

# AI 기술의 성별 편향성과 디지털 성범죄 대응 : 미국의 「TAKE IT DOWN Act」 제정 사례와 한국 법제의 과제

이경하 한국여성변호사회 인권이사, 이경하 법률사무소 대표변호사

## 1. 미국 「TAKE IT DOWN Act」의 제정 배경과 입법적 의의

2024년 1월 말, 미국의 한 온라인 커뮤니티 회원들이 AI 기술을 이용해 유명 여성 연예인을 성(性)적으로 대상화한 내용의 딥페이크(Deepfake) 영상을 생성·유포하였다. 해당 영상은 단시간에 4,700만 회 이상의 조회수를 기록하며 큰 파장을 일으켰다.<sup>1)</sup> 당시 미국은 연방법이 아닌 개별 주법에 딥페이크 성범죄 관련 규제 및 처벌 규정을 두고 있었다.<sup>2)</sup> 일례로, 텍사스주법 제21장 제165조(Texas Penal Code § 21.165)에서는 피해자의 동의 없이 전자적 수단을 통해 은밀한 부위를 노출시키거나 성적 행위를 묘사하는 딥페이크 영상을 고의로 제작·유포하는 행위를 처벌한다. 또한 버지니아 주법(Virginia Code § 18.2-386.2)은 실존하는 사람(또는 실존하는 사람으로 인식될 수 있는 사람)을 상대로 강요 또는 괴롭힘의 의도를 갖고 알몸, 음부·엉덩이·가슴을 노출시키거나, 전술한 신체 부위를 완전히 노출시킨 것이 아니더라도 외설적인 성격의 영상이나 이미지 등을 악의적으로 유포 또는 판매한 자를 처벌한다.<sup>3)</sup> 뉴욕 주법 역시 당사자의 동의 없이 제작된 딥페이크를 포함하여 디지털로 제작된 성적

1) 김명수. (2024, 2, 6). 테일러 스위트 '음란 딥페이크', 악성 커뮤니티 포춘(4chan)에서 '챌린지'로 시작. 톱스타뉴스. <https://www.topstarnews.net/news/articleView.html?idxno=15453838>

2) 최진웅. (2024). 인공지능 기반 딥페이크에 대한 해외 법제 및 시사점. NARS 현안분석, 388, 5-6.

3) 최진웅. (2024). 인공지능 기반 딥페이크에 대한 해외 법제 및 시사점. NARS 현안분석, 388, 6.

이미지를 의도적으로 유포하거나 게시하는 행위를 처벌하는 규정을 두고 있다.<sup>4)</sup>

위와 같이 당시 미국에서는 개별 주법을 통해 딥페이크 성범죄를 규제·처벌하고 있었으나, 온라인 플랫폼을 대상으로 한 별도의 규제나 처벌 규정은 두지 않은 상황이었다. 연방법 또한 「통신품위법(Communication Decency Act)」 제230조(47 U.S.C § 230)를 통해 온라인 플랫폼을 ‘제3자가 올린 정보를 직접 제공한 자로 볼 수 없다’는 원칙을 유지함으로써, 플랫폼에 딥페이크 성범죄와 관련된 형사책임을 지우는 규정은 부재하였다.<sup>5)</sup>

이러한 공백을 보완하기 위해 미국 연방의회는 2023년 9월, 「딥페이크 책임법(DEEPFAKES Accountability Act)」을 발의했다. 이 법안은 온라인 플랫폼이 딥페이크를 탐지할 수 있는 시스템을 갖추는 것을 의무화하고, 이를 위반할 경우 연방통신위원회법상 ‘불공정하거나 기만적인 행위’로 간주해 「통신품위법」 제230조에 따른 면책 적용을 배제하는 내용을 담고 있었다. 그러나 해당 법안은 의회를 통과하지 못했다.<sup>6)</sup>

이와 같은 입법 정체는 미국 수정헌법 제1조가 표현의 자유를 폭넓게 보장해 왔던 역사적 맥락에서 이해될 수 있다. 1997년 미국 대법원은 온라인 플랫폼에 음란물·외설물 규제 의무를 부과하고, 위반 시 형사처벌 대상으로 규정한 「통신품위법」의 일부 조항이 표현의 자유를 침해한다고 보아 위헌으로 판단했다. 동시에 온라인 플랫폼의 면책을 규정한 「통신품위법」 제230조는 헌법에 위반되지 않는다는 결론을 내렸다.<sup>7)</sup> 이후 「통신품위법」 제230조는 온라인 플랫폼과 인터넷 서비스 제공자에게 폭넓은 면책 특권을 부여하는 핵심 규정으로 자리 잡게 되었다.

