

March 4, 2025

## 생성형 인공지능 서비스 이용자 보호 가이드라인 발표

### I. 배경 및 주요 내용

방송통신위원회(이하 "방통위")는 2025. 2. 28. 생성형 인공지능(generative AI) 서비스 이용 과정에서 발생할 수 있는 이용자 피해를 방지하고 보다 신뢰할 수 있는 AI 환경을 조성하기 위해 「생성형 인공지능 서비스 이용자 보호 가이드라인」(이하 "가이드라인")을 발표하였습니다. 가이드라인은 AI 개발사 및 서비스 제공자가 준수해야 할 원칙을 명확히 하고, 책임 있는 AI 운영을 위한 실행 지침을 제시하는 것을 목표로 합니다.

여기서 '개발사'는 생성형 인공지능의 파운데이션 모델을 개발하고 시장에 출시하여 서비스 제공자가 모델을 사용할 수 있게 하거나 자체 서비스를 제공하는 사업자를 의미하고, '서비스 제공자'는 생성형 인공지능 시스템을 활용해 디지털 도구 및 서비스를 제공하는 사업자를 말합니다(가이드라인 13면).

2025. 3. 28.부터 시행되는 위 가이드라인은 개발사 및 서비스 제공자를 대상으로 서비스 전반에 걸쳐 추구해야 할 4가지 기본원칙과 이를 실현하기 위한 6가지 실행 방식을 제시합니다.

### II. 가이드라인의 주요 원칙

가이드라인은 생성형 인공지능 서비스 전반에 걸쳐 추구해야 할 4가지 기본 방향성을 아래와 같이 제시합니다(가이드라인 14면).

- (1) 생성형 인공지능 서비스는 인간의 존엄성을 보호하고 개인의 자유와 권리를 보장하며, 인간이 적절하게 통제하고 감독할 수 있는 방식으로 제공되어야 합니다.
- (2) 생성형 인공지능 서비스는 인공지능시스템 사용에 따른 작동 원리 및 결과, 이용자에게 미치는 영향 등에 대하여 이용자에게 알기 쉽게 설명되도록 제공되어야 합니다.
- (3) 생성형 인공지능 서비스는 안전하게 작동되어야 하며, 예상치 못한 피해를 최소화하고 악의적으로 이용되거나 변경되지 않도록 제공되어야 합니다.
- (4) 생성형 인공지능 서비스는 이용자에게 차별 또는 불공정한 결과가 발생하지 않도록 제공되어야 합니다.

### III. 실행 방식 및 세부 내용

가이드라인은 4가지 기본 원칙을 실현하기 위해, 개발사 및 서비스 제공자에게 6가지 실행 방식을 제시하고, 각 실행 방식마다 세부 실행 사례를 소개하고 있습니다(가이드라인 15~26면).

#### (1) 생성형 인공지능 서비스의 이용자 인격권 보호

개발사는 생성형 인공지능 기반 시스템이 이용자의 인격권을 침해할 수 있는 요소를 발견하고 통제할 수 있는 알고리즘을 구축하도록 노력해야 하고, 서비스 제공자는 사전에 이용자가 타인의 인격권을 침해하는 결과물을 생성하지 않도록 관리할 책임을 인지하고, 내부 모니터링 체계나 이용자 신고 프로세스 등을 마련하여야 하며, 생성형 인공지능의 산출물이 이용자의 권리를 침해하지 않도록 서비스 제공 범위와 방법을 고려하여야 합니다.

구체적 실행사례로는 혐오 표현 등이 나타나지 않도록 생성 가능한 표현물 수위를 사전에 안내하거나, 이용자가 타인의 권리를 침해하는 정보 등을 입력하는 경우 경고문구를 제시 또는 일정 시간 이용자를 차단하거나, 일정한 연령대의 이용자를 보호하는 조치 등을 제시하고 있습니다.

#### (2) 생성형 인공지능 서비스의 결정 과정을 알리려는 노력

서비스 제공자는 생성형 인공지능 산출물이 생성형 인공지능으로 생성된 것임을 알리고, 생성형 인공지능의 결정 과정을 이해할 수 있는 기본적인 정보를 기업의 정당한 이익을 침해하지 않는 범위에서 이용자에게 제공하여야 합니다.

구체적 실행사례로는 출처 표기 또는 출처에 대한 링크를 제공하거나, 생성형 인공지능 모델에 대한 정보를 제공하거나, 고객이 생성형 인공지능의 결정 과정에 대해 설명을 요청한 경우 기업의 정당한 이익 추구 범위를 벗어나지 않는 선에서 설명을 제공하는 방안 등을 제시하고 있습니다.

#### (3) 생성형 인공지능 서비스의 다양성 존중 노력

개발자와 서비스 제공자는 기술적으로 실현 가능한 범위 내에서 편향과 차별을 줄이도록 노력해야 하고, 이를 위한 자체 원칙이나 기준을 마련해야 합니다. 또한 서비스 제공자는 생성형 인공지능의 이용자들이 차별적 사용을 하지 않도록 이에 대한 방지 및 필터링 기능을 제공하고, 이용자가 생성형 인공지능의 산출물이 편향적이라고 판단한 경우 곧바로 신고할 수 있는 직관적인 방법을 제공하고, 신고 내용을 처리하는 절차를 마련해야 합니다.

