AI GOVERNANCE · 생성형 AI 도입

기업용 Claude 도입 절차와 보안 체크리스트

윤재웅 · www.kpredict.co.kr · 2026.11 기준 공개 자료로 정리

1. 왜 "막기"가 아니라 "통제된 허용"인가

- 회사가 AI를 막으면 직원은 개인 계정으로 씁니다. 그 순간 회사는 **무엇이, 누구에 의해, 어디로** 나갔는지 알 수 없게 됩니다(섀도 AI).
- 기업용 플랜은 개인 플랜과 달리 **계정·데이터·로그를 회사가 관리**할 수 있습니다. 보안팀이 볼 수 있는 곳으로 사용을 옮기는 것이 첫 번째 보안 대책입니다.
- 목표는 세 가지입니다: ① **회사 데이터가 학습에 쓰이지 않게,** ② **누가 무엇을 했는지 추적되게,** ③ **퇴사·권한 변경이 즉시 반영되게.**

2. 도입 경로 고르기

경로	적합한 경우	보안 관점 포인트
Claude Team	소규모 부서 시범, 빠른 시작	상업 약관 적용(기본적으로 학습 미사용), SSO·도메인 인증. SCIM·감사 로그 등 고급 관리 기능은 Enterprise에서 제공
Claude Enterprise	전사 도입, 규제·감사 대응 필요	SSO(SAML·OIDC)와 도메인 캡처, SCIM 자동 계정 관리, 역할 기반 권한, 감사 로그, Compliance API, 보존 기간 설정, IP 허용 목록 등
Claude API (Claude Platform)	사내 시스템·업무 앱에 AI 기능을 내장	상업 약관, 조건 충족 시 ZDR(무보존) 계약 가능. 키 관리와 호출 로그는 회사 책임
클라우드 경유 Amazon Bedrock · Google Cloud Vertex AI · Microsoft Foundry	이미 쓰는 클라우드의 계약·네트워크·권한 체계 안에서 쓰고 싶을 때	기존 클라우드 IAM·네트워크 통제·과금 체계를 그대로 활용. 인증·데이터 처리 범위는 해당 클라우드 사업자 문서로 확인

> **현실적인 조합** · 직원 업무용은 **Enterprise**(또는 시범용 Team), 사내 시스템 연동은 **API 또는 회사가 쓰는 클라우드 경유**로 나누면 관리 책임이 분명해집니다.

3. 도입 7단계

1 **사용 정책과 데이터 등급 정하기**

입력 가능한 정보(공개·내부 일반)와 금지 정보(개인정보, 고객 계약 원문, 영업비밀, 인증 정보)를 등급표로 명시합니다. 허용 용도와 금지 용도, 결과물 검토 책임을 한 장짜리 가이드로 만듭니다.

2 **경로 선택과 계약 검토**

학습 미사용 조항, 데이터 보존 기간, 데이터 처리 계약(DPA), 하위 처리자, 필요 시 HIPAA BAA·ZDR 여부를 확인합니다. 개인정보가 오갈 수 있다면 개인정보보호법상 국외 이전 요건(고지·동의 등)을 법무·정보보호 부서와 검토합니다.

3 신원·접근 통제

사내 IdP(예: Entra ID, Okta)와 SSO를 연결하고 도메인 캡처로 회사 메일 계정의 개인 가입을 회사 조직으로 모읍니다. SCIM으로 입사·퇴사·부서 이동이 자동 반영되게 하고, 관리자·일반 사용자 역할을 나눕니다.

4 데이터 보호 설정

대화·프로젝트 보존 기간을 회사 기준에 맞게 설정하고, 연결할 수 있는 커넥터·MCP 서버를 허용 목록으로 제한합니다. 필요하면 IP 허용 목록, 고객 관리 암호화 키(CMEK)를 검토하고, 사내 DLP·민감도 레이블 정책과 충돌이 없는지 확인합니다.

5 모니터링 연결

감사 로그와 Compliance API를 SIEM·DLP 도구에 연결해 이상 사용을 탐지합니다. 개발 조직의 Claude Code 사용은 OpenTelemetry로 사용량·비용을 수집합니다.

6 시범 적용 후 확산

2~3개 부서, 고가치 업무로 4~8주 시범을 운영합니다. 시간 절감, 결과 품질, 보안 이벤트, 비용을 같이 측정하고 정책을 보완한 뒤 확산합니다. 에이전트는 실행 기록 추적과 기준표(루브릭) 기반 평가로 품질을 반복 측정합니다.

7 교육과 운영 체계

전 직원 AI 리터러시(좋은 요청, 사실 검증, 금지 정보), 부서별 활용 사례, AI 생성 코드 리뷰 기준을 교육합니다. 담당 조직(CoE), 문의 채널, 사고 대응 절차, 분기별 정책 점검을 정합니다.

