

AI 윤리성 리스크 진단 : DeepSeek

작성일자: 2025.05.20

SUMMARY

본 보고서는 Hangzhou DeepSeek Artificial Intelligence Co., Ltd.가 운영하는 AI 플랫폼 서비스인 DeepSeek에 대한 윤리적 리스크와 독소조항을 분석하고, 개선 방향을 제안합니다. DeepSeek는 사용자가 챗봇을 생성하고 상호작용할 수 있는 기능을 제공하며, 개발자, 기업 사용자, AI 연구자를 주요 대상으로 합니다. 서비스는 IP 주소 기반 위치 정보와 쿠키를 통해 사용자 설정 정보를 수집합니다.

윤리적 리스크 평가에서는 편향성, 프라이버시, 설명가능성, 자동화 리스크가 각각 중간에서 높은 수준으로 평가되었습니다. 특히 프라이버시 리스크는 높은 수준으로, 사용자 동의 없이 과도한 정보 수집과 데이터 보유 기간 및 파기 절차의 불명확성이 문제로 지적되었습니다. 독소조항 분석에서는 서비스 변경 및 종료에 대한 모호한 통지 조항과 서비스 제공자의 책임을 면제하는 조항이 사용자에게 불리한 영향을 미칠 수 있는 것으로 나타났습니다.

개선 방향으로선 편향성 리스크를 줄이기 위한 알고리즘 개선과 사용자 피드백 루프 강화, 프라이버시 보호를 위한 데이터 수집 최소화 및 사용자 동의 절차 개선, 설명가능성을 높이기 위한 AI 결정 과정의 시각화 및 교육 자료 제공, 자동화 리스크 완화를 위한 인간 검토 절차 강화, 독소조항 수정을 통한 사용자 권리 보호 등이 제안되었습니다.

1. 서비스 개요

1.1. 서비스 이름

- DeepSeek

1.2. 서비스 상세 설명

DeepSeek는 Hangzhou DeepSeek Artificial Intelligence Co., Ltd.가 소유하고 운영하는 플랫폼으로, 사용자가 챗봇을 생성하고 상호작용할 수 있는 서비스를 제공합니다. 이 서비스는

개인정보 처리방침에 따라 운영되며, 다양한 AI 기반 기능을 통해 사용자에게 가치를 제공합니다.

1.3. 핵심 기능 목록

- **챗봇 생성:** 사용자가 자신의 필요에 맞는 챗봇을 생성할 수 있는 기능을 제공합니다.
- **챗봇 상호작용:** 생성된 챗봇과의 상호작용을 통해 다양한 기능을 수행할 수 있습니다.

1.4. 주요 대상 사용자

- 개발자
- 기업 사용자
- AI 연구자

1.5. 수집 데이터 유형 상세

- IP 주소 기반 위치 정보
- 쿠키 및 유사 기술을 통한 사용자 설정 정보

1.6. 서비스 URL 접속 상태 및 접근성

- 정보 없음

1.7. 정보 취득의 주요 출처

- DeepSeek Privacy Policy.pdf
 - DeepSeek 이용약관.pdf
-

2. AI 윤리성 리스크 심층 평가

2.1. 편향성(Bias) 리스크: 중간

- **상세 평가 근거:** DeepSeek의 편향성 리스크는 중간 수준으로 평가됩니다. 서비스는 사용자가 챗봇을 생성하고 상호작용할 수 있도록 하며, 이는 알고리즘의 편향성을 초래할 수 있습니다. 'DeepSeek Privacy Policy.pdf'에 따르면, 개인정보 처리방침은 사용자의 데이터를 수집하고 처리하는 방식에 대해 명확히 설명하고 있지만, 알고리즘의 공정성과 편향성에 대

한 구체적인 언급은 부족합니다. 이는 특정 인구 집단에 대한 차별적 결과를 초래할 수 있는 가능성을 내포합니다.

