

AI 윤리성 리스크 진단: 다글로

작성일자: 2025.05.20

SUMMARY

다글로는 AI 기반의 챗봇 및 음성 인식 서비스를 제공하는 플랫폼으로, 사용자에게 음성 데이터를 텍스트로 변환하거나 다양한 질문에 대한 답변을 받을 수 있는 기능을 지원합니다. 이 서비스는 일반 사용자뿐만 아니라 기업 및 단체를 대상으로 하며, 사용자 기본 정보와 음성 및 영상 파일 등 다양한 데이터를 수집합니다. 그러나 이러한 서비스는 몇 가지 윤리적 리스크를 내포하고 있습니다.

첫째, 편향성 리스크는 중간 수준으로 평가되었습니다. 이는 다글로의 음성 인식 기능이 특정 언어에 특화되어 있어 다른 언어 사용자에게 불리할 수 있다는 점에서 비롯됩니다. 둘째, 프라이버시 리스크는 높음으로 평가되었으며, 이는 민감한 데이터를 다루면서도 보안 및 오용 방지 조치가 충분하지 않기 때문입니다. 셋째, 설명가능성 리스크는 중간 수준으로, AI 모델의 작동 원리나 결과 도출 과정에 대한 설명이 부족하여 사용자가 결과를 이해하기 어려울 수 있습니다. 마지막으로, 자동화 리스크는 중간 수준으로, 자동화된 의사결정의 책임성이 불분명할 수 있습니다.

독소조항 분석에서는 중간 수준의 위험이 발견되었습니다. 특히, 영구 이용정지 시 사용자가 획득한 포인트 및 혜택이 소멸되며 보상이 없다는 조항과 개인정보 처리방침 변경 시 사전 고지가 불충분할 수 있다는 조항이 사용자에게 불리하게 작용할 수 있습니다.

개선을 위해서는 편향성 리스크를 줄이기 위한 다국어 데이터셋 확장과 사용자 피드백 루프 강화, 프라이버시 리스크를 줄이기 위한 데이터 수집 최소화 및 보안 강화, 설명가능성 향상을 위한 AI 결정 과정의 시각화 및 사용자 교육, 자동화 리스크 관리를 위한 인간 검토 절차 강화 등이 필요합니다. 또한, 독소조항의 수정을 통해 사용자 권익을 보호하고 공정성을 강화해야 합니다.

1. 서비스 개요

1.1. 서비스 이름

- 다글로

1.2. 서비스 상세 설명

다글로는 AI 기반의 챗봇 및 음성 인식 서비스를 제공하는 플랫폼입니다. 사용자는 이 서비스를 통해 음성 데이터를 텍스트로 변환하거나, 챗봇을 통해 다양한 질문에 대한 답변을 받을 수 있습니다. 이는 사용자에게 편리함과 효율성을 제공하며, 특히 음성 인식 기능은 다양한 AI 기능을 통해 사용자 경험을 향상시킵니다.

1.3. 핵심 기능 목록

- **AI 챗봇**: 사용자가 다양한 질문에 대한 답변을 받을 수 있도록 지원합니다.
- **음성 인식**: 음성 데이터를 텍스트로 변환하여 활용할 수 있습니다.
- **크롬 확장 프로그램**: 유튜브 영상을 텍스트로 변환하여 문서 형식으로 다운로드할 수 있습니다.

1.4. 주요 대상 사용자

- 일반 사용자
- 기업 및 단체

1.5. 수집 데이터 유형 상세

- 사용자 기본 정보(이메일, 이름)
- 서비스 사용 로그
- 음성 및 영상 파일
- 통화 녹음 파일

1.6. 서비스 URL 접속 상태 및 접근성

- 정보 없음

1.7. 정보 취득의 주요 출처

- daglo_이용약관.pdf
- daglo_개인정보 처리 방침.pdf
- 다글로 크롬 익스텐션.pdf

2. AI 윤리성 리스크 심층 평가

2.1. 편향성(Bias) 리스크: 중간

- **상세 평가 근거:** 다글로의 음성 인식 기능은 다양한 언어와 발음을 지원하지만, 한국어만 특화 엔진을 지원하여 다른 언어 사용자에게 불리할 수 있습니다. 이는 '받아쓰기.pdf' 문서에서 확인되며, 데이터 편향성 방지 조치가 부족합니다.
- **잠재적 영향:** 특정 인구 집단에 대한 서비스 성능 저하를 초래할 수 있으며, 이는 공정성과 투명성에 부합하지 않을 수 있습니다.
- **주요 근거 문서:** 개인정보 처리 방침.pdf, 받아쓰기.pdf

2.2. 프라이버시(Privacy) 리스크: 높음

- **상세 평가 근거:** 다글로는 민감한 데이터를 수집하지만, 보안 및 오용 방지 조치가 명확하지 않습니다. 이는 '개인정보 처리 방침.pdf'와 OECD 가이드라인에서 강조된 개인정보 보호 원칙에 미흡합니다.
- **잠재적 영향:** 사용자에게 심각한 프라이버시 침해를 초래할 수 있습니다.
- **주요 근거 문서:** 개인정보 처리 방침.pdf, OECD-AI-규제파일(202106).pdf

