

Enhancing Large Language Models' Data Analysis Capability with Planning-and-Execution and Code Generation Agents: A Use Case for Southeast Asia Real Estate Market Analytics

Authors : Kien Vu, Jien Min Soh, Mohamed Jahangir Abubacker, Piyawut Pattamanon, Soojin Lee, Suvro Banerjee

Abstract : Recent advances in Generative Artificial Intelligence (GenAI), in particular Large Language Models (LLMs) have shown promise to disrupt multiple industries at scale. However, LLMs also present unique challenges, notably, these so-called "hallucination" which is the generation of outputs that are not grounded in the input data that hinders its adoption into production. Common practice to mitigate hallucination problem is utilizing Retrieval Augmented Generation (RAG) system to ground LLMs' response to ground truth. RAG converts the grounding documents into embeddings, retrieve the relevant parts with vector similarity between user's query and documents, then generates a response that is not only based on its pre-trained knowledge but also on the specific information from the retrieved documents. However, the RAG system is not suitable for tabular data and subsequent data analysis tasks due to multiple reasons such as information loss, data format, and retrieval mechanism. In this study, we have explored a novel methodology that combines planning-and-execution and code generation agents to enhance LLMs' data analysis capabilities. The approach enables LLMs to autonomously dissect a complex analytical task into simpler sub-tasks and requirements, then convert them into executable segments of code. In the final step, it generates the complete response from output of the executed code. When deployed beta version on DataSense, the property insight tool of PropertyGuru, the approach yielded promising results, as it was able to provide market insights and data visualization needs with high accuracy and extensive coverage by abstracting the complexities for real-estate agents and developers from non-programming background. In essence, the methodology not only refines the analytical process but also serves as a strategic tool for real estate professionals, aiding in market understanding and enhancement without the need for programming skills. The implication extends beyond immediate analytics, paving the way for a new era in the real estate industry characterized by efficiency and advanced data utilization.

Keywords : large language model, reasoning, planning and execution, code generation, natural language processing, prompt engineering, data analysis, real estate, data sense, PropertyGuru

Conference Title : ICAIDM 2024 : International Conference on Artificial Intelligence and Data Mining

Conference Location : Kuala Lumpur, Malaysia

Conference Dates : August 22-23, 2024