

LLM-Powered Generative AI for Enhanced Real Estate Property

Sumit Ranjan
Associate Manager-Data Science
Majid Al Futtaim
Dubai, UAE
sumit.ranjan819@gmail.com

Lanwin Lobo
Director Data Science
MResult Services
Bangalore, India
lanwinlobo@gmail.com

Divya Chembachere
Lead Data Scientist
MResult Services
Bangalore, India
c.divya819@gmail.com

Abstract— This research paper introduces a novel approach to revolutionize the real estate property search process by leveraging Large Language Models (LLMs) [1] and advanced AI techniques. The traditional property search methods, often reliant on rigid keyword-based queries, can limit user experience and fail to capture nuanced preferences. This paper presents an innovative solution that harnesses LLMs to enable more natural language queries and deliver highly personalized results, enhancing the property discovery process.

Keywords— AI Large Language Models, LLM, AI, real estate, property search

I. INTRODUCTION

The real estate industry is evolving, driven by advancements in artificial intelligence and Large Language Models (LLMs) [1]. Property search, a crucial component of the real estate ecosystem, is poised for innovation. Conventional search methods often rely on rigid keyword-based queries, which can restrict user experience and overlook the intricacies of user preferences. This paper introduces an approach that utilizes LLMs to enhance property search by enabling more natural language queries and providing highly personalized results.

II. BACKGROUND

A. Traditional Property Search Methods

Historically, property searches have been conducted using rigid keyword-based queries that often fail to capture the full spectrum of user preferences. These methods limit the ability of users to articulate their requirements effectively.

B. Large Language Models (LLMs)

Large Language Models (LLMs) represent a class of artificial intelligence models designed for understanding and generating human-like text. Models like GPT-3.5 [2] have demonstrated exceptional capabilities in processing natural language, comprehending context, and generating coherent text. LLMs are well-suited for a wide range of natural language understanding and generation tasks. They empower users to express their criteria conversationally, making the property search process more user-friendly and effective. For example, users can input queries in a more natural manner, such as "Find homes with 3 bedrooms and a backyard close to a reputable school."

C. LangChain Framework

The LangChain framework [3] is an open-source tool designed to simplify the development of applications using LLMs, text embeddings, and text processing tasks. LangChain allows the chaining together of multiple prompts to an LLM, enabling the generation of complex and nuanced outputs. This framework finds applications in documentation, creating agents, chatbots, and various text-based tasks.

D. FAISS (Facebook AI Similarity Search)

FAISS, developed by Facebook AI Research, is an open-source library for efficiently searching for similar items or vectors in large collections of high-dimensional data. FAISS provides methods for indexing and searching vectors, making it ideal for applications requiring similarity search and clustering of dense vectors.

III. PROPOSED APPROACH

A. User Interface and Interaction

To enhance the property search experience, a user-friendly interface has been developed, allowing users to input their preferences in natural language. The LangChain framework interprets these conversational queries, providing a more intuitive and user-centric approach to property search.

B. Data Retrieval and Processing

The application employs web scraping techniques, utilizing BeautifulSoup, to gather property information from relevant real estate websites. This data is stored in an S3 bucket for efficient retrieval and analysis. To optimize search results, FAISS is used to construct a vector database of property descriptions and features. The OpenAI Embeddings model API, integrated with LangChain, converts this textual data into meaningful embeddings, further enhancing the search process.

C. Combining Retrieval and Generation

To provide intelligent and accurate responses to user queries, the approach employs LangChain's innovative approach combining retrieval-based and generation-based techniques. The FAISS vector database enables a similarity search to retrieve the most relevant property listings based on user queries. Subsequently, OpenAI's powerful language models, such as GPT-3.5, generate contextually rich and personalized responses based on the retrieved property data. This fusion of retrieval-augmented generation ensures that the application delivers accurate and engaging answers to user inquiries.

D. Architecture Details

The architecture of our real estate property search application is designed to ensure efficient data retrieval, processing, and personalized user interactions.

1) *User Interface and Filters*: The user interacts with our application through a user-friendly interface built using Streamlit. This lightweight framework was chosen for its ease of integration and user-friendliness. Users can apply various filters to narrow down their property search criteria using the Property Search app.

2) *Data Retrieval*: To fetch relevant property information from selected real estate websites, we employ web scraping

techniques utilizing BeautifulSoup. The scraped data is then processed for further analysis.

3) *Data Storage*: All scraped property data is stored efficiently in an Amazon S3 bucket [4], ensuring accessibility and durability. This storage approach facilitates data retrieval and analysis.

