

Daily 협업일지(02/11)

[1] 오늘 날짜 / 이름 / 팀명

- 날짜: 2026.02.11
- 이름: 김슬기
- 팀명: 6팀

[2] 오늘 맡은 역할 및 구체적인 작업 내용

오늘 당신이 맡았던 역할은 무엇이었고, 어떤 작업을 수행했나요?

(예: 모델 학습 파라미터 조정, 결측치 처리, 발표자료 구성 등)

👉 답변:

- 스마트 필터링 적용: "서식", "양식" 키워드가 있는 문서를 제거하려 했으나, 정답(예산표)까지 지워지는 문제 발생. -> "금액 패턴(Regex)이 있으면 무조건 보존"하는 로직 추가
- 리랭킹(Reranking) 테스트: BAAI/bge-reranker 모델을 적용해봤으나, 정확도 상승 대비 속도가 너무 느려(3배 지연) 도입 보류 결정
- 최종 설정 확정: configs/dev.yaml에 Chunk 500, K=10, Hybrid On 설정 적용. (최종 정확도 73% 달성)
- 보고서 작성: 중간 보고서 작성 및 팀 공유

[3] 오늘 작업 완료도 체크 (하나만 체크)

진척 상황을 정량적으로 표시하고, 간단한 근거도 작성하세요.

- ☐ 0% (시작 못함)
- ☐ 25% (시작은 했지만 진척 없음)
- ☒ 50% (진행 중, 절반 이하)
- ☐ 75% (거의 완료됨)
- ☐ 100% (완료 및 점검까지 완료)

👉 간단한 근거:

(50%) 파싱부터 검색까지 핵심 파이프라인 검증이 끝났고, 최적의 설정값을 찾음. 이제 고도화 및 UI 연동 단계로 넘어갈 준비 완료.

[4] 오늘 협업 중 제안하거나 피드백한 내용이 있다면?

오늘 회의나 메시지에서 당신이 제안하거나 팀에 피드백한 내용은 무엇인가요?

👉 답변:

-

[5] 오늘 분석/실험 중 얻은 인사이트나 발견한 문제점은?

EDA, 모델 실험 중 유의미한 점이나 오류가 있었다면 자유롭게 작성하세요.

👉 답변:

- 과유불급: 무조건적인 필터링이나 무거운 리랭킹 모델보다, 적절한 청크 사이즈와 기본 검색(Hybrid)의 조화가 가성비(속도/정확도) 면에서 최고임을 확인
- Regex의 힘: LLM에게 모든 걸 말하기보다, 정규식으로 중요 데이터(금액)를 보호하는 것이 Hallucination 방지에 효과적

[6] 일정 지면이나 협업 중 어려웠던 점이 있다면?

자기 업무 외에도 전체 일정이나 팀 내 협업에서 생긴 문제를 공유해 주세요.

👉 답변:

- 리랭킹 모델이 CPU 환경에서 너무 느려, 실서비스 적용이 불가능함을 확인하고 롤백함

[7] 오늘 발표 준비나 커뮤니케이션에서 기여한 부분은?

슬라이드 제작, 발표 연습, 질문 정리 등 발표와 관련된 활동을 썼다면 기록하세요.

👉 답변:

-

[8] 내일 목표 / 할 일

구체적인 개인 업무나 팀 목표 기반 계획을 간단히 적어주세요.

👉 답변:

- 중간 발표 및 팀별 역할 분담 및 고도화