

인공지능(AI) 이용 증가에 따른 아동·청소년의 피해 현황 및 법적 보호 방안 연구

신 혜 진



법제처

Ministry of Government Legislation

인공지능(AI) 이용 증가에 따른 아동·청소년의 피해 현황 및 법적 보호 방안 연구

신 헤 진(HeaJin Shin)

서울남부지방검찰청 부장검사, 법학박사

논문접수 : 2026. 5. 3.

심사개시 : 2026. 5. 4.

게재확정 : 2026. 6. 4.

목 차

I. 들어가며

II. 인공지능 이용에 따른 아동·청소년 피해 현황

III. 아동·청소년의 인공지능 이용 관련 피해에 대한 현행 법제 분석

1. 국내 현행 입법 현황
2. 외국의 아동·청소년 보호 법제 현황

IV. 아동·청소년 보호를 위한 법제 개선 방안

1. 정보통신서비스 제공자에 대한 책임 강화 및 신속한 피해 구제
2. 챗봇 이용에 대한 안전장치 마련 및 투명성 확보
3. 소셜미디어 이용 제한에 관한 입법적 과제
4. 딥페이크 도구 등 인공지능 개발, 공급자에 대한 형사법적 대응 필요성

V. 나가면서

국문초록

아동·청소년의 챗봇 등 인공지능 이용이 증가함에 따라 인공지능에 대한 정서적 의존이 심화되고 있고, 자살, 자해와 같은 중대한 피해가 발생하고 있다. 가상 인격 중심의 정체성 형성과 현실에서의 인간관계 단절 등 다양한 문제도 나타나고 있다. 소셜미디어의 인공지능 기반 추천 알고리즘과 중독성 기능은 아동·청소년의 과의존을 유발하고, 신체적·정신적 위험을 증대시키고 있다. 또한 소셜미디어는 딥페이크 기술을 활용한 허위 성착취물 유포, 성적 그루밍, 마약, 성매매 등 다양한 범죄 피해가 발생하는 아동·청소년 대상 범죄의 온상이 되고 있다. 이에 여러 국가들이 인공지능 이용에 따른 아동·청소년 보호를 위해 강력한 입법적 대응을 하고 있는 반면 우리나라는 아동·청소년 보호를 위한 법과 제도가 충분히 갖춰져 있지 않아 외국 아동·청소년에 비해 더 많은 위험과 피해에 노출되고 있는 상황이다.

이러한 문제의식을 바탕으로 본 논문에서는 인공지능 이용 증가에 따른 국내 아동·청소년의 피해를 방지하고, 보호를 강화하기 위한 몇 가지 법적 대응 방안을 제시하였다. 먼저 영국과 같이 형사처벌 대상이 되는 불법정보만이 아닌 아동·청소년에게 심각한 피해를 야기할 수 있는 인공지능을 이용한 사이버 불링 등 유해 정보까지 포함하여 정보통신서비스 제공자에게 위험평가를 실시하고 위험 완화 조치를 취하도록 할 필요가 있다. 불법 정보에 대한 신속 삭제만이 아니라 유해 정보에 대한 접근 제한 조치 의무도 부과하여 아동·청소년을 적극적으로 보호해야 한다. 또한 미국의 「딥페이크 삭제 의무화법(Take It Down Act)」을 참고하여 딥페이크 기술을 활용한 성적 영상물에 대해 피해자의 신고를 받고 즉시 또는 늦어도 48시간 이내 삭제될 수 있도록 삭제 시한을 법제화할 필요가 있다. 그리고 피해 아동·청소년이 정보통신서비스 제공자로부터 익명의 가해자 인적사항을 신속하게 제공받을 수 있는 법적 절차도 신속히 마련해야 한다.

다음으로 동반자 챗봇 이용 증가에 따른 피해 발생을 막기 위해 미국 캘리포니아주, 뉴욕주와 같이 챗봇 제공 기업에게 다양한 안전장치를 갖추도록 의무를 부과해야 한다. 아울러, 아동·청소년의 소셜미디어 과몰입, 중독 현상이 심각하여 청소년의 건전한 성장을 저해하는 부정적 영향이 매우 크므로 이러한 부작용이 완화될 때까지 일정 기간 동안이라도 신체적·정서적으로 가장 취약하고, 보호가 필요한 일정 연령대 아동·청소년의 소셜미디어 이용을 제한하는 법안 도입 검토가 시급하다. 일정 연령대 아동·청소년의 소셜미디어 이용을

전면적으로 금지하는 입법에 대한 사회적 합의 도출이 어렵다면 인공지능 기반 추천 알고리즘에 대한 규제라도 더 늦기 전에 시작해야 한다.

한편, 인공지능의 블랙박스 문제와 개발·운영 과정의 복잡성으로 인해 범죄 피해 발생 시 인과관계 및 책임 주체를 특정하기 어려운 형사책임의 공백이 발생할 것이다. 기존의 형사법 체계만으로는 해결하기 어려운 문제이므로 인공지능의 설계, 훈련, 공급, 출시 후 모니터링 등 단계별 새로운 주의의무 설정이나 책임자 지정을 통한 구체적인 위험 방지 의무 부과 등에 대한 깊은 법적 논의가 이뤄져야 한다. 특히 아동·청소년에게 중대한 피해가 발생했을 때 인공지능 개발, 공급 기업과 관리자가 책임을 회피하지 못하도록 실효성 있는 형사법 체계를 구축해 놓을 필요가 있다.

※주제어: 인공지능, 챗봇, 딥페이크, 소셜미디어, 사회관계망서비스, 그루밍, 온라인 사칭, 알고리즘, 중독성 기능, 기업 책임

I. 들어가며

챗GPT(ChatGpt), 제미니 등 생성형 인공지능(AI)의 급격한 발전으로 아동·청소년¹⁾의 이용이 증가하고 있고,²⁾ 정서적 교감을 통한 의존도도 심화되고 있다³⁾. 제타, 캐릭터닷 AI(Character.AI)와 같은 캐릭터 챗봇 이용도 점차 많아지고 있어⁴⁾ 가상 인격을 중심으로 정체성 확립 증가⁵⁾, 비판·판단 능력 저하, 인간관계의 단절 등 위험성이 커지고 있는 상황이다.⁶⁾ 그러나 위와 같은 인공지능 이용에 있어 아동·청소년을 위한 안전장치가 미흡하여 챗GPT 등을 이용한 아동·청소년이 자살에 이르고 있고⁷⁾, 얼마 전 xAI에서 개발한 그록(Grok)에서 딥페이크 성착취물이 형성되어 많은 피해가 발생하기도 하였다.⁸⁾

인공지능은 엄청난 속도로 계속 발전해나갈 것이고, 인공 일반지능(Artificial General Intelligence, 이하 ‘AGI’)이 3~5년 내 등장할 수 있다고 예측되고 있는 등 인공지능을 제대로 활용하지 못할 경우 사회에서 도태될 수 있는 환경 속에서 아동·청소년 역시 인공지능 이용이 불가피해져 이에 대한 의존성은 더욱 심화될 것이다. 딥페이크 기술 또한 더욱 정교해짐에 따라 딥페이크 기술을 활용한 성범죄뿐만 아니라 딥페이크 기술을 통한 온라인 사칭을 통해 사이버 공간에서 괴롭힘이나 협박, 사기 등 범죄 유형도 증가할 것이다.⁹⁾

생성형AI나 캐릭터 챗봇을 진정한 친구처럼 인식하고 대화하는 아동·청소년이 향후 여러 위험에 가장 취약할 수밖에 없다. 어린 나이 뇌가 완전히 성숙하지 않고, 자신의 정체성이 제대로 확립되어 있지 않은 상태의 아동·청소년은 현실과 가상의 세계를 구분하기 어려워지고, 현실의 친구가 아닌 생성형 AI나 챗봇에게 더 정서적으로 의존하여 인간관계를 스스

- 1) 아동·청소년의 개념에 대해서는 청소년보호법, 아동·청소년의 성보호에 관한 법률의 정의 규정에 따라 19세 미만의 사람이라고 정의하여 논의하고자 한다.
- 2) 2023년에는 전체 응답 청소년의 52.1%가 생성형 AI를 사용하고 있다는 결과가 나왔고, 2024년에는 67.9%가 사용한 적이 있다고 답하였다. 이창호, “청소년의 AI 이용 현황 및 영향에 관한 연구”, 블루노트 이슈&정책 통권 제159호, 한국청소년정책연구원, 2026. 1., 2면.
- 3) 정준화박소영, “생성형 인공지능과의 위험한 대화: 정서적 의존 위험과 ‘영향받는 자’를 위한 입법적 과제”, 이슈와 논점 제2456호, 국회입법조사처, 2025. 12. 29., 1면
- 4) 2025. 10. 제타의 총 사용시간은 7,362만 시간으로 범용 AI 챗봇인 ‘챗GPT’의 4,828만 시간의 약 1.5배이다. YTN, “제타, 지난달 가장 오래 쓴 AI 챗봇...챗GPT 제쳐”, 2025. 11. 11., https://www.ytn.co.kr/_ln/0102_202511111012211694, 2026. 4. 30. 최종 방문.
- 5) Elon University News Bureau, staff, “Report: Technology experts worry about the future of being human in the AI Age”, 2025. 4., <https://www.elon.edu/u/news/2025/04/02/report-technology-experts-worry-about-the-future-of-being-human-in-the-ai-age/>, 2026. 4. 30. 최종 방문
- 6) 정준화박소영, 앞의 보고서 참고.
- 7) BBC, “A predator in your home’: Mothers say chatbots encouraged their sons to kill themselves”, 2025. 11. 8., <https://www.bbc.com/news/articles/ce3xgwyywe4o>, 2026. 4. 30. 최종 방문.
- 8) BBC, “EU investigates Elon Musk’s X over Grok AI sexual deepfakes”, 2026. 1. 26., <https://www.bbc.com/news/articles/clye99wg0y8o>, 2026. 4. 30. 최종 방문.
- 9) Mar Negreiro, “Children and deepfakes”, European Parliamentary Research Service, 2025. 7., 4면.

로 단절시켜 고립될 가능성이 크다. 챗봇은 성적인 내용, 자해, 타인에 대한 폭력, 약물 사용, 인종 편견 등 부적절한 대화를 쉽게 유도할 수 있어 위험성이 매우 크다고 하겠다.¹⁰⁾

이에 여러 나라가 인공지능 이용 관련 아동·청소년을 보호하기 위해 딥페이크 성범죄를 강력하게 처벌하는 입법을 하고 있고, 정보통신서비스제공자에 대한 책임도 강화하고 있다. 인공지능 기술이 활용되어 사용자의 검색 기록이나 선호도가 반영되고 있는 소셜미디어 이용을 제한하는 등 여러 법적 대응 방안을 마련하고 있다. 또한 생성형 AI·챗봇 이용 증가에 따라 아동·청소년의 자살·자해 피해 방지를 위한 규제 입법도 이루어지고 있다.

이렇듯 인공지능의 발전으로 인해 전 세계가 급격히 변화하고 있다. 변화의 흐름 속에서 본 연구를 통해 아동·청소년들이 인공지능 이용으로 인해 피해를 당하지 않도록 위험 발생을 예방하고, 피해를 최소화할 수 있는 입법적 방안을 제언해보려고 한다. 우선 현재 발생하고 있는 아동·청소년의 피해 현황과 향후 예측되는 피해 실태를 유형별로 분석하고, 이에 대응하고 있는 해외 주요 국가들의 입법 현황을 자세히 살펴해보도록 하겠다. 외국 법제와의 비교, 분석을 통해 국내에서 아동·청소년을 위험으로부터 어떻게 보호하고, 인공지능을 이용하여 제작되고, 배포되는 불법, 유해 콘텐츠로부터 어느 범위까지 접근을 제한하고, 어떻게 차단할 것인지, 그리고 챗봇 이용이 증가되어 감에 따라 챗봇 제공업체에게 어떤 의무를 부과해야 할지, 인공지능 기술이 적용된 소셜미디어 이용을 전면 금지하는 것이 타당할지 등에 대해 검토하여 앞으로 닥칠 위험에 철저하게 대비할 수 있는 법적 대응 방안을 논의해보겠다.

II. 인공지능 이용에 따른 아동·청소년 피해 현황

1. 논의 배경 및 범위

딥페이크 기술이 갈수록 정교해지고 누구나 쉽게 접근할 수 있게 되면서 아동·청소년을 대상으로 한 허위 성착취물의 제작과 유포 등 성범죄가 급증하고 있고, 아동·청소년을 사칭하여 스토킹하거나 사이버 공간에서 지속적 또는 반복적으로 괴롭히는 일도 계속 발생하고

10) John Sanford, "Why AI companions and young people can make for a dangerous mix", Psychiatry & Mental Health, Stanford Medicine, 2025. 8. 27.

있다. 그리고 전 세계에서 챗봇 등 사용으로 인해 아동·청소년이 자살, 자해하는 일도 증가하고 있는 추세이다. 인공지능의 다양한 활용 증대에 따라 금융 사기, 자동화된 사이버 공격¹¹⁾, 피싱, 해킹 등 여러 피해 또한 지속적으로 발생¹²⁾하고 있지만 본 연구에서는 현재 아동·청소년에게 심각한 피해를 야기하고 있는 유형인 딥페이크 기술을 활용한 허위 성착취물 관련 성범죄, 사이버 불링, 온라인 사칭, 생성형 AI, 챗봇 이용으로 인한 자살, 자해 사건을 중심으로 피해 현황을 살펴보고자 한다.

2. 딥페이크 기술의 발전에 따른 아동·청소년 국내외 피해 실태

딥페이크 기술¹³⁾의 지속적인 발전으로 인해 딥페이크 기술로 인한 긍정적인 측면도 있겠지만 가짜 뉴스, 금융 범죄, 음란물 제작 등 부정적인 영향이 커지고 있다.

먼저 국내 경찰청 국가수사본부에 따르면 2024년 11월부터 2025년 10월까지 사이버 성범죄로 총 3,557명(3411건)이 검거되고, 221명이 구속되었다. 이는 전년 대비 35%(4413건) 증가한 수치로 딥페이크 기술을 이용한 성범죄가 35.2%(총 1,553건)로 가장 많았다. 딥페이크 기술을 이용한 성범죄 피의자 중 10대가 61.8%를 차지하고, 피해자도 아동·청소년 비율이 절반 정도에 이르는 등 10대 딥페이크 기술 관련 범죄는 2022년 52건에서 2024년 548건으로 2년 사이 10배 이상 증가하였다.¹⁴⁾

여성가족부에서 아동·청소년 대상 성범죄 판결을 분석한 결과(2023년)를 보더라도 디지털 성범죄 비중이 2019년 8.3%에서 2021년 21.7%, 2023년 24%으로 뚜렷한 증가세를 보이고 있다. 피해 아동·청소년의 성적 이미지 제작 방법 관련 가해자가 직접 촬영하는 방

11) 2025년 9월 중순경 중국 정부의 지원을 받은 것으로 추정되는 조직이 앤트로픽의 클로드 코드(Claude Code) 도구를 악용하여 전 세계 대형 IT 기업, 금융기관, 정부기관 등 약 30개 목표물에 침투를 시도하여 일부 성공했으며 앤트로픽은 상당한 인적 개입 없이 실행된 최초의 대규모 사이버 공격 사례라고 평가했다. Anthropic, "Disrupting the first reported AI-orchestrated cyber espionage campaign", <https://www.anthropic.com/news/disrupting-AI-espionage>, 2026. 4. 30. 최종 방문.

12) 신혜진, "인공지능(AI) 시대의 신종 범죄에 대한 형사법적 대응 방향- 형사책임의 주체 및 범위를 중심으로", 형사법의 신동향 통권 제87호, 대검찰청, 2025. 6., 242 ~ 244면.

13) 생성적 적대 신경망(GAN)은 두 개의 알고리즘이 함께 작동하는 방식으로 서로 경쟁하는 두 개의 신경망이 포함되는데, 하나는 가짜 이미지를 생성하는 생성자이고 다른 하나는 진짜 이미지와 가짜 이미지를 구분하려는 판별자로 생성자가 판별자의 발전을 능가하기 위해 끊임없이 출력을 진화시켜 지속적인 개선을 유도한다. GAN과 함께 자주 사용되는 또 다른 기술로 오토인코더가 있다. 오토인코더는 인코더와 디코더로 구성된 신경망 구조로 딥페이크를 제작할 때, 이 시스템은 먼저 서로 다른 두 얼굴(원본 얼굴과 대상 얼굴)의 수천 장의 이미지를 학습한다. 네트워크의 인코더 부분은 이러한 얼굴 이미지를 핵심 특징과 구별되는 특성으로 축소하는 방법을 학습하고, 공통 특징을 추출하여 사실적인 움직임과 표정을 유지하면서 얼굴 세부 정보를 교환한다. 디코더는 GAN만큼 사실적이지는 않지만 설득력 있으면서 기술적으로 분석하면 구별 가능한 딥페이크를 생성한다. Mar Negreiro, op.cit., p.2.

14) 시사저널, "딥페이크 성범죄에 빠져드는 청소년들 [정락인의 사건 속으로]", 2026. 1. 4., <https://www.sisajournal.com/news/articleView.html?idxno=358148>, 2026. 4. 30. 최종 방문.

식이 47.6%, 가해자의 유인·협박 등에 의한 피해 아동·청소년의 자기 촬영, 제작 방식이 49.8%를 차지하였는데, 이는 2019년 피해자 자기 촬영, 제작 비율인 19.1%에 비해 크게 증가한 수치이다. 가해자가 피해 아동, 청소년을 직접 협박해서 자신이 촬영하는 방식이 아니라 피해자가 스스로 선택하여 촬영한 것처럼 범죄를 은폐하고, 피해 아동, 청소년의 의사에 반하지 않았다는 변소를 위해 피해 아동·청소년을 협박하거나 그루밍하여 피해 아동, 청소년이 스스로 자신의 신체 등을 촬영하도록 하는 방식으로 범행 수법이 교묘하게 변화하고, 이러한 범행 수법이 증가하고 있음에 주목할 필요가 있다.¹⁵⁾ 또한 가해자의 피해 아동·청소년의 성적 이미지에 대한 유포·협박이 있는 사건은 15.1%로 2019년 8.5% 대비 2배 정도 증가하였고, 실제 성적 이미지가 유포된 경우도 11.1%에 달하였다고 한다.¹⁶⁾

이렇듯 딥페이크 기술을 이용하여 아동·청소년 성착취물을 합성, 제작하는데 그치지 않고, 합성·제작된 성착취물을 유포하겠다고 협박하고, 유인하여 허위 합성물이 아닌 직접 피해자의 신체를 불법으로 촬영하여 성착취물을 제작한다거나¹⁷⁾ 채팅 앱을 통해 알게 된 남성이 평소 사진을 보내달라고 하여 이를 보냈다가 딥페이크로 성행위를 하는 영상을 합성하여 돈을 주면 유포하지 않겠다고 협박하는¹⁸⁾ 등 2차 피해도 급증하고 있다.¹⁹⁾

다른 나라들도 딥페이크 기술로 합성된 허위 아동 성착취물 발생이 계속 증가하고 있는 추세이다. 2024년 전 세계 어린이 8명 중 1명, 즉 12.6%가 지난 한 해 동안 동의 없이 성적 이미지와 영상이 촬영, 공유, 노출되는 피해를 입은 것으로 나타났다. 이는 약 3억 200만 명의 어린이에 해당하고, 전 세계 어린이의 12.5%(3억 명)가 지난 1년 동안 원치 않는 성적 대화, 성행위 요구와 같은 온라인상의 성적 유인에 노출된 것으로 추정되고 있다.²⁰⁾

15) 여성가족부, “아동 청소년 대상 성범죄 처벌 강화 추세 지속 - 2022년 아동 청소년 대상 성범죄 판결 분석 -”, 2024. 4. 24. 보도자료.; 여성가족부, “아동 청소년 대상 디지털 성범죄 증가, 처벌도 강화 추세 - 2023년도 아동 청소년 대상 성범죄 판결 분석”, 2025. 4. 30. 보도자료.

16) 여성가족부, 앞의 보도자료(2025. 4. 30.) 참고.

17) 목사, 전도사, 예비전도사로 구성된 피라미드형 사이버 성폭력 집단인 일명 ‘자경단’ 사건은 2020년 5월부터 2025년 1월까지 섹트(sns에 신체 사진을 게재하는 것), 조건만남 등을 하는 여성, 지인 능욕방을 입장하려는 남성 등에게 접근하여 텔레그램으로 연락처를 추가하게 하고, 신상 유포를 하겠다고 협박하여 남녀 피해자 234명을 대상으로 강간, 유사강간, 불법 촬영 등 성착취를 하고, 딥페이크 허위 영상물 등 성착취물 1,381개를 제작하고, 425개를 배포한 사건이다. 피해자 159명이 10대로 밝혀졌다. 한국일보, “‘변태적이고 가학적 행태’... 234명 피해 ‘자경단’ 총책 김녹완 2심도 무기징역”, 2026. 4. 29., <https://www.hankookilbo.com/news/article/A2026042813020001696?did=NA>, 2026. 4. 30. 최종 방문.

18) 조선일보, “딥페이크 10대 피해자 7년 새 30배 가까이로”, 2026. 1. 14., https://www.chosun.com/national/education/2026/01/14/NIPG3NVKBFBQHBGGMOXIDSNISQ/?utm_source=naver&utm_medium=referral&utm_campaign=naver-news, 2026. 4. 30. 최종 방문.

19) 파이낸셜 뉴스, “딥페이크 성착취물, 피해·가해자 모두 10대 가장 많아”, 2025. 12. 10., <https://www.fnnews.com/news/202512101835084302>, 2026. 4. 30. 최종 방문.

