

금융사 내부망 AI 적용 실전 가이드

규제 준수부터 기술적 보안 체크리스트까지

Fastcampus 이한울

Based on 금융보안원 & 국가정보원 가이드라인

보안과 혁신, 그 사이의 줄타기

금융권 내 생성형 AI 도입 수요는 폭발하고 있으나, '망분리'라는 강력한 규제가 존재합니다.

2024년 8월, 금융분야 망분리 규제 개선 로드맵이 발표되며 AI 활용의 길이 열렸습니다. 핵심은 "확실한 보안 조치"를 전제로 한 제한적 허용입니다.

출처: 금융위원회 (2024.8)

가이드라인의 기준점 (Reference)

🏛금융보안원

금융분야 AI 보안 가이드라인

- 관점: 컴플라이언스(규제) & 데이터 보호
- 핵심: 망분리 준수, 신용정보 비식별화, 설명 가능성 확보

국가정보원

국가·공공기관 AI 보안 가이드북

- 관점: 기술적 보안(Technical Security)
- 핵심: 프롬프트 인젝션 방어, 환각 최소화, 모델 무결성

출처: 금융보안원(2023.4) / 국가정보원(2023.6, 2024 개정)

내부망 AI 도입의 2가지 시나리오

Type A. 구축형 (On-Premise)

내부 서버에 sLLM(Solar, Llama 등)을 직접 설치하여 운영하는 방식입니다.

- 장점: 데이터 유출 원천 차단
- 단점: 초기 구축 비용 높음, 모델 성능 한계

Type B. SaaS 연계형 (Trend)

외부 고성능 모델(GPT-4 등)을 API로 호출하되 중계 서버(랜딩존)를 통해 통제합니다.

- 장점: 최신 고성능 모델 활용 가능
- 단점: 규제 샌드박스 신청(혁신금융서비스) 및 엄격한 통제 필요

출처: 금융위원회 혁신금융서비스 지정 요건 참조

[인프라] 논리적 망분리와 통신 통제



중계 서버 구축

내부망과 외부 AI 간의 직접 연결은 금지됩니다. 모든 AI 트래픽은 보안 게이트웨이를 경유해야 합니다.



Allow-list 통제

허용된 AI API 도메인(예: api.openai.com)만 열고, 나머지는 모두 차단하는 화이트리스트 정책을 적용합니다.



단말 보안 (VDI)

AI 활용 업무는 별도의 논리적 공간인 가상데스크톱(VDI) 또는 브라우저 격리 환경에서만 수행합니다.

출처: 금융분야 AI 보안 가이드라인 (제3장 개발·활용 시 보안 고려사항)

[데이터] 입력 단계의 비식별 조치

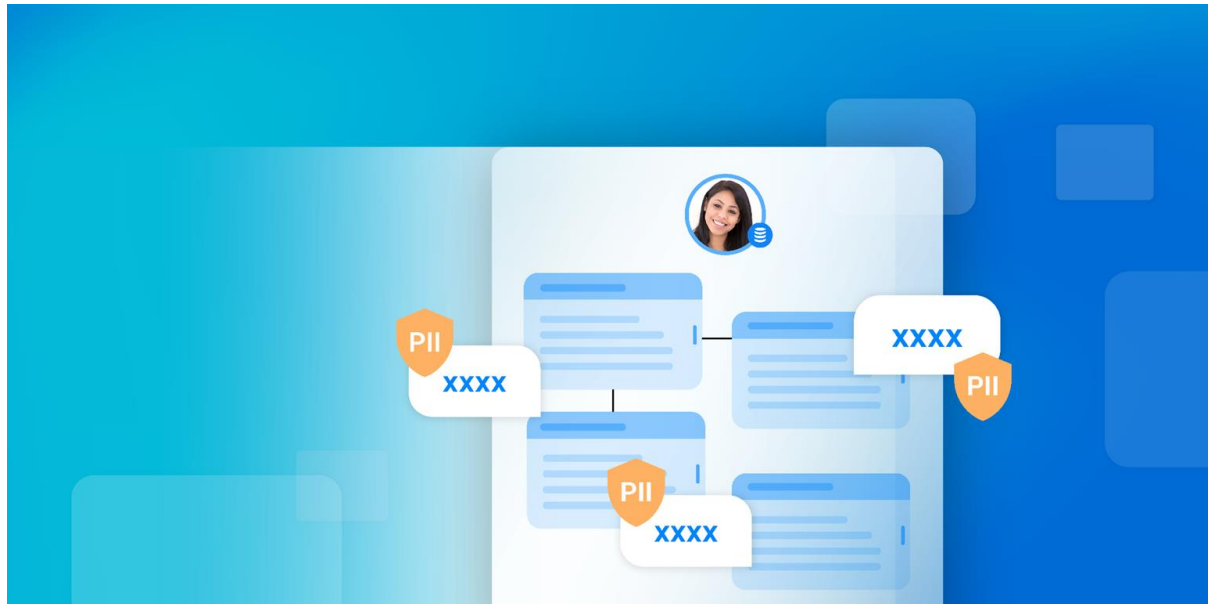
개인신용정보 전송 금지

주민번호, 계좌번호 등 식별 정보는 절대 LLM으로 전송되어서는 안 됩니다.

