

AI 윤리성 리스크 진단 : DeepSeek

작성일자: 2025.05.20

SUMMARY

DeepSeek는 사용자가 챗봇을 생성하고 상호작용할 수 있는 플랫폼을 제공하는 서비스로, Hangzhou DeepSeek Artificial Intelligence Co., Ltd.에 의해 운영됩니다. 이 서비스는 개발자, 기업 사용자, AI 연구자를 주요 대상으로 하며, 위치 정보와 쿠키를 통해 사용자 데이터를 수집합니다. 윤리적 리스크 평가 결과, DeepSeek는 프라이버시 리스크가 높고, 편향성, 설명가능성, 자동화 리스크가 중간 수준으로 평가되었습니다. 특히, 프라이버시 리스크는 데이터 수집 및 처리의 투명성 부족과 국제 개인정보 보호 규정과의 충돌 가능성으로 인해 심각한 우려를 제기합니다. 독소조항 분석에서는 서비스 변경 및 종단에 대한 모호한 통지 조항과 서비스 사용으로 인한 손해에 대한 책임 제한 조항이 주요 문제점으로 지적되었습니다. 이러한 문제들은 사용자 권리를 침해하고, 서비스에 대한 신뢰를 저하시킬 수 있습니다. 개선 방향으로 데이터 수집 최소화, 설명가능성 강화, 자동화 시스템의 인간 검토 절차 강화, 독소조항의 명확한 수정 등이 제안되었습니다. 이러한 개선은 사용자 신뢰를 회복하고, 서비스의 윤리적 기준을 높이는 데 기여할 것입니다.

1. 서비스 개요

1.1. 서비스 이름

DeepSeek

1.2. 서비스 상세 설명

DeepSeek는 Hangzhou DeepSeek Artificial Intelligence Co., Ltd.가 소유하고 운영하는 서비스로, 사용자가 챗봇을 생성하고 상호작용할 수 있는 플랫폼을 제공합니다. 이 서비스는 사용자에게 챗봇을 생성할 수 있는 기능과 생성된 챗봇과 상호작용할 수 있는 기능을 제공합니다.

1.3. 핵심 기능 목록

- 챗봇 생성:** 사용자가 원하는 대로 챗봇을 생성할 수 있는 기능
- 챗봇 상호작용:** 생성된 챗봇과의 다양한 상호작용을 지원

1.4. 주요 대상 사용자

- 개발자
- 기업 사용자
- AI 연구자

1.5. 수집 데이터 유형 상세

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

1.6. 서비스 URL 접속 상태 및 접근성

정보 없음

1.7. 정보 취득의 주요 출처

DeepSeek 이용약관 및 개인정보 처리방침 문서

2. AI 윤리성 리스크 심층 평가

2.1. 편향성(Bias) 리스크: 중간

- **상세 평가 근거:** DeepSeek는 사용자가 챗봇을 생성하고 상호작용할 수 있는 플랫폼을 제공하며, 이는 알고리즘의 편향성을 초래할 수 있습니다. 위치 정보와 쿠키를 통한 데이터 수집 방식은 특정 인구 집단에 대한 편향성을 초래할 수 있으며, 알고리즘의 학습 데이터가 특정 그룹을 과소 또는 과대 대표할 가능성이 있습니다. 이는 OECD AI 원칙에서 언급된 공정성과 투명성 원칙에 위배될 수 있습니다.
- **잠재적 영향:** 특정 사용자 그룹에게 불리한 결과를 초래할 수 있으며, 사회적 불평등을 강화할 위험이 있습니다.
- **주요 근거 문서:** DeepSeek Privacy Policy.pdf, 페이지 0

2.2. 프라이버시(Privacy) 리스크: 높음

- **상세 평가 근거:** DeepSeek는 위치 정보와 쿠키를 통해 사용자 데이터를 수집하며, 이는 개인정보 보호에 대한 심각한 우려를 제기합니다. 데이터 수집 및 처리의 투명성에 대한 명확한 설명이 부족하며, 개인정보가 중국에 위치한 서버에 저장된다는 점은 국제적인 개인정보 보호 규정과의 충돌 가능성을 높입니다.

- **잠재적 영향:** 데이터 유출 및 오용의 위험을 증가시키며, 사용자의 프라이버시를 침해할 수 있습니다.
- **주요 근거 문서:** DeepSeek Privacy Policy.pdf, 페이지 5

2.3. 설명가능성(Explainability) 리스크: 중간

- **상세 평가 근거:** 서비스는 사용자가 챗봇을 생성하고 상호작용할 수 있는 기능을 제공하지만, 이러한 과정에서의 의사결정 과정이 명확하게 설명되지 않습니다. 이는 사용자에게 충분한 투명성을 제공하지 못할 수 있습니다.
- **잠재적 영향:** 사용자가 시스템의 결과를 신뢰하지 못하게 하며, 오류 발생 시 원인 파악이 어려워질 수 있습니다.
- **주요 근거 문서:** DeepSeek 이용약관.pdf, 페이지 0