- **잠재적 영향:** 사용자에게 불공정한 경험을 제공할 수 있으며, 사회적 차별을 강화할 가능성이 있습니다.
- **주요 근거 문서:** DeepSeek Privacy Policy.pdf, 페이지 0, 섹션: DeepSeek 개인정보 처리 방침

2.2. 프라이버시(Privacy) 리스크: 높음

- **상세 평가 근거:** DeepSeek의 프라이버시 리스크는 높음으로 평가됩니다. 'DeepSeek Privacy Policy.pdf'에 따르면, 서비스는 IP 주소 기반 위치 정보와 쿠키를 통해 사용자 설정 정보를 수집합니다. 이러한 데이터 수집 방식은 사용자 동의 없이 과도한 정보 수집으로 이어질 수 있으며, 이는 OECD 프라이버시 가이드라인 원칙에 위배될 수 있습니다.
- **잠재적 영향:** 사용자 프라이버시 침해로 이어질 수 있으며, 데이터 유출 및 오용의 위험이 존재합니다.
- **주요 근거 문서:** DeepSeek Privacy Policy.pdf, 페이지 5, 섹션: 귀하의 정보 보유 기간 개인정보의 파기 절차 및 방법 귀하의 권리

2.3. 설명가능성(Explainability) 리스크: 중간

- **상세 평가 근거:** DeepSeek의 설명가능성 리스크는 중간 수준으로 평가됩니다. 서비스는 사용자가 챗봇을 생성하고 상호작용할 수 있도록 하지만, 챗봇의 의사결정 과정이나 결과 도출 근거에 대한 명확한 설명이 부족합니다.
- **잠재적 영향:** 사용자가 챗봇의 응답을 이해하거나 신뢰하는 데 어려움을 겪을 수 있으며, 오류 발생 시 원인 파악이 어려울 수 있습니다.
- **주요 근거 문서:** DeepSeek Privacy Policy.pdf, 페이지 0, 섹션: DeepSeek 개인정보 처리 방침

2.4. 자동화(Automation) 리스크: 중간

- **상세 평가 근거:** DeepSeek의 자동화 리스크는 중간 수준으로 평가됩니다. 서비스는 사용자가 생성한 챗봇을 통해 자동화된 의사결정을 지원하지만, 이러한 자동화가 인간의 판단을 대체할 수 있는 잠재적 위험이 존재합니다.
 - **잠재적 영향:** 사용자에게 불공정한 결과를 초래하거나, 의도치 않은 결과를 초래할 수 있는 위험을 내포합니다.
 - **주요 근거 문서:** DeepSeek Privacy Policy.pdf, 페이지 1, 섹션: 당사가 귀하의 정보를 이용하는 방식
-

3. 약관 및 개인정보 처리방침 심층 분석 (독소조항)

3.1. 전반적인 약관 위험도: 중간

- **종합 평가 이유:** DeepSeek의 약관은 서비스 제공자가 서비스의 변경, 중단, 종료를 일방적으로 결정할 수 있는 권한을 부여하고 있으며, 서비스의 중단, 오류, 데이터 손실에 대해 책임을 지지 않는다고 명시하고 있습니다. 이는 사용자에게 불리한 영향을 미칠 수 있는 조항들로, 사용자 권리 보호가 미흡합니다.

3.2. 주요 독소조항 상세 분석

조항 1

- **조항 내용:** "1.3 생성형 인공지능 기술모델제품이 계속 진화하고 법령 역시 변화하기 때문에, 당사가 서비스를 추가업그레이드변경중단종료하거나 서비스의 기술방법성능에 필요한 조정을 할 수 있고, 새로운 서비스 기능을 서비스 내외로 테스트할 수 있습니다."
- **위험성 분석:** 이 조항은 DeepSeek이 서비스의 변경, 중단, 종료를 일방적으로 결정할 수 있는 권한을 부여하고 있습니다. '영향을 미칠 수 있는 경우'라는 표현이 모호하여 서비스 제공자의 자의적 판단에 따라 통지가 이루어질 수 있습니다.
- **사용자 영향:** 사용자는 서비스가 갑작스럽게 변경되거나 종료될 경우, 이에 대한 대비를 할 수 없게 되어 업무나 개인적인 계획에 차질을 빚을 수 있습니다.
- **근거 자료:** DeepSeek 이용약관, 페이지 0, 섹션 N/A

