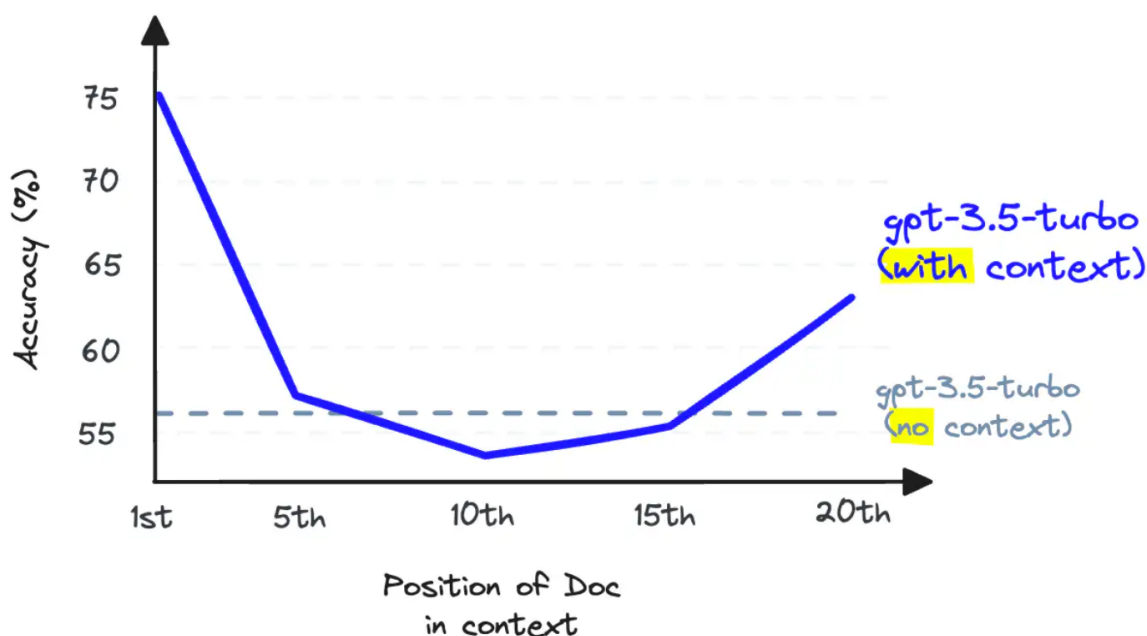




Reranking Method

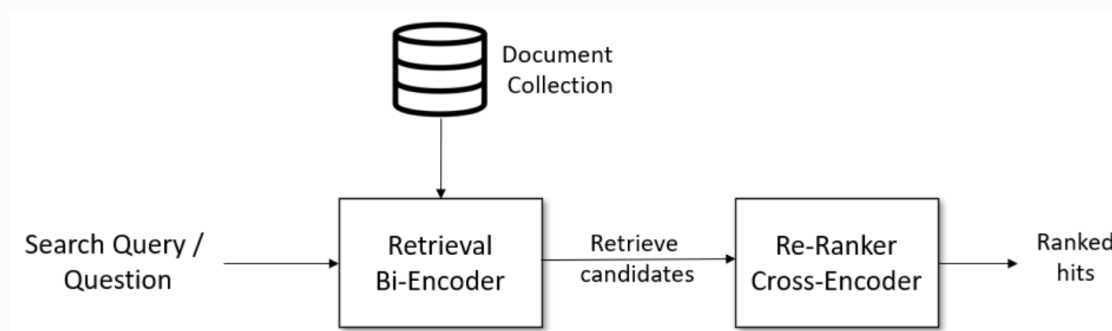
<https://www.pinecone.io/learn/series/rag/rerankers/>

1. Nguyên nhân



1. Lợi ích: Tuy việc đưa các tài liệu vào trong LLM sẽ giúp cho LLM có thể trả lời được kết quả tốt.
2. Hạn chế:
 - Việc đưa quá nhiều dữ liệu vào trong LLM làm Context có thể làm cho LLM lấy những thông tin bị nhiễu đi trả lời.
 - Và việc nhồi nhét quá nhiều dữ liệu cũng làm cho mô hình sẽ khó trả lời câu hỏi.

- Các mô hình ngôn ngữ có xu hướng xử lý và ghi nhớ thông tin ở đầu và cuối của số ngữ cảnh tốt hơn so với thông tin nằm ở giữa.



2. Một số model ranker

Tiếng Việt: <https://huggingface.co/itdainb/PhoRanker>

Đa ngôn ngữ: <https://huggingface.co/BAAI/bge-reranker-v2-m3>