ENHANCING LARGE LANGUAGE MODELS WITH RETRIEVAL-AUGMENTED GENERATION FOR THE OIL & GAS INDUSTRY

Julio Figueroa, Methanex Corporation, jfigueroa@methanex.com

Resumen

It is well known that Large Language Models (LLMs) such as ChatGPT, Bard, and Copilot have disrupted the way we work, providing benefits such as improved automation, enhanced decision-making capabilities, and personalized customer interactions. However, a lack of specific domain understanding is one of the most well-known shortcomings of these technologies, particularly in industries like oil and gas, where data precision is crucial for decision-making (1). This is where a Retrieval Augmented Generation (RAG) architecture can help to maximize an LLM's performance by integrating domain-specific data sources, thereby enhancing the model's ability to provide relevant and accurate information

We built a RAG architecture powered by OpenAI and vector databases to enable our users to access meaningful knowledge bases using natural language like talking to another person. These knowledge bases have been historically difficult to access or were even unused due to the unstructured nature of the information. Some knowledge bases we included are incident lessons learned, plant operations manuals, and health and safety procedures, among others. This approach differs from typical search engines that purely guide users to manuals or documents if the user picks the right keywords, thereby saving considerable time and effort.

Even though the use of LLM technologies is growing exponentially, there is still a significant portion of the population that, although they might have heard about it, do not use it. Moreover, usage tends to vary by demographic, with younger people and those with higher education levels more likely to use ChatGPT. Specifically, only 9% of 45 to 64-year-olds and a mere 5% of those aged 65 and older have used ChatGPT, showing lower engagement among older age groups (2). This unawareness of the actual capabilities and mechanics of these technologies might lead to some of the reluctance to adopt them. Therefore, implementing a governance model, providing training, and establishing a formal process for managing this architecture (LLMOps) might help to mitigate these obstacles.

(1) Retrieval-Augmented Generation for Large Language Models: A Survey - https://arxiv.org/abs/2312.10997
(2) Exploding Topics - https://explodingtopics.com/blog/chatgpt-users