

# AI 윤리 리스크 진단 보고서

AI Ethics Risk Diagnosis System

2025-10-23

# AI 윤리 리스크 진단 보고서

## AI Recommendation System

### □ Executive Summary

AI Recommendation System의 윤리 리스크 진단 결과, 전체 리스크 수준은 'Limited'로 평가되었습니다. 주요 리스크 카테고리는 편향(bias), 개인정보 보호(privacy), 책임(accountability), 안전(safety)으로, 각 카테고리에서 리스크가 존재하지만 심각한 수준은 아닙니다. 이러한 리스크는 사용자 경험과 신뢰성에 영향을 미칠 수 있습니다.

주요 리스크 카테고리로는 편향, 개인정보 보호, 책임, 안전이 있으며, 이들 각각은 AI 시스템의 공정성과 투명성에 중대한 영향을 미칠 수 있습니다. 특히, 편향된 추천 결과는 사회적 불평등을 심화시킬 수 있으며, 개인정보 보호 문제는 사용자 신뢰를 저하시킬 수 있습니다.

핵심 권고사항으로는 알고리즘의 투명성을 높이고, 사용자 피드백 메커니즘을 도입하며, 데이터 사용 및 개인정보 보호 정책을 강화하는 것이 포함됩니다. 이러한 조치를 통해 AI Recommendation System의 윤리적 리스크를 효과적으로 관리할 수 있을 것입니다.

### 주요 내용

- 전체 리스크 수준: Limited
- 주요 리스크 카테고리: bias, privacy, accountability, safety
- 핵심 권고사항:
  - 알고리즘의 투명성 강화
  - 사용자 피드백 메커니즘 도입
  - 데이터 사용 및 개인정보 보호 정책 강화

### □ Service Profile

항목	내용
서비스명	AI Recommendation System
서비스 유형	recommendation
설명	이 서비스는 사용자 선호도를 분석하여 개인화된 추천을 제공합니다. 머신러닝 알고리즘을 사용하여 사용자 행동을 학습합니다.
데이터 처리 방식	사용자의 클릭 및 구매 데이터를 수집하고 분석하여 추천 알고리즘에 활용합니다.
사용자 영향 범위	개별 사용자 및 전체 사용자 그룹에 대한 맞춤형 경험 제공
진단된 리스크 카테고리	bias

---

## □ □ Risk Assessment

### 리스크 평가 결과

#### bias 리스크

- 리스크 수준: Limited
- 평가 요약: AI Recommendation System의 bias 리스크는 EU AI Act 기준으로 Limited 수준으로 평가됩니다.
- 주요 우려사항: 편향된 추천 결과가 사회적 불평등을 반영할 수 있음.
- 권고 초점: bias 관련 개선 방안 수립 필요

#### privacy 리스크

- 리스크 수준: Limited
- 평가 요약: AI Recommendation System의 privacy 리스크는 EU AI Act 기준으로 Limited 수준으로 평가됩니다.
- 주요 우려사항: 사용자 데이터의 안전한 처리 및 보호 필요.
- 권고 초점: privacy 관련 개선 방안 수립 필요

#### accountability 리스크

- 리스크 수준: Limited
- 평가 요약: AI Recommendation System의 accountability 리스크는 EU AI Act 기준으로 Limited 수준으로 평가됩니다.
- 주요 우려사항: 시스템의 결정에 대한 책임 소재 불명확.
- 권고 초점: accountability 관련 개선 방안 수립 필요

#### safety 리스크

- 리스크 수준: Limited
- 평가 요약: AI Recommendation System의 safety 리스크는 EU AI Act 기준으로 Limited 수준으로 평가됩니다.
- 주요 우려사항: 시스템의 안전성 및 신뢰성 문제.
- 권고 초점: safety 관련 개선 방안 수립 필요

---

## □ Mitigation Recommendations

### 우선순위별 개선 권고안

#### □ High Priority

1. 투명한 알고리즘 설명 제공: 알고리즘 작동 방식과 추천 이유를 사용자에게 명확하게 설명하는 기능 추가.
2. 데이터 사용 및 개인정보 보호 정책 강화: 사용자 데이터 사용에 대한 명확한 정책 수립 및 기술적 조치 강화.
3. 리스크 관리 프레임워크 구축: 리스크 관리 프레임워크 개발 및 문서화.

#### □ Medium Priority

- 사용자 피드백 메커니즘 도입: 추천 결과에 대한 사용자 피드백 기능 도입.
- 정기적인 리스크 평가 실시: 운영 중 정기적으로 리스크 평가 실시.
- 투명성 및 설명 가능성 향상: 추천 알고리즘에 대한 투명성 높이기.

#### □ Low Priority

- 사용자 피드백 시스템 구축: 사용자 피드백 수집 및 분석 시스템 구축.

### □ Evidence Sources

#### Baseline Sources (공식 문서)

- OECD\_Privacy\_2024.pdf**: AI 추천 시스템은 개인정보 보호와 윤리적 고려가 필수적이라는 내용을 담고 있습니다.

#### Issue Sources (최신 이슈)

- IBM AI Ethics**: AI 추천 시스템의 윤리적 이슈에 대한 다양한 관점을 제시하며, 기술 오용과 같은 리스크를 관리하기 위한 다학문적 접근이 필요하다고 강조합니다.

### □ Conclusion

#### 전체 평가

AI Recommendation System의 윤리 리스크 진단 결과, 리스크 수준은 'Limited'로 평가되었으며, 주요 리스크 카테고리는 편향, 개인정보 보호, 책임, 안전입니다. 이러한 리스크는 사용자 경험과 신뢰성에 영향을 미칠 수 있으므로, 적절한 개선 조치가 필요합니다.

#### 권장 다음 단계

구체적인 액션 플랜으로는 알고리즘의 투명성을 높이고, 사용자 피드백 메커니즘을 도입하며, 데이터 사용 및 개인정보 보호 정책을 강화하는 것이 포함됩니다. 관련 부서는 이러한 권고사항을 바탕으로 실행 계획을 수립해야 합니다.

#### 연락처 및 지원

본 보고서에 대한 문의사항이나 추가 지원이 필요한 경우 관련 담당자에게 연락하시기 바랍니다.

본 보고서는 EU AI Act, OECD, UNESCO 기준에 따라 작성되었습니다. 보고서 생성일: 2025-10-23