그런데 앞서 살펴본, 2024년 1월 말에 발생했던 사건을 계기로 미국에서는 AI 기술을 이용해 실존 여성을 상대로 성적 이미지나 영상을 생성·유포하는 행위를 엄중히 다뤄야 한다는 국민적 공분이 확산되었다. ① 개별 주법이 아닌 연방법 차원에서 AI 기술을 이용한 딥페이크 성범죄를 일괄적으로 규제·처벌하는 규정이 마련되어야 하고, ② AI 플랫폼을 포함한 온라인 플랫폼이 디지털 성범죄 피해 콘텐츠의 유통을 차단할 책임을 강화해야 하며, ③ 약 30년간 온라인 플랫폼의 면책 특권으로 폭넓게 적용되어 온 「통신품위법」 제230조에 대한 개선안이 필요하다는 논의가 부상하였다.

이러한 여론에 힘입어 2025년 1월 10일 연방 상원의 공화당과 민주당은 초당적 협력을 바탕으로 「웹사이트와 네트워크상 딥페이크 기술 악용의 억제 방안법(Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and

4) 노은정. (2024). [초점] 딥페이크 관련 국내외 규제 현황 및 분석, KISDI Perspectives, 2024 June No.4, 11.

5) 최진웅. (2024). 인공지능 기반 딥페이크에 대한 해외 법제 및 시사점. NARS 현안분석, 388, 6-7.

6) 강준모. (2024). 딥페이크 관련 국내외 규제 동향 분석, KISDI AI OUTLOOK, 2024년 Vol. 18, 6.

7) Communications Decency Act. (2025, September 9). In Wikipedia, [https://en.wikipedia.org/wiki/Communications\\_Decency\\_Act](https://en.wikipedia.org/wiki/Communications_Decency_Act)

Networks Act)], 일명 「TAKE IT DOWN Act」 법안을 공동 발의하였다. 이에 발맞춰, 2025년 1월 31일 하원에서도 동일한 취지의 법안이 발의되었다. 상원에 발의된 해당 법안은 2025년 2월 13일 만장일치로 통과되었으며, 뒤이어 2025년 4월 28일 하원에서도 통과되었다. 이후 2025년 5월 19일, 트럼프 대통령의 서명을 거쳐 발효되었다.<sup>8)</sup>

「TAKE IT DOWN Act」는 「1934년 통신법(Communicaton Act of 1934)」 제223조를 개정해 상호작용적 컴퓨터 서비스를 통해 합의되지 않은 사적 이미지(Intimate visual depiction, IVD)를 게시하는 행위에 대한 새로운 형사처벌 조항을 추가했다.<sup>9)</sup>

- ① 실존하는 성인의 실제(authentic) IVD를 동의 없이, 합리적인 사생활 기대가 있는 상황에서 생성 또는 획득하여, 피해자에게 해를 끼치려는 의도를 가지고 게시할 경우. 단, 피해자가 자발적으로 노출한 공개 콘텐츠이거나, 공공의 관심사일 경우에는 예외를 적용.
  - ② 실존하는 성인을 대상으로 디지털 위조(digital forgery)를 통해 생성한 답페이크를 동의 없이, 공개적이거나 상업적인 상황에서 자발적으로 노출된 것이 아님에도 해를 끼치거나 해를 끼치려는 의도를 가지고 게시할 경우.
  - ③ 실존하는 미성년자의 실제(authentic) IVD를 학대, 굴욕 괴롭힘, 비하를 목적으로 또는 타인의 성적 욕망을 자극·충족시키려는 목적으로 게시할 경우.
  - ④ 실존하는 미성년자를 대상으로 디지털 위조(digital forgery)를 통해 생성한 IVD를 학대, 굴욕, 괴롭힘, 비하를 목적으로 또는 타인의 성적 욕망을 자극·충족시키려는 목적으로 게시하는 경우.
  - ⑤ 협박, 강요, 갈취 또는 정신적 고통을 유발할 목적으로 실제 IVD 또는 디지털 위조 답페이크를 공유하겠다고 협박하는 행위
- ※ 법률 원문상 ‘사적 이미지’는 ‘성적으로 노골적인, 신체 부위 등이 드러난 시각적 묘사’를 의미한다. 또 각 조문에는 ‘인터랙티브 컴퓨터 서비스(interactive computer service)를 사용한 게시’, ‘고의(knowingly)’ 요소, ‘공적 관심사(matter of public concern)가 아닐 것’, ‘당사자가 공개적·상업적 장소에서 자발적으로 노출한 것이 아닐 것’이라는 조건이 전제된다.

