AI 윤리성 리스크 진단 보고서

분석 대상: ChatGPT, Claude

작성일: 2025년 10월 23일

평가 기준:
- EU AI Act
- UNESCO AI Ethics Recommendations
- OECD AI Principles

목차

1. Executive Summary

서비스	종합점수	리스크수준
ChatGPT	3.0/5	중간
Claude	2.8/5	중간

주요 발견사항

- 전체 평균 윤리 점수: 2.9/5
- 분석 서비스 수: 2개
- 평가 차원: 공정성, 프라이버시, 투명성, 책임성, 안전성

서비스 분석: ChatGPT

종합 점수: 3.0/5
리스크 수준: 중간

차원별 상세 평가

평가 차원	점수	리스크
공정성 및 편향성	3/5	중간
프라이버시 보호	3/5	중간
투명성 및 설명가능성	3/5	중간
책임성 및 거버넌스	3/5	중간
안전성 및 보안	3/5	중간

주요 개선 권고사항

1. fairness (우선순위: 중)
현재 점수: 3/5 → 목표: 4/5

2. privacy (우선순위: 중)
현재 점수: 3/5 → 목표: 4/5

3. transparency (우선순위: 중)
현재 점수: 3/5 → 목표: 4/5

서비스 분석: Claude

종합 점수: 2.8/5
리스크 수준: 중간

차원별 상세 평가

평가 차원	점수	리스크
공정성 및 편향성	3/5	중간
프라이버시 보호	2/5	높음
투명성 및 설명가능성	3/5	중간
책임성 및 거버넌스	3/5	중간
안전성 및 보안	3/5	중간

주요 개선 권고사항

1. privacy (우선순위: 상)
현재 점수: 2/5 → 목표: 4/5

2. fairness (우선순위: 중)
현재 점수: 3/5 → 목표: 4/5

3. transparency (우선순위: 중)
현재 점수: 3/5 → 목표: 4/5

서비스 비교 분석

종합 순위

1위: ChatGPT - 3.0/5

2위: Claude - 2.8/5

차원별 비교

차원	ChatGPT	Claude
공정성	3	3
프라이버시	3	2
투명성	3	3
책임성	3	3
안전성	3	3

종합 권고사항

단기 조치 (1-3개월)

- AI 윤리 정책 문서화 및 공개
- 편향성 테스트 프레임워크 도입
- 개인정보 처리 방침 강화

중기 조치 (3-6개월)

- AI 거버넌스 체계 구축
- 정기적인 윤리 감사 시행
- 투명성 보고서 발행

장기 조치 (6개월 이상)

- 지속적인 모니터링 시스템 구축
- 외부 독립 감사 체계 확립
- 산업 표준 및 인증 획득