Headline

Shadow AI : 제조업 기밀 보호를 위한 탐지·통제·거버넌스 전략

사이버반도체사업그룹 반도체보안전략팀 송원준 수석

1. 제조업 환경에서 Shadow AI 가 제기하는 전략적 보안 위협

2023 년 이후, 대규모 언어모델(LLM) 기반 생성형 인공지능(Generative AI)의 상용화는 산업 전반에 걸쳐 업무 자동화, 엔지니어링 최적화, 지식 정제 등 다방면의 혁신을 가속화 시키고 있다. 특히 제조업에서는 제품 설계, 품질 관리, 공정 제어, 생산성 향상 등 다양한 기능 영역에서 AI 의 활용성이 빠르게 부각되고 있다. 그러나 이러한 기술적 진보는 새로운 보안 지형도 형성하고 있다. 그 중심에는 Shadow AI 라는 고위험 사용 패턴이 존재한다.

Shadow AI 는 조직의 공식 승인 없이, 개인 또는 부서 단위에서 AI 서비스를 비공식적으로 사용하는 행위를 의미한다. 특히 제조업과 같은 지식집약적 산업에서는 산업 기밀, 생산 레시피, 라우팅 정보, 설계 도면, 설비 로직 등 핵심 자산이 무단으로 외부에 전송될 경우, 중대한 보안 위협으로 작용할 수 있다. 이는 기존 보안 인프라(DLP, EDR, CASB 등)의 탐지 범위를 우회하고, 조직 내 보안 가시성(Visibility)을 현저히 저하시키는 결과로 이어질 수 있다.

예컨대, R&D 조직의 엔지니어가 LLM 기반 챗봇에 CAD 설계 내용을 설명하며 기술적 피드백을 요청하거나, 제조기술팀이 내부 공정 데이터를 기반으로 레시피 최적화 질문을 입력하는 사례다. 이 때 문제점은 입력값 대부분이 HTTPS 기반의 비정형 API 트래픽으로 전송돼, 조직 내부의 감사 및 접근 통제 범위를 벗어난다는 점이다. 해당 프롬프트가 외부 AI 서비스의 학습 데이터로 활용되거나 장기 저장될 경우, 향후 동일 산업군 AI 학습에 그대로 재사용될 위험이 있다.

또한 Shadow AI 는 정보 유출의 문제를 넘어 규제 불이행, 법적 분쟁, 산업 보호법 위반 등의 2 차적 리스크의 가능성을 초래한다. 특히 산업기술보호법, GDPR, ITAR 등 국내외 보안·기밀 관련 규제 체계에서는 기밀의 관리성 상실 자체를 보호 요건 박탈의 사유로 보기 때문에, 단 1 회의 외부 전송만으로도 특허 자산의 보호 지위를 상실할 수 있다.

따라서 Shadow AI 는 단순한 '사용자 행위 이슈'가 아닌, 산업지식 기반 보안 전략의 구조적 취약지점으로 간주해야 하며, 사전탐지, 행위제어, 프롬프트 단위의 리스크 평가체계 등 통합적인 거버넌스 모델의 수립이 절실하다.

본 Insight 에서는 Shadow AI 의 작동 구조와 제조업 특화 위협 시나리오를 고찰하고, 기술적 탐지 방안과 정책적 대응 전략을 포괄하는 실효성 있는 보안 모델을 제안하고자 한다. 더불어, 글로벌 규제 흐름과 대응 가이드라인을 기반으로, 실무 중심의 운영체계 수립을 위한 참조 프레임워크도 함께 제시한다.

2. Shadow AI 개념과 위협 모델 분석

2.1 Shadow AI 정의 및 행위 특성

Shadow AI 는 조직 내 보안·IT 관리 체계를 거치지 않고 개인 또는 부서 단위에서 비인가된 생성형 AI 도구(LLM, Vision AI, AutoML 등)를 활용하는 사용 행태를 의미한다. 이 과정에서 사용자는 보안이나 데이터 처리 규정을 충분히 인지하지 못한 채, 다음과 같은 행위를 무심코 수행하게 된다.

- 내부 문서, 도면, 공정정보 등을 외부 AI 에 프롬프트 형태로 직접 입력
- 외부 AI 가 생성한 코드·문서를 검증 없이 업무 시스템에 통합
- 기업 외부 서버에 민감 데이터를 자동 저장/캐싱되는 구조를 이해하지 못함

이러한 Shadow AI 의 사용은 표면적으로는 업무 생산성 향상 및 개인 업무 보조를 목적으로 하나, 보안적 관점에서는 비인가 채널을 통한 고위험 데이터 전송 행위로 간주된다.