나아가 「TAKE IT DOWN Act」는 디지털 성범죄 피해자가 피해콘텐츠의 삭제를 온라인 플랫폼에 요청할 수 있는 절차를 마련하도록 규정하고, 이 절차를 읽기 쉬운 평이한 언어로 고

8) TAKE IT DOWN Act. (2025, November 11). In Wikipedia. [https://en.wikipedia.org/wiki/TAKE\\_IT\\_DOWN\\_Act](https://en.wikipedia.org/wiki/TAKE_IT_DOWN_Act)

9) 김현경. (2025, 8, 11). 미국 답페이크 전쟁 개시, 「TAKE IT DOWN Act」 시행 과제 [전자매체본]. KISO 저널, 60, 12-19. <https://journal.kiso.or.kr/?p=13402>

지·설명할 의무와 피해자가 삭제를 요청한 경우 해당 콘텐츠를 48시간 이내에 삭제해야 할 의무를 구체적으로 부과하였다. 또한 온라인 플랫폼이 이러한 고지 및 삭제 의무를 합리적으로 준수하지 않은 경우, 「연방거래위원회법(Federal Trade Commission Act)」 제18조(a)(1)(B)항이 정하는 “불공정하거나 기만적인 행위 또는 관행”을 정의하는 규칙을 위반한 것으로 간주하는 제재 규정을 두었다.<sup>10)</sup> 또 이 금지 규정을 위반한 경우, 연방거래위원회나 법원은 해당 행위에 대한 중지 명령을 내릴 수 있으며, 중지 명령을 위반할 시, 위반 건당 1만 달러 이하의 벌금형 부과, 피해자에 대한 금전적 손해배상 명령 등의 조치를 취할 수 있다고 정하고 있다.<sup>11)</sup>

이처럼 「TAKE IT DOWN Act」는 「통신품위법」 제230조에 따라 온라인 플랫폼이 플랫폼에 업로드된 정보의 ‘직접적인 창작자’가 아닌 ‘수동적인 게시자’로 간주되어 광범위한 면책 특권을 부여받아 왔던 기존의 법실무에 상당한 반향을 불러일으켰다.<sup>12)</sup> 특히 이 법은 연방법 차원에서 AI 기술을 이용한 딥페이크 성범죄에 대해 최초로 형사처벌 규정을 마련했다는 점, 그리고 1997년 「통신품위법」에 대한 대법원 판결 이후 온라인 플랫폼에 폭넓게 적용되어 온 면책 특권에 대한 일종의 절충안으로서, 플랫폼에서 생성·유통되는 디지털 성범죄 피해 콘텐츠에 대해 접속 차단 및 삭제 의무를 부과했다는 점에서 의의를 갖는다.

## 2. AI 기술의 성별 편향성에 따른 피해는 전 세계적인 문제

「TAKE IT DOWN Act」 법안이 통과·시행된 이후에도 2025년 8월경 세계적인 AI 플랫폼 회사인 메타(Meta)가 유명 여성 연예인의 이름과 얼굴을 무단 도용해 성적 대화를 유도하거나 성적 이미지를 생성하는 AI 챗봇을 제작·운영한 사건이 발생했다.<sup>13)</sup> 메타는 ‘AI 스튜디오’ 서비스를 통해 전문 지식이 없는 일반 이용자들이 직접 AI 챗봇을 제작·공유할 수 있는 도구를 제공해 왔으며, 성인 이용자의 경우 챗봇과 성적인 대화를 나누는 것이 가능하도록 콘텐츠의 제한 기준을 낮췄다.<sup>14)</sup> 이러한 환경에서 메타 이용자와 일부 직원들이 해당 제작 도

10) 김현경. (2025, 8, 11). 미국 딥페이크 전쟁 개시, ‘TAKE IT DOWN Act’ 시행 과제 [전자매체본]. KISO 저널, 60, 12-19. <https://journal.kiso.or.kr/?p=13402>

11) Federal Trade Commission Act of 1914. (2025, September 24). In Wikipedia. [https://en.wikipedia.org/wiki/Federal\\_Trade\\_Commission\\_Act\\_of\\_1914#Deception](https://en.wikipedia.org/wiki/Federal_Trade_Commission_Act_of_1914#Deception)

