Q. 청킹 알고리즘은 어떤 걸 사용하는지?

A. 의미 기반, 고정 길이, Sliding Window 기반, 인접 문장 군집화 등 다양한 전략을 선택적으로 사용합니다.

Q. 자체 벡터 검색 엔진이 있는 건가요, 아니면 엘라스틱서치 기반인가요?

A. 엘라스틱서치를 기반으로 하되, 리트리버, 벡터 기반 서치, 리랭킹 기술과 관련한 자사 코드 기반 기술 및 노하우 등을 조합해 적용함. 프로젝트에 따라 알맞는 벡터 모델(M3, e5 등)을 기술 조사 및 검증을 통해해 적절히 선택하여 활용 중입니다.

6. 보안 및 개인정보 처리

#PII #비식별화 #암호화

6.1 민감 정보 처리 및 보안

Q. 검색 정보에 개인정보나 금융정보 처리는 어떻게 하는지?

A. PII(개인정보 식별자) 마스킹 및 비식별화 처리 후 검색 및 응답 제어 기능을 제공합니다. 데이터 수집단계에서 개인정보를 정규표현 및 패턴 기반으로 찾아서 마스킹 진행 하며 프롬프트에서 주요 개인 정보 금융정보를 노출 하지 않도록 프롬프트 엔지니어링을 적용합니다. 사후에도 프롬프트 및 조회 로그를 기반으로 개인정보 금융정보가 노출 되었는지 검증하는 방식으로 사업 진행을 하고 있습니다.

사업 규모 및 보안에 민감성에 따라서 전문 보안 솔루션(예시. 구간암호화, 데이터 비식별화 솔루션)과의 연계를 진행 하기도 합니다.

Q. 보안 공격 대응 (프롬프트 인젝션 등)은 가능한가?

A. 권한 기반 답변 제한, 프롬프트 패턴 필터링 등의 대응책이 포함되어 있습니다..