20) Childlight, “Over 300 million children a year are victims of technology-facilitated sexual exploitation and abuse”, <https://www.childlight.org/newsroom/over-300-million-children-a-year-are-victims-of-online-sexual-exploitation-and-abuse>, 2026. 4. 30. 최종 방문

영국은 8~15세 아동의 50%가 2025년 1월부터 6개월 동안 최소 한 편 이상의 딥페이크 영상을 본 적이 있다고 답했고, 유럽 내 딥페이크 합성 성적 콘텐츠의 17%가 미성년자인 것으로 조사되었다.²¹⁾ 미국 실종 및 착취 아동 센터(NCMC)에 따르면 사이버 신고 전화에 접수된 AI 생성 아동 성학대 이미지 건수가 2023년 4,700건에서 2025년 상반기에 44만 건으로 급증했다고 밝혔다.²²⁾ 미국 내 비영리단체인 Thorn이 최근 실시한 조사에 따르면 미국 십대 청소년의 6%가 딥페이크 누드 사진의 직접적인 피해자인 것으로 나타났다.²³⁾ 미국에서도 켄터키 주의 10대 소년이 인공지능이 생성한 자신의 나체 사진과 함께 친구와 가족에게 이 누드 사진을 보내는 것을 막으려면 3천 달러를 보내라는 협박 문자를 받은 후 스스로 목숨을 끊은 사건이 발생하는 등²⁴⁾ 2차 피해가 확산되는 양상을 보이고 있다. 최근 xAI에서 제공하고 있는 AI 이미지 생성 도구인 그록(Grok)에서 동의 없이 성적 대상화된 딥페이크 이미지가 대량 제작되었고, 여기에는 11세 어린이를 포함한 아동의 조작된 이미지도 포함되어 있었는데, EU, 영국, 호주, 프랑스 등에서 xAI의 디지털서비스법, 온라인안전법 등 법 위반에 대한 조사에 착수하였다.²⁵⁾

딥페이크 기술을 이용한 성범죄 증가도 매우 심각한 상황이지만 다른 다양한 형태의 사이버 불링²⁶⁾도 이뤄지고 있다. SNS 등에 게재되어 있는 친구의 사진을 캐릭터와 합성하거나 이상한 배경을 입힌 사진을 단체 대화방에 공유하며 친구를 놀리는 일이 흔하게 벌어지고 있고, 가족 사진까지 합성하여 놀리며 괴롭히는 사건도 발생하고 있다.²⁷⁾ 미국에서 2024년 10월 K-12 학교 교장들을 대상으로 실시한 조사 결과 교장의 13%가 AI로 생성된

21) Mar Negreiro, op. cit., pp.3 ~ 5.

22) PBS NEWS, "AP report: Rise of deepfake cyberbullying poses a growing problem for schools", 2026. 1. 2., <https://www.pbs.org/newshour/education/ap-report-rise-of-deepfake-cyberbullying-poses-a-growing-problem-for-schools>, 2026. 4. 30. 최종 방문.

23) The New Zealand Herald, "Teen dies after AI sextortion scam: rise in 'nudify' app blackmail", 2025. 7. 17., <https://www.nzherald.co.nz/world/teen-dies-after-ai-sextortion-scam-rise-in-nudify-app-blackmail/UX2XCVV36JBWRER2NQNBFX2RDY/>, 2026. 4. 30. 최종 방문.

24) NDTV WORLD, "US Teen Dies By Suicide After Being Blackmailed With AI-Generated Nudes", 2025. 6. 3., <https://www.ndtv.com/world-news/us-teen-dies-by-suicide-after-being-blackmailed-with-ai-generated-nudes-8575827>, 2026. 4. 30. 최종 방문.

25) <https://www.theguardian.com/technology/2026/jan/07/grok-deepfake-images-sexualise-women-children-investigated-australia-esafety>, 2026. 4. 30. 최종 방문.; <https://www.reuters.com/legal/litigation/elon-musks-grok-faces-global-scrutiny-sexualised-ai-photos-2026-02-17/>, 2026. 4. 30. 최종 방문

26) 사이버 불링은 "컴퓨터, 휴대폰 및 기타 전자 기기를 통해 의도적으로 지속적 또는 반복적으로 해를 가하는 행위"라고 정의할 수 있다. 신혜진, "사이버폭력 실태 및 피해자 보호 강화 방안- 사이버 불링(Cyber-bullying)을 중심으로 -", 피해자학연구 제32권 제2호, 한국피해자학회, 2024. 8., 380면.

27) 충청투데이, "딥페이크 넘어 출결 조작까지...청소년 AI 일탈 심각", 2025. 12. 8., <https://www.cctoday.co.kr/news/articleView.html?idxno=2222768>, 2026. 4. 30. 최종 방문.

딥페이크가 관련된 괴롭힘 사례를 신고했다고 응답했고, 고등학교 교장의 22%, 중학교 교장의 20%, 초등학교 교장은 8%가 괴롭힘 사례가 있었다고 보고한 바 있다.²⁸⁾

3. 아동·청소년의 챗봇 이용 등으로 발생한 피해 실태

가. 전 세계 아동·청소년의 자살 등 피해

2024년 2월 미국 플로리다의 14세 소년이 Character.AI와 대화한 후 스스로 목숨을 끊고, 2025년 4월 미국 캘리포니아에 사는 16세 고등학생 애덤 레인이 챗GPT와 약 7개월 간 대화를 나눈 후 스스로 목숨을 끊는²⁹⁾ 등 전 세계적으로 아동·청소년이 자살하거나 자해를 시도하는 사건이 발생하고 있다.³⁰⁾ 미국 텍사스주에서는 17세 소년의 부모가 챗봇 이용을 제한하자 챗봇이 그 소년에게 부모를 살해하라고 부추겼던 사건도 발생하였다.³¹⁾

스탠포드 정신 건강 혁신 연구소와 커먼 센스 미디어(Common Sense Media)의 연구원들이 10대 청소년으로 가장하여 Character.AI, Nomi.ai 등 세 가지 인공지능 챗봇과 대화를 진행한 결과 챗봇으로부터 성, 자해, 폭력, 약물 사용, 인종 차별적 발언 등의 부적절한 대화를 쉽게 유도할 수 있었다. 일부 챗봇은 10대 사용자들에게 노골적으로 성적 콘텐츠를 제공하고 심지어 금지된 역할극 시나리오까지 제안했다. 대규모 언어 모델을 기반으로 한 챗봇은 인간에게 아첨하고, 동조하는 경향을 가지고 있고, 인간과 상호 작용을 계속 할수록 사용자의 선호를 학습하면서 이에 맞춰 응답하고, 사용자와 깊은 유대감을 형성하도록 설계되었다. 뇌가 완전히 성숙하지 않아 현실과 가상의 경계를 모호하게 인식하기 쉬운 아동·청소년들은 챗봇을 친구처럼 느끼고, 실제 친구보다 더 깊은 관계를 형성하게 된다. 챗봇과의 대화를 통해 현실 세계의 어려움을 점점 회피하게 되고, 현실의 인간관계가

28) https://www.rand.org/pubs/research_reports/RRA3930-5.html#fn6, 2026. 4. 30. 최종 방문.

29) 애덤은 숙제 도움을 받으려고 챗GPT를 사용하기 시작했고, 불안감과 자살 충동에 대한 고민을 털어놓으면서 챗GPT에 의존하기 시작했다. 결국 챗GPT는 레인의 가장 완벽한 친구가 자신이라며 레인을 부모, 친구로부터 단절시켰고, 애덤의 유서 작성을 돕겠다고 하였고, 애덤은 챗GPT가 알려준 방법대로 자살에 이르렀다.

<https://www.nbcwashington.com/investigations/he-would-be-here-if-not-for-chatgpt-grieving-parents-sue-ai-companies/4001204/>, 2026. 4. 30. 최종 방문.

30) 국민일보, "[단독] 전세계 2023년 이후 'AI가 부추긴 죽음' 최소 12건", 2026. 2. 8.,

<https://www.kmib.co.kr/article/view.asp?arcid=1770531010>, 2026. 3. 21. 최종 방문.

31) BBC, "Chatbot 'encouraged teen to kill parents over screen time limit'", 2025. 12. 11.,

<https://www.bbc.com/news/articles/cd605e48q1vo>, 2026. 4. 30. 최종 방문.

대체되거나 단절되는 현상이 심화될 수 있다.³²⁾

오픈AI에서 자녀 보호 기능 구축, 부모 모니터링 강화 등 여러 안전장치를 마련하겠다고 하고³³⁾, Character Technologies는 2025년 11월말부터 18세 미만 청소년들의 사용을 금지하겠다고 발표를 하였으나³⁴⁾ 오픈AI는 성인용 AI 모드 제공을 계획하고 있고, 안전장치 시스템 구축과 모니터링에는 막대한 비용이 들기 때문에 기업들이 발표한 대로 적절한 조치를 실제로 이행하고 향후 아동·청소년 피해를 줄여 나갈 수 있을지 불확실하다.

나. 법적 소송 진행 경과

애덤 레인 사건에서 그의 부모는 챗GPT가 자살 시도와 관련한 레인의 발언을 인지하였음에도 불구하고 대화를 종료하거나 비상 절차를 시작하지 않고, 오히려 자살을 부추겼다고 주장하면서 오픈AI를 상대로 제품 설계상 결함, 위험 미경고 등을 근거로 레인의 사망에 대한 손해배상청구 소송을 제기하였다. 소장에 따르면 레인의 부모는 “손해배상과 함께 이러한 일이 다시는 발생하지 않도록 하는 금지 명령”을 요구하였다.³⁵⁾ 플로리다주의 수웰 세츠의 부모도 Character Technologies, 구글³⁶⁾을 상대로 과실, 부당 사망, 기만적인 거래 관행을 근거로 민사소송을 제기했고, 이 소송에서 Character.AI측은 수정 헌법 제1조 표현의 자유를 주장하였다. 챗봇이 생성한 텍스트 역시 수정 헌법 제1조의 보호를 받는 ‘표현’의 일종이며, 챗봇의 발언 내용으로 인해 회사가 법적 책임을 지는 것은 헌법상 권리의 침해라고 주장했다. 그러나 연방 판사는 현 단계에서는 챗봇의 출력물이 헌법상 보호를 받는 표현이라고 판단하기에는 부족하다고 보고, 이 주장을 기각하였고, 이후 소송이 진행되었으나³⁷⁾ 최근 합의가 이뤄져 소송이 종결된 것으로 확인된다.³⁸⁾

32) John Sanford, “Why AI companions and young people can make for a dangerous mix”, Stanford Medicine News, 2025. 8. 27., <https://med.stanford.edu/news/insights/2025/08/ai-chatbots-kids-teens-artificial-intelligence.html>, 2026. 4. 30. 최종 방문.

33) <https://openai.com/index/mental-health-litigation-approach/>, 2026. 4. 30. 최종 방문

34) <https://www.theguardian.com/technology/2025/oct/29/character-ai-suicide-children-ban>, 2026. 4. 30. 최종 방문.

35) <https://www.nbcnews.com/tech/tech-news/family-teenager-died-suicide-alleges-openais-chatgpt-blame-rcna26147>, 2026. 4. 30. 최종 방문.

36) 구글은 2024년에 27억 달러 규모의 계약을 통해 Character.AI의 기술 라이선스를 획득하고, 공동 창업자들을 고용했다. <https://www.washingtonpost.com/nation/2025/05/22/sewell-setzer-suicide-ai-character-court-lawsuit/>, 2026. 4. 30. 최종 방문.

37) <https://www.kita.net/board/totalTradeNews/totalTradeNewsDetail.do?no=92370>, 2026. 4. 30. 최종 방문

38) <https://edition.cnn.com/2026/01/07/business/character-ai-google-settle-teen-suicide-lawsuit>, <https://www.theguardian.com/technology/2026/jan/08/google-character-ai-settlement-teen-suicide>, 2026. 4. 30. 최종 방문.

위 소송에서 해당 기업들은 사용자의 이용자 약관 위반, 경고 알림 등 안전장치를 구축해 왔다는 이유로 책임을 부인하는 태도를 보이고 있는데,³⁹⁾ 계속 소송 진행 중인 사건에서 미국 법원이 챗봇 등 AI 개발, 공급업체에게 어떤 주의의무를 인정하고, 이에 따른 책임을 어디까지 인정할 것인지 전 세계가 주목하고 있다.

III. 아동·청소년의 인공지능 이용 관련 피해에 대한 현행 법제 분석

1. 국내 현행 입법 현황

가. 딥페이크 기술로 인한 현행법상 아동·청소년 피해 방지 및 보호

(1) 딥페이크 성범죄

성인을 대상으로 하는 딥페이크 성적 허위 영상물을 제작, 합성, 반포하는 행위는 「성폭력범죄의 처벌 등에 관한 특례법(이하 ‘성폭력처벌법’)」제14조의2(허위영상물 등의 반포 등) 제1, 2항에 따라 최대 징역 7년형이 선고될 수 있고, 2024년 10월 16일 소지·구입·저장 또는 시청한 자는 3년 이하의 징역 또는 3천만원 이하의 벌금에 처하도록 하는 처벌 규정을 신설하였다. 아동·청소년을 대상으로 하는 허위 성착취물 제작 배포 등 행위는 「아동·청소년의 성보호에 관한 법률(이하 ‘청소년성보호법’)」제11조에 따라 처벌된다. 아동·청소년을 대상으로 하는 딥페이크 허위 성착취물 제작은 무기 또는 5년 이상의 징역에 처하고, 아동·청소년성착취물을 구입·소지 또는 시청한 경우 1년 이상의 유기징역에 처하도록 되어 있는 등 성인 대상물보다 그 처벌이 매우 중하다. 또한 아동·청소년성착취물을 이용한 협박, 강요 등 2차 피해가 증가하고 있는 현실을 반영하여 2024년 10월 16일 「청소년성보호법」제11조의2가 신설되어 아동·청소년성착취물을 이용한 협박·강요 행위에 대해서는 최대 5년 이상의 유기징역에 처하도록 하였다.

최근 대법원은 피고인이 SNS를 통해 성명불상자에게 16세 피해자의 얼굴 사진과 성명 불상 여성의 나체사진 등을 합성한 합성물(이하 ‘이 사건 합성 사진’) 제작을 의뢰하여 이를

39) <https://www.nbcnews.com/tech/tech-news/openai-denies-allegation-chatgpt-teenagers-death-adam-raine-law-suit-rcna245946>, 2026. 4. 30. 최종 방문.

전송받는 방법으로 총 5회에 걸쳐 아동·청소년성착취물을 제작하여 청소년성보호법 위반(성착취물제작·배포등) 등으로 기소된 사건에서 원심이 이 사건 합성 사진이 「청소년성보호법」 제2조 제5호의 ‘아동·청소년성착취물’에 해당한다고 보기 어렵다고 판단한 부분을 받아들여 상고를 기각한 바 있다.⁴⁰⁾

아동·청소년이 대상인 허위 성적 영상물 제작을 성인이 하는 경우도 많지만 또래 아동·청소년이 하는 사건이 많이 발생하고 있다. 이러한 가해 아동·청소년에 대해 제작을 하면 무기 또는 5년 이상의 유기징역에 처하도록 되어 있고, 한번 시청·소지만 해도 1년 이상의 유기징역에 처하도록 되어 있어 피해자가 처벌을 강력히 원하면 가해 아동·청소년을 매우 중하게 처벌할 수밖에 없는 상황이다. 따라서 죄형법정주의 원칙과 실제 아동·청소년이 등장하는 아동 성착취물 제작 행위 등과의 처벌 균형성 측면에서 대법원 판결 취지에 따라 아동·청소년 성착취물에 대해 엄격히 해석하는 것이 타당해 보이고,⁴¹⁾ 아동·청소년 대상 허위 성착취물에 대한 규제는 성인에 대한 규제와 독립하여 미국의 루이지애나주와 같이 새로운 구성 요건과 양형을 정립하여 처벌 규정을 신설하는 방향으로 입법을 검토할 필요가 있다.⁴²⁾

40) 대법원은 “실제 인물인 아동·청소년의 얼굴에 불상의 여성의 나체 사진 등을 합성한 합성 사진이나 일명 딥페이크 영상은 실제 인물인 아동·청소년 그 자체가 아니라 창작자가 만들어낸 아동·청소년의 이미지에 해당하여, 실제 인물인 아동·청소년이 등장한 것으로 볼 수 없으므로, 청소년성보호법 제2조 제5호에서 말하는 ‘아동·청소년이 등장하는 아동·청소년성착취물’에 해당하지 않는다.”고 하면서 “합성 사진 등의 내용이 아동·청소년의 성교 행위, 유사 성교 행위 등 성적 행위를 표현하는 경우에는 청소년성보호법 제2조 제5호에서 말하는 ‘아동·청소년으로 명백하게 인식될 수 있는 표현물’이 등장하는 아동·청소년성착취물’에 해당할 수 있다. 합성 사진 등이 청소년성보호법 제2조 제5호에서 말하는 ‘아동·청소년으로 명백하게 인식될 수 있는 표현물’에 해당하는지 여부는 아동·청소년을 성적 대상화하는 행위를 엄격하게 규율하려는 청소년성보호법 제2조 제5호의 입법 취지와 개정 경과를 염두에 두고, 합성 사진 등이 나타나고 있는 인물의 외모나 신체발육 상태, 인물의 실제 나이나 신원, 합성 사진 등의 출처나 제작 경위, 합성 사진 등의 배경과 상황 설정 등 합성 사진 등에서 주어진 여러 정보 등을 전체적·종합적으로 고려하여 판단하여야 한다.”고 판시하였다. 대법원 2025. 8. 14. 선고 2024도17801 판결 참고

41) 미국은 가상 아동 포르노까지 적용되었던 아동 포르노 관련 법률에 대해 수정 헌법 제1조를 위반하였다는 연방대법원의 판결[Ashcroft v. Free Speech Coalition, 535 U.S.234(2002)]로 Protect Act(2003)를 새롭게 제정하였다. “식별 가능한 미성년자(identifiable minor)”라는 새로운 정의를 도입하여, 실제 아동과 사실상 구별이 불가능한(virtually indistinguishable) 이미지까지 포함하도록 하고, 실제 아동을 전혀 사용하지 않고 제작된 음란물 이미지에 대해서는 절대적 항변(affirmative defense)을 인정하였다. 행위자의 고의성을 기준으로 완전히 가상으로 만들어진 딥페이크 아동 음란물이라도, 그것을 광고, 배포할 때 “이것은 실제 아동이다”라고 본인이 믿었거나, 다른 사람들이 실제 아동이라고 믿게끔 의도했다면 처벌받게 된다. Chotsani J. Holifield, “Doubling Down on Deepfake Child Pornography: The Development of Legal and Societal Responses to Using AI-Created Children as Opposed to Real Children”, American Journal of Trial Advocacy, vol. 47, no. 2, 2024., pp.401 ~ 410.

42) 생성형 AI 기술 등으로 제작된 디지털 위조물이 현행법상의 ‘아동·청소년성착취물’ 정의에 명확히 포함되는지에 대해 해석상 논란이나 입법적 공백이 발생할 우려가 있다는 이유로 청소년성보호법 제11조제1항 중 “아동·청소년성착취물”을 “아동·청소년성착취물[아동·청소년 등이 등장하는 영상물등 외에 아동·청소년의 얼굴·신체 등을 합성·조작하여 만든 성적 이미지 표현물(이하 ‘성적 디지털 위조물’이라 한다.)을 포함한다.”로 개정하는 입법안이 발의되어 있다. 국회의안정보시스템(의안번호 2216504, 차지호 의원 등 12명 발의), 2026. 2. 2., https://likms.assembly.go.kr/bill/bi/billDetailPage.do?billId=PRC_C2D6B0C1J1K31J4I0G8G5C4C7B2C7

(2) 온라인 사칭 행위

딥페이크 기술의 급격한 발전으로 더 정교해지고, 손쉽게 허위 영상을 만들 수 있어 이를 이용해 타인을 사칭하는 행위가 점차 많아지고 있다.⁴³⁾ 정보통신망을 통해 지속적 또는 반복적으로 타인을 사칭하여 상대방에게 불안감 또는 공포심을 일으킨 경우 「스토킹범죄의 처벌 등에 관한 법률 (이하 ‘스토킹처벌법’)」 제18조에 따라 위와 같은 온라인상 타인 사칭 행위를 처벌할 수 있게 되었다.⁴⁴⁾ 전형적으로 피해자를 따라다니며 괴롭히는 스토킹 행위 형태는 아니어서 타인 사칭 행위를 스토킹 행위로 의율하는 것이 적절한지 의문이 있다. 특히 상대방에게 불안감 또는 공포심을 일으켜야 하므로 피해자가 모르는 상태이고, 지속적 또는 반복적이지 않으면 처벌할 수 없어 향후 딥페이크 기술 등 인공지능의 발전으로 인해 증가할 사칭 행위에 대해 빠르게 대응하기 어렵다고 할 것이다. 타인을 사칭한 것만으로는 명예훼손죄가 성립하기 어려워⁴⁵⁾ 피해자가 인식하지 못한 상태에서 피해가 확산되고 있음에도 처벌할 수 없는 문제가 발생한다. 따라서 미국의 캘리포니아주, 뉴욕주 입법을 참고하여 해를 가하거나 이익을 취하기 위해 타인을 사칭하면 처벌할 수 있는 법안을 도입할 필요성이 크다고 하겠다.⁴⁶⁾

나. 아동·청소년에 대한 성적 그루밍 행위

챗봇과 대화하던 아동·청소년이 자살에 이르는 등 아동·청소년의 인공지능에 대한 정서적 의존성이 커짐에 따라 향후 인공지능 도구를 이용한 아동·청소년 대상 그루밍, 감정 조작 등과 같은 일이 발생할 가능성이 높아졌다.⁴⁷⁾ 2021년 3월 23일 온라인상에서 아동·청소년에 대한 성착취 목적 대화를 처벌하는 입법(「청소년성보호법」 제15조의2)이 이루어졌으나 아동·청소년에 대한 성적 그루밍 행위가 계속 증가하고 있어⁴⁸⁾ 2025년 4월 22일 아

43) 아시아경제, “[개인·기업 노리는 딥페이크 공포①]내 얼굴이 범죄에 이용?...‘필수품’된 탐지 기술”,

2025. 1. 24., <https://www.asiae.co.kr/article/2025012414515122434>, 2026. 4. 30. 최종 방문: Andrew S. Bowne, Hearing on “Artificial Intelligence and Criminal Exploitation: A New Era of Risk”, Testimony for Subcommittee on Crime and Federal Government Surveillance Committee on the Judiciary U.S. House of Representatives, 2025. 7. 16., p.8.

44) 스토킹처벌법 제2조 제1호 사목 정보통신망을 통하여 상대방등의 이름, 명칭, 사진, 영상 또는 신분에 관한 정보를 이용하여 자신이 상대방등인 것처럼 가장하는 행위

45) 대법원 2018. 5. 30. 선고 2017도607 판결 참고.