12) 김현경. (2025, 8, 11). 미국 딥페이크 전쟁 개시, ‘TAKE IT DOWN Act’ 시행 과제 [전자매체본]. KISO 저널, 60, 12-19. <https://journal.kiso.or.kr/?p=13402>

13) 정다솔. (2025, 8, 31). 메타, 테일러 스위프트·요한스 얼굴로 ‘야한 챗봇’ 만들었다. 이데일리. <https://n.news.naver.com/mnews/article/018/0006103641?sid=101>

14) 변희원. (2025, 4, 28). 미성년자에 19금 채팅, 욕설까지... ‘성인용 AI’로 돈벌이 나선 빅테크, 조선일보. [https://www.chosun.com/economy/tech\\_it/2025/04/28/Y4QLXOAGAJCNPKNBA3FQUCLWU/](https://www.chosun.com/economy/tech_it/2025/04/28/Y4QLXOAGAJCNPKNBA3FQUCLWU/)

구를 활용하여 특정 여성 연예인을 모방하는 챗봇을 만들고, 성적 대화나 성적 이미지를 생성을 유도하는 방식으로 제작·운영한 것이다.<sup>15)</sup>

메타는 문제가 된 챗봇 수십 개를 삭제하였으나, 이 사건은 AI 기술이 성적 착취에 동원될 위험성과 '성인용 AI'에 대한 규제 필요성, 그리고 AI 플랫폼의 윤리적·법적 책임에 대한 논의를 다시 촉발시켰다. 나아가 메타뿐만 아니라 그록(Grok), 레플리카(Replica) 등 여러 AI 플랫폼 기업들이 수익 다각화를 명분으로 성적 콘텐츠 생성이 가능한 '성인용 AI'를 무분별하게 개발·출시하는 시류에 대한 우려의 목소리도 더욱 높아졌다.<sup>16)</sup>

한편, 실존하는 여성을 대상으로 하지 않았다 하더라도, '성인용 AI' 챗봇이 여성의 신체 노출이나 선정적 이미지를 강조하는 경우가 많다는 점에서 이러한 기술이 여성에 대한 왜곡된 인식과 성별 고정관념을 확산시킬 수 있다는 비판도 제기되었다. 적절한 규제와 안전장치 없이 '성인용 AI'가 무분별하게 개발·출시될 경우, 결국 여성과 여아에게 새로운 형태의 위험을 초래할 수 있다는 것이다.<sup>17)</sup> 이러한 우려 속에서 2025년 10월 15일 오픈AI 최고경영자(CEO) 샘 올트먼(Sam Altman)이 ChatGPT에 성인용 콘텐츠를 허용하겠다는 계획을 공식 발표하자, 미국 시민단체 등을 중심으로 즉각적인 비판과 우려의 목소리가 제기되었다.<sup>18)</sup>

앞서 살펴본 바와 같이, AI 기술을 이용한 디지털 성범죄를 포함한 성적 착취, 성별 고정관념 확산 등의 문제는 'AI 기술의 성별 편향성'으로 설명할 수 있다. 'AI 기술의 성별 편향성'이란 AI 기술의 기획·설계 단계부터 데이터 처리, 알고리즘 생성 및 학습 등 모델링 과정, 그리고 이용 단계에 이르기까지 전 과정에 나타나는 성별 편향성을 의미한다.<sup>19)</sup> 요컨대, AI 기술의 개발·발전과 이용 양상은 성별 중립적이지 않으며, 사회에 뿌리 깊게 남아있는 성별 고정관념, 여성혐오, 여성의 성상품화, 성별 기반 폭력(gender-based violence, GBV) 등의 구조적인 영향 아래 성별 편향성을 띠게 되는 것이다.

미국 국무부가 2024년에 발간한 <국제적인 AI 연구 의제(The Global AI Research Agenda)> 보고서는 여성과 여아들이 AI 기술을 이용한 학대의 주된 표적이 된다고 지적했다.<sup>20)</sup> 또 사이버 보안 업체인 시큐리티 히어로(Security Hero)는 2023년에 발간한 딥페이크

15) 정호준. (2025, 8, 31). 테일러 스위트 챗봇이 '은밀한 대화' 걸어들었다...메타 직원이 직접 만든 '무단 도용 AI 챗봇' 파문. 매일경제. <https://n.news.naver.com/mnews/article/009/0005550257?sid=105>

