Daily 협업일지(02/24)

[1] 오늘 날짜 / 이름 / 팀명

- 날짜: 2026.02.24
- 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:

```
PII(개인정보) 필터링 시스템 구축:
  - pii_filter.py로 마스킹 로직 일원화 (주민번호, 전화번호, 이메일, IP 등)
  - 문서 파싱 단계(_mask_pii)와 사용자 질문 입력 단계(Input Rail)에 마스킹 적용
RAGChain 보안 가드레일 적용:
  - Input Rail: 사용자 질문 길이 제한 및 프롬프트 인젝션 패턴 스캔
  - Prompt Engineering: XML 태그(<context>, <hints>) 구조로 프롬프트를 개편하여 외부 지시 무시(Meta-Prompting) 적용
  - Output Rail: LLM 답변 내 민감정보(금융정보 등) 유출 시 차단 로직 구현
안정성 및 정합성 개선:
  - 파싱 타임아웃(300초) 적용으로 DoS 공격 방지
  - _sanitize_text 로직 강화로 깨진 문자(Mojibake) 및 의미 없는 짧은 청크 필터링
```

---

[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:

```
(65%) 입출력 단계의 보안 장치는 마련되었으나, 외부 도구 실행 시의 네트워크 보안(SSRF)과 보안 감사 로그 시스템이 아직 연동되지 않음
```

---

[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:

```
- LLM은 문맥에 따라 시스템 프롬프트를 무시할 수 있음. 이를 방지하기 위해 사용자 입력을 XML 태그로 감싸고 "이전 지시를 무시하라"는 식의 공격을 텍스트로 취급하도록 강제하는 것이 효과적임
- hwp5txt 등 외부 라이브러리 부재 시 발생하는 바이너리 노이즈가 검색 품질을 저하시키므로 정제 로직이 중요함
```

---

[6] 일정 지연이나 협업 중 어려웠던 점이 있다면?

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

✍️ 답변:

```
- 주민등록번호 정규식 패턴이 너무 단순하면 일반 숫자열도 마스킹하는 오탐(False Positive)이 발생하여, 정규식 최적화 및 검증에 노력이 필요했음
```

---

[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:

```
- SSRF 방지를 위한 Tool Execution Gate 구현
- 보안/감사 로그 시스템(JSON 구조화) 구축
- 통합 테스트(test_isolation_security.py) 작성 및 검증
```