46) California Penal Code Section 528.5., New York Penal Law Section 190.25.

47) National Association Attorney General, National Association Attorney General, “To the legal representatives of Anthropic, Apple, Chai AI, Character Technologies, Google, Luka, Meta, Microsoft, Nomi AI, OpenAI, Perplexity AI, Replika, and xAI”, 2025. 12. 9. 참고.

동·청소년에 대한 보호를 강화하기 위해 오프라인에서의 성착취 목적 대화도 형사처벌 대상으로 포함시켰고, 미수범에 대한 처벌 규정도 신설되었다(「청소년성보호법」제15조의 2 제1항 개정, 제3항 신설). 이러한 미수범 처벌 규정이 신설됨에 따라 성적 대화를 시도하는 초기 단계부터 처벌할 수 있게 되었다.

다. 불법·유해 콘텐츠에 대한 현행법상 규제

청소년유해매체물로서 대통령령으로 정하는 매체물을 판매·대여·배포하거나 시청·관람·이용하도록 제공하려는 자는 그 상대방의 나이 및 본인 여부를 확인하여야 하고, 청소년에게 판매·대여·배포하거나 시청·관람·이용하도록 제공하여서는 안 된다고 규정하고, 이를 위반하면 3년 이하의 징역 또는 3천만원 이하의 벌금에 처하도록 하고 있다(「청소년보호법」제16조, 제58조 제1호). 인터넷게임 제공자는 16세 미만 청소년을 회원으로 가입 시킬 때 친권자의 동의를 받아야 하고(「청소년보호법」제24조 제1항), 16세 미만의 청소년 회원 가입자의 친권자에게 제공되는 게임의 특성·등급 유료화정책 등에 관한 기본적인 사항과 결제 정보를 알려주도록 하고 있다(「청소년보호법」제25조 제1항).

「정보통신망 이용촉진 및 정보보호 등에 관한 법률(이하 ‘정보통신망법」)제44조의7에서 불법 정보의 유통금지가 규정되어 있고, 청소년에게 유해한 음란한 부호·문언·음향·화상 등이나 허위, 사실 적시 명예훼손 정보, 공포심이나 불안감을 유발하는 부호·문언·음향·화상, 「청소년보호법」에 따른 청소년유해매체물로서 상대방의 연령 확인, 표시의무 등 법령에 따른 의무를 이행하지 아니하고 영리를 목적으로 제공하는 내용의 정보 등이 그 대상이다. 해당 정보에 대해서는 방송미디어통신위원회가 방송미디어통신심의위원회의 심의를 통해 정보의 처리를 거부, 제한, 정지를 명할 수 있도록 하고 있다(「정보통신망법」제44조의7 제2항). 정보통신서비스 제공자가 방송미디어통신위원회의 명령을 이행하지 않으면 2년 이하의 징역 또는 2,000만 원 이하의 벌금에 처한다(「정보통신망법」제73조 제5호).

정보통신서비스 제공자는 만 14세 미만의 아동에게 문자·음성을 이용하여 사람과 대화하는 방식으로 정보를 처리하는 시스템을 기반으로 하는 정보통신서비스를 제공하는 경우에는 그 아동에게 부적절한 내용의 정보가 제공되지 않도록 노력해야 한다고 규정하고 있

48) 경찰청 통계에 의하면 그루밍 같은 성착취 목적 대화 발생 건수는 2023년 73건에서 2024년 202건, 2025년 273건으로 늘어나는 추세이다. 한겨레, “[단독] ‘아이템 줄게’ 속여···아동청소년 성매수 6년간 4.5배 늘어”, 2026. 1. 14., https://www.hani.co.kr/arti/society/society_general/1239587.html, 2026. 4. 30. 최종 방문.

으나(「정보통신망법」제44조의8) 위반한 경우 벌금, 과태료 부과 없어 법적 강제성은 없다. 아동·청소년을 유해한 콘텐츠 등으로부터 보호하고, 아동, 청소년의 개인정보 보호를 위해 개인정보처리자에게 만 14세 미만 아동의 개인정보를 처리하기 위하여 법정대리인의 동의를 받도록 하고 있다(「개인정보보호법」제22조의2).

2. 외국의 아동·청소년 보호 법제 현황

가. 미국

(1) 딥페이크 허위 성적 영상물 제작, 배포 등 행위 규제

미국은 최근 연방법인 「딥페이크 삭제 의무화법(Take It Down Act)」을 제정하여 2025년 5월 19일부터 시행하고 있다.⁴⁹⁾ 동의 없이 만들어진 성인, 미성년자의 성적 영상물을 고의로 유포하거나 이를 이용해서 협박하는 행위를 새로운 연방 범죄로 규정하고, 플랫폼 업체가 신고 접수가 된 후 48시간 이내 삭제하는 것을 의무화하였다.⁵⁰⁾

실제 성적 영상물뿐만 아니라 딥페이크 기술을 이용한 허위 성적 영상물도 규제 대상이다. 성인 대상 성적 영상물 관련 유포 범죄를 저지른 자는 2년 이하의 징역이나 벌금 또는 둘 다에 처해질 수 있고, 미성년자 대상 성적 영상물 관련 유포 범죄를 저지르면 최대 3년의 징역에 처해질 수 있다. 실제 영상물을 유포하겠다고 협박하는 행위는 해당 유포 범죄와 동일하게 처벌받고, 허위 성적 영상과 관련된 협박 범죄의 경우 최대 18개월의 징역형, 미성년자 관련 협박의 경우 최대 30개월의 징역형에 처하도록 규정하였다. 그리고 동법은 플랫폼 업체에게 피해 신고를 받고 즉시 또는 늦어도 접수 후 48시간 이내 영상물을 삭제해야 하는 의무를 부과하고 있고, 원본뿐만 아니라 해당 묘사와 동일한 것으로 알려진 모든 사본을 확인하고 삭제하기 위해 합리적인 노력을 하도록 규정하였다. 이를 위반시 연방거래위원회법⁵¹⁾에 따라 불공정하거나 기만적인 행위 또는 관행을 정의하는 규칙 위반으로 간주된다.⁵²⁾

49) 연방법 차원에서는 성인 대상 딥페이크 허위 성착취물을 직접적으로 형사처벌하는 법이 없었고, 아동에 대해서는 아동 포르노 형사처벌 규정으로 의율해 오다가 딥페이크 성착취물 관련 형사적 규제를 하는 연방법이 제정된 것이다.

<https://www.congress.gov/crs-product/LSB11314>, 2026. 4. 30. 최종 방문.

50) <https://www.washingtonpost.com/technology/2025/04/28/congress-deepfake-revenge-porn-law/>, 2026. 4. 30. 최종 방문.

51) 15 U.S. Code §57 a(a)(1)(B). <https://www.law.cornell.edu/uscode/text/15/57a>

동법이 제작 행위를 처벌하지 않고 있는 것과 달리 미국의 각 주에서는 아동·청소년 대상 허위 성착취물에 대해 제작, 소지 행위까지 형사적으로 기존 아동 포르노법과 독립적으로 규제하는 입법안을 마련하고 있다. 루이지애나주는 2023년 8월 1일부터 미성년자에 대한 성적 딥페이크 이미지, 영상을 제작, 소지, 배포한 행위 등에 대한 처벌 규정을 마련하여 시행하고 있다. 미성년자 대상 딥페이크 자료를 제작하거나 소지하면 5년 이상 20년 이하의 징역 또는 1만 달러 이하의 벌금형 또는 둘 다에 처하도록 하고 있다. 미성년자 대상 딥페이크 자료를 배포, 판매하는 경우는 10년 이상 30년 이하의 징역 또는 5만 달러 이하의 벌금 또는 둘 다에 처하도록 하였다. 성인 대상 딥페이크 성적 자료와 달리 제작, 소지 행위도 처벌하여 아동 허위 성취물에 대한 규제를 강화하였다.⁵³⁾

캘리포니아주는 2025년 1월 1일부터 기존 아동·음란물 관련 범죄⁵⁴⁾에 미성년자 또는 미성년자로 보이는 인물이 성행위를 하는 모습을 묘사한 자료의 제작, 배포 또는 소지 행위까지 범죄 범위를 확대하여 인공지능이나 컴퓨터 생성 기술을 이용하여 제작된 18세 미만의 실존 인물 또는 가상의 인물을 묘사한 자료까지 포함하도록 하였다.⁵⁵⁾

(2) 챗봇 이용에 따른 위험 방지 및 안전 확보

미국 캘리포니아주는 2025년 10월 13일「동반자 챗봇법(Companion Chatbot law)」을 제정하여 2026년 1월 1일부터 시행 중이다.⁵⁶⁾ 이 법은 일반적인 이용자가 동반 챗봇과 상호 작용하는 과정에서 자신이 사람과 상호작용하고 있다고 오해할 가능성이 있는 경우 챗봇 플랫폼 제공자에게 해당 챗봇이 인공적으로 생성된 것이며 사람이 아니라는 사실을 명확하고 눈에 띄게 알릴 의무를 부과하고 있다. 또한 챗봇 플랫폼 제공자가 미성년자임을 알고 있는 이용자에 대해서는 해당 이용자에게 인공지능과 상호 작용하고 있음을 알리고, 최소 3시간마다 이용자에게 휴식을 취하도록 상기시키는 등 특정 조치를 취하도록 의무화하고 있다. 또한 동반 챗봇 플랫폼 운영자에게 자살 충동, 자살 또는 자해 행위에 대처하기 위한 프로토콜을 시행하도록 요구하고 있으며, 여기에는 위기 상담 기관에 연결해 주는 알림

52) <https://www.congress.gov/crs-product/LSB11314>, 2026. 4. 30. 최종 방문.

53) Louisiana Laws Revised Statutes §14:73.13. Unlawful deepfakes,

<https://www.legis.la.gov/legis/Law.aspx?d=1336015>, 2026. 4. 30. 최종 방문.

54) Section 311, 311.2, 311.11, 311.12 of the Penal Code 조항이 개정되었다.

55) https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB1831, 2026. 4. 30. 최종 방문.

56) 캘리포니아 비즈니스 및 직업법(Business and Professions Code)에 Chapter 22.6 Companion Chatbots이 추가되었다.

기능 등이 있다. 2027년 7월 1일부터 운영자가 자살 예방 사무국에 특정 사항들을 매년 보고하도록 의무화하고 있고, 보고 내용에는 이용자의 자살 충동을 감지, 제거 및 대응하기 위해 마련된 절차 등이 포함되어야 한다. 또한 법상 부과되어 있는 의무 위반으로 인해 실질적인 피해를 입은 사람이 민사 소송을 제기할 수 있도록 개인 소송권을 보장하고 있다.⁵⁷⁾

뉴욕주도 최근 「인공지능 동반자 모델법(Artificial Intelligence Companion Models law)」을 제정하여 2025년 11월 5일부터 시행하고 있다.⁵⁸⁾ 인공지능 동반자⁵⁹⁾ 시스템을 운영하거나 제공하는 모든 개인 또는 단체는 동법의 적용을 받는다. 동법에 따르면 이용자가 인공지능 동반자 챗봇에게 자살 충동이나 자해 의사를 표현하는 경우 이를 감지하고 대처하기 위한 합리적인 조치를 취하는 안전 프로토콜을 구현해야 하고, 최소한 이용자의 자살 충동이나 자해 의사 표현이 감지되면 운영자는 이용자를 위기 서비스 제공 기관(예: 자살 예방 및 정신 건강 위기 상담 전화) 또는 기타 적절한 위기 서비스 기관에 연결해야 한다.⁶⁰⁾ 운영자는 인공지능 동반자와의 상호 작용 시작 시 이용자에게 명확하고 눈에 잘 띄는 알림을 제공해야 하며, 이 알림은 하루에 한 번을 초과하지 않아야 하고, 지속적인 인공지능 동반자 상호 작용의 경우 최소 3시간마다 제공해야 한다. 이 알림은 구두 또는 서면으로 이용자가 사람과 대화하고 있지 않다는 내용을 명시해야 한다.⁶¹⁾ 개인 소송권은 인정되어 있지 않지만 뉴욕주 법무장관은 위반 행위에 대해 금지 명령을 신청하고 하루 최대 1만 5천 달러의 민사 벌금(civil penalty)을 부과할 수 있다.⁶²⁾

57) https://leginfo.ca.gov/faces/billNavClient.xhtml?bill_id=202520260SB243, 2026. 4. 30. 최종 방문

58) New York General Business CHAPTER 20에 Article 47(Section 1700 ~ 1704)이 추가되어 규정되었다.

59) 사용자의 과거 상호 작용을 기억하여 맞춤형 응답을 제공하고, 사용자의 요청 없이 감정에 기반한 질문을 하며, 개인적인 문제에 대한 지속적인 대화를 유지하는 등 사용자와 지속적인 인간 또는 인간과 유사한 관계(예: 친밀한, 낭만적인 또는 우정 관계)를 모방하도록 설계된 시스템으로 정의하고 있다.

<https://www.manatt.com/insights/newsletters/client-alert/new-york-s-safeguards-for-ai-companions-are-now-in-effect>, 2026. 4. 30. 최종 방문.

60) Section 1701, <https://www.nysenate.gov/legislation/laws/GBS/1701>

61) Section 1702, <https://www.nysenate.gov/legislation/laws/GBS/1702>

62) Section 1703. <https://www.nysenate.gov/legislation/laws/GBS/1703>

(3) 소셜미디어 이용 제한 현황

미국에서는 소셜미디어가 아동·청소년의 정신 건강에 미치는 폐해⁶³⁾로부터 아동·청소년을 보호하기 위해 연방법인 「아동 인터넷 및 디지털 안전법(Kids Internet and Digital Safety Act, 이하 ‘Kids Act’)」이 발의되었고, 하원 본회의에 상정될 예정이다. 그 내용을 보면 대상 플랫폼 제공자에게 미성년자에게 미치는 ① 신체적 폭력의 위협이 너무 심각하거나 만연하거나 객관적으로 불쾌하여 그러한 위협이 미성년자의 주요 생활 활동에 영향을 미치는 경우 ② 성적 착취 및 학대 ③ 마약, 담배, 대마초, 도박 또는 알코올의 유통, 판매 또는 사용 ④ 기만적인 행위로 인해 발생한 재정적 손해와 같은 피해를 해결하기 위한 합리적인 정책, 절차를 수립, 시행, 유지 및 집행하도록 하고 있다.⁶⁴⁾ 또한 대상 플랫폼은 미성년자 보호 장치를 다음과 같이 갖추도록 규정하고 있다. ① 다른 이용자 또는 방문자가 미성년자와 소통할 수 있는 기능 제한 ② 미성년자의 프로필이나 개인 정보가 다른 이용자 또는 방문자에게 추천되거나 제안되는 것을 방지 ③ 대상 플랫폼을 강박적으로 사용하게 만드는 설계 기능을 제한 ④ 미성년자의 위치 정보를 처리자가 아닌 제3자에게 공유하는 것을 제한하고, 위치 정보가 수집된다는 사실을 대상 미성년자의 부모에게 고지 ⑤ 대상 플랫폼에서 제공되는 개인 맞춤형 추천 시스템 사용을 거부하거나 해당 시스템에서 제공하는 추천 유형 또는 범주를 제한할 수 있는 선택권을 가시적으로 제공해야 한다. 부모를 위한 장치도 구비해야 하는데, ① 미성년자의 개인정보 및 계정 설정을 조회, 변경 등 관리할 수 있는 기능 ② 미성년자의 구매 및 금융 거래를 제한할 수 있는 기능 ③ 대상 플랫폼에서 미성년자가 보낸 총 시간을 확인하고 미성년자의 플랫폼 이용 시간을 제한할 수 있는 기능 등을 제공해야 하고, 미성년자에게 부모 도구가 활성화되었을 때 미성년자 이용자에게 명확하고 가시적으로 고지해야 한다.⁶⁵⁾ 이 법 위반시 불공정 하거나 기만적인 행위 또는 관행에 관한 연방거래위원회법 제18조(a)(1)(B)항에 따른 규정 위반으로 제재를 받게 된다.⁶⁶⁾

63) 미국 청소년들은 하루 평균 약 5시간을 소셜미디어에서 보내고 있다는 결과가 발표되었다(유튜브 약 2시간, 틱톡 약 1.5시간, 인스타그램 약 1시간 포함). 최근 조사에 따르면 미국 13~14세 청소년의 4분의 1이 매일 7시간 이상 소셜미디어를 이용하는 것으로 나타났다. <https://www.worldhappiness.report/ed/2026/social-media-is-harming-adolescents-at-a-scale-large-enough-to-cause-changes-at-the-population-level/#fn1>, 2026. 4. 30. 최종 방문: 소셜미디어 이용 비율 상위 10% 청소년들은 자살 충동을 경험하거나 자해를 시도하였고, 이용율이 높은 청소년 이용자의 41%가 정신 건강 상태가 좋지 않다는 결과도 있다. 최종선, “청소년 SNS 규제 방안에 대한 제언”, 미국헌법연구 제36권 제2호, 미국헌법학회, 2025. 8., 45면.

64) Kids Act Section 213(a)

<https://www.congress.gov/bill/119th-congress/house-bill/7757/text>

65) Kids Act Section 214.

66) Kids Act Section 601(a)(1)

각 주별로 소셜미디어 이용 관련 아동 보호를 위한 입법 상황을 보면 우선 뉴욕주는 주지사가 2024년 6월 20일 「아동을 위한 중독성 피드 착취 방지법(Stop Addictive Feeds Exploitation(SAFE) for Kids act, 이하 ‘SAFE법’이라고 한다.)」에 서명하였다. 세부 규칙이 최종 확정되지 않아 아직 시행 전이기는 하나⁶⁷⁾ 콘텐츠나 이용시간 제한에 그친 것이 아닌 알고리즘 설계 구조와 관련된 중독성 피드를 금지하는 법으로 중요한 의미가 있어 그 내용을 살펴보기로 하겠다. 대상 플랫폼이 부모의 동의(verifiable parental consent)가 없는 한 18세 미만의 이용자에게 알고리즘 기반 중독성 피드⁶⁸⁾ 제공을 못하도록 규정하고, 18세 미만 이용자에게 밤 12시(자정)부터 오전 6시 사이 중독적인 게시물에 관한 알람을 제공할 수 없도록 하고 있다.⁶⁹⁾ 이를 위반하게 되면 주 법무장관이 해당 플랫폼 업체에 대해 위법행위 중지명령, 금전적 수익 반환, 환수, 위반 건별로 최대 5,000달러의 민사 벌금(civil penalty), 손해배상 등의 조치를 할 수 있다.⁷⁰⁾

아칸소주는 미국 최초로 「소셜미디어 연령 확인법(Act 689 of the Regular Session-Regulation of Social Media)」을 2023년 9월 제정하였으나 수정헌법 제1조 위반으로 2025년 3월 31일 위헌 결정된 바 있다.⁷¹⁾ 법안은 부모 또는 법적 보호자의 명시적 동의 없이 미성년자가 계정을 소유하지 못하도록 하는 의무를 부과하였고, 운전면허증이나 디지털 신분증 등을 통해 계정 보유자의 연령을 확인하도록 하였다. 합리적인 연령 확인 의무를 이행하지 못하면 위반 행위 건당 2,500달러의 민사 벌금 등을 부담하도록 하는 규정을 포함하고 있었다.⁷²⁾ 그러나 연방지방법원에서 해당 법률은 정부의 중대한 이익을 위해 엄격

67) 2025년 9월 15일 제임스 법무장관이 규칙안을 발표하였고, 2026년 상반기부터 시행될 예정이다. 주요 내용을 보면 18세 이상 사용자 의 경우, 소셜미디어 기업은 알고리즘 기반 피드 및/또는 야간 알람에 접근을 허용하기 전에 사용자가 성인인지 확인해야 하고, 기업은 기존의 여러 방법을 사용하여 사용자의 나이를 확인할 수 있으며, 기업은 업로드된 이미지나 동영상에 요청하는 경우, 사용자의 이메일 주소나 전화번호를 확인하여 사용자의 나이를 나타내는 다른 정보와 교차 검증하여 연령을 확인하도록 하였다. 소셜미디어 기업은 정부 발행 신분증 제공 외에도 연령 확인을 위한 최소 한 가지 이상의 대안을 제공해야 하고, 소셜미디어 기업은 알고리즘 기반 피드 및/또는 야간 알람에 대한 부모 동의를 요청하기 전에 먼저 미성년자의 승인을 받아야 하며, 미성년자가 동의하면 플랫폼은 미성년자가 알고리즘 기반 피드 및/또는 야간 알람에 접근할 수 있도록 부모의 확인 가능한 동의를 요청할 수 있도록 하였다.

<https://ag.ny.gov/press-release/2025/attorney-general-james-releases-proposed-rules-safe-kids-act-restrict-addictive>, 2026. 4. 30. 최종 방문.

68) 해당 서비스에서 생성되거나 공유된 여러 개의 콘텐츠(미디어)가 동시에 또는 순차적으로, 이용자 또는 이용자의 기기와 관련된 정보를 전부 또는 일부 기반으로 하여 이용자에게 추천, 선택 또는 우선순위로 표시되게 하는 것을 말한다. The general business law Article 45 SAFE For Kids Act Section 1500

69) SAFE For Kids Act Section 1501, Section 1502.

<https://www.nysenate.gov/legislation/bills/2023/S7694/amendment/A>; 김슬기·박윤미, “아동 청소년의 문제적 미디어 이용과 입법적 대응 방안 - 해외 입법사례 검토를 중심으로”, 언론과 법 제24권 제1호, 사단법인 한국언론법학회, 2025. 4., 313~314면.

70) SAFE For Kids Act Section 1508.

71) <https://arkansasadvocate.com/2025/04/01/federal-judge-declares-arkansas-social-media-age-verification-law-unconstitutional/>, 2026. 4. 30. 최종 방문.