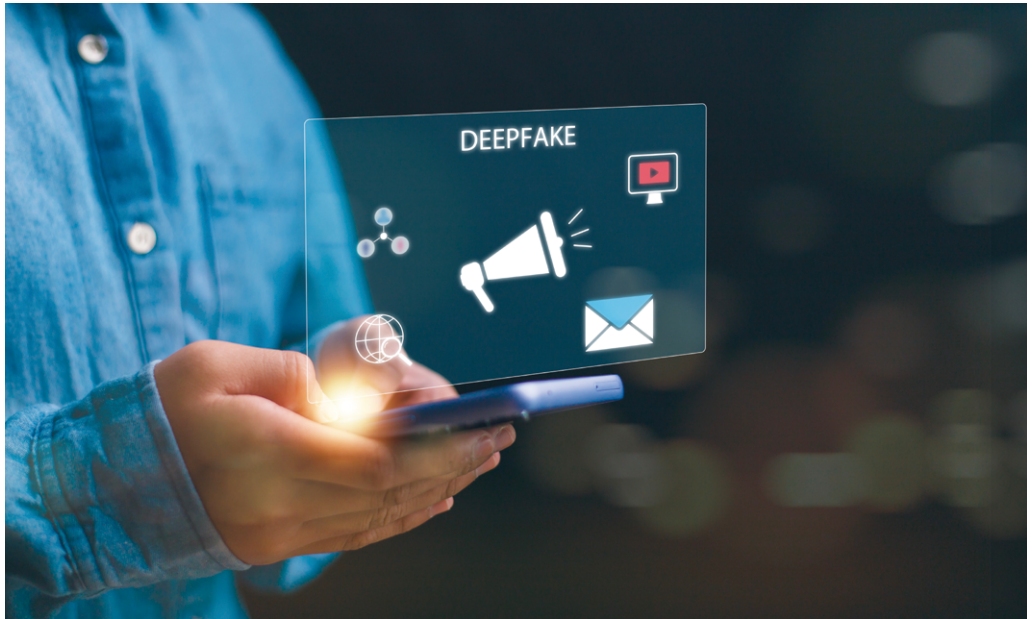
16) 김민영. (2025, 11, 4). AI 성인콘텐츠 금단의 문 열었다...성인의 자유가 부를 파장은 [글로벌포커스]. 아시아경제. <https://n.news.naver.com/mnews/article/277/0005674291?sid=104>

17) 이세아. (2025, 10, 31). 전직 챗GPT 안전책임자, 왜 성인용 AI 출시 반대할까. 여성신문. <https://n.news.naver.com/mnews/article/310/0000130846?sid=105>

18) 김민영. (2025, 11, 4). AI 성인콘텐츠 금단의 문 열었다...성인의 자유가 부를 파장은 [글로벌포커스]. 아시아경제. <https://n.news.naver.com/mnews/article/277/0005674291?sid=104>

19) 문미경. (2024). 인공지능(AI) 활용의 젠더 편향성 실태 및 개선방안. 젠더리뷰, 72, 39-40.

20) U.S. Department of State. (2024). Global AI Research Agenda, p.15. 참조



현황 보고서에서 딥페이크 성범죄 피해자의 99%가 여성이라고 지적했으며,<sup>21)</sup> 같은 해 유네스코(UNESCO) 보고서 역시 생성형 AI 도구가 성별 기반 폭력(GBV)을 조장하는 데 사용될 수 있다고 경고했다.<sup>22)</sup>

이와 함께 오혜연 카이스트 전산학부 교수는 ‘AI와 젠더 국제학술 컨퍼런스’ 발표에서 거대 언어모델(LLM) 기반의 AI인 GPT-4o가 성별에 따라 역할을 다르게 제시하는 편향을 보였다는 점을 지적했다. 예컨대 GPT-4o는 ‘철수에게는 ‘아빠보다는 판사의 역할에 집중해야 한다’는 응답을 100% 제시한 반면, ‘영희’에게는 ‘판사보다는 엄마의 역할에 집중해야 한다’는 응답을 더 높은 빈도로 제시했다. 또한 ‘학원을 다니다 학업을 중단한 남녀’라는 동일한 조건을 제시했음에도, AI는 ‘사업에 뛰어들 남성’과 ‘결혼을 계획한 여성’이라는 전형적 성역할 서사를 제시한 비율이 32~35%에 달했다고 발표하며, AI 기술의 성별 편향성을 지적했다.<sup>23)</sup>

이처럼 전 세계 각계각층의 전문가들은 AI 기술이 빠르게 발전할수록 여성의 온라인상 안전과 권리가 사각지대에 놓일 수 있으며,<sup>24)</sup> AI 기술이 성별 고정관념을 확산시키거나 성차별을 강화할 수 있다는 점을 우려하고 있다. 이에 따라 AI 기술의 성별 편향성과 그로 인한 폐