2.2 Shadow AI 위협 모델 분류 (제조업 중심)

다음은 Shadow AI 의 제조업 환경 내 위협 유형을 행위/위험/영향/예시 중심으로 설명해본다.

① 설계/기술 문서 유출

요소	설명
행위	CAD 도면 설명, 제품 설계 구조 요약 요청 등
위험	설계 노하우, 부품 규격, 포지셔닝 정보가 LLM 에 노출됨
영향	유사 제품 모방, 경쟁사 기술 학습에 악용 가능
예시	“이 구조가 전체적으로 문제가 없나요?”라는 질문에 도면 전체 구조 포함

② 제조 레시피 및 공정 조건 노출

요소	설명
행위	AI 에게 공정 조건 조정, 수율 개선 방식 문의
위험	생산 온도, 속도, 재료 비율 등의 내부 변수 전송
영향	품질 경쟁력 상실, OEM/ODM 경쟁자에 정보 이전
예시	“이 원료 비율에서 결함이 생기는 원인을 분석해줘” 등

③ 품질 데이터 기반 민감 정보 유출

요소	설명
행위	결함 발생 DB, 검사 이미지, 불량 유형 등을 AI 에 입력
위험	제품 결함 데이터와 구조적 약점 정보가 외부에 학습됨
영향	취약 제품 식별 → 악의적 리콜 유도 가능
예시	“이 사진은 왜 B 등급 불량이었는지 설명해줘” 등

④ 자동화 코드/시퀀스 유출 및 오염

요소	설명
행위	PLC 제어 코드, 시퀀스 논리 등을 AI 에 진단 요청
위험	코드 로직 유출 또는 AI 생성 코드에 보안 미흡 로직 존재
영향	설비 정지, 안전사고 유발, OT 공격 확산 가능
예시	AI 가 생성한 코드에 인증 절차 누락 → 외부 명령 삽입 가능

⑤ 사용자 인증정보/시스템 정보 간접 유출

요소	설명
행위	LLM 에 개발 코드나 API 예시 제공
위험	토큰, 계정명, 시스템 포트 구조 노출
영향	공격자에게 내부 API 맵 제공하는 결과 초래
예시	“이 API 로 품질 관리 시스템 연동하는 방법 알려줘” 등

⑥ 학습 재사용에 의한 정보 역추출

요소	설명
행위	LLM 에 반복적으로 내부 정보 포함한 프롬프트 입력
위험	향후 타 사용자 프롬프트에 의해 해당 정보가 생성 응답으로 재노출
영향	정보가 공개된 것과 동일한 효과, 기밀성 상실
예시	“지난번에 입력한 공정 레시피를 다시 보여줘”와 유사한 요청

⑦ 규제 및 컴플라이언스 위반

요소	설명
행위	AI 가 위치한 국외 서버로 기밀 전송 (GDPR, ITAR, 산업기술보호법 등 위반 소지)
위험	규제 미준수, 법적 소송, 인증 취소 위험
영향	기업 이미지 훼손 및 대외 계약 상실
예시	국방 부품 제조사의 설계 도면이 OpenAI 에 전송됨

위 사례 등으로 볼 때 Shadow AI 는 다음과 같은 복합적인 성격을 지닌다.

- Low-intent, High-impact : 사용자 의도는 선의일 수 있으나, 결과는 치명적일 수 있음
- Technical Undetectability : 프롬프트 내 정보는 구조적으로 식별·분류하기 어려움
- Governance 외부성 : 전통적인 정보보호 관리체계 밖에서 행위 발생
- Attack Surface 확장자 : 외부 API/모델 호출이 사실상 새로운 경계면이 됨

2.3 문제점 도출

제조업 조직에서 Shadow AI 는 단순한 직원의 부주의가 아니라, 보안 거버넌스 체계의 구조적 결핍을 드러내는 경고 신호다. 공격자 없이도 내부에서 자산이 탈취되고, 누출된 정보를 되찾을 수 없다는 점에서 이는 비가역적(irreversible) 보안 사고로 간주해야 한다.

따라서 Shadow AI 에 대한 탐지, 예방, 억제 및 사고 대응은 선택이 아닌 디지털 제조 시대의 필수적 보안 전략 항목으로 자리잡아야 한다.