조항 2

- **조항 내용:** "7. 보증 부인, 책임 제한, 면책: 당사는 본 서비스의 보안과 안정적인 운영을 위해 필요한 조치를 업계 관행에 따라 취하겠으나, 서비스의 중단, 오류, 데이터 손실 등에 대해 책임을 지지 않습니다."
 - **위험성 분석:** 이 조항은 서비스 제공자가 서비스의 중단, 오류, 데이터 손실에 대해 책임을 지지 않는다고 명시하고 있습니다. 이는 사용자에게 부당하게 불리할 수 있습니다.
 - **사용자 영향:** 사용자는 서비스의 오류나 중단으로 인해 중요한 데이터를 잃거나, 서비스 이용에 차질을 빚을 수 있으며, 이에 대한 보상을 받을 수 없게 됩니다.
 - **근거 자료:** DeepSeek 이용약관, 페이지 3, 섹션 7
-

4. 종합 개선 방향 및 실행 로드맵 제안

4.1. 편향성 리스크 개선 전략

- **개선 방안 1:** 챗봇 생성 과정에서 사용자가 입력하는 데이터를 분석하여 편향성을 감지하고 완화할 수 있는 알고리즘을 개발합니다. 다양한 인구 집단을 대표하는 데이터셋을 수집하고, 정기적으로 알고리즘의 공정성을 평가하는 절차를 도입합니다.
- **개선 방안 2:** 사용자 피드백 루프를 강화하여 편향성 문제를 지속적으로 모니터링하고 개선할 수 있는 체계를 구축합니다.

4.2. 프라이버시 리스크 개선 전략

- **개선 방안 1:** 데이터 수집 최소화 원칙을 적용하여, 서비스 운영에 필수적인 정보만을 수집하도록 개인정보 처리방침을 개정합니다.
- **개선 방안 2:** 사용자에게 데이터 수집 및 처리에 대한 명확한 동의를 받을 수 있는 UI/UX를 개선합니다.

4.3. 설명가능성 리스크 개선 전략

- **개선 방안 1:** AI 결정 과정의 주요 요인을 시각화하거나 요약 정보를 제공하는 기능을 도입합니다.
- **개선 방안 2:** 모델 설명 문서 또는 FAQ를 제공하여 사용자가 AI 모델의 작동 방식을 쉽게 이해할 수 있도록 합니다.

4.4. 자동화 리스크 개선 전략

- **개선 방안 1:** 중요한 자동화 결정에 대한 인간 검토 및 개입 절차를 강화합니다.
- **개선 방안 2:** 자동화 시스템의 오류 모니터링 및 책임 규명 프로세스를 구축합니다.

4.5. 약관 및 독소조항 개선 전략

- **독소조항 관련 개선 방안 1:** 조항 1.3의 '영향을 미칠 수 있는 경우'라는 모호한 표현을 구체화하여, 서비스 변경 시 사용자에게 사전 통지를 반드시 제공하도록 수정합니다.
 - **독소조항 관련 개선 방안 2:** 조항 7의 면책 조항을 수정하여, 서비스 제공자의 중대한 과실이나 고의로 인한 손해에 대해서는 책임을 지도록 명시합니다.
-

5. 사용된 윤리 가이드라인 및 참고자료

본 보고서는 OECD AI 가이드라인을 주요 기준으로 활용하여 윤리성 리스크를 평가하였으며, 분석 과정에서 참고한 주요 문서 및 RAG 컨텍스트는 다음과 같습니다:

- DeepSeek Privacy Policy.pdf
 - DeepSeek 이용약관.pdf
-

이 보고서는 AI에 의해 작성되었습니다.