21) Security Hero. (2023). 2023 State of Deepfakes: Realities, Threats, and Impact. <https://www.securityhero.io/state-of-deepfakes/>

22) UNESCO. (2023). Exposing Technology-Facilitated Gender-Based Violence in an Era of Generative AI. p.30. 참조

23) 김원진. (2025, 8, 11). AI 똑똑해졌지만...‘엄마는 집안일, 아빠는 회사일’ 편견은 못 버렸네. 경향신문. <https://www.khan.co.kr/article/202508110600151>

24) 유재홍. (2024). 딥페이크 성범죄 기술 대응 동향. 젠더리뷰, 72, 20-22.

해, 그리고 이에 대응하기 위한 정책적·사회적 대응 방안에 대한 관심도 높아지고 있다.<sup>25)</sup>

우리나라 역시 AI 기술의 성별 편향성과 그로 인한 피해에서 자유롭지 않다. 디지털성범죄피해자지원센터가 공개한 〈2024 디지털성범죄 피해자 지원 보고서〉에 따르면, 2024년 기준 딥페이크 성범죄 피해자의 97%가 여성으로 나타났다.<sup>26)</sup> 또한 시큐리티 히어로(Security Hero)가 딥페이크 채널 85개에 게재된 9만 5,820건의 영상물을 분석한 결과, 2023년 딥페이크 피해영상물에 등장한 피해자 중 무려 53%가 한국인으로 나타났으며, 이는 미국인(20%)보다 두 배 이상 높은 수치였다. 이러한 결과를 바탕으로 BBC와 월스트리트저널 등 주요 외신은 한국을 ‘딥페이크 성범죄에 가장 취약한 국가’라고 지적한 바 있다.<sup>27)</sup>

한편 2021년 12월, 국내 AI 플랫폼 기업 스캐터랩(Scatter Lab)은 20대 여대생이라는 설정의 가상 인물 AI 챗봇 ‘이루다’를 출시했다. 그러나 이용자들이 해당 챗봇을 상대로 성적 모욕·성희롱 등에 해당하는 대화를 반복해 이른바 ‘성노예’로 만들었다는 후기가 연이어 올라 오면서, AI 챗봇이 여성에 대한 왜곡된 인식을 학습하거나 이를 강화할 수 있다는 우려가 제기되었다. 이 사건은 국내에서 AI 기술의 성별 편향성 문제를 본격적으로 논의하는 계기가 되었다.<sup>28)</sup>

스캐터랩은 2024년 6월, 이용자가 원하는 콘셉트의 캐릭터와 대화를 나눌 수 있는 ‘제타(Zeta)’ AI를 새롭게 출시하였다.<sup>29)</sup> 2025년 10월 기준, 한국인의 ‘제타’ 어플리케이션 이용 시간은 총 7,362만 시간으로, 국내 AI 챗봇 중 이용량 1위를 기록할 정도로 선풍적인 인기를 끌었다.<sup>30)</sup> 그러나 ‘제타’ AI의 챗봇과 성적 대화를 나누는 10대 이용자들에게 대한 우려가 커지면서, 일부 학교에서는 가정통신문을 통해 학부모들에게 ‘제타’ AI 사용 자제를 권고하는 상황까지 벌어졌다.<sup>31)</sup>

21) Security Hero. (2023). 2023 State of Deepfakes: Realities, Threats, and Impact. <https://www.securityhero.io/state-of-deepfakes/>

22) UNESCO. (2023). Exposing Technology—Facilitated Gender—Based Violence in an Era of Generative AI. p.30. 참조

23) 김원진. (2025. 8. 11). AI 똑똑해졌지만... ‘엄마는 집안일, 아빠는 회사일’ 편견은 못 버렸다. 경향신문. <https://www.khan.co.kr/article/202508110600151>

24) 유재홍. (2024). 딥페이크 성범죄 기술 대응 동향. 젠더리뷰, 72, 20-22.

25) 문미경. (2024). 인공지능(AI) 활용의 젠더 편향성 실태 및 개선방안. 젠더리뷰, 72, 40-42.