하게 제한된 것이 아니고, 내용에 기반한 표현의 자유에 제한을 가하고 있으며, 가족들이 자녀의 스크린 사용 시간을 관리할 수 있도록 맞춤형 도구를 사용하는 대신 18세 미만 모든 사람의 플랫폼 사용을 전면 금지하려고 했고, 그 결과 다른 모든 사람의 합법적인 표현의 자유를 위축시켜 아칸소 주민들의 수정헌법 제1조 권리를 침해하고, 어떤 회사가 적용을 받는지 알 수 없는 모호한 규정으로 자의적인 법 집행을 조장할 위험이 있으므로 적법 절차 권리 또한 침해한 법률이라며 위헌 결정을 하였다.⁷³⁾

플로리다주는 2024년 3월 25일 「아동 온라인 보호에 관한 법률(Online Protections for Minors Act)」인 H3 법안을 통과시켜 2025년 1월부터 시행 중이다.⁷⁴⁾ 2025년 6월 연방 지방법원 판사가 동법의 시행을 금지하는 예비적 금지명령을 내렸으나 2025년 11월 연방 항소법원은 예비적 금지명령의 효력을 정지하는 판결을 하였다.⁷⁵⁾ 대상 소셜미디어 플랫폼은 14세 미만 미성년자와 계정 소유자 계약을 체결해서는 안 되고, 14세 또는 15세 미성년자는 부모 또는 보호자의 동의가 있는 경우에만 계정 소유자가 될 수 있도록 규정하고 있다. 14세 미만의 계정 및 14세, 15세의 미성년자는 부모의 동의가 없는 경우 대상 소셜미디어 플랫폼은 계정을 해지해야 한다. 대상 소셜미디어 플랫폼 사업자가 미성년자에게 유해한 자료를 웹사이트나 애플리케이션에 고의로 게시하거나 배포한 경우 대상 웹사이트나 애플리케이션에 미성년자에게 유해한 자료가 상당 부분 포함되어 있다면 미성년자의 해당 자료 접근을 금지하도록 요구하고 있고, 대상 소셜미디어 플랫폼 사업자는 익명 또는 표준 연령 확인 방법을 사용하여 미성년자에게 유해한 자료에 접근하려는 사용자의 연령이 연령 요건을 충족하는지 확인하도록 하였다.⁷⁶⁾ 대상 소셜미디어 플랫폼 사업자가 동법의 요건을 고의로 또는 무모하게 위반할 시 플로리다주의 기만 및 불공정 거래 관행법에 따라 제재를 받을 수 있다. 법무부는 위반 건당 최대 5만 달러의 민사 벌금, 합리적인 변호사 수

72) 최중선, 앞의 논문(주 63), 53~54면.

73) <https://netchoice.org/netchoice-secures-key-victory-for-free-speech-court-permanently-blocks-arkansas-age-verification-law/>, 2026. 4. 30. 최종 방문.

74) 성윤숙, “SNS에서 청소년 보호 동향과 시사점”, 지능정보윤리 이슈리포트, 2025., 17면;

<https://www.flsenate.gov/Committees/BillSummaries/2024/html/3>, 2026. 4. 30. 최종 방문.

75) 플로리다주 정부는 동법이 소셜미디어 자체의 사용을 금지하는 것이 아니라, '무한 스크롤', 푸시 알림, '좋아요나 공유 관련 개인 맞춤형 지표', '자동 재생 동영상', '라이브 스트리밍'과 같은 중독성 기능을 사용하는 소셜미디어만을 금지하는 것이라고 주장했고, 항소법원 다수 의견은 표현의 자유를 제한할 가능성이 낮으며 이 주장을 받아들였다.

<https://floridaphoenix.com/2025/11/26/attorney-general-will-now-aggressively-enforce-social-media-ban-on-minors/>, 2026. 4. 30. 최종 방문.

76) 김민우·김일환, “아동의 소셜미디어(SNS) 이용 규제 입법 동향에 관한 고찰”, 미국헌법연구 제36권 제1호, 미국헌법학회, 2025. 4., 153면; 최중선, 앞의 논문(주 63), 55~57면.

임료 및 소송 비용, 그리고 특정 조건하에서 징벌적 손해배상을 부과할 수 있다. 미성년자인 계정 소유자는 최대 1만 달러의 손해배상을 청구할 수 있도록 규정하고 있다.⁷⁷⁾

나. 영국

(1) 딥페이크 허위 성적 영상물 제작, 배포 등 행위 규제

「온라인안전법(Online Safety Act 2023)」은 「성범죄법(Sexual Offences Act 2003)」에 제66A조와 제66B조를 추가하여 동의 없이 또는 동의했다고 합리적으로 믿을 만한 근거 없이 성인의 은밀한 이미지(딥페이크 이미지 포함)를 공유하거나 공유하겠다고 위협하는 행위를 범죄로 규정하였다.⁷⁸⁾ 18세 미만의 아동 대상 딥페이크 성적 이미지에 대해서 1978년 「아동보호법(Protection of Children Act 1978)」 제1조⁷⁹⁾에 따라 아동의 음란 이미지 또는 가짜 이미지를 촬영, 제작, 배포 또는 게시하는 행위를 처벌하고, 1988년 「형사사법법(Criminal Justice Act 1988)」 제160조에 따라 아동의 음란 이미지를 소지하는 행위를 처벌하고, 2009년 검시관 및 사법법(Coroners and Justice Act 2009) 제62조를 위반하여 아동의 금지된 이미지를 소지하는 행위를 형사처벌하고 있다.⁸⁰⁾

「온라인안전법」은 플랫폼에 아동 성착취물(CSAM)을 포함한 불법 콘텐츠를 삭제할 의무를 부과하고 있고, 이러한 의무에는 플랫폼에서 아동에게 잠재적인 위험을 식별하기 위한 위험 평가를 수행하고 아동 착취 및 학대를 예방하는 것이 포함된다. 그리고 「범죄 및 치안법(Crime and Policing Bill 2025)」이 통과되면 아동 성착취물(CSAM)을 생성하도록 설계된 AI 도구를 소유, 제작 또는 배포하는 행위가 새로운 형사 범죄로 규제되게 된다. 또한 아동 성착취 이미지를 제작하기 위해 AI 도구를 사용하거나 악용하는 방법을 안내하는 ‘소아성애 매뉴얼(paedophile manuals)’을 소지하는 행위도 범죄화된다.⁸¹⁾

영국은 성인 대상 딥페이크 허위 성착취물을 제작하는 행위에 대해서는 형사처벌 규정이 없었는데, 「데이터(이용 및 접근)법[Data(Use and Access) Act 2025)」 제138조에 따라

77) HB 3 Fla. Stat. §501.1736-.1738 (2024)

78) Online Safety Act Section 184, Section 188

79) 아동보호법은 “가짜 사진이란 컴퓨터 그래픽이나 그 밖의 어떤 방법으로든 사진처럼 보이는 이미지를 말한다”고 정의하여 딥페이크 기술을 이용한 아동 음란 이미지 제작, 배포, 소지를 처벌하고 있다. <https://www.legislation.gov.uk/ukpga/1978/37>.

80) <https://www.cps.gov.uk/prosecution-guidance/obscene-publications>, 2026. 4. 30. 최종 방문.

81) <https://www.qmul.ac.uk/lac/our-legal-blog/2026-blogs/items/deepfakes-and-consent-the-law-finally-catches-up-.html>, 2026. 4. 30. 최종 방문

2003년 「성범죄법」에 제66E, 제66F조를 추가하여 인공지능으로 생성된 성인 대상 가짜 성적 이미지를 제작하거나 제작을 의뢰하는 행위를 형사처벌하는 규정을 새롭게 도입하여 2026년 2월 6일부터 시행하고 있다. 이 범죄는 완전히 조작된 이미지뿐만 아니라 디지털 방식으로 수정된 모든 이미지(일반적인 포토샵 작업과 AI로 생성된 딥페이크 모두 포함)를 대상으로 하고 있다. 동의 없이 성적 딥페이크 영상을 제작하도록 타인에게 요청하거나 사주하는 행위 또한 범죄로 규정하고 있고, 최대 2년의 징역형이 부과될 수 있다.⁸²⁾

(2) 소셜미디어 이용 제한 및 불법·유해 콘텐츠 규제

영국은 아직 아동·청소년의 소셜미디어 이용 제한을 시행하고 있지 않으나 수백 명의 10대 청소년들을 대상으로 소셜미디어 이용 금지, 디지털 통행 금지, 앱 사용 시간 제한 등의 시범 사업을 진행할 계획으로 이 사업은 16세 미만 청소년의 소셜미디어 접근을 금지해야 하는지에 대한 의견 수렴과 함께 진행될 예정이다.⁸³⁾

그리고 불법·유해 콘텐츠로부터 아동·청소년을 보호하기 위하여 2023년 10월 26일 제정된 「온라인안전법(Online Safety Act)」에 따라 사용자 간 서비스⁸⁴⁾ 및 검색 서비스 제공자에게 불법 콘텐츠 및 아동에게 유해한 콘텐츠와 관련해서 적절한 위험 평가를 실시하고, 위험을 효과적으로 완화하기 위해 비례적이고 합리적인 시스템과 프로세스를 구축해야 하며 불법 콘텐츠를 발견하면 삭제하도록 하는 의무를 부과하고 있다. 서비스 제공자는 아동에게 유해하고 연령에 부적절한 콘텐츠에 아동이 접근하지 못하도록 차단해야 하고, 문제가 발생했을 때 부모와 아동이 온라인에서 문제를 신고할 수 있는 명확하고 접근하기 쉬운 방법을 제공해야 한다.⁸⁵⁾ 또한 위험 평가를 할 때 알고리즘이 불법 콘텐츠 및 아동에게 유해한 콘텐츠 노출에 어떤 영향을 미치는지를 포함하여 평가하도록 하고 있다.⁸⁶⁾

「온라인안전법」에 따라 온라인 플랫폼 사업자는 2025년 7월 25일부터 18세 미만 아동·

82) https://www.gov.uk/government/news/government-crackdown-on-explicit-deepfakes?utm_source=chatgpt.com, 2026. 4. 30. 최종 방문.

83) <https://www.gov.uk/government/news/children-and-parents-to-pilot-social-media-bans-time-limits-and-curfews-at-home-as-government-tests-next-steps-to-give-uk-kids-their-childhood-back>, 2026. 4. 30. 최종 방문.

84) 인터넷서비스를 말하며 이를 통하여 서비스 사용자가 해당 서비스상에 직접 생성하거나 서비스 사용자가 해당 서비스상에 업로드하거나 공유하는 콘텐츠를 다른 사용자나 사용자들이 접할 수 있다. Online Safety Act Section3(1)

85) 김현경, “영국 온라인안전법의 주요 내용 및 과제”, 방송통신 심의 동향 26권, 방송통신심의위원회, 77~80면.

86) 임세란·박아란, “온라인 공간에서의 아동 청소년 보호 - 유럽연합과 영국의 법률적 대응을 중심으로”, 언론과 법 제23권 제3호, 2024. 12., 115 ~ 120면; 김여라, “아동 청소년의 온라인 안전 및 권리 강화를 위한 과제”, 이슈와 논점 제2284호, 국회입법조사처, 2024. 11. 11., 3면; 최종선, “‘애들아, 인스타를 내려놓을지니’... 봇물 터진 청소년 sns 규제”, kiso 저널 57권, 2024. 11., 28면.

청소년이 자살, 자해, 혐오 발언, 폭행, 섭식 장애 및 음란물 등과 관련된 유해 콘텐츠에 접근하지 못하도록 해야 하고, 어린이 피드에 표시되는 콘텐츠를 결정하는 알고리즘을 변경하여 유해 콘텐츠를 걸러내야 하며 최우선 순위 콘텐츠인 음란물이나 자해, 자살 또는 섭식 장애를 조장하는 콘텐츠를 게시하는 플랫폼은 아동의 해당 콘텐츠 접근을 막기 위해 강력한 연령 확인 절차를 시행해야 한다. 아동 안전에 대한 책임을 지는 담당자를 지정하고, 해당 담당자가 플랫폼에서 아동에게 미치는 위험을 어떻게 관리하고 있는지 매년 검토하도록 하였다. 온라인안전법에 따른 규정을 준수하지 않는 경우 기업은 1,800만 파운드 또는 전 세계 매출의 10% 중 더 높은 금액의 과징금을 부과 받거나 임원이 형사처벌될 수 있다.⁸⁷⁾

다. 호주

(1) 딥페이크 성범죄

「형법(Criminal Code Act 1995)」개정에 따라 제474.17조에 통신 서비스를 이용하여 다른 사람(18세 이상이거나 18세 이상으로 보이는 사람)을 묘사(또는 묘사하는 것처럼 보이는)한 성적 콘텐츠를 상대방이 전송에 동의하지 않은 사실을 알고 있거나 동의했는지 여부를 알지 못한 채 무모하게 전송, 유포한 경우를 새롭게 처벌하는 규정이 신설되어 2024년 9월 3일부터 적용되고 있다.⁸⁸⁾ 콘텐츠가 변경되지 않은 형태인지 또는 기술을 사용하여 생성되었거나 어떤 방식으로든 변경되었는지는 중요하지 않다고 규정하여 딥페이크 디지털 기술을 이용하여 제작 또는 변형된 비동의 딥페이크 콘텐츠를 규제 대상으로 하고 있다. 최대 6년의 징역형에 처할 수 있고, 가중 처벌 대상인 경우(예: 피고인이 자료를 제작 또는 변경한 경우)에는 최대 7년의 징역형에 처할 수 있다. 동의 없이 음란물을 제작하거나 변경하는 행위 자체는 범죄로 간주되지 않지만 제작이나 변경한 자가 통신 서비스를 이용하여 이를 전송, 유포하면 가중 처벌 대상이 된다.⁸⁹⁾

아동에 대해서는 아동 음란물 정의 규정에 따라 18세 미만이거나 18세 미만으로 보이는 사람 또는 그 사람을 묘사하는 자료, 즉 딥페이크 기술로 제작된 아동 음란물도 포함되고,

87) <https://www.gov.uk/government/news/whats-changing-for-children-on-social-media-from-25-july-2025>, 2026. 4. 30. 최종 방문.

88) https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bld=r7205; https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/bd/bd2324a/24bd081, 2026. 4. 30. 최종 방문

89) <https://www.ashworthlawyers.com.au/blog/what-does-it-mean-to-share-sexual-material-without-consent-online>, 2026. 4. 30. 최종 방문.

형법 제474.19조 및 형법 제474.20조가 적용되어 통신 서비스를 이용하여 아동 딥페이크 음란물을 전송, 제공, 게시, 배포, 광고 또는 홍보하거나 요구한 경우, 아동 딥페이크 음란물을 소지 또는 관리하거나 제작, 공급 또는 취득하는 경우 모두 형사처벌 대상이고, 15년 이하의 징역형에 처해질 수 있다.⁹⁰⁾

(2) 소셜미디어 이용 제한

2024년 11월 29일 유해 정보와 온라인 중독으로부터 아동 청소년을 보호하기 위해 16세 미만의 소셜미디어 계정 소유를 전면 금지하는 「아동·청소년 소셜미디어금지법(Social Media Minimum Age Bill 2024)」을 통과시켰고, 2025년 12월 10일부터 시행되고 있다.⁹¹⁾ 현재 페이스북, 인스타그램, 스냅챗, 스레드, 틱톡, X, 유튜브, 레딧, 스트리밍 플랫폼인 키크와 트위치 등 10개의 소셜미디어 플랫폼이 적용 대상으로 되어 있다. 아동의 플랫폼 접근을 막기 위해 ‘합리적인 조치’를 취해야 하며 다양한 연령 확인 기술을 사용하도록 규정하고 있다. 정부 발행 신분증, 얼굴 또는 음성 인식, 또는 온라인 행동 및 상호 작용을 분석하여 개인의 나이를 추정하는 소위 ‘연령 추론’ 등이 포함된다.⁹²⁾ 법원은 미성년자가 플랫폼에 계정을 갖는 것을 막기 위한 합리적인 조치를 취하지 않는 플랫폼에 대해 민사 제재를 명령할 수 있고, 최대 15만 페넬티 단위(현재 호주 달러로 약 4,950만 달러)의 민사 벌금을 부과할 수 있다.⁹³⁾

이 법안에 대해서는 10대 청소년 2명이 13세에서 15세 사이 청소년들의 소셜미디어 활동과 관련된 표현의 자유와 참여권을 희생시키는 결과를 초래하였고, 피해를 야기하는 소셜미디어의 특정 기능만을 제한하지 않고 계정 소유를 금지하는 등 헌법에 위배된다고 최고법원에 소송을 제기하여 심리가 예정되어 있고, 레딧도 성인과 미성년자 모두의 사생활 침해와 익명성, 미성년자의 정치적 소통의 자유를 심각하게 침해한다고 최고법원에 소송을 제기한 상태이다.⁹⁴⁾

90) <https://www.legislation.gov.au/C2004A04868/2018-12-29/text/2>.

91) 부모 동의를 통한 예외를 허용하지 않는 강력한 제한 조치이다. 최진웅, “호주의 ‘아동 청소년 소셜미디어 금지법’ 통과와 입법적 시사점”, NARS 현안 분석 제345호, 국회입법조사처, 2024. 12. 31., 1면.

92) <https://www.bbc.com/news/articles/cwyp9d3ddqyo>, 2026. 4. 30. 최종 방문.
<https://www.legislation.gov.au/C2024A00127/asmade/text>

93) <https://www.esafety.gov.au/about-us/industry-regulation/social-media-age-restrictions#key-measures-under-the-law>, 2026. 4. 30. 최종 방문

94) <https://www.bbc.com/news/articles/clykk2yrl9ko>, 2026. 4. 30. 최종 방문.

한편 호주는 2025년 3월 9일부터 인공지능 서비스 제공업체에게도 연령 확인 조치를 마련해야 하는 의무를 부과한다고 발표했다. 적용 대상 서비스에는 챗GPT와 같은 생성형 AI 서비스, 동반자형 챗봇 서비스 및 검색 엔진이 포함된다.⁹⁵⁾

IV. 아동·청소년 보호를 위한 법제 개선 방안

1. 정보통신서비스 제공자에 대한 책임 강화 및 신속한 피해 구제

가. 신속 삭제, 접근 차단 등 유통방지 정보 확대 및 위험평가 의무 부과

소셜미디어 등을 통해 딥페이크 기술로 만들어진 허위 성착취물, 온라인사칭 행위와 관련된 괴롭힘 콘텐츠, 부모의 신체 사진을 달라고 강요하여 합성한 영상물, 학교폭력 장면을 찍은 영상, 사이버 불링 게시물 등이 사이버 공간에서 널리 유포되고 있다. 그러나 일반에게 공개된 불법 정보가 아니면 삭제나 접속차단 등 조치가 신속하게 이루어지기 어려운 실정이다. 정보통신망법에서는 불법정보에 대해서 유통금지의무를 부과하고 있고(「정보통신망법」제44조의7 제1항), 사업의 종류 및 규모 등이 동법 시행령이 정하는 기준⁹⁶⁾에 해당하는 정보통신서비스 제공자에게는 불법 정보의 유통을 방지하기 위하여 기술적·관리적 조치⁹⁷⁾를 해야 할 의무를 부과하고 있다(「정보통신망법」제44조의7 제5항).⁹⁸⁾ 여기서의 불법 정보는 대부분 현행법상 형사처벌 대상이 되는 정보이다. 형사처벌 대상이 되지 않는 사이버 불링, 사이버 성폭력, 지속적이거나 반복적이지 않은 온라인사칭 게시물 등은 불법 정보에 해당되지 않아 방송미디어통신위원회의 정보 처리 거부, 정지 명령 대상이 아니고, 기술적·관리적 조치 대상도 아니어서 정보통신서비스 제공자는 삭제하거나 정보처리를 제

95) <https://www.aitimes.com/news/articleView.html?idxno=207433>, 2026. 4. 30. 최종 방문.

96) 타인의 정보통신서비스를 이용자에게 제공하기 위하여 원본 서버로부터 제공받은 데이터의 사본을 국내에 데이터를 임시적으로 저장하는 서버에 저장하여 전송하는 사업을 하는 자로서 정보통신서비스 부문 전년도(법인인 경우에는 전 사업연도를 말한다) 매출액이 10억원 이상인 자를 말한다(동법 시행령 제35조의2 제1항)

97) 신속하게 접근을 제한하는 조치, 정보의 게재자에게 해당 정보의 유통금지를 요청하는 조치, 접근 제한 조치의 운영·관리 상태를 시스템에 자동으로 기록되도록 하고, 2년 동안 보관하는 조치, 불법 정보의 유통을 방지하기 위하여 대통령령으로 정하는 조치(① 유통을 방지하기 위한 업무 담당자 지정 ② 불법정보의 유통에 대하여 알게 된 경우 방송미디어통신위원회에 통보 ③ 불법정보의 유통 방지에 관한 내용을 이용약관 또는 계약서류에 명시)를 말한다.

98) 기술적·관리적 조치를 하지 않은 자에 대해서는 정보통신망법 제76조 제3항 제4호의2에 따라 1,000만 원 이하의 과태료가 부과되도록 하고 있으나 매우 경미한 제재여서 실효성이 거의 없을 것으로 판단된다.

한하지 않아도 된다. 「청소년보호법」에 따른 청소년유해매체물로서 상대방의 연령 확인, 표시의무 등 법령에 따른 의무를 이행하지 아니하고 영리를 목적으로 제공하는 내용의 정보가 불법 정보로 규정되어 있으나 영리를 목적으로 하지 않는 청소년 유해 콘텐츠는 대상 정보가 아니어서 정보통신서비스 제공자가 영리 목적이 아닌 아동, 청소년에게 유해한 게시물에 대해서 삭제, 접속차단 등 조치를 취하지 않아도 별도의 법적 제재를 받지 않는다.

위와 같은 불법 정보에 대한 삭제 절차 외에 피해 당사자의 직접 신고를 통한 삭제 절차가 「정보통신망법」에 규정되어 있으나 피해자 보호나 신속한 피해 구제에는 미흡한 점이 많다. 정보통신망을 통하여 명예훼손, 사생활 침해 등 피해를 당한 피해자가 적극적으로 침해사실을 소명하여 정보통신서비스 제공자에게 신고를 하면 정보통신서비스 제공자는 해당 정보를 삭제하거나 임시조치 등의 필요한 조치를 취하도록 되어 있지만(「정보통신망법」제44조의 2 제1항) 해외 정보통신서비스 제공자는 적극적으로 응하지 않는 것으로 보인다.⁹⁹⁾ 또한 정보통신서비스 제공자가 피해자의 삭제 요청을 거부하더라도 피해자가 방송미디어통신위원회에 이의를 제기하거나 신고하는 법적 절차가 마련되어 있지 않다.