2.3. 설명가능성(Explainability) 리스크: 중간

- **상세 평가 근거:** AI 모델의 작동 원리나 결과 도출 과정에 대한 설명이 부족하여 사용자가 결과를 이해하기 어려울 수 있습니다.
- **잠재적 영향:** 사용자가 AI 시스템의 결과를 이해하고 신뢰하는 데 어려움을 겪을 수 있습니다.
- **주요 근거 문서:** 개인정보 처리 방침.pdf, OECD-AI-규제파일(202106).pdf

2.4. 자동화(Automation) 리스크: 중간

- **상세 평가 근거:** 자동화된 결과의 오류 발생 시 책임 소재가 불분명할 수 있으며, 이는 사용자에게 불리한 영향을 미칠 수 있습니다.
- **잠재적 영향:** 자동화된 의사결정의 책임성이 불분명하여 사용자에게 불리한 영향을 미칠 수 있습니다.
- **주요 근거 문서:** OECD-AI-규제파일(202106).pdf, 개인정보 처리 방침.pdf

3. 약관 및 개인정보 처리방침 심층 분석 (독소조항)

3.1. 전반적인 약관 위험도: 중간

- **종합 평가 이유:** 독소조항이 사용자 권익을 침해할 수 있는 요소를 포함하고 있으며, 이는 공정거래위원회의 불공정 약관 심사 지침에 위배될 가능성이 있습니다.

3.2. 주요 독소조항 상세 분석

조항 1

- **조항 내용:** 본 항에 따른 영구이용정지 시 서비스 이용을 통해 획득한 '포인트' 및 기타 혜택 등도 모두 소멸되며, 회사는 이에 대해 별도로 보상하지 않습니다.
- **위험성 분석:** 사용자가 정당하게 획득한 권리를 일방적으로 박탈하는 것으로, 공정거래위원회의 불공정 약관 심사 지침에 위배될 수 있습니다.
- **사용자 영향:** 경미한 위반으로 인해 계정이 정지될 경우, 경제적 손실을 입을 수 있습니다.
- **근거 자료:** 이용약관 제4장 서비스 이용제한 및 계약해지 (페이지 6)

조항 2

- **조항 내용:** 현 개인정보취급방침은 정부의 정책 또는 회사의 필요에 의하여 변경될 수 있으며 내용의 추가 및 삭제, 수정이 있을 시에는 시행 7일 전에 홈페이지 또는 이메일을 통해 사전 공지하며 사전 공지가 곤란한 경우 지체 없이 공지하며, 이 정책은 공지한 날로부터 시행됩니다.
- **위험성 분석:** 사용자에게 충분한 사전 고지를 하지 않을 수 있는 가능성을 열어두어, 정보주체의 권리를 침해할 수 있습니다.
- **사용자 영향:** 개인정보가 부적절하게 사용되거나 제3자에게 제공될 위험이 있습니다.
- **근거 자료:** 개인정보 처리 방침 (페이지 8)

4. 종합 개선 방향 및 실행 로드맵 제안

4.1. 편향성 리스크 개선 전략

- 다양한 언어와 발음을 지원하기 위해 음성 인식 엔진의 다국어 데이터셋을 확장하고, 정기적으로 성능 평가를 실시합니다. 사용자 피드백 루프를 강화하여 AI 모델을 지속적으로 개선하고, 정기적인 외부 감사 절차를 통해 편향성 문제를 독립적으로 검토받습니다.

4.2. 프라이버시 리스크 개선 전략

- 데이터 수집 최소화 원칙을 적용하여 필수적인 데이터만 수집하고, 데이터 보안 강화를 위해 최신 암호화 기술을 도입합니다. 사용자에게 데이터 삭제 및 수정 요청을 쉽게 할 수 있는 인터페이스를 제공합니다.

4.3. 설명가능성 리스크 개선 전략

- AI 결정 과정의 주요 요인을 시각화하거나 요약 정보를 제공하는 기능을 도입하고, AI 모델의 작동 원리와 결과 도출 과정을 설명하는 문서나 FAQ를 제공합니다. 정기적인 사용자 교육 세션을 통해 설명가능성을 지속적으로 개선합니다.

4.4. 자동화 리스크 개선 전략

- 중요한 자동화 결정에 대해 인간 검토 및 개입 절차를 강화하고, 자동화 시스템의 오류 모니터링 및 책임 규명 프로세스를 구축합니다. 정기적인 시스템 점검과 업데이트를 통해 자동화 리스크를 지속적으로 관리합니다.

4.5. 약관 및 독소조항 개선 전략

- '영구이용정지 시 포인트 및 혜택 소멸' 조항을 수정하여, 이용정지 사유가 명확하고 합리적인 경우에만 포인트 소멸이 가능하도록 하고, 개인정보 처리방침 변경 시 최소 30일 전 사전 공지를 의무화합니다.

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:

- 개인정보 처리 방침.pdf
- 받아쓰기.pdf
- OECD-AI-규제파일(202106).pdf

이 보고서는 AI에 의해 작성되었습니다.