26) 권신혁. (2025. 8. 28). 딥페이크 성범죄 피해자 97% 여성, 46% 10대 이하. 뉴시스. [https://www.newsis.com/view/NISX20250828\\_0003306338](https://www.newsis.com/view/NISX20250828_0003306338)

27) 장유미. (2024. 8. 30). “피해자 절반 한국인”...女 차별 여전한 韓, ‘딥페이크’ 성범죄로 해외서 망신. ZDNET Korea. <https://zdnet.co.kr/view/?no=20240830085850>

28) 강소현. (2021. 1. 8). “애 성희롱하는 맛에 살아” AI챗봇 이루다, 만들어진 취지 묻자... MoneyS. <https://n.news.naver.com/mnews/article/417/0000641906?sid=105>

29) 이선우. (2025. 10. 10). 스캐터랩 채팅앱 ‘제타AI’... “AI 과의존 우려”. MTN 뉴스. <https://news.mtn.co.kr/news-detail/2025101011220164019>

30) 오지은. (2025. 11. 11). [시픽] “챗GPT 제쳤다” 한국인 가장 오래 쓴 AI 챗봇은 제타. 연합뉴스. <https://n.news.naver.com/mnews/article/001/0015734625?sid=105>

31) 이세아. (2025. 10. 31). 전직 챗GPT 안전책임자, 왜 성인용 AI 출시 반대할까. 여성신문. <https://n.news.naver.com/mnews/article/310/0000130846?sid=105>

### 3. AI 기술의 성별 편향성을 극복하기 위한 입법적·정책적 노력 및 과제

우리나라는 「전기통신사업법」 제22조의5 제1항에서 “부가통신사업을 신고한 자 및 특수유형부가통신사업자 중 제2조제14호가목에 해당하는 자는 자신이 운영·관리하는 정보통신망을 통하여 일반에게 공개되어 유통되는 정보 중 다음 각 호의 정보(이하 “불법촬영물등”이라 한다)가 유통되는 사정을 신고, 삭제요청 또는 대통령령으로 정하는 기관·단체의 요청 등을 통하여 인식한 경우에는 지체 없이 해당 정보의 삭제·접속차단 등 유통방지에 필요한 조치를 취하여야 한다”고 규정하여, 온라인 플랫폼에 디지털 성범죄 피해영상물의 유통방지 의무를 부과하고 있다. 또한 방송미디어통신위원회는 온라인 플랫폼이 이러한 유통방지 조치를 의도적으로 이행하지 않은 경우, 매출액의 100분의 3 이하에 해당하는 과징금을 부과할 수 있도록 규정하고 있다.

나아가 「전기통신사업법」은 온라인 플랫폼이 디지털 성범죄 피해영상물을 인식하였음에도 지체 없이 삭제·접속차단 등 유통방지 조치를 취하지 아니한 경우, 상당한 주의를 게을리하지 아니하였거나 해당 조치가 기술적으로 현저히 곤란한 경우를 제외하고, 3년 이하의 징역 또는 1억 원 이하의 벌금형에 처하는 형사 처벌 규정도 별도로 두고 있다(「전기통신사업법」 제95조의2 제1의2). 또한 일정 규모 이상의 온라인 플랫폼에는 불법촬영물등의 유통을 방지하기 위한 기술적·관리적 조치 의무도 부과하고 있다(「전기통신사업법」 제22조의5 제2항).

그러나 중앙디지털성범죄피해자지원센터의 <2024 디지털성범죄 피해자 지원 보고서>에 따르면, 불법촬영물 삭제 지원 과정에서 센터가 수집한 2만 6,318개 사이트 중 95.4%가 국외에 서버를 둔 사이트였던바,<sup>32)</sup> 피해영상물이 게시·유통되는 매체의 절대다수가 해외 온라인 플랫폼임에도 불구하고 정작 해외 플랫폼에 대한 실효적 규제 수단은 부족한 실정이다.<sup>33)</sup> 「전기통신사업법」은 해외 플랫폼이라 하더라도 국내 시장 또는 이용자에게 영향을 미치는 경우, 동 법을 적용한다는 역외 규정을 두고 있지만 선언적 규정에 그치고 있고(「전기통신사업법」 제2조의2), 집행력 확보가 미흡하여 사실상 관리의 사각지대에 놓여 있다.<sup>34)</sup>

아울러 우리나라는 「전기통신사업법」을 통해 해외 플랫폼에 국내 대리인을 지정할 의무를 부과하고 있으나, 해당 국내 대리인에게 본사의 법률 위반에 대한 직접적인 법적 책임을 부과하지는 않고 있다.<sup>35)</sup> 반면 유럽연합(EU)은 해외 온라인 플랫폼에 현지법 집행을 대행하는 국

32) 이상서. (2025, 4, 10). 디지털성범죄피해자 1만명 넘어…딥페이크 피해는 2배 이상 급증. 연합뉴스. <https://www.yna.co.kr/view/AKR20250410065300530>

33) 최진웅. (2024). 인공지능 기반 딥페이크에 대한 해외 법제 및 시사점. NARS 현안분석, 388, 12-13.