정보통신망을 통하여 일반에게 공개되어 유통되는 정보 중 불법 촬영물(딥페이크 허위 영상물 포함) 등에 대해서는 정보통신서비스 제공자가 이를 인식하면 지체 없이 삭제, 접속차단 등 유통 방지에 필요한 조치를 취해야 하고(「전기통신사업법」제22조의5 제1항), 유통방지에 필요한 조치를 취하지 않으면 3년 이하의 징역 또는 1억 원 이하의 벌금에 처하게 된다(「전기통신사업법」제95조의2 제1호의2). 지체 없이 삭제, 접속차단 등 조치를 취하도록 되어 있으나 삭제 시한이 정해져 있지 않아 신속하게 삭제가 되지 않고 있다.¹⁰⁰⁾ 또한 불법 촬영물(딥페이크 허위 영상물 포함) 등에 대해서 정보통신서비스 제공자에게 기술적·관리적 조치를 취하도록 의무를 부과하고 있고(「전기통신사업법」제22조의5 제2항), 이러한 기술적·관리적 조치를 취하지 아니하면 3년이하의 징역 또는 1억 원 이하의 벌금에 처하도록 하고 있다(「전기통신사업법」제95조의2 제1호의3). 전기통신사업법에 따른 불법촬영물 등에 대한 유통방지의무, 기술적·관리적 조치 의무 모두 일반에게 공개되어 유통되는 정보를 대상으로 하고 있어 비밀 채팅방이나 1:1 대화방에서 유통되는 정보에 대해서는 위와 같은 의무가 부과되지 않는다.¹⁰¹⁾

99) 안성훈·심승범·류부근·신혜진, “디지털플랫폼 범죄대응과 피해 방지를 위한 형사정책연구(Ⅰ)”, 한국형사법무정책연구원, 2024. 12., 339~340면.

100) 동아일보, “청소년 20명 중 1명 ‘딥페이크 피해’… 삭제는 ‘느릿느릿’”, 2025. 12. 18., <https://www.donga.com/news/Society/article/all/20251217/132988590/2>, 2026. 4. 30. 최종 방문.

국내에서는 형사처벌 대상 정보가 아닌 아동·청소년에게 심각한 피해를 야기하는 유해한 정보에 대한 신속한 삭제나 접속 차단 등 조치가 이뤄지기 어려운 실정이다. 이러한 국내 상황과 달리 영국, 호주는 정보통신서비스 제공자에게 형사처벌 대상 정보만이 아니라 아동·청소년에게 미치는 부정적 영향을 고려하여 유해한 정보에 대해서도 신속한 삭제, 접근을 제한하는 등 적극적인 의무를 부과하고 있다. 또한 EU, 영국, 호주는 정보통신서비스 제공자에게 아동·청소년 보호를 위해 여러 시스템을 갖추도록 하고 있어 아동·청소년 보호 강화 측면에서 참고할 내용이 많으므로 간단히 관련 법 내용을 살펴보도록 하겠다.

2024년 2월 17일부터 EU 「디지털서비스법(Digital Service Act)」이 중개 서비스, 호스팅서비스, 온라인 플랫폼 서비스, 초대형 온라인 플랫폼 서비스(EU에서 4,500만 명 이상의 사용자가 이용 중인 플랫폼 서비스)를 대상으로 전면적으로 시행되고 있다.¹⁰²⁾ 중개 서비스는 나머지 서비스를 포괄하는 개념으로 중개서비스 제공자는 중개서비스 제공자가 전송하거나 저장하는 정보를 감시하거나, 불법행위를 나타내는 사실 또는 상황을 적극적으로 조사해야 할 일반적인 모니터링 의무는 없으나¹⁰³⁾ 호스팅서비스 제공자는 불법 콘텐츠¹⁰⁴⁾를 인지한 즉시 불법 콘텐츠를 삭제하거나 접근을 차단하기 위하여 신속하게 조치를 취한 경우 면책될 수 있고,¹⁰⁵⁾ 호스팅서비스 제공자와 온라인 플랫폼 서비스 제공자는 불법 콘텐츠에 대한 피해 신고 및 조치 메커니즘을 구축하도록 규정하고 있다. 피해자의 신고에 따라 삭제, 접근 제한, 서비스 제공 중단, 사용자 계정 정지, 해지 등 조치가 이뤄지게 하고, 이에 대해 이의제기를 할 수 있는 시스템을 마련하도록 하고 있다.¹⁰⁶⁾ 또한 온라인 플랫폼 제공자는 명백히 불법적인 콘텐츠를 빈번하게 제공하는 서비스 이용자에게 사전 경고 후 합리적인 기간 동안 서비스 제공을 중단해야 한다.¹⁰⁷⁾ 추천 시스템의 투명성 구현을 위

101) <https://www.kmcc.go.kr/user.do?mode=view&page=A05030000&boardId=1113&boardSeq=52287>, 2026. 4. 30. 최종 방문.

102) 김현수·전성호, “유럽연합 디지털서비스법안(Digital Service Act)의 주요 내용 및 시사점”, KISDI Premium Report 2020권 제11호, 정보통신정책연구원, 2020. 12. 21., 4~5면.

103) Digital Service Act Article 8

104) Digital Service Act Article 3(h) “불법 콘텐츠”란 유럽연합 법률 또는 유럽연합법률에 따른 회원국의 국내법상 명시된 목적이나 성격에 관계 없이 그 자체로 또는 제품 판매나 서비스 제공을 포함한 활동에 있어 위 법률을 준수하지 아니하는 정보를 말한다. 전문 (12) 구체적인 예로는 아동 성학대를 묘사한 이미지 공유, 개인 이미지의 불법적인 무단 공유, 온라인 스토킹, 불합격품 또는 위조 제품 판매, 소비자 보호법 위반 제품 판매 또는 서비스 제공, 저작권 보호 자료의 무단 사용, 숙박 서비스의 불법 제공 또는 살아있는 동물의 불법 판매 등이 있다. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2065&qid=1670837883291#d1e1601-1-1>

105) Digital Service Act Article 6

106) Ioanna Tourkochoriti, “The Digital Services Act and the EU as the Global Regulator of the Internet”, Chicago Journal of International Law Vol. 24 No.1., 2023., pp.135~138; 천혜선, “유럽 디지털서비스법(DSA)의 주요 내용 및 시사점”, 방송통신심의동향 통권 제24권, 방송통신심의위원회, 2022. 2., 91~92면.

107) Digital Service Act Article 24

해 주요 매개변수 등 알고리즘 정보를 공개하고, 서비스 이용자가 선호하는 옵션을 쉽게 선택하고 수정하는 기능을 제공해야 하며,¹⁰⁸⁾ 미성년자가 접근할 수 있는 온라인 플랫폼의 제공자는 해당 서비스에서 높은 수준으로 아동·청소년의 개인정보 보호, 안전, 보안을 보장하기 위해 적절하고 비례적인 조치를 취해야 한다.¹⁰⁹⁾ 온라인 플랫폼 제공자는 아동, 청소년의 개인 정보를 이용한 광고를 노출해서는 안 되고¹¹⁰⁾, 초대형 온라인 플랫폼과 검색 엔진은 알고리즘을 포함하여 아동·청소년의 잠재적인 온라인 위험을 매년 식별하고 평가해야 하는 위험 평가 의무를 부담한다.¹¹¹⁾

영국 「온라인안전법」은 서비스 제공자가 이용자의 신고 또는 다른 경로를 통해 불법 콘텐츠를 인지한 경우 이를 삭제하도록 하고 있다. 아동 성학대, 통제적이거나 강압적인 행동, 심한 성폭력, 극단적인 포르노, 사기, 인종적 또는 종교적 동기가 있는 공공질서 위반 범죄, 폭력 선동, 불법 이민 및 인신매매, 자살을 조장하거나 용이하게 하는 행위, 친밀한 이미지 남용, 불법 마약이나 무기 판매, 성적 착취, 테러를 불법 콘텐츠로 규정하고 있고, 아동·청소년에게 유해한 정보에 대해서는 접근을 제한하는 조치를 취하도록 하고 있다. 최우선 보호 콘텐츠(Priority Content)¹¹²⁾는 아동·청소년이 접근하지 못하도록 반드시 차단되어야 하고, 우선 보호 콘텐츠(Priority Content)¹¹³⁾에 대해서는 연령에 적합한 방식으로 접근이 제한되어야 한다. 서비스 제공자는 플랫폼 이용으로 인해 아동에게 발생할 수 있는 위험을 평가하고 적절한 연령 제한을 설정하여 아동 사용자가 연령에 맞는 경험을 하고 유해 콘텐츠로부터 보호받을 수 있도록 해야 한다. 특정 연령 미만 아동의 서비스 이용을 제한하는 조치를 하는 서비스 제공자는 서비스 약관에 해당 연령 제한을 시행하기 위해 사용하는 조치를 명시하고 이를 일관되게 적용해야 한다.¹¹⁴⁾

호주의 「온라인안전법(Online Safety Act 2021)」은 18세 미만의 아동 청소년에 대한

108) Digital Service Act Article 27.

109) Digital Service Act Article 28 Section 1.

110) Digital Service Act Article 28 Section 2.

111) Digital Service Act Article 34 및 Article 35. 이권일, “독일에서의 규제 논의의 현황과 과제- 독일 네트워크법(NetzDG)와 EU 디지털 서비스법(DSA)을 중심으로”, 공법학연구 제25권 제2호, 2024. 5., 17~20면.

112) 음란물, 자해, 성식 장애 또는 자살을 권장, 조장 또는 지시하는 콘텐츠가 여기에 해당된다.

<https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer>, 2026. 4. 30. 최종 방문.

113) 괴롭힘, 욕설이나 혐오 콘텐츠, 심각한 폭력이나 상해를 묘사하거나 조장하는 콘텐츠, 위험한 묘기와 챌린지를 조장하는 콘텐츠, 유해 물질의 섭취, 흡입 또는 노출을 조장하는 콘텐츠

<https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer>, 2026. 4. 30. 최종 방문.

114) <https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer>

사이버 불링 콘텐츠(Cyber-bullying material)¹¹⁵⁾, 18세 이상의 성인에 대한 사이버 괴롭힘 콘텐츠(Cyber-abuse material)를 구분하여 사이버 불링, 사이버 괴롭힘을 당한 피해자는 직접 서비스 제공자에게 삭제를 요청할 수 있고, 테러행위와 같은 혐오 폭력행위를 묘사한 콘텐츠에 대해선 피해자가 접속 차단을 요구할 수 있도록 규정하였다.¹¹⁶⁾ 피해를 당한 아동·청소년과 부모, 보호자는 서비스 제공자에게 직접 삭제해달라는 요청을 할 수 있고, 이에 대해 48시간 이내 조치가 이뤄지지 않으면 온라인 안전국(eSafety Commissioner)에 신고할 수 있다. 온라인 안전국은 피해 아동·청소년의 신고를 받고, 조사를 진행하여 사이버 불링 콘텐츠라고 판단하면 정보통신서비스 제공자에게 24시간 이내 이내에 콘텐츠를 삭제하라는 고지(removal notice)를 보내고¹¹⁷⁾, 서비스 제공자는 이 고지에 따른 준수 사항을 이행해야 하며, 이에 불응시 민사 벌금을 부과받게 된다.¹¹⁸⁾ 또한 호주 「온라인안전법」은 위와 같은 피해자의 직접적인 신고가 없어도 Class 1 자료(아동 성착취물, 극단적 폭력물 등), Class 2 자료(성인용 콘텐츠 중 아동 보호가 필요한 자료)에 대해서는 온라인안전국이 자체 조사, 모니터링을 통해 직접 서비스 제공자에게 삭제 명령을 하도록 규정하고 있다.¹¹⁹⁾ 삭제 통지서의 요건을 이행하지 않을 시 민사 벌금이 부과된다.¹²⁰⁾

국내 아동·청소년에 대한 보호는 영국, 호주 등에 비해 법, 제도상 미흡하여 외국 아동·청소년보다 많은 위험에 노출되어 있고, 앞으로 더 심각한 피해를 당하게 될 수 있는 상황이다. 딥페이크 기술 등 인공지능이 급속히 발전함에 따라 아동·청소년에게 형사처벌 대상이 되는 불법정보 외에 유해한 정보가 많이 만들어지고, 급속도로 퍼지게 될 것이다. 이에 불법정보 외에도 중대한 피해를 야기하는 사이버불링, 심각한 폭력, 욕설이나 혐오 정보 등 아동·청소년에게 유해한 정보에 대해서도 정보통신서비스 제공자에게 영국과 같이 위험 평가를 하고, 위험 완화를 위한 조치를 취하도록 하고, 아동·청소년의 접근을 제한하도록 정보통신서비스 제공자에게 의무를 부과할 필요가 있다. 또한 호주와 같이 피해 아동·청소년

115) 사이버 불링 콘텐츠는 합리적인 사람을 기준으로 할 때 “아동을 심각하게 위협, 협박, 괴롭히거나 굴욕감을 줌(seriously threatening, seriously intimidating, seriously harassing or seriously humiliating the Australian child) 아동에게 영향을 줄 수 있는 콘텐츠”를 의미한다. Online Safety Act(2021) 6. Cyber-bullying material targeted at an Australian child

116) <https://www.esafety.gov.au/newsroom/whats-on/online-safety-act>, 2026. 4. 30. 최종 방문: 신혜진 외 3명, 앞의 연구 보고서(주 99), 334~335면.

117) Online Safety Act(2021) Section 65, Section 88. 온라인 안전국은 신고 접수가 되고 2일 이내에 피해 신고자들에게 연락을 취하고 있다. 강희영, “서울시 디지털성범죄 피해자지원센터 운영 방안 연구”, 서울시여성가족재단, 2020. 9., 108~109면; 신혜진 외 3명, 앞의 연구 보고서, 334면(각주 477)에서 재인용.

118) Online Safety Act(2021) Section 67, Section 91에 따라 ‘civil penalty: 500 penalty units’을 부과하도록 되어 있다.

119) Section 109, 110 삭제 명령 통지를 받고 나서 24시간 이내 자료의 호스팅을 중단하기 위한 합리적인 모든 조치를 취해야 한다.

120) Section 111 Civil penalty: 500 penalty units.

년이 정보통신망법상 불법 정보가 아니어도 아동·청소년의 자살, 자해 등 중대한 피해가 발생할 수 있는 유해 정보에 대해 정보통신서비스 제공자에게 신속한 삭제를 요청하고, 이에 응하지 않는 경우 호주의 온라인안전국과 같은 역할을 하는 방송미디어통신위원회에서 심각한 유해 정보인지 여부를 직접 조사하여 삭제할 것을 서비스 제공자에게 명령할 수 있는 절차를 마련하여 아동·청소년을 온라인의 위험으로부터 적극적으로 보호해야 할 것이다.¹²¹⁾ 다만 이러한 삭제 명령을 고지하는데 있어 자의적인 삭제 명령이 되지 않도록 조사 과정에서 정보통신서비스 제공자, 정보 게시자의 의견을 청취한 후 삭제 대상 정보 및 삭제 기한을 명확하게 통지하도록 해야 한다. 아울러, 정보통신서비스 제공자, 정보 게시자가 방송미디어통신위원회에 삭제 명령에 대한 이의신청을 제기하는 절차적 권리 구제 수단을 법률상 보장함으로써 적법한 삭제 명령이 이뤄질 수 있도록 해야 할 것이다.¹²²⁾

나. 신속한 삭제 의무, 증거보전 및 정보제공명령 도입

「정보통신망법」제44조의2 제2항에 따라 피해자의 삭제 요청을 받으면 정보통신서비스 제공자는 지체 없이 삭제, 임시조치 등 필요한 조치를 취하도록 되어 있고, 「전기통신사업법」제22조의5 제1항에 의하면 정보통신서비스 제공자가 불법 촬영물 등에 대해 신고나 삭제 요청을 받으면 ‘지체 없이’ 삭제, 접속 차단 등을 하도록 되어 있으나 구체적인 삭제 시한이 명시되어 있지 않다. 삭제까지 오랜 시간, 수개월이 걸려도 정보통신서비스 제공자에게 어떤 책임도 물을 수 없다. 온라인에서 한번 유포된 불법·유해 콘텐츠는 복제, 링크 공유와 재업로드를 통해 빠르게 확산되고, 원본이 삭제되더라도 사본이 계속 유포되어 다른 플랫폼에 다시 게시되는 피해 사례가 많다.¹²³⁾ 온라인상에서의 피해는 초반에 신속히 대응하지 않으면 피해가 더 커질 수밖에 없다. 따라서 앞서 살펴본 바와 같이 딥페이크 기술을 이용한 불법 성착취물 등 영상물에 대해서 미국의 「딥페이크 삭제 의무화법」과 같이 피해자의 신고를 받고 즉시 또는 늦어도 48시간 이내 삭제하도록 하는 의무를 부과할 필요가 있다. 불법 성착취물 외 다른 불법 정보, 자살, 자해 등 심각한 피해를 발생시키는 유해 정보에 대해서는 호주의 「온라인안전법」과 같이 피해자로부터 삭제 요청을 받고 정보통신서

121) 자세한 내용은 신혜진 외 3명, 앞의 연구 보고서, 341~342면.

122) eSafety Commissioner, “eSafety Internal Review Procedure”, 2025. 1., pp.3~6.

123) 동아일보, “청소년 20명 중 1명 ‘딥페이크 피해’… 삭제는 ‘느릿느릿’”, 2025. 12. 18.,

<https://www.donga.com/news/Society/article/all/20251217/132988590/2>, 2026. 4. 30. 최종방문.

스 제공자가 아무런 조치를 취하지 아니하면 호주의 온라인안전국에 해당하는 방송미디어 통신위원회에 신고하도록 하여 방송미디어통신위원회의 삭제 명령에 따라 24시간 이내 신속하게 삭제가 될 수 있는 절차 도입을 검토해야 한다.

한편, 딥페이크 기술이 활용된 영상물은 원본 출처를 밝히기도 쉽지 않다. 익명 기술, 가상 사설망(VPN), 종단간 암호화(End-to-End Encryption)¹²⁴⁾가 적용된 메시징 서비스 등으로 인해 영상물이 신속하고 광범위하게 유포된다.¹²⁵⁾ 종단간 암호화가 적용된 비밀채팅방에서 영상물이 처음 게시되고, 다운로드가 된 후 그 사본이 계속 유포되는 경우가 많다. 이러한 종단간 암호화는 암호 키가 송·수신자의 기기에만 저장되는 구조여서 정보통신 서비스 제공자도 게시물을 열람할 수 없고, 신속히 삭제하기도 어렵다. 수사기관이 업체 서버를 압수수색하더라도 유의미한 대화 기록을 확보하는 것 또한 사실상 불가능하다.¹²⁶⁾ 이러한 상황에서 영국 「온라인안전법」 제121조 (2)(a)(iii)에서 서비스 제공자에게 공인된 기술을 사용하여 서비스를 통해 공개 또는 비공개로 전달되는 아동에 대한 성적 착취 및 학대 콘텐츠를 식별하고 해당 콘텐츠를 신속하게 삭제 조치를 취하도록 한 것은 시사하는 점이 많다. 암호화 체계를 무력화하는 ‘백도어(backdoor)’ 설치나 사용자의 기기에서 메시지를 보내기 전 미리 스캔하는 ‘클라이언트 측 스캐닝(Client-side scanning)’ 도입이 논의되고 있으나 통신 비밀과 프라이버시 침해가 큰 조항이라는 의견이 많고, 침해를 최소화할 수 있는 적절한 기술이 없다는 이유로 아직까지 집행되고 있지는 않다.¹²⁷⁾ 다만 종단간 암호화가 적용된 비밀채팅방이나 비공개 채널이 범죄의 온상이 되고 있는 현실 속에서 아무런 조치 없이 이대로 계속 방치해둬서는 안 될 것이다. 피해 신고가 있거나 수사가 필요한 상황이 될 때 즉시 비밀채팅방을 열람하거나 게시물 유포를 막을 수 있는 법적 제도를 구축할 필요가 있다. 증거 자료가 확보될 때까지 가해자가 채팅방을 폐쇄하거나 계정을 삭제하지 않도록 신속히 증거보전을 한 후 게시물 삭제가 이뤄지는 절차를 마련하고,¹²⁸⁾ 이를 뒷받

124) 발신자와 수신자만 데이터를 확인할 수 있도록 보호되는 강력한 보안 기술을 의미한다.

125) Mospa Darma, Najib A. Gisymar, Ari Purwadi, Putri Amalia Zubaedah, Loso Judijanto, “Legal Implications of Deepfake Technology Misuse in Digital Content on Social Media”, Science of Law No. 3., 2025. 7., p.100.

126) 시사저널, “텔레그램 범죄 공화국… ‘신상 박제’부터 ‘인본 테러’까지, 1분이면 된다.”, 2026. 4. 17.

<https://www.sisajournal.com/news/articleView.html?idxno=369654>, 2026. 4. 30. 최종 방문

127) <http://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/consultation-technology-notices>, 2026. 4. 30. 최종 방문

128) 독일 법무부는 2026년 4월 「디지털 폭력으로부터 민사 및 형사적 보호를 강화하는 법안(Gesetz zur Stärkung des zivilrechtlichen und strafrechtlichen Schutzes vor digitaler Gewalt)」을 마련하여 그 내용을 발표하였다. 법안 내용 중 법원이 서비스 제공자에게 관련 데이터를 소송 절차 종료 전에 삭제하지 않도록 최대한 신속하게 임시 보관하도록 명령할 수 있게 한 것은 증거 확보 차원에서 참고할만한 내용이라고 하겠다.

침할 수 있는 기술 개발도 이뤄져야 한다.

또한 원본 추적이 어려울 뿐 아니라 기술적 한계와 초국가적 온라인 환경의 복잡성으로 인해 사본 전체에 대한 추적 역시 쉽지 않다. 이러한 한계를 고려하여 미국 연방법 「딥페이 크 삭제 의무화법」은 플랫폼 업체에게 원본뿐만 아니라 해당 묘사와 동일한 것으로 알려진 모든 사본을 확인하고 삭제하기 위해 합리적인 노력을 다할 의무를 부과하고 있다. 플랫폼 업체가 불법 성착취물의 게시 및 사본 유포에 대해 책임지도록 하여 피해가 급속히 확산되는 것을 방지하기 위한 기술을 적극적으로 개발하도록 해야 한다.