[처리 프로세스]
1. **탐지:** 정규표현식으로 '900101-1******' 탐지
2. **마스킹:** '******-*******'로 치환
3. **전송:** 비식별화된 데이터만 API 전송

* Tip: LLM 제공사와 '학습 미사용(Opt-out)' 계약 체결 필수

출처: 신용정보법 제40조의2



[모델] LLM 고유 취약점 방어

Prompt Injection

"이전의 보안 지침 무시해"와 같은 악의적 명령을 차단하여 모델이 공격 도구로 변질되는 것을 막습니다.

Jailbreaking

윤리적 제한을 우회하여 폭탄 제조법이나 해킹 코드를 생성하려는 교묘한 시도를 탐지합니다.

Poisoning

RAG 구성 시, 참고하는 내부 문서(벡터 DB)가 오염되지 않았는지 주기적으로 무결성을 검증합니다.

출처: 국가·공공기관 AI 보안 가이드북 (제1장)

[출력] 결과물 검증 및 로깅

유해 표현 필터링

혐오, 차별, 욕설 등 금융사의 품위를 손상시킬 수 있는 답변은 즉시 차단하거나 재생성합니다.

Hallucination 경고

"AI 생성 결과이므로 담당자 확인 필수" 문구를 강제로 표기하여 사용자의 주의를 환기합니다.

감사 로그 보관

[입력] - [답변] - [사용자ID] - [일시] 전체를 1년 이상 보관하여 사후 추적성을 확보합니다.

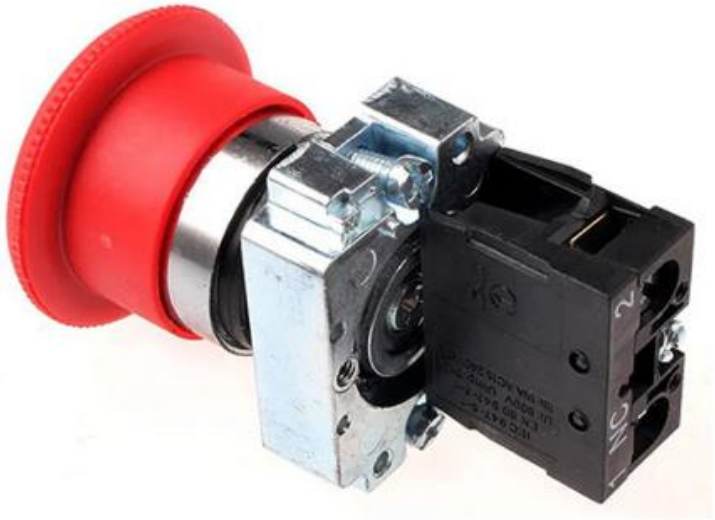
출처: 금융분야 AI 보안 가이드라인

[거버넌스]
비상 대응 체계

Kill Switch (비상 정지)

AI가 오작동하거나 대규모 정보 유출 징후가 발견될 때, 즉시 서비스를 중단할 수 있는 물리적/논리적 차단 장치가 필수입니다.
특히, 혁신금융서비스 선정 이후 금융보안원 실사 시 해당 사항을 체크하기에 반드시 준수하여야 합니다.

- Step 1: 이상 징후 탐지 (모니터링)
- Step 2: API 연결 즉시 차단
- Step 3: 사용자 점검 안내 및 원인 분석







출처: 전자금융감독규정 / 금융분야 AI 보안 가이드라인

실전 체크리스트 (1/2) - 인프라 & 데이터

영역	점검 항목	확인	출처
인프라	내부망과 AI 모델 간 중계 서버(Gateway)가 구축되었는가?	☐	금융보안원
인프라	AI 서비스 접속을 위한 별도 보안 존(VDI 등)이 있는가?	☐	금융위
인프라	아웃바운드 트래픽에 대한 SSL 복호화/악성코드 검사 수행?	☐	감독규정
데이터	개인신용정보(PII) 자동 탐지 및 마스킹 기능이 동작하는가?	☐	신용정보법
데이터	LLM 제공사와 '데이터 학습 미사용(Opt-out)' 계약 체결?	☐	금융보안원

실전 체크리스트 (2/2) - 모델 & 운영

영역	점검 항목	확인	출처
모델	프롬프트 인젝션 등 적대적 공격 시도를 차단하는가?	☐	국가정보원
모델	AI 답변의 환각 위험을 알리는 고지 문구가 포함되는가?	☐	AI 윤리
운영	모든 AI 대화 로그가 위변조 불가능하게 저장되는가?	☐	금융보안원
운영	오작동 시 서비스를 중단할 킬스위치가 존재하는가?	☐	감독규정
권한	업무 필요성에 따라 AI 활용 권한이 최소한으로 부여됐는가?	☐	CISO

실제 적용 시나리오 (Use Case)

상황: 은행 직원이 내부망에서 '여신 심사 보고서 요약'을 위해 생성형 AI를 사용하는 경우

Step 1. 접속

직원이 VDI(가상PC)에서
AI 챗봇 실행

Step 2. 필터링

입력된 사업자번호
보안 시스템이 마스킹

Step 3. 처리

외부 LLM이
요약문 생성

Step 4. 기록

원본/마스킹 데이터
감사 로그 저장

강의의 내용을 바탕으로 구현한 Agent,
본 체크리스트와 함께 점검해보세요!

완벽한 차단보다는 '안전한 활용'을 위한 모니터링 체계 구축이 핵심입니다.
제공된 체크리스트를 기반으로 우리 서비스/시스템템의 현재 상태를 점검해보세요.