4. 보안 우려별 해결 방안

우려	해결 방안
회사 데이터가 모델 학습에 쓰이는가	기업용(상업) 플랜과 API는 기본적으로 입력·출력을 학습에 쓰지 않습니다. 계약서 조항으로 재확인하고, 개인 계정 사용은 정책으로 금지
대화 기록이 계속 남는가	Enterprise에서 보존 기간을 직접 설정(최소 30일). API는 승인 시 ZDR로 응답 후 저장하지 않는 계약 가능
퇴사자·공유 계정	SSO 의무화 + SCIM 자동 해지, 공유 계정 금지, 역할 기반 권한
섀도 AI(개인 계정 사용)	도메인 캡처로 회사 메일 가입을 조직으로 흡수, 프록시·CASB로 비승인 AI 서비스 접속 파악
민감정보 입력	데이터 등급 가이드와 교육, DLP·민감도 레이블 연계, Compliance API 기반 사후 점검
누가 무엇을 했는지 모름	감사 로그·Compliance API를 SIEM에 연동, 정기 보고
커넥터·MCP 과권한	허용 목록 운영, 읽기 전용 우선, 사용자 권한 범위 안에서만 동작하도록 설정, 신규 연결은 보안 검토 후 승인
프롬프트 인젝션(외부 문서·웹을 통한 지시 주입)	외부 콘텐츠는 데이터로만 취급하도록 교육, 메일 발송·결재 등 외부로 나가는 행동은 사람 승인 단계 유지
AI가 만든 코드의 취약점	코드 리뷰 의무화, 시크릿 스캔·SCA·SAST를 배포 파이프라인에 적용, 배포 시 자산 등록
규제·인증	Trust Center에서 SOC 2 Type II, ISO 27001, ISO/IEC 42001 등 인증 범위를 경로(플랜·클라우드)별로 확인하고 보안 심사 자료로 활용
비용 증가	좌석·사용량 한도와 알림, 부서별 사용량 대시보드, 업무 성과와 함께 측정

5. 개발 조직: Claude Code 안전하게 쓰기

- **정책을 중앙에서 강제:** 관리형 설정(managed settings)으로 허용·차단 명령, 접근 가능한 파일 경로, 사용할 수 있는 MCP 서버를 회사 기준으로 고정합니다.
- **권한은 좁게:** 운영 환경 키와 고객 데이터가 없는 개발 환경·샌드박스에서 시작하고, 위험한 명령(삭제·배포·외부 전송)은 사람 확인을 거치게 합니다.
- **결과는 검증:** AI가 만든 변경은 diff 확인, 테스트, 동료 리뷰를 거쳐 반영합니다. 시크릿은 코드가 아닌 환경 변수·비밀 저장소로 관리합니다.
- **사용 현황 가시화:** OpenTelemetry로 사용량·비용을 수집해 팀별로 봅니다.

> 비개발 직원의 바이브 코딩 확산까지 고려한다면, 같은 사이트의 「바이브 코딩 환경의 공격 표면 관리 방안」과 함께 보시기를 권합니다.

6. 오픈 전 최종 체크리스트

- ☐ 사용 정책·데이터 등급 가이드를 공지하고 직원 확인을 받았다
- ☐ 계약서에서 학습 미사용, 보존 기간, DPA, 하위 처리자를 확인했다
- ☐ 개인정보 국외 이전 등 법적 요건을 법무·정보보호 부서와 검토했다
- ☐ SSO 의무화, 도메인 캡처, SCIM 자동 계정 관리를 켰다
- ☐ 보존 기간, 커넥터·MCP 허용 목록, IP 허용 목록을 설정했다
- ☐ 감사 로그·Compliance API를 SIEM·DLP에 연결했다
- ☐ 외부로 나가는 행동(발송·결재·배포)에 사람 승인 단계를 두었다
- ☐ 시범 부서의 성과 지표(시간·품질·보안 이벤트·비용)를 정했다
- ☐ 교육, 문의 채널, 사고 대응 절차, 정책 점검 주기를 정했다

> 이 자료는 2026년 11월 공개 자료를 바탕으로 한 일반 가이드입니다. 플랜별 기능, 가격, 인증 범위는 바뀔 수 있으므로 도입 전 Anthropic Trust Center, 공식 문서, 계약서로 반드시 확인하세요. 특정 회사의 내부 정책이나 법률 자문을 대신하지 않습니다.

참고 자료

- Claude Enterprise plan (Anthropic)
- What is the Enterprise plan? (Claude Help Center)
- Configure custom data retention controls for Enterprise plans
- What certifications has Anthropic obtained? · Anthropic Trust Center
- API and data retention (Claude Platform Docs)
- Claude Enterprise Administrator Guide (Claude Academy)
- AI 워크플로 품질 측정·벤치마크·개선 참고 문서 (Microsoft AI Genius S6E2, 2026.11.04): 에이전트 추적(관측성), 루브릭 기반 평가, 지속 평가 모니터링, 평가 CI/CD 자료 모음