3. 기술적 대응 전략 : 탐지, 통제, 차단 체계 (Detection, Control, Mitigation)

3.1 탐지 전략 (Detection)

Shadow AI 탐지에는 가시성 확보가 핵심이다. HTTPS 기반 AI API 호출, 동적 도메인, 비정형 프롬프트를 효과적으로 식별하기 위해 아래 대응체계를 권장한다.

① 설계/기술 문서 유출

- AI 플랫폼 호출은 SNI(Server Name Indication), User-Agent, DNS(Domain Name System) 요청 패턴으로 식별 가능
- 고급 CASB(Cloud Access Security Broker) 솔루션을 통해 외부 LLM(Large Language Model) API 호출에 대한 실시간 탐지 및 정책 제어 적용
- 단, 기존 CASB 는 Shadow AI 용 탐지가 미비하므로, AI 흐름 포착이 가능한 DSPM(Data Security Posture Management) 기능이 필요

② 프롬프트 컨텐츠 기반 이상 요청 탐지

- “설계”, “기밀”, “공정”, “매출” 등 고위험 키워드 검출을 적용 등 민감한 데이터가 포함 정책 적용
- 자연어 비정형 요청에 대한 AI-aware DLP(Data Loss Prevention)나 프롬프트 인젝션 탐지 규칙 적용

③ Shadow AI 도구 인텔리전스 및 목록화

- 비공식 ChatGPT, Gemini 외에도 Perplexity, DeepSeek 같은 신규 툴 탐지 강화
- DNS, IP, User-Agent 기반 블랙리스트 및 APT(Advanced Persistent Threat) 예방 수준의 탐지 베이스 구축

3.2 통제 전략 (Control)

탐지 이후에는 엄격한 접근 제어 정책과 내부 대체 모델 제공을 통해 Shadow AI 사용을 구조적으로 관리

① RBAC 기반 AI 사용 권한 제한

- RBAC(Role-Based Access Control) 방식으로 부서별 사용 권한 차등 적용
- 설계/R&D 직무는 대외 AI 사용을 차단하고, 마케팅 등은 안전한 요약 기능만 허용
- ‘최소권한’ 원칙에 준하는 정책 수립 및 관리 자동화

② 프록시 차단 및 AI SaaS 블랙리스트

- SaaS(Software as a Service) 형태의 AI 서비스 접근을 HTTPS 프록시에서 차단
- 신규 AI 서비스 자동 탐지 기능으로 가시성 및 통제 범위 확대

③ 사내 프라이빗 LLM 환경 운영

- Azure OpenAI Private Endpoint 등 내부 모델로 전환 유도
- 외부 접속을 사전 통제하여 인프라 경계 내에서 보안 거버넌스 체계 유지

④ AI-aware DLP 기반 민감정보 실시간 필터링

- PII(Personally Identifiable Information), IP(Intellectual Property), mCAD(manufacturing CAD data) 등 민감 정보 탐지
- AI 특화 DLP 제품 등 활용 등

3.3 차단·억제 전략 (Mitigation)

차단 단계는 Zero Trust 기반 데이터 흐름 통제, 사용자 인식 제어 강화, 그리고 사후 대응 체계 수립 포함

① Zero Trust 기반 프롬프트 경로 통제

- ZTNA(Zero Trust Network Access) 기반 인증·허용 방식으로 외부 LLM 요청 경로 제어
- 내부망 → 인터넷 출구 → 외부 AI 서비스까지 '데이터 전송' 단위로 분석 및 제한

② Security Nudging: 경고 UI 및 정책 기반 인식

- KPI(Key Performance Indicator): 부서별 AI 사용 건수, 탐지 횟수, 정책 위반 상승 추이
- 정기 보고로 조직 내 인식 강화 및 책임 공유 체계 확립

③ 탐지 지표 KPI 기반 모니터링/경영 보고

- RBAC(Role-Based Access Control) 방식으로 부서별 사용 권한 차등 적용
- 설계/R&D 직무는 대외 AI 사용을 차단하고, 마케팅 등은 안전한 요약 기능만 허용
- '최소권한' 원칙에 준하는 정책 수립 및 관리 자동화

④ 사고 대응 준비 – 프롬프트 저장, 백업, 분석

- SOC(Security Operation Center) 내 Prompt/Response 로그 분석을 위한 연동
- 사고 발생 시 누출 범위, API 사용 기록, 사용자 ID 추적 등 포함

4. 거버넌스 및 정책 중심의 조직 대응 체계

Shadow AI 의 위협은 기술적 대응만으로는 불완전하다. 사내 구성원의 무지 또는 관행적 사용, 정책 부재 등이 복합적으로 얽혀 있기 때문에, 조직 전반의 보안 의사결정 체계(거버넌스)를 통한 전사적 관리체계 구축이 필요하다.

