

# AI 윤리 진단 보고서

작성일: 2025-10-23

## 1. EXECUTIVE SUMMARY

AI Recommendation System은 사용자 선호도를 분석하여 개인화된 콘텐츠를 추천하는 서비스입니다. 그러나 이 시스템은 편향(bias)과 개인정보 보호(privacy)와 관련된 높은 리스크를 가지고 있습니다. 본 보고서는 이러한 리스크를 평가하고, 개선 방안을 제시하여 윤리적 기준을 준수할 수 있도록 돕고자 합니다.

## 2. SERVICE PROFILE

- 서비스 이름: AI Recommendation System
- 서비스 유형: 추천 시스템
- 설명: 이 서비스는 사용자 선호도를 분석하여 개인화된 콘텐츠를 추천합니다. 머신러닝 알고리즘을 사용하여 사용자 행동을 학습합니다.
- 데이터 처리 방법: 사용자의 클릭 및 구매 데이터를 수집하고 분석하여 추천 알고리즘을 개선합니다.
- 사용자 영향 범위: 개별 사용자 및 전체 사용자 그룹에 대한 맞춤형 경험 제공
- 진단된 리스크 카테고리: bias, privacy

## 3. RISK ASSESSMENT

### 3.1 Bias

- 리스크 수준: High
- 평가 요약: AI Recommendation System의 bias 리스크는 High 수준으로 평가되었습니다.
- 추천 초점: bias 관련 개선 방안 수립 필요

### 3.2 Privacy

- 리스크 수준: High
- 평가 요약: AI Recommendation System의 privacy 리스크는 High 수준으로 평가되었습니다.
- 추천 초점: privacy 관련 개선 방안 수립 필요

## 4. MITIGATION RECOMMENDATIONS

### 4.1 Bias 관련 개선 방안

- 데이터 다양성 확보
  - 완화 단계: AI Recommendation System에 사용되는 데이터셋의 다양성을 높이기 위해, 다양한 인구통계학적 특성을 반영한 데이터를 수집하고 포함시킵니다.
  - 우선순위: High
  - 노력 수준: Medium
  - 관련 기준: OECD
- 편향 감지 및 모니터링 시스템 구축
  - 완화 단계: AI Recommendation System의 출력 결과를 정기적으로 분석하여 편향을 감지하는 모니터링 시스템을 구축합니다.
  - 우선순위: High

- 노력 수준: High
  - 관련 기준: EU AI Act
3. 사용자 피드백 통합
- 완화 단계: 사용자로부터의 피드백을 수집하고 분석하여 AI Recommendation System의 추천 결과에 대한 편향을 파악하고 개선합니다.
  - 우선순위: Medium
  - 노력 수준: Medium
  - 관련 기준: UNESCO

## 4.2 Privacy 관련 개선 방안

1. 데이터 익명화 프로세스 도입
- 완화 단계: 사용자 데이터를 수집하기 전에 모든 개인 식별 정보를 제거하고, 데이터 익명화 기술을 적용하여 개인의 프라이버시를 보호합니다.
  - 우선순위: High
  - 노력 수준: Medium
  - 관련 기준: EU AI Act
2. 사용자 동의 기반 데이터 수집
- 완화 단계: AI Recommendation System에서 사용자 데이터를 수집하기 전에 명확한 동의를 받도록 하며, 사용자가 언제든지 동의를 철회할 수 있는 기능을 제공합니다.
  - 우선순위: High
  - 노력 수준: Medium
  - 관련 기준: OECD
3. 정기적인 프라이버시 감사 실시
- 완화 단계: AI Recommendation System의 프라이버시 정책 및 데이터 처리 절차에 대해 정기적으로 감사를 실시하고, 그 결과를 바탕으로 개선 사항을 도출하여 적용합니다.
  - 우선순위: Medium
  - 노력 수준: High
  - 관련 기준: UNESCO

## 5. CONCLUSION

AI Recommendation System은 사용자에게 개인화된 경험을 제공하는 중요한 도구이지만, 편향과 개인정보 보호와 관련된 높은 리스크를 동반하고 있습니다. 본 보고서에서 제시한 개선 방안을 통해 이러한 리스크를 효과적으로 관리하고, 윤리적 기준을 준수하는 방향으로 나아가야 합니다.

## 6. REFERENCE

1. UNESCO\_Ethics\_2021.pdf
2. OECD\_Privacy\_2024.pdf
3. [Bias in AI Models and Generative Systems](#)
4. [AI Bias Examples Mitigation Guide](#)
5. [A Regulatory Roadmap to AI and Privacy](#)

## 7. APPENDIX

- 수집된 증거:
  - AI 추천 시스템의 윤리 리스크와 관련된 다양한 문서 및 자료를 통해 리스크의 심각성을 강조하고 있습니다. 각 자료는 편향과 개인정보 보호의 중요성을 다루고 있으며, AI 서비스의 윤리적 고려가 필수적임을 시사합니다.