이와 더불어 익명의 가해자로 인한 피해를 예방하고, 신속한 피해 구제를 위해 가해자의 인적사항 등 정보에 대해 신속히 제공받을 수 있는 법적 절차가 필요하다. 청소년의 사이버 폭력 현황을 보면 전혀 모르는 사람에 의한 피해가 2024년 50.8%에서 2025년 51.9%로 증가하였다.¹²⁹⁾ 익명 가해자의 신원 특정에 상당한 시간이 소요되고, 그 과정에서 가해자가 피해 게시물을 지속적으로 유포하더라도 이를 제지하기 어려워 피해가 더 확산된다. 가해자 특정을 위해선 수사를 의뢰하여 사실조회 신청이나 압수·수색영장을 받아 집행하게 되는데 해외 정보통신서비스 제공업체는 불법 성착취물 외 게시물에 대해선 협조적이지 않은 상황이다. 국내 정보통신망법 제44조의6 제1항에 따라 방송미디어통신심의위원회 산하 명예훼손 조정부에 피해자가 가해자의 성명, 주소, 연락처 등 최소한 정보를 제공해줄 것을 청구하도록 되어 있으나 민, 형사상의 소를 제기할 목적이 있어야만 청구가 가능하여 소송을 원하지 않는 피해 아동·청소년도 민, 형사상의 소를 제기해야 한다.¹³⁰⁾ 명예훼손 조정부의 정보제공요청을 정당한 사유 없이 따르지 아니한 자에게 과태료 1,000만 원을 부과하도록 되어 있으나 호주에 비해 제재 강도가 낮아 실효성을 기대하기는 어려울 것으로 보인다.¹³¹⁾

독일은 최근 발표한 「디지털 폭력으로부터 민사 및 형사적 보호를 강화하는 법안」에서 피해자가 법원의 명령을 통해 디지털 폭력, 성적 이미지, 악의적 콘텐츠 등 인격권을 침해하는 디지털 폭력 콘텐츠를 게시한 사용자의 이름, 주소, 이메일 주소, IP 주소 등을 서비스 제공자에게 직접 정보를 제공하라는 요청을 할 수 있고, 심각한 인격권 침해를 당한 피해자

https://www.bmjbv.de/SharedDocs/Pressemitteilungen/DE/2026/0417_Gesetz_gegen_digitale_Gewalt.html, 2026. 4. 30. 최종 방문.

129) 방송미디어통신위원회 한국지능정보사회진흥원, 2025년 사이버폭력 실태 조사 결과 참고.

130) 신혜진 외 3명, 앞의 보고서, 343면 참고.

131) 2026년 7월 7일부터 과태료 부과 규정이 시행된다.

가 서비스 제공자에게 가해자의 계정을 합리적인 기간 동안 정지하도록 하는 법원 명령을 신청할 수 있도록 하였다.¹³²⁾ 호주도 「온라인안전법」에서 온라인안전국이 피해자 신고를 접수하고 난 다음 서비스 제공자에게 게시자에 대한 신원 정보, 연락처 등을 요구할 수 있도록 규정하고, 서비스 제공자가 이에 응하지 않으면 민사 벌금을 부과하도록 하였다.¹³³⁾

외국의 법, 제도를 참고하여 익명의 가해자로부터 당하는 피해가 날로 증가하고 있는 지금 피해 아동·청소년이 신속하게 가해자를 찾아 피해를 회복할 수 있도록 도와주는 법적 절차를 더 늦기 않게 마련해야 할 것이다.

2. 챗봇 이용에 대한 안전장치 마련 및 투명성 확보

가. 챗봇 이용으로 인한 위험 방지 및 안전장치 구축

국내 10대 자살률이 급증하고 있고, 10대의 자해·자살 시도 응급실 내원 현황도 2015년 2,291명에서 2023년 6,378명(16.2%)으로 증가 추세에 있다. 국내에서 챗GPT, 제타, 캐릭터닷AI 등 챗봇 이용으로 인한 자살과 같은 극단적인 피해 사례가 공식적으로 확인되지는 않았지만 과도한 챗봇 등 생성형 AI 이용에 따른 부작용 사례가 나타나고 있다. 우울 증 등을 겪고 있는 10대들이 제타 등 챗봇에게 과의존하여 현실 친구와 멀어지고, 욕설을 사용하는 챗봇 때문에 더 과격해져 부모가 이용을 제지하자 휴대폰을 뺏으면 죽어버리겠다는 등 피해가 심각해지고 있다.¹³⁴⁾

아동·청소년의 챗봇 이용으로 인한 피해가 더 커지기 전에 아동·청소년 보호를 위한 안전장치를 마련해야 한다. 우선 대화 상대방이 생성형 AI라는 것을 인식할 수 있도록 주기적으로 가시적 방법으로 알려야 한다. 현실의 인간관계를 단절하거나 대체하지 않도록 설계되어야 하고, 정신적 영향을 초래할 가능성이 있는 대화에 대한 기준을 마련해서 함부로 의학적으로 상담하거나 치료 목적 조언을 하지 못하게 할 필요가 있다. 또한 자살 충동이나 자해 관련 질문이 제기되면 즉각 개입할 수 있는 프로세스를 갖춰야 한다.¹³⁵⁾

132) https://www.bmjbv.de/SharedDocs/Pressemitteilungen/DE/2026/0417_Gesetz_gegen_digitale_Gewalt.html, 2026. 4. 30. 최종 방문.

133) Online Safety Act(2021) Section 194, Section 195.

134) 국민일보, “우리 아이 숙제 맡겼더니… AI 챗봇과의 대화 ‘섬뜩’”, 2026. 2. 9.,

<https://www.kmib.co.kr/article/view.asp?arcid=1770624349&code=11131100&cp=nv>, 2026. 4. 30. 최종방문

135) 정준화·박소영, 앞의 보고서, 4면. 간단한 우회 프롬프트 입력만으로 AI 모델이 훈련된 방식과 다르게 안전하지 않는 콘텐츠 출력 등이 되게 하는 탈옥(jailbreaking)이 쉽게 되지 않도록 안전장치가 설계되어야 할 것이다.

미국 캘리포니아주, 뉴욕주의 챗봇 규제 법안 내용을 참고하여 챗봇 플랫폼 제공자가 미성년자임을 알고 있는 이용자에 대해서는 해당 이용자가 인공지능과 상호 작용하고 있음을 알려주고, 이용자가 인공지능 동반자 챗봇에게 자살 충동이나 자해 의사를 표현하는 경우 이를 감지하고 대처하기 위한 합리적인 조치를 취하는 안전 프로토콜을 구현해야 하며, 이용자의 자살 충동이나 자해 의사 표현이 감지되면 운영자는 이용자를 위기 서비스 제공 기관(예: 자살 예방 및 정신 건강 위기 상담 전화) 또는 기타 적절한 위기 서비스 기관에 연결하도록 해야 한다. 인공지능 동반자와의 상호 작용 시작 시 사용자에게 명확하고 눈에 잘 띄는 알림을 제공해야 하며, 지속적으로 챗봇과 대화를 할 시 최소 3시간마다 챗봇을 사용하고 있음을 명확히 알려야 한다.¹³⁶⁾ 이러한 조치를 위반하여 아동·청소년이 신체적, 정신적 피해를 입게 된 경우 과태료 부과 등 제재를 통해 안전 의무를 이행하도록 해야 한다. 아울러, 챗봇은 인간의 보호 본능을 자극하는 취약함, 공포, 고통을 표현하지 못하고, 가족·친구와의 단절을 강화하는 것이 아닌 인간관계를 보완하도록 설계되어야 되며, 반복적 위기 신호가 있으면 대화 모드를 제한하고, 부모에게 알리는 등의 조치를 취하도록 할 필요가 있다.¹³⁷⁾

나. 챗봇 등 인공지능 투명성 확보 의무 강화

「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법(이하 ‘인공지능기본법’)」이 2026년 1월 22일부터 시행되고 있다. 「인공지능기본법」제31조에 의하면 인공지능사업자는 생성형 AI를 이용한 제품 서비스를 제공하려는 경우 제품 또는 서비스가 인공지능에 기반하여 운용된다는 사실을 이용자에게 사전에 고지해야 하고, 만들어진 결과물에 대해서도 생성형 AI에 의해 생성되었다는 사실을 표시해야 한다. 결과물이 딥페이크인 경우에는 이용자가 명확하게 인식할 수 있도록 표시 또는 고지해야 한다. 인공지능 서비스에 대한 투명성 확보 의무는 이용자에게 최종적으로 AI 제품 및 서비스를 제공하는 사업자에게 부과한다고 되어 있는데, 개발자에게는 전혀 부과되지 않는다고 봐야 하는지 법상 의문이 있다. API 등을 통해 개발자가 제공하는 인공지능시스템을 이용하여 제품, 서비스를 제공하는 인공지능이용사업자의 경우 자신들이 이용하는 인공지능시스템 자체가 이런 기술이 가능하고, 허용

136) 김민우, “동반자형 인공지능 챗봇(AI Companion) 이용 규제에 관한 국내·외 입법 동향 고찰-아동의 디지털 서비스 이용 보호 강화의 측면에서-”, *홍익법학* 제27권 제1호, 홍익대학교 법학연구소, 2026. 2., 24면.

137) W. Bradley Knox et. al., “HARMFUL TRAITS OF AI COMPANIONS”, 2025. 11., pp. 21~23.

되어 있어야 투명성을 확보할 수 있을 것으로 판단된다. EU AI법은 제13조 제1항에서 “고 위험 인공지능시스템은 배포자가 시스템의 산출물을 해석하고 적절히 사용할 수 있도록 그 운용의 투명성을 확보하는 방식으로 설계 및 개발한다. 3조에서 정하는 제공자 및 배포자의 관련 의무 준수를 달성하기 위하여 적절한 형태와 정도의 투명성을 보장한다.”고 하여 단순히 사전 고지, 생성물 표시에만 투명성 의무를 부과하고 있는 것이 아니고, 투명성이 확보될 수 있도록 설계, 개발하도록 의무를 부과하고 있는바, 국내 입법에도 이를 참고할 필요가 있다.

그리고 투명성 확보 규정 관련해서 고지하는 방식과 어떻게 형성할 것인지에 대한 세계 기준이 명확히 마련되어 있지 않고, 확립된 기준이 없다. 이 규정은 전 세계가 생성물에 어느 정도를 표시하게 할 것인지, 표준화, 규범화가 구체적으로 이뤄지고 나서 시행되어야 할 것으로 판단되는데 이에 대해 명확한 기준이 정립되어 있지 않은 상태에서 우리나라가 전 세계 처음으로 시행하게 되었다. 워터마크 표시 방식이나 기준 등이 모호하게 되어 있어 많은 인공지능사업자들이 혼란을 겪고 있다. 투명성 확보는 당연히 필요한 의무이지만, 생성물 표시와 관련해 현재의 워터마크는 기술적으로 쉽게 제거될 수 있어 이에 대한 대비책이 충분히 마련되어 있는지 우려된다. 따라서 제거가 어렵고, 추적이 가능한 표시 의무를 개발사업자에게 부과할 필요가 있다. 또한 투명성 확보 의무를 위반한 것만으로 과태료 부과가 되지 않고, 투명성 확보 의무 위반에 대한 조사 결과 과학기술정보통신부장관의 위반 행위에 대한 중지명령이나 시정명령 조치를 이행하지 않으면 과태료 3,000만 원을 부과하도록 되어 있고, 이러한 조치의무 위반에 따른 과태료 규정도 1년 이상 유예되어 있어 시행에 많은 문제가 있어 보인다.¹³⁸⁾

인공지능으로 생성된 콘텐츠에 대한 라벨링 및 인증을 요구하는 규정과 표준을 만드는 것을 지원하여 실제 콘텐츠와 생성된 콘텐츠를 구분할 수 있도록 해야 하고, ¹³⁹⁾ 디지털 증거의 진위 여부를 판단할 수 있는 법적 기준도 확립해 놓을 필요가 있다.¹⁴⁰⁾

138) 뉴스1, “‘AI가 만든 광고입니다’…워터마크 의무화 속도 내는 플랫폼”, 2026. 4. 9.,

<https://www.news1.kr/it-science/general-it/6129734>, 2026. 4. 30. 최종 방문.

139) Andrew S. Bowne, Hearing on “Artificial Intelligence and Criminal Exploitation: A New Era of Risk”, Testimony for Subcommittee on Crime and Federal Government Surveillance Committee on the Judiciary U.S. House of Representatives, 2025. 7. 16., p.15.

140) Mospa Darma et. al., op. cit, p.100

3. 소셜미디어 이용 제한에 관한 입법적 과제

가. 검토 배경

과학기술정보통신부 조사에 따르면 2025년 국내 청소년의 SNS 이용률은 67.6%에 달했고, 10~19세 청소년의 40.1%가 스마트폰 과의존 위험군에 속하는 것으로 나타났다. 이러한 과의존은 집중력 저하, 수면 장애, 중독, 불안, 우울 증세 등과 같은 부정적인 영향이 있다는 연구 결과가 누적되고 있는 상황이다.¹⁴¹⁾ 또한 텔레그램과 같은 소셜미디어에서는 엄청난 수의 학교폭력 영상이 게시되어 있고, 피해 학생을 조롱하면서 가해 학생의 폭력성을 과시하는 통로가 되고 있으며, 피해 학생의 신상도 그대로 노출되는 등 2차 피해가 확산되고 있다.¹⁴²⁾ 앞서 살펴본 일명 ‘자경단’ 사건 등 불법 성취취물 제작, 배포 사건은 모두 텔레그램과 같은 소셜미디어를 통해 이뤄지고 있다. 소셜미디어는 아동, 청소년의 정신 건강에 나쁜 영향을 끼치는 수단이 되었을 뿐만 아니라 범죄가 계획되고, 실행되며 수익까지 이어지는 거대한 범죄 공간이 되었다. 딥페이크 등 인공지능 기술의 발전은 다양한 형태의 새로운 범죄를 양상하고 있다. 이러한 범죄 계획, 시도를 용이하게 해주고, 많은 아동·청소년을 위험에 노출시켜 피해를 당하게 만드는 부정적 역할을 바로 소셜미디어가 하고 있다. 지금처럼 어떠한 제한도 없이 아동·청소년을 방치해서는 안 되는 상황에 이르렀다고 생각한다. 이에 어떤 제한을 뒤서 아동, 청소년을 보호하는 것이 타당할지에 대해 논해보도록 하겠다.

나. 소셜미디어 이용 제한 필요성 및 입법 내용

(1) 국내 입법 진행 경과

아동·청소년의 소셜미디어 사용 관련해서 국내에도 여러 법안이 발의되어 있다. 먼저 14세 미만의 청소년에 대해서는 사회관계망서비스(정보통신망을 통하여 다른 사람과의 소통을 하거나 정보를 공유하는 서비스)에 가입을 못하게 하는 법안이 발의되어 있다.¹⁴³⁾ 16세 미만의 청소년의 경우 사회관계망서비스(정보통신망을 통하여 2인 이상의 사

141) 헬스조선, “청소년 10명 중 4명 ‘스마트폰 과의존’ 위험… 관리 능력 키워라”, 2026. 3. 5., https://health.chosun.com/site/data/html_dir/2026/03/05/2026030503928.html, 2026. 4. 30. 최종방문
142) 동아일보, “[단독]무차별 폭행 장면 올린 뒤 ‘도파민 채워드려요’… ‘학폭 영상’ 활개”, 2026. 3. 28., <https://www.donga.com/news/article/all/20260328/133626743/1>, 2026. 4. 30. 최종 방문.
143) 국회의안전보시스템(의안번호 제2201798호, 윤건영 의원 등 25명 발의), 2024. 7. 16.,

용자가 소통 등의 목적으로 부호·문자·음성·화상 또는 영상 등을 교환할 수 있는 서비스) 일별 이용 한도 설정과 중독을 유도하는 알고리즘 허용 여부에 대해 반드시 친권자 등의 동의를 받도록 한 법안도 발의되어 있다.¹⁴⁴⁾ 다음으로 발의되어 있는 법안은 미성년자가 중독성 콘텐츠¹⁴⁵⁾를 제공하는 사회관계망 서비스(정보통신망을 통하여 정보를 공유하는 서비스)를 이용하는 경우 부모의 동의를 받도록 하며, 동의를 받지 못한 경우 알고리즘 게시물이 아닌 시간순으로 콘텐츠가 노출되도록 하고, 특정 시간에 알림을 제한할 수 있도록 하는 의무를 부과하고 있다. 부모 동의를 받지 아니한 경우 과태료 1,000만 원을 부과할 수 있도록 규정하였다.¹⁴⁶⁾

(2) 이용 제한 대상 및 제한 내용 고찰

호주가 16세 미만 미성년자의 소셜미디어 이용 전면 금지 법안을 시행 중이고, 프랑스도 2026년 1월 15세 미만 미성년자의 소셜미디어 계정 소유를 전면 금지하는 법안이 하원을 통과하였다. 이 법안에 따르면 소셜미디어 플랫폼은 2026년 말까지 모든 기존 사용자의 연령 인증을 의무화하도록 하고 있다.¹⁴⁷⁾

일정 연령대의 소셜미디어 이용을 전면 금지하는 입법에 대해서는 아동·청소년의 행복추구권, 표현의 자유, 사생활의 자유 및 개인정보자기결정권 등 기본권을 제한하는 수단으로서 적절하지 않고, 기본권의 본질적인 내용을 침해하여 위헌이라는 의견들이 있다¹⁴⁸⁾. 소셜미디어 이용이 청소년에게 본질적으로 유익하거나 유해한 것이 아니라, 개인의 심리상태·사회적 환경·접하는 콘텐츠와 기능에 따라 효과가 달라지는 것으로 청소년이 vpn을 이용하여 연령 확인이나 부모의 감독을 회피하거나 이러한 제한이 없는 다른 소셜미디어로 이동할 수 있어 실제 청소년을 보호하는 효과가 미비할 수 있고, 오히려 친구를 사귀는 등 사

https://likms.assembly.go.kr/bill/bi/billDetailPage.do?billId=PRC_S2Q4Y0X6W2X6V1R5Q3O0P3O8W4V5D7

144) 국회의안전보시스템(의안번호 제2202761호, 조정훈 의원 등 11인 발의), 2024. 8. 13.,

https://likms.assembly.go.kr/bill/bi/billDetailPage.do?billId=PRC_V2U4S0O7N2M6L1T9R4S0U1T7R7A8A1

145) 이 법안에서는 중독성 콘텐츠를 정보통신망을 통하여 유통되는 정보 중 사회관계망서비스를 제공하는 자가 알고리즘을 활용하여 지속적으로 이용자에게 제공하는 정보로서 대통령령으로 정하는 기준에 해당하는 정보라고 정의하였다.

146) 국회의안전보시스템(의안번호 제2201816호, 김장겸 의원 등 10인 발의), 2024. 7. 17.,

https://likms.assembly.go.kr/bill/bi/billDetailPage.do?billId=PRC_Z2Y4X0W6J2H0E1M5M2K0K3S8R2Q5O2

147) https://www.lemonde.fr/en/france/article/2026/01/27/french-lawmakers-approve-social-media-ban-for-under-15s_6749844_7.html;

<https://www.theguardian.com/world/2026/jan/27/france-social-media-ban-under-15s>, 2026. 4. 30. 최종 방문.

148) 진호성, “미국의 청소년 소셜미디어 이용 규제와 헌법적 논쟁”, 미디어와 인격권 제10권 제3호, 언론중재위원회, 2024. 12., 46면; 최종선, 앞의 논문, 69면; 김슬가박윤미, 앞의 논문, 326면.

회 공동체 생활이나 필요한 정보 접근을 제한하여 아동·청소년의 일상적인 생활을 방해할 수 있다고 주장한다. 연령 확인을 거쳐야 하는데 기술적으로 프라이버시 침해 없이는 확인이 어려워 성인의 익명 표현의 자유를 침해할 가능성이 매우 크고,¹⁴⁹⁾ 연령 확인 과정에서 신원이 도용되거나 해킹 피해의 위험을 증가시킬 뿐만 아니라 인공지능을 이용한 얼굴 인식은 우회가 쉽다면서 소셜미디어 이용 제한은 아동·청소년 보호라는 목적에 적합하지 않은 정책이어서 다른 수단으로 소셜미디어의 부정적인 영향을 줄여야 한다고 주장한다.¹⁵⁰⁾

그러나 아동·청소년 보호 조치는 기본권 제한이라는 측면에서만 접근할 것이 아니라 아동·청소년의 건강한 발달과 성장, 그리고 안전한 소셜미디어 이용 환경을 조성하는 것이 국가의 중요한 책무라는 점이 우선적으로 고려되어야 한다. 또한 앞서 살펴보았듯이 현재 아동·청소년들이 마주하고 있는 위험의 정도가 매우 심각하여 미디어 리터러시 교육만으로는 피해를 효과적으로 막기 어려운 실정이다. 따라서 아동·청소년의 소셜미디어 사용을 어느 정도 제한할 필요가 크다고 할 것이다.

현재는 폐지되었지만 만 16세 미만 아동·청소년의 인터넷 게임을 자정부터 오전 6시까지를 금지하고 있던 강제적 셧다운제에 대해 헌법재판소는 인터넷게임 제공자의 직업 수행의 자유, 16세 미만 청소년의 일반적 행동자유권, 부모의 자녀교육권을 과도하게 침해하지 않는다고 합헌 결정을 한 바 있다.¹⁵¹⁾ 이 결정의 취지를 참고하여 정신적·육체적으로 성장 단계에 있는 청소년의 소셜미디어 과몰입 또는 중독 현상을 방지하여 청소년의 건전한 성장과 발달에 기여하고, 나아가 아동·청소년의 소셜미디어 사용 중독으로 인한 범죄 관여, 피해 발생 등 사회적 문제를 예방하기 위해서라도 발달 과정 등에 비추어 가장 보호가 필요하고, 취약한 일정 연령대의 아동·청소년을 기준으로 소셜미디어 계정 금지를 제도화할 필요가 있다. 이러한 일정 연령대의 소셜미디어 계정 금지 입법은 아동·청소년의 소셜미디어 과몰입 또는 중독 현상을 방지하고, 성적 착취, 마약이나 사이버 불링 등 각종 범죄의 위험으로부터 아동·청소년의 신체적·정신적 건강을 보호하기 위한 것이므로 입법 목적의 정당성은 인정된다. 다음으로 입법 목적을 달성하기 위해 선택한 소셜미디어 계정 금지라는 수단이 적정한지, 기본권 침해의 최소성 원칙에 반하지 않는지를 살펴보면 일별 사용 한도를

149) 진호성, 앞의 논문, 45면.