4.1 Shadow AI 정책 수립 프레임워크

① AI Usage Policy (생성형 AI 사용 정책)

- 모든 구성원이 명확히 이해할 수 있도록 Shadow AI 금지/허용 기준을 문서화
- 필수 포함 항목으로 어떤 AI 도구를 사용할 수 있는가? (허용/제한/금지 목록) / 어떤 데이터가 입력되어서는 안 되는가? (PII, CAD, 설계서, 소스코드 등 예시 제공) / 위반 시의 제재 수준 및 예외 승인 절차 등을 수립

② AI Risk Classification (업무 기반 리스크 등급화)

- 부서/직무/업무 프로세스별로 AI 사용 위험도 등급을 부여하여 관리
- 등급별로 차등적 승인/통제 체계를 수립 (RBAC + AI Usage Scope Matrix)

③ AI 사용 승인 프로세스

- 신규 AI 도구 사용 요청 시, 보안팀 또는 AI 거버넌스 위원회의 사전 심의 필요
- API 통신, 브라우저 확장 기능, 내부망 접속 요청에 대한 기술적 심사 프로세스 운영
- 예외 사용 시 관리자 승인 절차 필수화

4.2 교육 및 조직 인식 제고 전략

Shadow AI 의 80% 이상은 '보안 인식을 하지 못한 채 무심코 사용'한 것으로 분석된다. 따라서 단순한 차단이 아닌 전사 차원의 인식 개선 프로그램이 필수적이다.

① AI 보안 인식 교육 프로그램

- 정기 교육 (반기 1 회 이상) 및 전파용 콘텐츠 제작

② 프롬프트 작성 가이드 배포

- '절대 입력 금지 문장 예시' 중심의 현장 실무형 가이드 문서화

③ 보안 책임제 제도 운영

- 각 부서 내 보안 리더 지정 → AI 사용 모니터링, 캠페인 진행, 이슈 리포트
- 보안팀과의 소통 창구 역할 수행
- 보안 이슈 공유 채널 상시 운영

4.3 AI 거버넌스 조직 모델

① AI 사용 통제 전담 TF (AI Risk Control Taskforce)

- 구성 : 보안팀(CISO), 정보팀(CIO), 법무, 내부통제, 각 실무 부서 대표
- 역할 : 사내 AI 도구 화이트리스트 관리 / Shadow AI 탐지 리포트 주간 공유 / 신규 정책 및 위반 대응 협의

② AI Risk Steering Committee

- 경영진 보고 체계로 기능, 리스크 등급 상승 시 빠른 의사결정 구조
- KPI : Shadow AI 탐지율, 위반 건수, 보안 가이드 수강률 등

③ 감사 및 내부통제 연계 안 책임제 제도 운영

- AI 사용 행위에 대한 내부 감사 항목 도입
- 보안 로그, 프롬프트 사용 이력, 외부 접속 이력 등 정기 리포트화

4.4 산업 기준 및 컴플라이언스 대응

제조업 내 보안 거버넌스를 강화하는 동시에, 국내외 법적·산업 기준과의 연계도 필수적이다.

규제 기준	적용 항목	대응 방안
ISO/IEC 42001	생성형 AI 운영 시 거버넌스 체계 수립	AI 위험 등급 분류, 위원회 운영
NIST AI RMF	AI 리스크 관리 프레임워크	Shadow AI 리스크 대응 포함
KISA AI 보안 가이드	국내 산업 기반 AI 보안 권고안	AI 프롬프트 필터링, 민감정보 탐지 포함
GDPR/개인정보보호법	자동화 처리 및 민감정보 유출	AI 입력 사전 탐지 및 마스킹 도입

5. 결론 및 대응 로드맵 제안

5.1 결론

Shadow AI 는 단순한 IT 통제를 넘어, 조직의 기밀정보와 경쟁력을 직접 위협하는 신종 보안 리스크로 부상했다. 특히 제조업 기반 기업의 경우, 설계도면, 공정 노하우, 원가 자료와 같은 산업기밀(Intellectual Property)이 LLM(Large Language Model) 기반 AI 도구를 통해 외부로 유출될 가능성이 높아지는 상황이다.

