Juan David Pérez Olivares
20231020028

T9: What are Vector Database and why are they crucial for LLM development?

Vector databases (VDMs) and Large Language Models (LLMs) like the GPT series are gaining popularity since 2023. In the context of AI and machine learning, the matrices used in VDMs, in their fields represent complex objects like words, sentences, and various media files, in an embedding. In the framework of LLMs, embeddings represent text as a dense vector of assigning semantic meaning to words together or similar features in almost any other type of data. This is used in search engines, an as i said at the beginning, in the GPT series, which embeddings, to query them quickly, more specifically these databases contain matrices of numbers grouped based on their similarity, which can be queried with ultra-low latency.

In other words, vector databases index vectors to make them easier to search and retrieve by comparing values and finding the ones that are most similar to each other. This makes vector databases ideal for AI-driven applications. This proves that data reigns supreme in computational advancements, thus dictating technological trends.

VDMs are crucial for LLMs, since in this model of embeddings to find relationships with the data used in the databases, low latency is required in queries to related indexes, in order to deliver specific answers given a base text or file for processing.