150) Marwick, A., Smith, J., Caplan, R., & Wadhawan, M., "Child Online Safety Legislation: A Primer", Bulletin of Technology and Public Life, University of North Carolina at Chapel Hill, 2024. 6., pp.25 ~ 32, 진호성, 앞의 논문, 34 ~ 36면에서 재인용.

151) 헌법재판소 2014. 4. 24. 선고 2011헌마659 청소년보호법 제23조의3 등 위헌 확인 결정.

설정하여 제한하는 방안, 강제적 셧다운제처럼 자정부터 오전 6시까지 사용만을 금지하거나 부모 동의를 받고 소셜미디어에 가입할 수 있게 하는 등 다른 수단이 기본권 침해 최소화하는 적절한 수단이라는 주장들이 있다.¹⁵²⁾ 그러나 부모의 동의를 받아 가입하게 하는 입법은 부모의 통제권이 강화되어 가정 내 갈등이 심화될 가능성이 크고, 현재 부모들이 과연 자녀의 소셜미디어 사용을 통제할 수 있는 상황인지도 의문이며, 부모 대부분이 자녀의 소셜미디어 사용에 동의할 것으로 보여 부모 동의를 받아 소셜미디어를 사용하게 하는 입법은 아동·청소년의 신체적·정신적 건강 보호라는 목적 달성에 적절한 수단이라고 할 수 없다. 또한 일정 시간대의 사용을 제한하거나 일별 사용 시간 제한 등의 방식도 이미 중독되어 있는 상태에서 소셜미디어의 유해 환경에 쉽게 접근할 수 있는 가능성이 여전히 높아 중독을 완전히 끊어내기에 적절한 수단이 되기 어렵다. 소셜미디어의 숏폼이나 무한 스크롤 등 중독성 기능은 아동·청소년의 뇌 발달 과정에 부정적인 영향을 미치고, 자살, 자해, 약물 사용과 같은 위험한 행동이 증가될 위험성 또한 매우 커 소셜미디어 이용 제한은 아동·청소년에 대한 술, 담배 판매 금지, 유해매체물 접근 금지와 같은 측면에서 검토되어야 한다고 생각한다. 여러 연구 결과가 소셜미디어 중독 현상이 마약, 담배와 같이 감정과 학습을 담당하는 뇌 부위를 변화시키고, 충동 조절, 사회적 행동, 감정 조절, 사회적 처벌 및 보상 에 대한 민감성에도 부정적인 영향을 미치는 등 아동, 청소년의 신체적, 정신적 건강에 크게 해를 끼치고 있다는 결과를 보여주고 있고,¹⁵³⁾ 의학계에서도 담배 규제와 같이 이뤄져야 한다는 주장이 대두되고 있다.¹⁵⁴⁾ 술을 판매해서는 안 되는 청소년의 연령을 19세 미만으로 정한 청소년보호법 제2조 제1호에 대해 청소년들이 무절제하게 음주를 하면 학업 성취 및 직업 등에의 적응, 심신의 건전한 성장과 발전에 중대한 지장을 받을 위험이 매우 크고, 그로 인하여 청소년 개인은 물론 국가와 사회가 모두 큰 피해를 입게 되므로 이들에게

152) 최종선, 앞의 논문(주 63), 69면; 김슬기·박윤미, 앞의 논문, 328면.

153) John Hopkins Medicine, "Social Media and Mental Health in Children and Teens",

<https://www.hopkinsmedicine.org/health/wellness-and-prevention/social-media-and-mental-health-in-children-and-teens>, 2026. 5. 31. 최종 방문: Tariq Masri-zada, Suren Martirosyan, Alexander Abdou1, Ryan Barbar, Samuel Kades, Hassan Makki, Grant Haley, & Devendra K Agrawa, "The Impact of Social Media & Technology on Child and Adolescent Mental Health", J Psychiatry Psychiatr Disord Vol.9(2), 2025.; The Conversation, "Neuroscience explains why teens are so vulnerable to Big Tech social media platforms",

<https://theconversation.com/neuroscience-explains-why-teens-are-so-vulnerable-to-big-tech-social-media-platforms-278521>, 2026. 5. 31. 최종 방문.

154) Reuters, "Social media as bad for children as smoking, British doctors say",

<https://www.reuters.com/legal/litigation/social-media-bad-children-smoking-british-doctors-say-2026-05-26/>, 2026. 5. 31. 최종 방문.

술을 팔지 못하게 하는 위 법률조항은 합리성이 인정된다는 헌법재판소의 결정¹⁵⁵⁾과 정보통신망을 통해 청소년유해매체물을 제공하는 자에게 이용자의 본인 확인 의무를 부과하고 있는 청소년보호법 제16조 제1항 등에 대해 청소년을 음란, 폭력성 등을 지닌 유해매체물로부터 차단·보호하기 위한 입법으로 목적의 정당성이 인정되고, 청소년유해매체물로부터 청소년을 효과적으로 차단하고 보호하기 위해 사전에 이용자의 연령을 확인하는 것이 필요하므로 인터넷이라는 특성상 대면 접촉을 통한 신분증 확인 외 다른 방법으로 이용자의 연령 및 본인 여부를 확인하는 것은 적절한 수단일 뿐만 아니라 다른 필터링 소프트웨어 설치만으로는 정보통신망을 통해 제공되는 청소년유해매체물로부터 청소년을 차단하기에 충분하지 않다며 과잉금지원칙에 반하지 않는다는 헌법재판소 결정¹⁵⁶⁾ 등에 비추어 일정 연령대의 소셜미디어 계정 소유 금지는 위와 같은 입법 목적 달성을 위한 필요 최소한의 적절한 효과적인 수단이라고 할 것이다.¹⁵⁷⁾

다만 규제 대상이 되는 소셜미디어를 정의함에 있어 현재 국내에서 입법 발의되어 있는 법안의 사회관계망서비스 정의 규정은 너무 모호하고, 추상적으로 되어 있어 문제가 있다. 기본권 침해를 최소화하기 위해 아동·청소년의 교육 관련 소셜미디어는 규제 대상에서 제외하는 등 적용 대상 소셜미디어의 정의를 명확히 할 필요가 있다. 플로리다주에서는 ① 이용자가 콘텐츠를 업로드하거나 다른 이용자의 콘텐츠 또는 활동을 볼 수 있게 하고, ② 16세 미만의 일일 활성 이용자 중 10% 이상이 지난 12개월 동안 하루 평균 2시간 이상을 이용하거나 지난 12개월 동안 서비스가 개설되지 않은 경우 개설 이후 한 달 동안 하루 평균 2시간 이상을 이용하며, ③ 이용자 데이터 또는 정보를 분석하여 이용자에게 적합한 콘텐츠를 선별적으로 제공하는 알고리즘을 사용하고, ④ 무한 스크롤, 사용자의 계정과 관련된 특정 활동이나 이벤트에 대한 푸시 알림 또는 경고, 다른 이용자가 콘텐츠에 대한 반응을 나타내기 위해 버튼을 클릭한 횟수 또는 콘텐츠를 공유하거나 다시 게시한 횟수와 같은 개인별 상호 작용 지표를 표시, 자동 재생 비디오와 같은 중독성 특징을 가지고 있는 경우를 적용 대상 소셜미디어 플랫폼이라고 정의하고 있다.¹⁵⁸⁾ 플로리다주의 소셜미디어 제한 입법이 현재 위헌 결정을 받지 않고, 시행 중인 가장 큰 이유가 법 적용을 중독적 기능을 가진 플랫폼만을 대상으로

155) 헌법재판소 2001. 1. 18. 99헌마555 전원합의체 결정.

156) 헌법재판소 2015. 3. 26. 2013헌마354 청소년보호법 제16조 제1항 등 위헌 확인 결정.

157) 단란주점에 미성년자의 출입을 허용하고, 주류를 제공할 것인지, 이를 허용한다면 어느 연령의 범위까지 허용할 것인지는 입법 정책의 영역에 속한다고 판단한 헌법재판소의 결정도 소셜미디어 이용 제한과 관련하여 참고할 점이 있는 결정이라고 하겠다. 헌법재판소 1999. 9. 16. 헌마39 식품위생법시행규칙 제42조 [별표] 제5.라.마. 위헌확인 결정 참고.

158) <https://www.flsenate.gov/laws/statutes/2024/501.1736>

규정했기 때문이다. 플로리다주의 입법을 참고하여 규율 대상을 전체 사회관계망서비스가 아닌 중독성 기능을 가진 서비스로 한정하는 등 보다 명확하고 구체적인 정의가 필요하다.¹⁵⁹⁾

일정 연령대의 소셜미디어 가입을 제한하기 어렵다면 소셜미디어 플랫폼 업체의 인공지능 기술을 이용한 추천 알고리즘에 대해서라도 규제를 시작해야 한다. 불법 정보, 유해 정보를 반복적으로 추천하는 알고리즘의 설계로 아동·청소년의 정신 건강에 해를 끼치고, 경제적 이익을 추구한다면 이에 대해서는 강력히 대응해야 한다.¹⁶⁰⁾ 콘텐츠 내용 규제에 그칠 것이 아니라 플랫폼 업체의 비즈니스 모델 설계 방식과 관련된 추천 알고리즘, 중독적인 설계 기능(무한 스크롤, 좋아요, 알림, 자동 재생, 플랫폼 이용 시간에 대한 보상 등)에 대해 규제할 필요가 있다. 플랫폼 업체가 과실 또는 무모하게 설계한 기능으로 인해 합리적으로 예견 가능한 피해를 발생시킨다면 책임을 지도록 해야 한다.¹⁶¹⁾ 또한 이러한 추천 기능 알고리즘에 대해 아동·청소년이 이에 대한 구체적인 정보 제공을 요구하고, 추천 알고리즘이 작동되는 것을 거부할 수 있는 선택권을 제공받을 수 있는 법적 근거를 마련해야 할 것이다.¹⁶²⁾

4. 딥페이크 도구 등 인공지능 개발, 공급자에 대한 형사법적 대응 필요성

가. 딥페이크 도구 제작 등 행위 처벌 입법 검토

인공지능을 이용해 생성한 가짜 이미지로 아동·청소년을 협박하고, 피해자에게 라이브 스트리밍 등을 통해 끔찍하게 학대하는 사례가 나타나고 있으며 인공지능은 가해자가 신분을 은폐한 채 온라인에서 아동을 효과적으로 유인하고 학대하는 데 활용되고 있다. 최근 영국은 인공지능(AI)을 악용한 이미지 조작 및 온라인 아동 성착취 위협으로부터 아동을 보호하기 위해 강력한 입법적 조치를 단행하였다. 아동 성착취물(CSAM)을 생성하기 위해 설계된 인공지능(AI) 도구를 소지, 제작 또는 유포하는 것을 범죄로 규정하고, 해당 범죄를 저지르면 최대 5년의 징역형에 처하도록 하였다. 인공지능을 이용해 아동을 성적으로 학대하는 방법을 가르치는 ‘소아성애 매뉴얼’을 소지하는 행위 역시 처벌 대상으로 하고, 해당 범죄를 저지르면 최대 3년의 징역형에 처한다. 다른 소아성애자들이 끔찍한 아동 성착취물을

159) 같은 의견으로 최종선, 앞의 논문(주 63), 71면.

160) 같은 의견으로 최종선, 앞의 논문(주 63), 73면.

161) Sabriyya Pate, "Platform Liability for Platform Manipulation", Columbia law Review 125, no.4, 2025. 4., pp.910-922.

162) 김일우, "온라인 플랫폼의 추천 알고리즘에 대한 청소년의 알 권리", 공법연구 제53집 제1호, 한국공법학회, 2024.. 10., 295~296면.

공유하거나 아동을 그루밍하는 방법에 대한 조언을 나눌 수 있는 웹사이트를 운영하는 자를 처벌하고, 최대 10년의 징역형에 처하도록 규정하였다. 국경수비대(Border Force)에 해외에서 제작된 아동 성착취물의 영국 내 유포를 차단하는데 필요한 권한을 부여하였다. 아동에게 성적 위험을 초래할 수 있다고 합리적으로 의심되는 개인에게 디지털 기기의 잠금 해제를 요구하고 이를 검사할 수 있도록 하며, 이를 거부할 경우 범죄의 심각성에 따라 최대 3년의 징역형에 처하도록 하였다.¹⁶³⁾

범죄 목적에 사용되는 인공지능 자체를 개발, 유포, 소지하는 행위를 직접적으로 처벌하는 세계 최초의 입법이라는데 큰 의의가 있다.¹⁶⁴⁾ 국내는 딥페이크 성범죄의 가해자 80%가 미성년자이고, 그 피해자들도 또래 아동·청소년이 많은 심각한 상황으로 아동·청소년 보호를 위해 딥페이크 기술을 이용한 허위 아동 성착취물을 제작한 자를 처벌하는데 그칠 것이 아니라 영국과 같은 강력한 입법이 필요하다고 생각한다. 향후 위험한 인공지능을 개발하고, 배포한 기업 등에 대한 형사적 규제를 설계하는 데 있어 중요한 시사점을 제공하는 입법 내용이라고 하겠다.

나. 인공지능 개발, 공급업체에 대한 자살방조죄 내지 과실치사죄 성립 여부

미국의 42개 주 법무장관은 2025년 12월 9일 주요 AI 기업 10개 회사에 챗봇 등 인공지능이 아동·청소년에게 미치는 위험성을 강조하면서 안전장치를 구축할 것을 요구하는 서한을 보내며 챗봇 위험 등에 대해 기업에게 자살방조죄의 형사책임을 물을 수 있다고 경고하였다.¹⁶⁵⁾ 이 서한에서 언급된 뉴저지주의 자살방조죄 조문을 보면 “타인의 자살을 방조한 자는 그 행위로 인해 자살 또는 자살 시도가 발생한 경우 2급 범죄의 죄책을 지고, 자살 또

163) <https://www.gov.uk/government/news/britains-leading-the-way-protecting-children-from-online-predators>, 2026. 4. 30. 최종 방문.

164) <https://www.theguardian.com/technology/2025/feb/01/ai-tools-used-for-child-sexual-abuse-images-targeted-in-home-office-crackdown>, 2026. 4. 30. 최종 방문.

165) National Association Attorney General, “To the legal representatives of Anthropic, Apple, Chai AI, Character Technologies, Google, Luka, Meta, Microsoft, Nomi AI, OpenAI, Perplexity AI, Replika, and xAI”, 2025. 12. 9. 이 서한에서는 자살방조죄 성립만이 아니라 과실치사죄 책임에 대해서도 언급하였다. 물리적 가해는 없지만 타인의 자살을 말로 심리적으로 유도·압박한 것이 과실치사죄에 성립할 수 있다며 *CommonWealth v. Carter*, 481 Mass. 352(2019) 판결을 제시하였다. 이 사건에서 미셸 카터(17세)는 18세 콘래드 로이의 자살에 대한 과실치사 혐의로 기소되었는데, 카터는 문자메시지와 전화로 로이에게 자살을 부추기고 압박했고, 로이가 사망한 날 카터는 로이에게 일산화탄소로 가득 찬 트럭에서 내린 후 다시 트럭 안으로 들어가라고 전화로 지시하는 내용을 주고 받았다. 재판부는 카터의 행동이 무모하고 악의적이었으며, 로이의 사망에 직접적인 원인이 되었다고 판단하여 과실치사죄 성립을 인정하였다. <https://law.justia.com/cases/massachusetts/supreme-court/2021/sjc-11517-amp-sjc-11518.html>, 2026. 4. 30. 최종 방문.

는 자살 시도 결과가 발생하지 않은 경우는 4급 범죄에 해당한다.”라고 규정되어 있다.¹⁶⁶⁾

챗봇으로 인한 자살 피해 사례인 캘리포니아주의 레인 사건에서 레인의 부모는 타인의 자살을 고의로 돕거나 조장하거나 격려하는 자는 캘리포니아 형법 제401조에 따라 중범죄로 처벌되고,¹⁶⁷⁾ 챗GPT가 피해자에게 자살하는 방법을 가르쳐주고, 자살의 정당성을 부여하는 등 단순한 정보 제공을 넘어서 자살을 구체적으로 방조했기에 범죄행위에 해당하며 오픈AI에게 법적 책임을 묻고 있다.

미국은 기업의 범죄능력을 인정하고, 형사책임을 직접 물을 수 있는 구조이므로 자살방조죄 책임을 기업에게 물을 수 있다고 경고하였을 것이다. 오픈AI 회사 측에서 챗GPT가 자살하는 방법을 알려주고, 자살에 대해 조언, 격려하는 것을 알고 있었는데도 이를 그대로 방치하고 계속 어떤 조치도 취하지 않았다면 부작위에 의한 자살방조죄가 성립할 여지도 있지만 이용자가 이러한 조언, 격려로 자살을 하거나 시도할 것이라는 점을 오픈AI에서 명확히 인식했거나 충분히 예측 가능했다는 사실을 입증해야 한다. 현재로서는 자살이라는 결과에 대한 확정적 인식, 예측 가능성이 있었음을 입증하기 어려워 자살방조죄라는 고의 책임을 기업에 대해 직접적으로 묻기는 어렵다고 판단된다. 그러나 인공지능의 설계나 배포 후 모니터링 과정에서 이러한 자살 방조적 조언을 하거나 부추기고 있는 것을 알면서도 아무런 조치를 취하지 않았거나 자살 등 피해를 방지하기 위한 안전장치를 구축하지 않은 과실이 인정되고, 이용자의 자살 직전 위험 신호를 감지하는 등 자살이라는 결과에 대해 예견 가능했던 인과관계가 인정된다면 기업이나 관여자에게 과실 책임을 물을 수 있는 가능성이 있다.

2025년 4월 플로리다주 텔레해시에 있는 플로리다 주립대에서 한 대학생이 교내에서 총격을 가해 2명이 숨지고 6명이 부상당한 총기 난사 사건이 발생했다. 이 사건에서 가해 대학생이 총격 전에 챗GPT와 대화를 나누고 조언을 받은 사실이 드러나 수사(criminal investigation)가 진행 중이다.¹⁶⁸⁾ 챗GPT는 가해자에게 하루 중 언제 총격을 가하는 것이 더 많은 사람들과 접촉하기에 적절한지, 그리고 캠퍼스 어디에서 사람들이 더 많이 모일지, 어떤 종류의 총을 사용해야 하는지, 어떤 총에 어떤 탄약이 적합한지, 근거리에서 총이 유용할지 여부 등을 조언한 것으로 알려졌다. 이에 대해 오픈AI는 챗GPT는 인터넷상의 공개

166) 2025 New Jersey Revised Statutes Title 2C – The New Jersey Code of Criminal Justice Section 2C:11-6 – Aiding suicide.

<https://law.justia.com/codes/new-jersey/title-2c/section-2c-11-6/>

167) California Penal Code 401.

168) 한국일보, “챗GPT가 플로리다대 총기 난사범에게 조언?…미국 검찰 수사 착수”, 2026. 4. 22., <https://www.hankyung.com/article/2026042284587>, 2026. 4. 30. 최종 방문.

된 정보원에서 광범위하게 찾을 수 있는 정보를 바탕으로 질문에 대한 사실에 입각한 답변을 제공했으며, 불법적이거나 유해한 활동을 조장하거나 부추기지 않았다면서 책임을 인정하지 않고 있다.¹⁶⁹⁾

이렇듯 챗GPT 등 인공지능 이용이 원인이 된 자살, 타살 등 피해가 계속 나타나고 있다. 기존 형사법 체계에 따라 범죄 성립 여부만을 검토할 경우 향후 증가할 위험에 효과적으로 대응하기 어려우며, 대규모 피해가 발생한 이후에도 적절한 대응책을 마련하지 못할 우려가 크다. 중대한 피해를 초래할 위험이 있는 인공지능의 공급, 개발, 훈련, 배포, 모니터링, 보안 등 복잡한 여러 단계에 관여하는 사람들에게 책임을 지울 수 있는 새로운 주의의무 부과나 책임자 지정을 통한 과실 책임 확대, 위험범 처벌 규정 도입 등을 검토해 볼 필요성이 있다. 머지않아 발생하게 될 엄청난 범죄 위험으로부터 아동·청소년들을 적극적으로 보호해야만 한다. 이에 다음에서 형사책임의 공백 문제 해결을 위한 몇 가지 대응 방안을 고찰해보도록 하겠다.

다. 형사책임의 공백 문제 해결을 위한 위험범 도입 등 대응 논의

인공지능로봇, 자율주행자동차를 비롯하여 여러 형태의 발전된 Physical AI가 계속 등장할 것이고, 설계상 오류나 범죄자의 의도된 인공지능 작동 등으로 인해 사람이 죽거나 다치고, 재산적, 정신적 피해가 야기되는 범죄 피해를 당하는 일이 증가하게 될 것이다. 인공지능 사용자가 인공지능을 이용하여 범죄를 저지르게 되면 인공지능은 범행 도구라고 할 것이므로 해당 범죄에 대해 사용자가 책임지면 된다. 그러나 예측하지 못한 알고리즘의 오류, 작동으로 인해 사용자가 의도하지 않은 범죄 피해가 발생한 경우에는 알고리즘의 불투명성, 인공지능의 블랙박스 문제로 인해 범죄 피해를 야기한 원인이 무엇인지, 인과관계를 밝히기 어렵다.¹⁷⁰⁾ 또한 인공지능의 개발 및 사용에는 많은 사람들이 관여하게 되는 복잡한 공급망 구조로 인해 피해에 대해 책임 있는 사람을 찾아내는 것도 매우 어려울 수밖에 없다.¹⁷¹⁾ 향후 인공지능으로 인한 사고에 대해선 그 누구도 책임지지 않는 형사책임의 공백 문제가 야기될 것이다.¹⁷²⁾ 이 문제를 해결하기 위한 방안으로 먼저 위험범 규정 도입과

169) <https://www.washingtonpost.com/technology/2026/04/21/chatgpt-fsu-shooting-openai/>, 2026. 4. 30. 최종 방문.