이 Insight 에서는 이러한 Shadow AI 위협에 대해 기술적, 정책적, 거버넌스 차원의 대응 전략을 통합적으로 제시하였다. 대응의 핵심은 다음과 같다.

- 탐지 : AI-aware DLP, CASB, DSPM 등을 활용한 AI 사용 가시성 확보
- 통제 : AI 사용 정책 수립, 프록시 차단, 역할기반 권한관리(RBAC)
- 차단 : Zero Trust 기반 경로 통제, 경고 UI 및 사후대응 체계
- 거버넌스 : 전사 정책 수립, 부서 리스크 등급화, 교육 및 내부감사 체계화

이러한 대응은 단발성 정책이 아닌, 조직문화와 보안 거버넌스 체계 내재화를 목표로 추진되어야 한다.

5.2 대응 로드맵 제안

다음은 Shadow AI 대응을 위한 3 단계 실행 로드맵이다.

[1 단계 : 가시화 및 인식 재고]

- 목표 : Shadow AI 의 존재와 위협 파악
- 주요 조치 사항
  → Shadow AI 의 존재와 위협 파악
  → 사내 Shadow AI 사용 현황 조사
  → Shadow AI 사고 사례 교육자료 배포
  → 부서별 민감정보 분류 체계 수립

[2 단계 : 정책 및 기술적 통제 수립]

- 목표 : Shadow AI 사용 통제 및 최소화
- 주요 조치 사항
  → AI 사용 정책 수립 및 공지
  → RBAC 기반 AI 사용 권한 설정
  → 민감정보 DLP 정책 적용 및 테스트
  → 프록시 기반 LLM 접속 차단 체계 구축

[3 단계 : 조직 내재화 및 거버넌스]

- 목표 : 대응 체계의 조직 내 정착
- 주요 조치 사항
  → AI 거버넌스 위원회 및 보안 책임제 제도 운영
  → AI 사용 모니터링 및 정기 리포트화
  → AI 보안 인식 정기 교육
  → AI 관련 컴플라이언스 대응체계 정비 (ISO, NIST 등)

5.3 향후 과제 및 제언

- 사내 LLM 도입 검토 : 보안 위험 없이 생성형 AI 를 활용할 수 있는 프라이빗 LLM 환경을 구축하고, 이를 통해 외부 Shadow AI 사용을 줄일 수 있음
- AI 특화 보안 솔루션 확대 도입 : 기존 보안 장비만으로는 LLM 의 비정형성 탐지가 어렵기 때문에, AI-aware DLP, Prompt 보안 필터링, 데이터 흐름 탐지 기술 도입이 필요
- 보안팀의 역할 전환 : Shadow AI 대응은 단순 감시가 아닌 'AI 활용 가이드 및 보안 조언자'로서의 역할 확대가 중요함
- 법·규제 대응 체계 고도화 : 생성형 AI 관련 국내외 규제가 빠르게 정비되고 있으므로, 대응 전담조직 및 감사 항목 통합이 요구됨

Shadow AI 는 단순한 기술 도입의 문제가 아닌, 기밀 보호와 조직의 생존을 좌우하는 보안 과제다. 기술, 정책, 문화가 함께 작동하는 다층적 대응이 필요한 시점이다.

Shadow AI 로부터 조직의 산업기밀을 지키기 위한 보안 정책 수립이 필요하다면, SK 쉴더스가 보유하고 있는 기술 및 정책 노하우를 기반으로 AI 보안 거버넌스를 시작하길 바란다.

■ 참고문헌

[1] Structured, Shadow AI – The Hidden Threat to Governance & Compliance, 25.04

[2] Inteleca, Shadow AI in the Workplace: The Hidden Security and Compliance Risks, 25.03

[3] CIODIVE, AI-generated code leads to security issues for most businesses, 24.01

[4] Nightfall AI, The Nightfall Approach: 5 Ways Our Shadow AI Coverage Differs from Generic DLP, 25.07

[5] NIST AI RISK MANAGEMENT FRAMERK (AI RMF), 23.01

■ 참고 자료

[1] Paloalto, What Is Shadow AI? How It Happens and What to Do About It (Cyberpedia)

[2] ISO/IEC 42001:2023, Information technology — Artificial intelligence — Management system

[3] 행정안전부, 공공기관 AI 보안 가이드라인, 23.10

[4] NIPA, 산업별 생성형 AI 활용과 보안 위협 보고서, 2024

[5] SK쉴더스, EQST Insight 블로그 시리즈 (2023~2024)