4) *Data Optimization*: To optimize the quality and relevance of the search results, we leverage Large Language Models (LLMs) [5]. The scraped property data [6] is passed through the OpenAI Embeddings model API to convert textual information into meaningful embeddings. These embeddings are used to enhance the quality of the search results and provide context for user queries.

5) *Efficient Retrieval*: Efficient retrieval of property listings is a critical aspect of our application. To achieve this, we utilize FAISS (Facebook AI Similarity Search) [7], a robust library for similarity search and clustering of dense vectors. FAISS constructs a vector database of property descriptions and features, making it possible to rapidly retrieve relevant results based on user queries.

6) *Intelligent Responses*: Our application employs an innovative approach called LangChain to provide intelligent and accurate responses to user queries. LangChain combines retrieval-based and generation-based techniques. Utilizing the FAISS vector database, LangChain performs a similarity search to retrieve the most relevant content based on user queries. OpenAI's powerful language models, including GPT-3.5, are then employed to generate contextually rich and personalized responses based on the retrieved text. This fusion of retrieval-augmented generation ensures that the application delivers accurate and engaging answers to user inquiries.

7) *Containerization and Deployment*: For ease of deployment and management, our application is hosted as a containerized application on Amazon Web Services (AWS) [8]. This containerization solution ensures a seamless and hassle-free user experience while providing scalability and reliability.

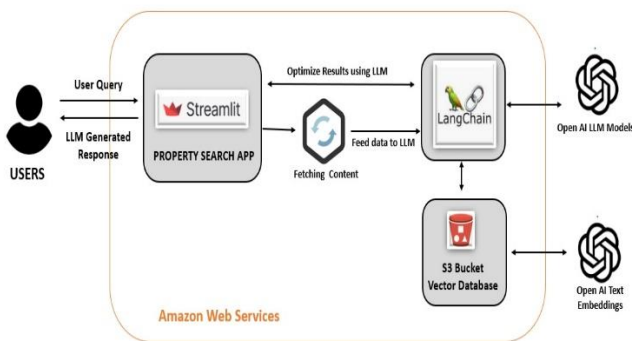


Fig. 1. Architecture Diagram

Fig. 1., presents our architecture diagram, offering an insightful view of our system's intricate workflow. At the forefront, users engage with the Streamlit frontend [9] to initiate property searches, creating an intuitive and user-friendly experience. The system, acting swiftly, retrieves property data and methodically feeds it into LangChain,

accompanied by vector databases [10] for comprehensive analysis.

Within LangChain, a sophisticated data transformation and refinement process takes place, enhancing the quality and relevance of the information. This refined data serves as the foundation for generating contextually rich outputs that represent the most pertinent property listings matching user preferences. These intelligently crafted outputs are seamlessly returned to users through the application interface, ensuring a smooth and user-centric property search journey that empowers users to find their ideal property effortlessly.

IV. CHALLENGES AND CONSIDERATIONS

A. Domain-Specific Terminology and Regional Nuances

One of the primary challenges in implementing this approach is ensuring that LLMs [11] understand domain-specific real estate terminology and regional nuances. Fine-tuning LLMs to accommodate these intricacies is essential for accurate search results.

B. User Data Privacy

The Protecting user data privacy is of utmost importance. Implementing robust data security measures and providing clear transparency regarding the AI algorithm's functioning are critical aspects of building user trust.

V. CONCLUSION

The integration of Large Language Models into real estate property searches has the potential to redefine how individuals find their dream homes. By enabling more natural language queries and harnessing contextual understanding, LLMs bridge the gap between user intent and search outcomes. This paper advocates for the incorporation of LLM-powered [12] property search platforms, enhancing the accuracy, efficiency, and personalization of property recommendations. By embracing this technology, the real estate industry can empower users in their quest for the perfect home.

VI. FUTURE DIRECTIONS

The real estate industry is poised for continued transformation, driven by advancements in technology and AI. As we look ahead, several promising avenues for research and development emerge:

A. Domain-Specific LLM Fine-Tuning

Large Language Models (LLMs) [1] have proven their effectiveness in processing natural language and understanding context. However, fine-tuning these models specifically for the real estate domain presents an exciting opportunity. Future research can focus on enhancing LLMs' [13] proficiency in recognizing domain-specific real estate terminology and regional nuances. This fine-tuning process would lead to more accurate and context-aware property recommendations.

B. Advanced User Personalization

Personalization is a key driver of user satisfaction in the property search process. Continuous learning mechanisms that adapt to users' evolving preferences and feedback can significantly enhance the user experience. Research efforts should explore methods to implement dynamic personalization, taking into account user behaviour, preferences, and changing market conditions.