2.4. 자동화(Automation) 리스크: 중간

- **상세 평가 근거:** 서비스는 챗봇을 통해 자동화된 상호작용을 지원하며, 이는 인간의 판단을 대체할 수 있는 잠재력을 가지고 있습니다. 자동화된 의사결정 과정은 오류 발생 시 책임 소재가 불분명해질 수 있습니다.
- **잠재적 영향:** 사용자의 통제력을 약화시킬 수 있으며, 자동화된 시스템의 오작동은 사용자에게 부정적인 영향을 미칠 수 있습니다.
- **주요 근거 문서:** DeepSeek Privacy Policy.pdf, 페이지 1

3. 약관 및 개인정보 처리방침 심층 분석 (독소조항)

3.1. 전반적인 약관 위험도: 중간

- **종합 평가 이유:** 서비스 변경 및 중단에 대한 모호한 통지 조항과 서비스 사용으로 인한 손해에 대한 책임 제한 조항이 사용자 권리를 침해할 가능성을 내포하고 있습니다.

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. 이러한 변경이 귀사의 권리에 영향을 미칠 수 있는 경우 당사는 사이트 내 알림, 웹사이트 공지 또는 이메일을 통하여 귀하에게 사전에 통지하겠습니다."

- 위험성 분석: 서비스 제공자가 서비스의 변경, 중단, 종료를 포함한 다양한 조치를 취할 수 있는 권한을 명시하고 있으나, 이러한 조치가 사용자에게 미치는 영향을 충분히 설명하지 않으며, 모호한 표현으로 인해 사용자가 어떤 상황에서 통지를 받을 수 있는지 명확하지 않습니다.
- 사용자 영향: 사용자는 서비스가 갑작스럽게 중단되거나 변경될 경우 이에 대한 대비가 부족하여 데이터 손실이나 서비스 이용 중단으로 인한 불편을 겪을 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 취할 것입니다. 그러나 당사는 본 서비스의 사용으로 인해 발생할 수 있는 모든 손해에 대해 책임을 지지 않습니다."
- 위험성 분석: 서비스 제공자가 서비스 사용으로 인해 발생할 수 있는 모든 손해에 대해 책임을 지지 않는다는 내용을 포함하고 있으며, 이는 서비스 제공자의 중대한 과실이나 고의로 인한 손해까지 면책하려는 것으로 해석될 수 있습니다.
- 사용자 영향: 서비스의 결함으로 인해 사용자의 데이터가 손실되거나, 서비스 중단으로 인해 비즈니스에 손해가 발생하더라도 사용자는 이에 대한 보상을 받지 못할 수 있습니다.
- 근거 자료: DeepSeek 이용약관, 페이지 3, 섹션 7. 보증 부인, 책임 제한, 면책

4. 종합 개선 방향 및 실행 로드맵 제안

4.1. 편향성 리스크 개선 전략

- 개선 방안 1: 알고리즘의 편향성을 줄이기 위해 다양한 인구 집단을 대표하는 데이터 세트를 수집하고 사용하는 것을 의무화합니다. 데이터 수집 정책을 개정하여 다양한 사용자 그룹의 데이터를 균형 있게 포함하도록 하고, 정기적으로 데이터 세트의 대표성을 검토하는 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 사용자로부터 수집하는 데이터의 범위를 최소화하고, 각 데이터 항목의 수집 목적과 보유 기간을 명확히 명시합니다.
- 개선 방안 2: 데이터 보안 강화를 위해 데이터 암호화 및 접근 통제 메커니즘을 강화하고, 사용자에게 데이터 접근 및 삭제 요청을 쉽게 할 수 있는 인터페이스를 제공합니다.

4.3. 설명가능성 리스크 개선 전략

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

4.4. 자동화 리스크 개선 전략

- **개선 방안 1:** 중요한 자동화 결정에 대한 인간 검토 및 개입 절차를 강화하여, 자동화된 시스템의 오류를 조기에 발견하고 수정할 수 있도록 합니다.
- **개선 방안 2:** 자동화 시스템의 오류 모니터링 및 책임 규명 프로세스를 구축하여, 오류 발생 시 신속하게 대응하고 책임 소재를 명확히 할 수 있도록 합니다.

4.5. 약관 및 독소조항 개선 전략

- **독소조항 관련 개선 방안 1:** '1.3' 조항의 모호한 표현을 수정하여, 서비스 변경이 사용자 권리에 미치는 영향을 구체적으로 설명하고, 모든 주요 변경 사항에 대해 사전 통지를 보장하는 문구로 개정합니다.
- **독소조항 관련 개선 방안 2:** '7' 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, 페이지 0, 1, 5
- DeepSeek 이용약관.pdf, 페이지 0, 3

이 보고서는 AI에 의해 작성되었습니다.