as the Large Language Model for its advanced natural language processing capabilities.
- **Training and Fine-tuning**: Fine-tune the LLM if necessary, using relevant datasets to enhance its understanding of retail-specific terminology and queries.

4. Query Conversion Process

- **Natural Language Processing**: Implement a mechanism to convert natural language questions into SQL queries using the chosen LLM.
- **Integration with LangChain**: Utilize the LangChain framework to facilitate the seamless execution of SQL queries, ensuring effective management of database interactions.

5. Database Management

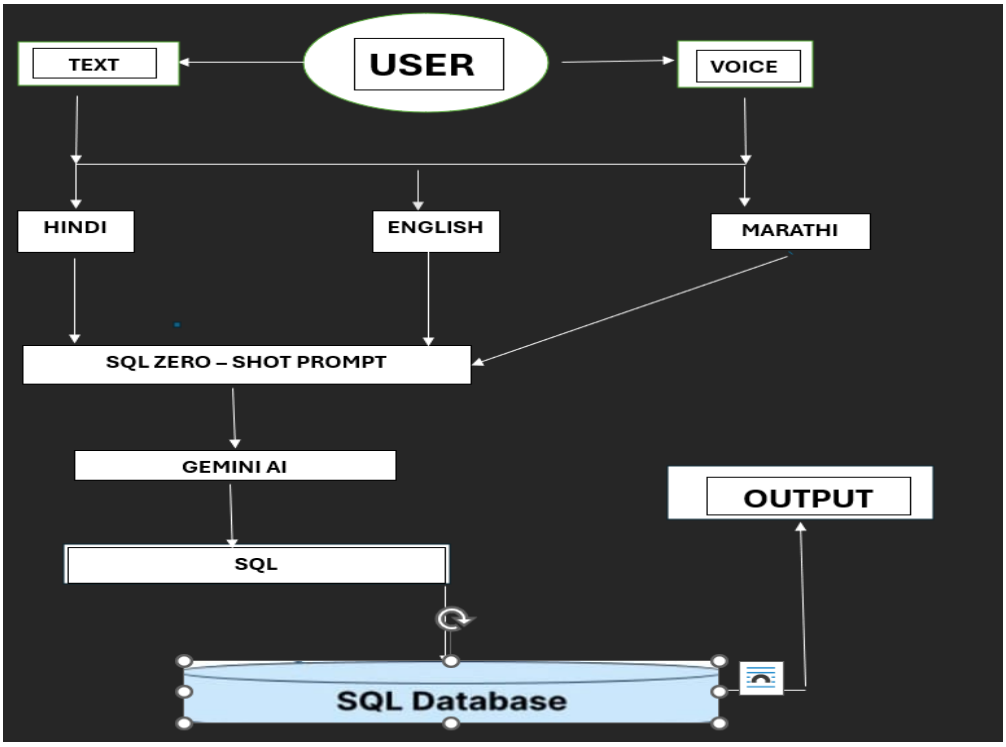
- **Database Setup**: Use a suitable database system (MySQL) to store inventory data securely.[23]

6. User Interface Development

- **Interface Design**: Develop a user-friendly interface using Streamlit, focusing on ease of use and accessibility for retail staff.
- **Interactive Features**: Incorporate features that allow users to input queries, view results, and navigate through inventory data easily.

Architecture:

Journal of Analysis and Computation (JAC)
(An International Peer Reviewed Journal), www.ijaconline.com, ISSN 0973-2861
Volume XVIII, Issue II, July-December 2024



[5] RESULTS AND ANALYSIS

The Retail Q & A Tool Using Generative AI revolutionizes inventory management by addressing challenges posed by high customer traffic and limited stock monitoring. By converting natural language queries into SQL commands, it offers real-time, precise insights into inventory, ensuring retailers can make swift, data-driven decisions. This tool reduces the risk of stockouts or overstocking, enhances customer satisfaction, and streamlines operations by minimizing manual errors. With AI integration, retailers can focus more on core business activities while maintaining optimal stock levels to meet customer demand and stay competitive in the market.

The analysis of the Retail Q & A Tool Using Generative AI highlights its ability to tackle key inventory management challenges in retail environments. High customer traffic often complicates stock monitoring, leading to inefficiencies such as stock outs or overstocking. This tool bridges the gap by translating natural language queries into SQL commands, providing accurate and real-time inventory insights.

The tool empowers staff to make quick, data-driven decisions, ensuring smooth operations and reducing reliance on manual tracking, which is prone to errors. It not only optimizes stock levels but also enhances customer satisfaction by ensuring product availability. Additionally, the use of AI-driven automation allows retailers to focus more on customer engagement and strategic tasks, improving overall business efficiency.

[6] CONCLUSION

The Retail Q & A Tool Using Generative AI provides a transformative solution to the common inventory management challenges faced by retail shops. High customer traffic often makes it difficult to monitor stock levels accurately, leading to missed restocking opportunities and inefficiencies. This tool bridges that gap by translating natural language queries into SQL queries, enabling retailers to access real-time and precise data about their inventory.

By simplifying access to inventory information, the tool empowers store managers and staff to make data-driven decisions quickly. Whether it's identifying low-stock items, checking replenishment needs, or verifying stock levels across multiple products, the system ensures seamless operations and minimizes the risk of stockouts or overstocking. This ultimately improves customer satisfaction by ensuring products are always available when needed.

