Daily 협업일지(02/23)

[1] 오늘 날짜 / 이름 / 팀명

- 날짜: 2026.02.23
- 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:

```
테넌트 데이터 격리 구현:
  - DocumentStore 경로를 data/{tenant_id}/ 구조로 변경하여 물리적 파일 격리
  - Vector DB 인덱싱 시 tenant_id, user_id, group_id 메타데이터 주입 로직 구현

검색 접근 제어(ACL) 적용:
  - HybridRetriever 리팩토링: 레거시 코드를 제거하고 검색 시 ACL 필터($and 조건)가 강제 적용되
도록 수정
  - 타 테넌트 데이터 검색 원천 차단 로직(Query-time filtering) 적용

파일 업로드 보안 정책 수립:
  - 확장자 화이트리스트(.pdf, .hwp 등) 및 블랙리스트(.exe, .sh) 검사 로직 추가
  - 파일 헤더(Magic Number) 검사를 통한 확장자 위변조 탐지 기능 구현
  - 파일 크기 제한(50MB) 설정
```

---

[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:

```
(55%) 데이터 저장 및 검색 시 격리 구조는 완성되었으나, 실제 데이터 내용에 대한 보안(PII 마스킹) 및
LLM 연동 시의 가드레일은 아직 미구현 상태임
```

---

[4] 오늘 협업 중 제안하거나 피드백한 내용이 있다면?

> 오늘 회의나 메시지에서 당신이 제안하거나 팀에 피드백한 내용은 무엇인가요?

✍️ 답변:

```
역할 분담 의견 전달
```

---

[5] 오늘 분석/실험 중 얻은 인사이트나 발견한 문제점은?

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

✍️ 답변:

```
- 단순히 파일 확장자만 검사해서는 보안이 취약함을 확인. 파일의 첫 몇 바이트(Magic Number)를 읽어 실제 파일 형식을 검증하는 것이 필수적임
- 기존 검색 로직에서 테넌트 필터가 누락될 경우 데이터 유출 위험이 있어, Retriever 초기화 단계에서 필터를 강제하는 방식이 안전함
```

---

[6] 일정 지연이나 협업 중 어려웠던 점이 있다면?

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

✍️ 답변:

```
- HybridRetriever 내부에 사용되지 않는 레거시 코드(_fetch_all_docs)가 많아 이를 정리하고 ACL 로직을 통합하는 데 시간이 소요됨
```

---

[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:

```
- 개인정보(PII) 보호를 위한 필터링 로직 구현
- RAGChain 입력/출력 가드레일(Prompt Injection 방어 등) 적용
```