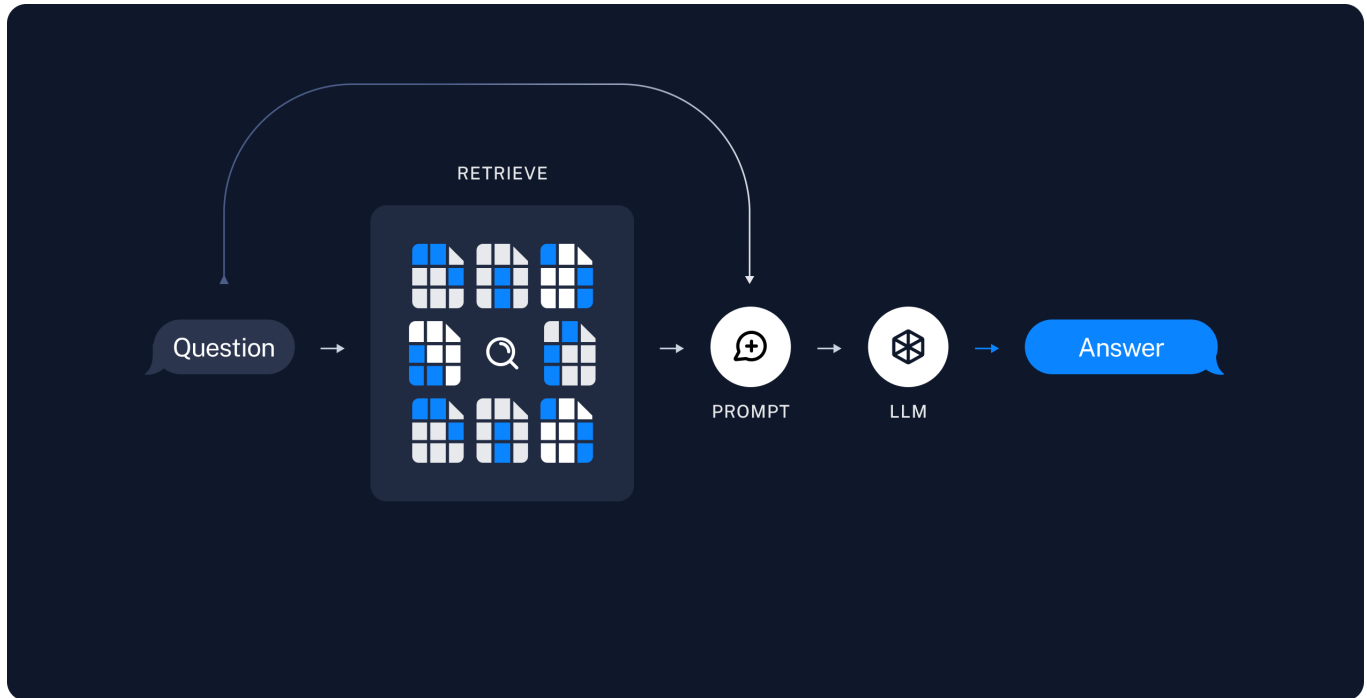


# LLM (Large Language Model)



LLM(Large Language Model) 단계는 Retrieval-Augmented Generation(RAG) 시스템의 일곱 번째 단계로, 이전 프롬프트 단계에서 구성된 입력을 기반으로 대규모 언어 모델을 활용하여 응답을 생성하는 과정입니다. 이 단계는 RAG 시스템의 핵심적인 부분으로, 언어 모델의 능력을 최대한 활용하여 사용자의 질문에 대해 정확하고 자연스러운 답변을 생성합니다.

## LLM의 필요성

1. **사용자 의도 이해:** LLM은 다양한 언어의 구조와 의미를 깊이 이해하고 있으며, 이를 바탕으로 복잡한 질문에 답할 수 있습니다. 자연어 이해(NLU)와 자연어 생성(NLG) 능력이 결합되어, 보다 자연스럽고 유익한 응답을 제공할 수 있습니다.
2. **문맥적 적응성:** LLM은 주어진 문맥을 고려하여 응답을 생성합니다. 이는 사용자의 질문에 더욱 정확하게 대응할 수 있습니다. 또한, 사전학습된 지식외

사용자가 제공한 정보에 기반한 답변을 문맥을 참고하여 답변하도록 합니다.

## LLM의 중요성

LLM 단계는 사용자의 질문에 대한 답변의 질과 자연스러움을 결정짓는 핵심 요소입니다. 이 단계에서 LLM은 지금까지의 모든 데이터와 정보를 종합하여 사용자의 질문에 최적화된 답변을 생성합니다. **LLM의 성능은 RAG 시스템의 전체적인 성능과 사용자 만족도에 직접적으로 영향을 미치며**, 이는 RAG 시스템을 사용하는 많은 응용 분야에서 매우 중요한 역할을 합니다.

## 코드

OpenAI GPT-4o 활용

```
# 단계 7: 언어모델(LLM) 생성
# OpenAI 의 GPT-4o 모델 을 생성합니다.
llm = ChatOpenAI(model_name="gpt-4o")
```

Anthropic Claude3 Sonnet 활용

```
from langchain_anthropic import ChatAnthropic

# 단계 7: 언어모델(LLM) 생성
# Anthropic 의 Claude 모델 을 생성합니다.
llm = ChatAnthropic(model="claude-3-sonnet-20240229")
```

로컬 모델(llama3-8b) 활용

```
from langchain_community.chat_models import ChatOllama

# 단계 7: 언어모델(LLM) 생성
# LangChain이 지원하는 Ollama(로컬) 모델을 사용합니다.
llm = ChatOllama(model="llama3:8b")
```

# 참고

- [OpenAI GPT 모델](#)
- [LangChain LLMs](#)