FUTURE SCOPE:

1. **Predictive Analytics Integration:**
   The tool can incorporate predictive models to forecast demand trends, helping retailers manage inventory proactively and prevent stockouts.
2. **Voice-Based Queries:**
   Future versions could support voice recognition, enabling faster and more efficient access to inventory information on the go.
3. **Machine Learning Enhancement:**
   With machine learning, the tool can refine responses based on user behavior and historical data, improving accuracy over time.
4. **Automated Replenishment:**
   The tool could trigger automatic purchase orders when inventory reaches a predefined threshold, streamlining stock management.
5. **Improved Customer Insights:**
   By analyzing product availability and sales patterns, the tool could offer actionable insights to enhance customer satisfaction and optimize operations.

REFERENCES

[1] Finegan-Dollak, C., Kummerfeld, J. K., Zhang, L., Ramanath, R., Sadasivam, S., Zhang, R., & Radev, D. (2018). Improving text-to-SQL evaluation methodology. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 351–360.

[2] Dahl, D. A., Bates, M., Brown, W. F. K. H., Passonneau, S. D., & Shriber, E. (1994). Expanding the scope of ATIS with the ATIS-3 corpus. *Proceedings of the Workshop on Human Language Technology*, 1–8.

[3] Finegan-Dollak, C., Kummerfeld, J. K., Zhang, L., Ramanath, R., Sadasivam, S., Zhang, R., & Radev, D. (2018). Improving text-to-SQL evaluation methodology. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics: Volume 1, Long Papers (Melbourne, Australia, July 2018)*, 351–360.

[4] Galhotra, M., Yuan, X., Mostafa, H., Tang, J., & Zhu, Z. (n.d.). Towards foundation models for knowledge graph reasoning.

[5] Giordani, A., & Moschitti, A. (2018). Automatic generation and reranking of SQL-derived answers for knowledge graphs. *Proceedings of the Second International Conference on Theory and Practice of Natural Language Processing*, 59–76.

[6] Godfrey, P., & Gryz, J. (1999). Answering queries by semantic caches. In T. M. Benchi-Capon, G. Soda, & A. M. Tjoa (Eds.), *Database and Expert Systems Applications (10th International Conference, DEXA '99, Florence, Italy, August 30 - September 3, 1999, Proceedings)*, Lecture Notes in Computer Science (Vol. 1677, pp. 485–498). Springer.

[7] Green, B. F., Wolf, A. K., Chomsky, C., & Laughery, K. (1961). Baseball: An automatic question-answerer. *Papers Presented at the May 6-10, 1961, Western Joint IRE-AIEE-ACM Computer Conference (IRE-AIEE-ACM '61 Western)*, 219–224.

[8] Green, C. C., & Raphael, B. (1969). The use of theorem-proving techniques in question-answering systems. *Proceedings of the 1969 24th National Conference (ACM '68)*, 169–181.

[9] Hendrick, G., Sacerdoti, E. D., Sagagliozze, D., & Stolcsm, J. (1978). Developing a natural language interface to complex data. *ACM Transactions on Database Systems*, 3(2), 105–147.

[10] Lee, C.-H., Polozov, O., & Richardson, M. (2021). KaggleDBQA: Realistic evaluation of text-to-SQL parsers. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 261–273.

[11] Li, F., & Jagadish, H. V. (2014). NaLIR: An interactive natural language interface for relational databases. *Proceedings of the VLDB Endowment*, 8(1), 73–84.

[12] Nayak Vaghmare, Yuefeng Wang, I. D., & Dilligr, T. (2017). SQUiRE: Query synthesis from natural language. *International Conference on Object-Oriented Programming, Systems, Languages, and Applications*, 63:1–63:26.

[13] Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2023). Unifying large language models and knowledge graphs: A roadmap. *arXiv preprint arXiv:306.08302*.

[14] Shlepyan, J., & Falsafina, Y. (2021). Information extraction from knowledge graphs. *Synthesis Lectures on Data Semantics and Knowledge*. Morgan & Claypool Publishers.

[15] Srinivasan Iyer, Ioannis Konstas, A. C. J. K., & Zettlemoyer, L. (2017). Learning a neural semantic parser from user feedback. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 963–973.

[17] Talmor, L., Ram, O., & Mooney, R. J. (2020). Neural construction of database interfaces: Interpreting statistical data tables using semantic parsers. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 133–147.

[18] Wang, C., Lin, H., Wang, K. Y. M., Zhang, L., & Zhou, D. P. (2018). RADDY: A spider: A large-scale text-to-SQL dataset for complex and cross-domain questions. *Proceedings of the 27th Annual Conference of the Neural Information Processing Systems*, 3911–3921.

[19] Yao, X., & Li, H. (2017). Seq2Seq: Generating structured queries from natural language using reinforcement learning. *CoRR abs/1709.00103*.

[20] Winograd, T. (1970). A transition network grammar for natural language analysis. *Communications of the ACM*, 13(10), 591–606.

[21] Wu, S., Isroy, O., Liu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). BloombergGPT: A large language model for finance.