

고객사 및 내부 질의응답 중심

1. LLM 관련 질문

1.1 LLM 관련 성능 및 구성

#LLM #PrivateLLM #모델선택

Q. LLM 모델 관련 사양은 무엇이 주요 지표인가?

LLM(대규모 언어 모델)은 파라미터 수와 컨텍스트 윈도 크기(최대 토큰 수)가 주요 사양입니다.

- LLM의 사양과 성능 간의 상관관계

일반적으로 파라미터 수가 많을수록 성능이 좋아지지만, 일정 수준 이상에서는 데이터 품질, 학습 방법이 더 중요해집니다.

즉, 크기가 성능에 영향을 주긴 하지만, '무조건 클수록 좋은 건 아님'이 최근 트렌드입니다.

- 사양과 비용 간의 상관관계

파라미터 수가 많을수록 연산량과 메모리 사용량이 급증해 학습·추론 비용도 높아집니다. 따라서 고사양 모델은 높은 성능을 주지만, 실사용 시 운영비용과 응답 속도 문제도 함께 고려해야 합니다.

Q. LLM 관련하여 파라미터 수란 무엇인가?

파라미터 수 = LLM 모델 사양(크기)

LLM의 크기는 내부에 존재하는 파라미터 수로 결정되며, 파라미터가 많을수록 더 복잡한 문맥과 개념을 학습할 수 있습니다.

즉, 파라미터 수는 모델이 얼마나 정교하고 깊이 있게 사고할 수 있는지를 나타냅니다.

Q. LLM 관련하여 컨텍스트 윈도우란 무엇인가?

- 컨텍스트 윈도우 (Context Window)

컨텍스트 윈도우는 한 번에 처리할 수 있는 입력 길이로, 대화나 문맥을 얼마나 길게 기억할 수 있는지를 뜻합니다.