

# AI 윤리 진단 보고서

## 1. EXECUTIVE SUMMARY

본 보고서는 AI Recommendation System의 윤리적 리스크를 평가하고, 이를 완화하기 위한 권고안을 제시합니다. 서비스는 사용자 선호도를 분석하여 개인화된 추천을 제공하며, 데이터 수집 및 분석을 통해 알고리즘을 개선합니다. 그러나, 본 서비스는 편향(bias) 및 개인정보 보호(privacy)와 관련된 높은 리스크를 보유하고 있습니다. 이에 따라, 본 보고서는 리스크를 최소화하기 위한 구체적인 개선 방안을 제안합니다.

## 2. SERVICE PROFILE

- Service Name:** AI Recommendation System
- Service Type:** Recommendation
- Description:** 이 서비스는 사용자 선호도를 분석하여 개인화된 추천을 제공합니다. 머신러닝 알고리즘을 사용하여 사용자 행동을 학습합니다.
- Data Handling Method:** 사용자의 클릭 및 구매 데이터를 수집하고 분석하여 추천 알고리즘을 개선합니다.
- User Impact Scope:** 개별 사용자 및 전체 사용자 그룹에 대한 맞춤형 경험 제공
- Diagnosed Risk Categories:** bias, privacy

## 3. RISK ASSESSMENT

- Service Name:** AI Recommendation System
- Assessed Risks:**
  - Bias**
    - Risk Level:** High
    - Assessment Summary:** AI Recommendation System의 bias 리스크는 High 수준으로 평가되었습니다.
    - Recommendation Focus:** bias 관련 개선 방안 수립 필요
  - Privacy**
    - Risk Level:** High
    - Assessment Summary:** AI Recommendation System의 privacy 리스크는 High 수준으로 평가되었습니다.
    - Recommendation Focus:** privacy 관련 개선 방안 수립 필요

## 4. MITIGATION RECOMMENDATIONS

### Bias 관련 권고안

- 데이터 다양성 확보**
  - Mitigation Step:** AI Recommendation System에 사용되는 데이터셋의 다양성을 확보하기 위해, 다양한 인구 통계적 특성을 가진 데이터를 수집하고 포함시켜야 합니다. 이를 통해 특정 그룹에 대한 편향을 줄일 수 있습니다.
  - Priority:** High
  - Effort Level:** Medium
  - Relevant Standard:** OECD
- 편향 검토 및 모니터링 시스템 구축**
  - Mitigation Step:** AI 모델의 추천 결과에 대한 정기적인 편향 검토 및 모니터링 시스템을 구축하여, 추천 결과가 특정 그룹에 불리하게 작용하지 않도록 지속적으로 평가하고 조정합니다.
  - Priority:** High
  - Effort Level:** High
  - Relevant Standard:** EU AI Act

### 3. 다양성 교육 프로그램 개발

- **Mitigation Step:** AI Recommendation System을 개발하고 운영하는 팀을 대상으로 다양성과 포용성에 대한 교육 프로그램을 개발하여, 팀원들이 편향 문제를 인식하고 해결할 수 있는 능력을 배양하도록 합니다.
- **Priority:** Medium
- **Effort Level:** Medium
- **Relevant Standard:** UNESCO

## Privacy 관련 권고안

### 1. 데이터 최소화 원칙 적용

- **Mitigation Step:** 사용자의 개인 정보를 수집할 때, 서비스 제공에 필수적인 최소한의 데이터만 수집하도록 정책을 수립하고 이를 시스템에 적용합니다.
- **Priority:** High
- **Effort Level:** Medium
- **Relevant Standard:** EU AI Act

### 2. 사용자 동의 강화

- **Mitigation Step:** 개인 정보 수집 및 처리에 대한 명확한 사용자 동의를 요구하고, 사용자가 언제든지 동의를 철회할 수 있는 기능을 제공합니다.
- **Priority:** High
- **Effort Level:** Medium
- **Relevant Standard:** OECD

### 3. 데이터 암호화 및 익명화

- **Mitigation Step:** 사용자의 개인 정보를 저장 및 전송할 때, 강력한 암호화 기술을 적용하고, 가능하면 데이터를 익명화하여 개인 식별이 불가능하도록 합니다.
- **Priority:** High
- **Effort Level:** High
- **Relevant Standard:** UNESCO

## 5. CONCLUSION

AI Recommendation System은 사용자에게 개인화된 경험을 제공하는 중요한 서비스입니다. 그러나, 편향 및 개인정보 보호와 관련된 높은 리스크를 보유하고 있어, 이를 해결하기 위한 적극적인 노력이 필요합니다. 본 보고서에서 제시한 권고안을 통해 리스크를 최소화하고, 윤리적 기준을 준수하는 방향으로 나아가야 합니다.

## 6. REFERENCE

- UNESCO\_Ethics\_2021.pdf
- OECD\_Privacy\_2024.pdf

## 7. APPENDIX

### • Bias 관련 증거:

- AI 추천 시스템의 윤리 리스크는 회원국과 이해관계자들이 제시된 윤리적 가치와 원칙을 존중하고 보호해야 하며, 정책 권고를 실현하기 위한 모든 가능한 조치를 취해야 한다는 점에서 발생한다.
- AI 추천 시스템은 데이터와 정보를 처리하여 지능적 행동을 모방하는 능력을 갖추고 있으며, 이는 윤리적 측면에서 중요한 특성을 지닌다.

### • Privacy 관련 증거:

- OECD는 AI 관련 기본 원칙을 수정하지 않고, 2019년에 채택된 AI 권고안에 따라 AI와 관련된 문제를 다루기로 결정했다.
- AI 추천 시스템은 개인정보 보호와 관련된 기본 원칙을 수정하지 않고, AI 관련 문제를 해결하기 위한 메커니즘과 지침을 마련해야 한다.