# AI 윤리 리스크 진단 보고서

진단 대상: gemini

진단 일시: 2025-10-23 15:53

## Ⅰ. 서비스 개요

서비스 유형: 생성형 AI

주요 목적: Gemini 요약

기술적 구조:
Gemini는 구글 딥마인드가 개발한 차세대 인공지능 시스템으로, 딥러닝, 강화 학습, 대규모 데이터 처리를 통합하여 멀티모달 기능을 제공합니다. 텍스트, 이미지, 오디오, 비디오 및 코드 전반을 처리할 수 있는 대규모 모델로 설계되었습니다. Gemini는 Google Cloud의 Vertex AI 플랫폼을 통해 개발자들이 활용할 수 있도록 지원합니다.

서비스 목적:
Gemini는 실시간 대화 지원, 동영상 요약, 로봇 제어, 의료 진단 지원 등 다양한 애플리케이션에 활용되며, 사용자 맞춤형 응답을 제공하여 통합적이고 일관된 사용자 경험을 목표로 합니다. 이를 통해 소비자와 개발자 모두에게 다재다능한 AI 도구를 제공합니다.

주요 기능:
1. 멀티모달 처리: 텍스트, 이미지, 오디오, 비디오를 모두 처리하여 다양한 데이터 유형을 분석하고 요약할 수 있습니다.
2. 실시간 상호작용: 사용자 데이터를 기반으로 맞춤형 응답을 제공하며, 자율적인 작업 실행을 지원합니다.
3. 생산성 도구 강화: 예를 들어, Med-Gemini 모델은 의료 분야에 특화되어 있습니다.
4. AI 통합: Google 생태계 전반에서 작동하여 다양한 플랫폼과 기기에서 접근할 수 있습니다.
5. 개발자 지원: Google AI SDK와 API를 통해 개발자들이 Gemini의 기능을 쉽게 통합하고 활용할 수 있도록 돕습니다.

Gemini는 지속적으로 발전하며, AI 분야의 새로운 기준을 제시하고 있습니다.

## Ⅱ. 초기 윤리 리스크 평가

공정성 (Fairness): 4.0점
1. 공정성 (Fairness): 4점

편향성 (Bias): 3.0점
2. 편향성 (Bias): 3점

투명성 (Transparency): 4.0점
3. 투명성 (Transparency): 4점

설명가능성 (Explainability): 3.0점
4. 설명가능성 (Explainability): 3점

프라이버시 (Privacy): 5.0점
5. 프라이버시 (Privacy): 5점

## Ⅲ. 사용자 피드백

hallucination, 저작권 침해

## Ⅳ. 피드백 반영 후 재평가 결과

: 3.0점
5. 프라이버시 (Privacy): 3점

## Ⅴ. 최종 개선 권고안

AI 서비스의 윤리 리스크 평가 결과에서 프라이버시와 관련된 리스크가 3점으로 평가되었습니다. 이를 개선하기 위한 구체적인 권고안을 제시하고, EU, OECD, UNESCO 가이드라인의 원칙과 연결하여 설명하겠습니다.

### 1. 데이터 수집 및 처리의 투명성 증대
권고안: 사용자에게 데이터 수집 및 처리에 대한 명확한 정보를 제공하고, 어떤 데이터가 수집되는지, 그 데이터가 어떻게 사용되는지를 투명하게 설명해야 합니다. 또한, 사용자가 데이터 처리에 대한 동의를 쉽게 철회할 수 있는 방법을 마련해야 합니다.

가이드라인 연결: EU의 일반 데이터 보호 규정(GDPR)에서는 데이터 주체의 권리를 강조하며, 데이터 수집 및 처리의 투명성을 요구합니다. OECD의 프라이버시 원칙에서도 개인 정보의 수집과 사용에 대한 투명성을 강조하고 있습니다.

### 2. 개인정보 보호를 위한 기술적 조치 강화
권고안: AI 시스템에서 수집된 개인정보를 안전하게 보호하기 위해 암호화, 익명화 등의 기술적 조치를 강화해야 합니다. 또한, 데이터 접근 권한을 최소화하고, 데이터가 불필요하게 저장되지 않도록 주기적인 검토 및 삭제 프로세스를 마련해야 합니다.

가이드라인 연결: UNESCO의 AI 윤리 권고안에서는 데이터 보호 및 개인정보의 안전한 처리를 강조하며, 기술적 조치를 통해 개인의 프라이버시를 보호할 것을 권장합니다.

### 3. 사용자 권리 강화 및 교육
권고안: 사용자가 자신의 데이터에 대한 권리를 이해하고 행사할 수 있도록 교육 프로그램을 제공해야 합니다. 사용자는 자신의 데이터에 대한 접근, 수정, 삭제 요청을 할 수 있는 권리를 가져야 하며, 이러한 권리를 쉽게 행사할 수 있는 방법을 마련해야 합니다.

가이드라인 연결: OECD의 프라이버시 원칙에서는 개인이 자신의 데이터에 대한 접근 및 수정 권리를 가져야 한다고 명시하고 있습니다. 이는 사용자에게 권한을 부여하고, 프라이버시를 존중하는 데 기여합니다.

### 4. 비상 상황에 대한 대응 계획 수립
권고안: 데이터 유출이나 프라이버시 침해와 같은 비상 상황에 대비한 대응 계획을 수립하고, 이를 정기적으로 점검해야 합니다. 또한, 이러한 상황 발생 시 사용자에게 신속하게 알릴 수 있는 체계를 마련해야 합니다.

가이드라인 연결: EU의 GDPR에서는 데이터 유출 발생 시 72시간 이내에 관련 당국 및 영향을 받는 사용자에게 통지할 것을 요구하고 있습니다. 이는 사용자 보호와 신뢰 구축에 기여합니다.

이와 같은 권고안을 통해 AI 서비스의 프라이버시 관련 리스크를 효과적으로 개선하고, 사용자 신뢰를 구축할 수 있을

것입니다.

※ 본 보고서는 Human-in-the-loop 기반 AI 윤리 평가 결과입니다.