C. Multi-Modal Approaches

Incorporating image analysis into property recommendations introduces a multi-modal dimension to the search process. Beyond textual descriptions, images of properties can convey essential information and aesthetics. Future research can delve into the integration of computer vision and image recognition technologies [14] to analyze property images. This would enable the system to recommend properties not only based on textual criteria but also on visual attributes, such as architecture, interior design, and property condition.

D. Expanding Data Sources

To provide users with the most comprehensive and up-to-date property listings, it is essential to expand data sources. Research efforts should focus on integrating a wider range of data streams, including real-time market data, local government records, and even user-generated content. By aggregating data from various sources, property search platforms can offer users a more holistic view of the real estate landscape.

E. Enhanced Data Privacy and Security

As AI-driven property search platforms collect and process user data, ensuring robust data privacy and security measures remains paramount. Future research should explore innovative encryption and privacy-preserving techniques to safeguard user information. Additionally, transparent data handling practices and clear communication regarding data usage will be crucial in building and maintaining user trust.

F. Sustainable Real Estate

The growing emphasis on sustainability and eco-friendly living presents opportunities for research in the real estate sector. Future directions could include the integration of environmental impact data into property listings, helping users make informed choices based on ecological considerations.

G. Augmented Reality (AR) and Virtual Reality (VR)

AR and VR technologies [15] can revolutionize the way users explore and visualize properties remotely. Future developments may include immersive property tours and virtual staging, allowing users to experience properties in a more engaging and interactive manner.

These future directions represent just a glimpse of the exciting possibilities within the real estate industry. As technology continues to advance, the potential for innovation in property search and real estate-related AI applications is boundless.

ACKNOWLEDGMENT

The authors would like to acknowledge the support and contributions of the LangChain framework, FAISS, and OpenAI's language models in making this research possible.

REFERENCES

- [1] Yiheng Liu, Tianle Han, Siyuan Ma, Jiayue Zhang, Yuanyuan Yang, Jiaming Tian, Hao He, Antong Li, Mengshen He, Zhengliang Liu, Zihao Wu, Lin Zhao, Dajiang Zhu, Xiang Li, Ning Qiang, Dingang Shen, Tianming Liu, Bao Ge “Summary of ChatGPT-Related Research and Perspective Towards the Future of Large Language Models”, *Meta-Radiology* (2023)100017 : <https://arxiv.org/abs/2304.01852>
- [2] OpenAI. (2021). GPT-3.5. <https://www.openai.com/research/gpt-3>
- [3] LangChain Documentation. https://python.langchain.com/docs/get_started/introduction
- [4] Amazon Web Services. AWS: Containerization Solutions. <https://aws.amazon.com/what-is/containerization/>
- [5] Noah Ziems, Wenhao Yu, Zhihan Zhang, Meng Jiang “Large Language Models are Built-in Autoregressive Search Engines”: <https://arxiv.org/abs/2305.09612>
- [6] Beautiful Soup Documentation. <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
- [7] FAISS: Facebook AI Similarity Search. <https://github.com/facebookresearch/faiss>
- [8] Amazon Web Services. AWS: Containerization Solutions. <https://aws.amazon.com/what-is/containerization/>
- [9] Streamlit. (2021). Streamlit: The fastest way to create data apps. <https://streamlit.io/>
- [10] Jianguo Wang*, Xiaomeng Yi, Rentong Guo, Hai Jin, Peng Xu, Shengjun Li, Xiangyu Wang, Xiangzhou Guo, Chengming Li, Xiaohai Xu, Kun Yu, Yuxing Yuan, Yinghao Zou, Jiquan Long, Yudong Cai, Zhenxiang Li, Zhifeng Zhang, Yihua Mo, Jun Gu, Ruiyi Jiang, Yi Wei, Charles Xie “Milvus: A Purpose-Built Vector Data Management System” : https://www.cs.purdue.edu/homes/csjpgwang/pubs/SIGMOD21_Milvus.pdf
- [11] Sofia Eleni Spatharioti, David M. Rothschild, Daniel G. Goldstein, Jake M. Hofman “Comparing Traditional and LLM-based Search for Consumer Choice”: A Randomized Experiment By Microsoft Research, New York City: <https://arxiv.org/ftp/arxiv/papers/2307/2307.03744.pdf>
- [12] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, Ji-Rong Wen “A Survey of Large Language Models”: <https://arxiv.org/abs/2303.18223>
- [13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. “Attention Is All You Need (2017)” : <https://arxiv.org/abs/1706.03762>
- [14] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale” : <https://arxiv.org/abs/2010.11929>
- [15] A. Alaseeri; A. Janahi; F. Al Khalifa “The application of virtual and augmented reality technologies in the real estate industry in bahrain” : <https://ieeexplore.ieee.org/document/9545695>