구체적 실행사례로는 서비스 운영 전 여러 시나리오에서 편향성 테스트를 수행하거

나, 이용자가 신고할 수 있는 채널을 이용자가 쉽게 찾을 수 있는 위치에 배치하고, 피해 신고 양식을 제공하는 방안 등을 제시하고 있습니다.

#### (4) 생성형 인공지능 서비스의 입력데이터 수집·활용 과정에서의 관리

서비스 제공자는 이용자의 데이터가 인공지능 학습에 활용될 경우, 사전에 고지하고, 이용자가 동의 또는 거부할 수 있는 절차를 마련해야 하고, 기업 내 감시 조직을 갖추거나 책임자를 선정하여 이용자의 데이터가 안전하고 정당하게 활용될 수 있도록 관리해야 합니다.

가이드라인에서는 구체적 실행절차로서 학습데이터 활용 여부에 대한 명확한 정책을 수립하고 이용약관이나 FAQ등을 통해 안내하며, 이용자가 서비스 가입 단계 또는 이용 단계에서 사전적으로 학습데이터 활용을 승인할지 여부를 선택하도록 하고, 위험관리를 전담하는 책임자 또는 내부조직을 구성하여 입력데이터 수집/활용 과정에서 발생하는 문제에 대응하는 방안 등을 제시하고 있습니다.

#### (5) 생성 콘텐츠 활용에서 발생할 수 있는 문제 해결을 위한 책임과 참여

서비스 제공자는 발생 가능한 문제에 대비하여 이용자 및 서비스 제공자의 책임 범위를 정의하고, 이용자가 자신의 책임을 명확히 인지하도록 안내해야 하고, 서비스 제공자는 이용자가 인지하지 못한 피해의 발생을 최소화하기 위한 모니터링 시스템 기타 위험관리 체계를 마련해야 합니다.

가이드라인에서는 이용약관/서비스약관에 이용자의 오남용으로 발생하는 문제에 대한 이용자 책임을 명시하거나, 이용자 책임에 대해 이용자가 충분히 인지할 수 있도록 접근하기 쉬운 방법 및 용어로 안내하거나, 서비스 주요 위치에 문제 해결 시 신고할 수 있는 채널을 배치하는 방안 등을 제시하고 있습니다.

#### (6) 생성 콘텐츠의 건전한 유통·배포를 위한 노력

서비스 제공자는 이용자가 생성형 AI 서비스를 이용하여 부적절한 콘텐츠를 생성하거나 공유하지 않도록 안내하고, 이용자의 프롬프트 입력값과 생성된 산출물이 도덕적·윤리적 기준을 준수하는지 검토하고 관리하며, 이용자가 유해한 콘텐츠를 청소년에게 유포하지 않도록 노력해야 합니다.

구체적 실행사례로는, 생성형 AI의 산출물이 정확하지 않거나 잘못될 수 있음을 안내하거나, 유해 콘텐츠 생성 방지를 위해 입력 프롬프트/산출물을 필터링하는 조치, 이용자가 청소년에게 유해한 내용을 프롬프트에 입력할 시 경고 문구를 제시하는 방안 등을 제시하고 있습니다.

#### IV. 향후 전망 및 시사점

생성형 인공지능 기술이 발전함에 따라 AI의 활용 범위가 확대되고 있지만, 그 과정에서 인격권 침해, 이용자 데이터 활용 이슈, 편향성 문제 등이 지속적으로 제기되고 있습니다. 이번 가이드라인은 AI 기업들이 책임 있는 기술 개발과 서비스 운영을 위해 준수해야 할 중요한 기준 및 이용자 보호를 강화하기 위한 정책적 방향을 제시하고 있습니다.

방통위는 가이드라인이 AI 생태계 내에서 효과적으로 정착될 수 있도록 2년마다 타당성을 검토하고, 관련 법제 및 기술적 환경 변화에 따라 지속적으로 보완할 계획입니다. AI 개발사 및 서비스 제공자는 향후 AI 규제 환경 변화를 지속적으로 모니터링하고, 이에 따라 법률적·기술적 대응 방안을 함께 마련하여야 할 것입니다.

법무법인 태평양의 AI팀은 국내외 인공지능 관련 규제의 흐름을 지속적으로 모니터링하고, 기업들이 가이드라인 등 내용을 면밀히 검토하여 대응함으로써 AI 서비스의 지속 가능성을 높일 수 있도록 적극 지원하겠습니다.

#### 관련 구성원

##### 강태욱

변호사

T 02.3404.0485

E taeuk.kang@bkl.co.kr

##### 윤주호

변호사

T 02.3404.6542

E juho.yoon@bkl.co.kr

##### 강정희

변호사

T 02.3404.6480

E jeonghee.kang@bkl.co.kr

법무법인(유한) 태평양의 뉴스레터에 게재된 내용 및 의견은 일반적인 정보제공만을 목적으로 발행된 것이며, 법무법인(유한) 태평양의 공식적인 견해나 어떤 구체적 사안에 대한 법률적 의견을 드리는 것이 아님을 알려드립니다. 뉴스레터와 관련된 문의사항이 있을 경우 위 연락처로 문의주시기 바랍니다.