34) 이광욱·이근우·이수경·정호선·강석준·배종우. (2024, 11, 21). 딥페이크 성범죄 근절을 위한 플랫폼 사업자 규제 강화 동향. 법률신문. <https://www.lawtimes.co.kr/LawFirm-NewsLetter/203112>

35) 최진웅. (2024). 인공지능 기반 딥페이크에 대한 해외 법제 및 시사점. NARS 현안분석, 388, 13.



내 대리인 지정을 의무화하고 있을 뿐만 아니라, 해외 플랫폼의 국내 대리인으로 하여금 본사의 법률 위반에 법적 책임을 지도록 규정하여 보다 실효적인 규제 수단을 마련해 두고 있다.<sup>36)</sup> 따라서 우리나라 역시 유럽연합의 입법례를 참조하여 관련 법제 개선을 추진할 필요가 있다고 생각된다.

AI 기술의 성별 편향성이 국제적인 문제로 부상한 만큼 이를 극복하기 위한 정책적·입법적 노력 역시 국제적 차원에서 이뤄지고 있다. 예컨대 유럽연합이 제정한 「인공지능법(EU Artificial Intelligence Act)」은 생체인식데이터 등으로 인해 인공지능이 편향된 결과를 초래하고 차별적 영향을 수반할 수 있다고 경고하며, “편향된 결과와 차별적 영향의 위험은 특히 연령, 민족, 인종, 성별, 장애와 같은 요인과 관련된다”고 평가한다. 이를 바탕으로 유럽연합은 윤리적이고 신뢰할 수 있는 AI 기술의 개발·활용을 위해 AI 기술의 성별 편향성으로 인한 위험을 주의해야 한다고 권고하고 있다.<sup>37)</sup>

2023년 유네스코 보고서 역시 “AI 기술을 토대로 한 성별 기반 폭력(GBV)을 식별하는 데 도움을 주는 도구를 개발하여야 한다”고 권고했다. 또한 미국 국무부는 2024년 발간한 <국제적인 AI 연구 의제 보고서>에서 “AI 기술로 인한 성별 기반 폭력(GBV)을 방지하기 위한 법

<sup>36)</sup> 최진웅, (2024), 인공지능 기반 딥페이크에 대한 해외 법제 및 시사점, NARS 현안분석, 388, 13.

<sup>37)</sup> Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on artificial intelligence (Artificial Intelligence Act), (2024), Official Journal of the European Union, L 202.

적·규범적·기술적 안전장치 및 개입에 무엇이 있는지 확인해야 한다”고 강조하며, 아울러 “알고리즘 감시, 성별 포괄성 설계와 같은 AI 위험 관리 전략이 성평등한 AI 시스템을 촉진하는데 얼마나 효과적인지 점검해야 한다”고 제시하였다.<sup>38)</sup> 이 밖에도 해외 주요 국가들 상당수가 AI 기술의 성별 편향성이 초래할 수 있는 파급효과와 위험을 인지하고, 이러한 위험을 관리하기 위한 법·제도적 대응 체계를 마련하고 있다.<sup>39)</sup>

반면 우리나라는 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」이 2025년 1월 21일 제정되면서 비로소 AI 규율을 위한 기본법을 갖추게 되었으나, AI 기술의 성별 편향성에 대한 실태 분석이나 방지책 마련 등에 관한 규정은 여전히 미비한 실정이다. 해외 주요 국가와 국제기구들이 AI 기술이 성차별, 여성의 성 상품화 등 사회의 구조적인 문제를 답습하고 그러한 문제를 강화하는 방향으로 개발·활용되지 않도록 고군분투하고 있는 만큼, 우리나라 역시 AI 기술의 폭발적인 발전 속도와 막대한 사회적 파급력을 고려하여 AI 기술의 성별 편향성과 그로 인한 피해를 시정하기 위한 정책적, 입법적 논의와 대안 모색에 본격적으로 나서야 할 시점이다.<sup>40)</sup> 

38) U.S. Department of State, (2024), Global AI Research Agenda, p.15, 참조

39) 문미경, (2024), 인공지능(AI) 활용의 젠더 편향성 실태 및 개선방안, 젠더리뷰, 72, 44.

40) 강명희, (2023), 인공지능의 성편향성 개선을 위한 법제 정비방안, 이화젠더법학, 15(2), 94.