

목차

고객사 및 내부 질의응답 중심.....	6
1. LLM 관련 질문.....	6
1.1 LLM 관련 성능 및 구성.....	6
Q. LLM 모델 관련 사양은 무엇이 주요 지표 인가?.....	6
Q. LLM 관련하여 파라미터 수란 무엇인가?.....	6
Q. LLM 관련하여 컨텍스트 윈도우 란 무엇인가?.....	6
Q. 대표적인 LLM 모델 종류와 관련 사양은 어떻게 되는가?.....	7
Q. LLM 품질 및 정확도는 어떤 방식으로 측정하는지?.....	8
Q. ChatGPT와 LLama 기반 Private LLM 성능 차이는 얼마나 나나?.....	9
Q. 다양한 LLM 사용 중 어떤 LLM이 가장 좋다고 평가하는지?.....	9
Q. 모델 선별 기준이란 무엇인가?.....	9
Q. ChatGPT가 아닌 Local LLM을 썼을 때 효과 중 환경 보안 강화의 의미는?	9
Q. 온프레미스란??.....	10
Q. LLM 생성 데이터를 HWP나 Office 형태 문서로 출력 지원 가능한지?....	10
Q. 나중에 LLM 버전이 개선되면 변경 가능한지? (Multi LLM).....	10
Q. Private LLM 기반으로 구축한 사업 경험이 있는가?.....	10
Q. OpenAI 모델을 그대로 사용하는지, 로컬 모델도 가능한지?.....	11
Q. 파인튜닝(모델 재학습)과 기본 모델 활용 방식의 차이는 무엇인가?.....	11
2. RAG 관련 질문.....	11
2.1 RAG 구성 및 운영.....	11
Q. RAG란 무엇인가? 구성과 기술 요소는?.....	11
Q. Local LLM을 내부적으로 사용 중이며, RAG에도 로컬 LLM 연동이 가능한가?.....	12
Q. RAG 관련 성능 평가를 받은 지표가 있는가?.....	13
Q. RAG 구성을 위한 하드웨어 스펙이나 아키텍처 구성은 어떻게 되는가?... ..	13
Q. RAG 기반 챗봇 개발 기간은 대략 얼마나 되는가?.....	13
Q. RAG 말고 DB 데이터를 조회하여 처리 지원 가능한지? (Text to SQL 관련). ..	13
Q. 프롬프트 질문은 RAG 구성에서 어디에 전달되는가?.....	14
Q. RAG 챗봇 사례 중 실제 구현 예시는?.....	14
1.3 임베딩 및 벡터 DB.....	15
Q. VectorDB란 무엇인가?.....	15
Q. Vector Embedding Model이란 무엇인가?.....	15
Q. Vector DB는 어떤 걸 사용하는지?.....	15