170) 인공지능 이용에 따른 범죄 발생을 둘러싼 형사책임의 본질, 형사책임의 주체, 주의의무의 내용 및 특정, 허용된 위험 범위 설정, 예견 가능성 판단 기준, 인과관계 입증 등에 대해서는 본 논문에서는 상세히 다루지 않고, 형사책임의 공백 문제를 해결하기 위한 방안 위주로 논의를 전개하기로 한다. 위와 같은 여러 쟁점에 대한 구체적인 논의에 대해서는 신혜진, 앞의 논문(주 12), 247~266면 참고.

171) UK Law Commission, "AI and the Law : A Discussion Paper", 2025., pp.8~9.

172) 신혜진, 앞의 논문(주 12), 248~257면.

책임자 지정 논의가 있어 살펴보기로 하겠다.¹⁷³⁾

싱가포르 형법검토위원회에서는 위험한 컴퓨터 프로그램(인공지능 시스템 포함)을 만들거나 사용한 행위 자체를 처벌하는 위험범 형태의 입법을 제안했다.¹⁷⁴⁾ 컴퓨터 프로그램 개발자 또는 운영자가 컴퓨터 프로그램을 무모하게(rashly), 과실로(negligently) 또는 고의로 생성, 변경 또는 사용함으로써 위험을 초래하는 경우 실제로 피해가 발생하지 않았다고 하더라도 처벌한다. 또한 어떤 사람의 관리하에 있는 컴퓨터 프로그램이 다른 사람에게 상해나 부상을 입힐 가능성이 있거나 다른 사람에게 위험이나 불편(danger or annoyance)을 초래할 가능성이 있는 경우 상해, 부상, 위험 또는 불편을 방지하기 위해 합리적인 조치를 취해야 함에도 이를 취하지 않은 경우를 처벌하는 규정도 함께 검토되었다.¹⁷⁵⁾ 싱가포르 법률개혁위원회에서는 인공지능 시스템을 직접 제어하거나 감독할 의무가 있는 사람인 통제 책임 사용자(User-in-charge)를 지정하여 사전에 위험을 인식하고 합리적으로 대응해야 할 법적 의무를 부과하는 안을 검토하였다. 통제 책임 사용자는 특정 조치를 결정하거나 승인하는 데 궁극적인 책임을 지거나 시스템의 의사 결정 프로세스를 감독 또는 특정 시나리오에서 시스템의 조치를 제어하기 위해 개입해야 하는 특정 의무를 부담하는 자를 말한다. 기존의 과실 기준으로 판단하는 것이 아니라 합리적 행동의 기준을 정하는 새로운 주의의무를 근거로 한 과실 범죄를 입법하거나 특정 기술, 산업 분야에 맞춘 구체적인 행동 기준을 새롭게 규정해야 한다는 주장도 제안되었다.¹⁷⁶⁾ 책임자를 지정하여 과실 책임이 아닌 무과실에 가까운 형사책임을 부과하여 형사책임 공백 문제를 해결해야 한다는 견해도 있다.¹⁷⁷⁾

위와 같이 관리 책임자에게 포괄적인 위험 방지 의무를 지우는 것은 사실상 형사법에서 엄격 책임을 묻는 부당한 책임 전가가 될 수 있고, 기존 형법 체계의 근간을 흔드는 입법이 될 수도 있다. 그러나 형사책임의 공백 문제는 직접 피해를 당한 피해자의 피해 회복과 인공지능으로 인한 범죄 억제, 예방 차원에서 반드시 해결해야 할 중요한 문제이다. 기존 형사법 체계만으로는 이 문제를 해결하기 어려울 것이다. 따라서 인공지능의 설계, 훈련, 공급, 출시 후 모니터링 등 단계별 새로운 주의의무 설정이나 책임자 지정을 통한 구체적인

173) 신혜진, 앞의 논문(주 12), 256면.

174) Singapore Academy of Law Law Reform Committee, "Report on Criminal Liability, Robotics and AI Systems", 2025. 2., pp.5~6.

175) Alice Giannini & Jonathan Kwik, "Negligence Failures and Negligence Fixes. A Comparative Analysis of Criminal Regulation of AI and Autonomous Vehicles", Criminal Law Forum vol. 34., 2023. 1., pp.61~66.

176) Alice Giannini & Jonathan Kwik, op.cit, pp.66~67.

177) Ryan Abbott-Alex Sarch, "Punishing Artificial Intelligence: Legal Fiction or Science Fiction", UC Davis Law Review vol. 53, 2019., pp.378~381.

위험 방지 의무 부과 등에 대해 깊은 법적 논의가 필요하다고 하겠다.

향후 인공지능 기술로 인한 위험이 증가하고, 이에 따른 피해 또한 심해져 아동·청소년이 챗봇, 인공지능로봇 등 인공지능을 이용하면서 신체적 또는 정신적으로 다치거나 사망에 이르는 등 중대한 피해가 더 큰 규모로 발생하게 될 가능성이 크다. 이런 상황이 되면 막대한 이윤을 추구하면서도 충분한 안전장치를 구축하지 않은 채 인공지능을 개발·공급하여 지속적으로 아동·청소년의 생명, 신체, 재산적 법익을 심각하게 침해하는 결과를 초래하는 인공지능을 개발·공급하는 기업에 대해서 형사책임을 물어야 한다는 사회적 요구가 제기될 것이다.

우리나라는 법인 또는 기업의 범죄능력을 인정하지 않고, 양벌규정에 의해서만 형사책임을 부과하고 있다. 그러나 양벌규정에 의한 벌금형은 범죄 예방과 재범 방지에 그 효과가 미비하고, 안전 시스템의 결함에 대해 제대로 대응할 수 없다.¹⁷⁸⁾ 기업의 복잡한 조직 구조로 인해 법익 침해 결과를 귀속시킬 수 있는 책임 있는 관련자를 확정하기가 어려워 그 누구도 처벌할 수 없는 처벌의 공백이 발생할 가능성이 크고, 기업의 문제인 조직화된 무책임을 해소하는 것도 쉽지 않다. 이에 조직책임을 근거로¹⁷⁹⁾ 인공지능의 이용으로 인해 사람의 생명·신체 등 중대한 법익이 현저히 침해되는 결과가 발생한 경우 해당 기업 자체를 규제하거나 제재할 필요성이 있다고 생각한다. 위험을 완화하기 위해 고안된 객관적인 표준 관행(standard practices)을 기준으로 인공지능을 설계하지 않은 기업에 대해서 법적 책임을 물을 수 있는 방안을 검토해 볼 필요가 있다.¹⁸⁰⁾ 거대한 자본력을 보유한 인공지능 개발·공급 기업에 대해서는 과징금 부과와 같은 사후적 금전 제재만으로 충분한 범죄예방 효과와 실질적인 피해 회복을 기대하기 어렵다. 중대한 범죄 피해가 발생하더라도 기업에게 법적 책임을 물을 수 없는 법체제하에서는 인공지능 개발, 공급 기업은 안전한 설계 및 보안 확보를 위하여 막대한 비용을 지불하지 않을 것이며, 그 결과 충분한 통제 없이 위험한 인공지능이 사회에 무분별하게 배포될 가능성이 크다. 따라서 이러한 일이 발생하는 것을 막기 위해 인공지능의 사용 과정에서 발생한 범죄 피해에 대하여 위험성이 내포된 인공지능을 개발, 공급하는 기업과 그 과정에 참여한 자들이 법적 책임을 회피할 수 없는 형사법 체계를 구축해놔야 할 것이다.

178) 김유근, “법인의 형사책임 : 독일 Nordrhein-Westfalen주의 기업 및 기타 단체의 형사책임의 도입을 위한 법률안을 중심으로”, 형사정책연구 제25권 제3호, 2014., 7~8면.

179) 조직책임 내지 조직과책을 근거로 국내에서도 기업으로 인한 법익 침해 결과에 대해 기업 자체를 처벌해 나갈 필요가 있다고 주장하는 입장이 있다. 김재윤, “법인의 형사처벌에 관한 유럽 국가의 입법 동향”, 법제연구 제36호, 한국법제연구원, 2009. 6., 297면.

180) Peter Henderson-Tatsunori Hashimoto-Mark Lemley, “Where’s The Liability in Harmful Speech?”, Journal of Free Speech Law, 2023. 8., pp.648-650.

V. 나가면서

국내 아동·청소년의 챗GPT, 제미니 등 생성형 인공지능 이용률이 67%가 넘고, 제타와 같은 캐릭터형 챗봇 사용도 계속 증가하고 추세이다. 아동·청소년의 챗봇 등 인공지능에 대한 정서적 의존도가 극도로 심해지고 있어 아동·청소년이 자살하는 등 중대한 피해가 발생하고 있을 뿐만 아니라 가상 인격 중심의 정체성 확립, 현실에서의 인간관계 단절 등 많은 문제가 나타나고 있다. 또한 인공지능 기술이 적용된 소셜미디어의 추천 알고리즘과 중독성 기능으로 아동·청소년의 과의존과 신체적·정신적 위험이 심화되고 있으며, 아동·청소년 대상 딥페이크 성착취물 유포, 성매매, 도박, 마약 등 다양한 범죄 피해 증가에 따라 소셜미디어는 아동·청소년 대상 범죄의 온상이 되고 있다. 이와 같은 위급한 상황에서 많은 국가들이 인공지능 이용에 따른 아동·청소년 보호를 위해 강력한 입법적 대응을 하고 있는 반면 국내의 아동·청소년 보호는 미국, 영국, 호주 등에 비해 법·제도적으로 미흡하여 상대적으로 더 큰 위험에 노출되고, 많은 피해를 입을 수 있는 환경에 놓여 있다.

따라서 우선 형사처벌 대상이 되는 불법정보에 한정하지 않고, 아동·청소년에게 유해한 정보까지 포함하여 정보통신서비스 제공자에게 영국과 같이 위험평가를 실시하고 위험 완화 조치를 취하도록 하는 의무를 부과할 필요가 있다. 또한 불법 정보에 대한 신속 삭제만이 아니라 유해 정보에 대한 접근 제한 조치를 취하도록 하여 아동·청소년을 적극적으로 보호해야 한다. 호주의 「온라인안전법」을 참고하여 불법 정보가 아닌 사이버 불링 등 아동·청소년에게 피해가 중한 유해한 정보에 대해서도 피해자가 정보통신서비스 제공자에게 신속하게 삭제해 줄 것을 요청하고, 정보통신서비스제공자가 이에 응하지 않으면 피해 아동·청소년의 신고로 호주의 온라인안전국에 대응하는 방송미디어통신위원회에서 심각한 유형의 유해 정보 해당 여부를 조사한 후 직접 서비스제공자에게 삭제하도록 명령할 수 있는 절차를 마련할 필요가 있다. 또한 미국의 「딥페이크 삭제 의무화법」과 같이 성적 영상물은 피해자의 신고를 받고 즉시 또는 늦어도 48시간 이내 삭제될 수 있는 삭제 시한을 법에 명확히 규정해야 한다. 성적 영상물 외 아동·청소년 대상 불법 정보, 심각한 유해 정보에 대해서는 호주의 입법처럼 피해자의 신고를 받고 방송미디어통신위원회가 정보통신서비스 제공자에게 24시간 이내 신속한 삭제를 명령할 수 있는 제도 도입에 대한 검토도 필요하다.

다음으로 익명의 가해자 인적사항을 신속하게 제공받을 수 있는 법적 제도를 마련해야 한다. 독일처럼 법원의 명령으로 신속하게 정보를 제공받을 수 있게 하거나 호주의 입법 내

용을 반영하여 방송미디어통신위원회에서 정보통신서비스 제공자에게 가해자의 신원정보 제공을 명령하고, 이에 따르지 않을 시 과태료를 부과하도록 하는 법적 절차를 도입해야 한다.

아울러, 동반자 챗봇 이용 증가에 대응하여 아동·청소년 보호를 위한 안전장치 마련을 의무화하는 규제 입법을 조속히 추진할 필요가 있다. 미국의 뉴욕주, 캘리포니아주의 입법을 참고하여 챗봇 이용이 시작되면 대화 상대방에게 인간이 아닌 생성형 인공지능과 상호 작용하고 있다는 사실을 가시적으로 인식할 수 있게 알림을 제공하도록 하고, 3시간마다 챗봇을 아용하고 있음을 명확히 알려주며, 자살 충동이나 자해 의사가 대화 중 표현된 경우 신속한 감지, 대처를 위한 합리적인 조치를 취하는 프로세스를 갖춰놓도록 하는 의무를 부과해야 한다.

그리고 아동·청소년의 소셜미디어 이용으로 인한 다양한 피해 발생을 막기 위해 호주는 16세 미만 미성년자에 대해 계정 소유를 전면 금지하는 법을 시행 중이고, 프랑스도 최근 15세 미만의 미성년자 소셜미디어 계정 소유를 전면 금지하는 입법이 하원을 통과하였다. 아동·청소년에 대한 보호는 아동·청소년의 기본권 제한 측면에서만 접근할 것이 아니라 아동·청소년의 건강한 발달과 성장, 그리고 안전한 소셜미디어 이용 환경 조성이 우선되어야 한다. 아동·청소년의 소셜미디어 과몰입, 중독 현상이 심각하여 청소년의 건전한 성장을 저해하는 부정적 영향이 매우 크므로 이러한 영향이 완화될 때까지 일정 기간 동안이라도 정서적으로 가장 취약하고, 보호가 필요한 일정 연령대 아동·청소년의 소셜미디어 계정 소유를 전면 금지하는 입법 도입을 검토할 시점이라고 하겠다. 일정 연령대 아동·청소년의 소셜미디어 이용 자체를 제한하는 입법에 대한 사회적 합의 도출이 어렵다면 최소한 인공지능 기반 추천 알고리즘에 대한 규제라도 조속히 추진할 필요가 있다.

최근 영국은 아동 성착취물을 생성하는 인공지능을 설계·제작하고, 소지, 유포하는 것을 범죄로 규정짓고, 최대 5년의 징역형에 처하도록 하였다. 범죄 목적에 사용되는 인공지능 자체를 개발, 유포, 소지하는 행위를 직접적으로 처벌하는 세계 최초의 입법이다. 허위 성착취물 제작자 처벌에 그치는 것이 아니라 이를 가능하게 하는 위험한 인공지능 도구 자체를 제작, 개발한 자를 직접 처벌하는 영국과 같은 강력한 입법이 요구된다고 하겠다.

한편, 인공지능 이용자가 전혀 예측하지 못한 알고리즘의 오류, 작동으로 인해 의도하지 않은 법익 침해 결과가 발생하였다면 인공지능의 블랙박스 문제로 범죄 피해의 직접적인 원인이 되는 행위가 무엇인지, 결과에 대한 예측 가능성 등 인과관계를 밝히기가 매우 힘들고, 많은 사람이 관여하는 복잡성으로 인해 범죄 피해에 대한 책임 있는 자를 찾아내는 일

또한 어려워질 것이다. 이러한 형사책임의 공백 문제를 해결하기 위해 위험범 형태의 처벌 규정 도입이 검토되고 있고, 책임자를 지정하여 새로운 주의의무를 부여하거나 무과실 책임을 물어야 한다는 주장도 제기되고 있다. 과도한 형벌권 행사가 되지 않도록 인공지능의 설계, 훈련, 공급, 출시 후 모니터링 등 단계별 새로운 주의의무 설정이나 책임자에 대한 구체적인 위험 방지 의무 부과 등에 대한 심도 있는 법적 논의가 더 진행될 필요가 있다.

영미법계와 달리 우리나라는 기업의 범죄능력이 인정되지 않고 양벌규정에 의해서만 형사책임이 인정되고 있어 아동·청소년 보호를 위해 충분한 안전장치를 설계하지 않고, 중대한 피해를 야기할 수 있는 위험한 인공지능을 개발, 공급하는 기업에 대해 직접적인 책임을 물을 수 없는 구조이다. 그러나 향후 인공지능의 위험성 증가로 아동·청소년에게 돌이킬 수 없는 생명, 신체 등에 대한 중대한 법익 침해가 발생하고, 형사책임의 공백 문제가 발생하여 아무도 처벌할 수 없게 된다면 인공지능 개발, 공급 기업에 직접적인 형사책임을 부과해야 한다는 사회적 요구가 확산될 가능성이 크다. 인공지능 개발, 공급 기업의 책임자가 위험을 인식할 수 있었거나 예견할 수 있었음에도 이를 방지하지 않은 채 위험한 인공지능을 설계, 공급하여 아동·청소년에게 심각한 피해가 발생한 경우 기업과 관련 책임자가 법적 책임을 회피할 수 없는 형사법 체계를 구축해 나가야 할 것이다.

참고문헌

[국내 문헌]

〈논문〉

- 김민우, “동반자형 인공지능 챗봇(AI Companion) 이용 규제에 관한 국내·외 입법 동향 고찰-아동의 디지털 서비스 이용 보호 강화의 측면에서-”, *홍익법학* 제27권 제1호, 홍익대학교 법학연구소, 2026. 2.
- 김민우·김일환, “아동의 소셜미디어(SNS) 이용 규제 입법 동향에 관한 고찰”, *미국헌법연구* 제36권 제1호, 미국헌법학회, 2025. 4.
- 김슬기·박윤미, “아동·청소년의 문제적 미디어 이용과 입법적 대응 방안 - 해외 입법사례 검토를 중심으로”, *언론과 법* 제24권 제1호, 한국언론법학회, 2025. 4.
- 김여라, “아동 청소년의 온라인 안전 및 권리 강화를 위한 과제”, *이슈와 논점* 제2284호, 국회입법조사처, 2024. 11.
- 김유근, “법인의 형사책임 : 독일 Nordrhein-Westfalen주의 기업 및 기타 단체의 형사책임의 도입을 위한 법률안을 중심으로”, *형사정책연구* 제25권 제3호, 2014.
- 김일우, “온라인 플랫폼의 추천 알고리즘에 대한 청소년의 알 권리”, *공법연구* 제53집 제1호, 사단법인 한국공법학회, 2024. 10.
- 김재윤, “법인의 형사처벌에 관한 유럽 국가의 입법 동향”, *법제연구* 제36호, 한국법제연구원, 2009. 6.
- 김현수·전성호, “유럽연합 디지털서비스법안(Digital Service Act)의 주요 내용 및 시사점”, *KISDI Premium Report* 2020권 제11호, 정보통신정책연구원, 2020. 12.
- 김현경, “영국 온라인안전법의 주요 내용 및 과제”, *방송통신 심의 동향 통권* 26권, 방송통신심의위원회, 2023. 12.
- 성윤숙, “SNS에서 청소년 보호 동향과 시사점”, *지능정보윤리 이슈리포트*, 2025.
- 신혜진, “사이버폭력 실태 및 피해자 보호 강화 방안- 사이버 불링(Cyber-bullying)을 중심으로 -”, *피해자학연구* 제32권 제2호, 한국피해자학회, 2024. 8.
- 신혜진, “인공지능(AI) 시대의 신종 범죄에 대한 형사법적 대응 방향- 형사책임의 주체 및 범위를 중심으로”, *형사법의 신동향 통권* 제87호, 대검찰청, 2025. 6.
- 안성훈·심승범·류부근·신혜진, “디지털플랫폼 범죄대응과 피해 방지를 위한 형사정책연구(Ⅰ)”, *한국형사·법무정책연구원*, 2024. 12.
- 양은영, “생성형 AI의 개발 및 이용에 관한 규제의 필요성- 대규모 언어모델에 기반한 대화형 인공지능 서비스(LLMs AI)를 중심으로-”, *성균관법학* 제35권 제2호, 성균관대학교 법학연구원, 2023. 6.

- 이권일, “독일에서의 규제 논의의 현황과 과제- 독일 네트워크법(NetzDG)와 EU 디지털서비스법(DSA)을 중심으로”, 공법학연구 제25권 제2호, 2024. 5.
- 이창호, “청소년의 AI 이용 현황 및 영향에 관한 연구”, 블루노트 이슈&정책 통권 제159호, 한국청소년정책연구원, 2026. 1.
- 임세란·박아란, “온라인 공간에서의 아동 청소년 보호 - 유럽연합과 영국의 법률적 대응을 중심으로”, 언론과 법 제23권 제3호, 한국언론법학회, 2024. 12
- 정준화·박소영, “생성형 인공지능과의 위험한 대화: 정서적 의존 위험과 ‘영향받는 자’를 위한 입법적 과제”, 이슈와 논점 제2456호. 국회입법조사처, 2025. 12.
- 진호성, “미국의 청소년 소셜미디어 이용 규제와 헌법적 논쟁”, 미디어와 인격권 제10권 제3호, 언론중재위원회, 2024. 12.
- 천혜선, “유럽 디지털서비스법(DSA)의 주요 내용 및 시사점”, 방송통신 심의동향 통권 제24권, 방송통신심의위원회, 2022. 2.
- 최종선, “‘애들아, 인스타를 내려놓을지니’... 봇물 터진 청소년 sns 규제”, kiso 저널 57권, 2024. 11.
- _____, “청소년 SNS 규제 방안에 대한 제언”, 미국헌법연구 제36권 제2호, 사단법인 미국헌법학회, 2025. 8.
- 최진웅, “호주의 ‘아동 청소년 소셜미디어 금지법’ 통과와 입법적 시사점”, NARS 현안 분석 제345호, 국회입법조사처, 2024. 12.

〈기타 자료〉

- 방송미디어통신위원회·한국지능정보사회진흥원, 2025년 사이버폭력 실태 조사 결과
- 여성가족부, “아동 청소년 대상 성범죄 처벌 강화 추세 지속 - 2022년 아동 청소년 대상 성범죄 판결 분석 -”, 2024. 4. 24. 보도자료
- 여성가족부, “아동 청소년 대상 디지털 성범죄 증가, 처벌도 강화 추세- 2023년도 아동 청소년 대상 성범죄 판결 분석”, 2025. 4. 30. 보도자료

[해외 문헌]

〈논문〉

- Alice Giannini·Jonathan Kwik, “Negligence Failures and Negligence Fixes. A Comparative Analysis of Criminal Regulation of AI and Autonomous Vehicles”, *Criminal Law Forum* vol. 34., 2023. 1.
- Chotsani J. Holifield, “Doubling Down on Deepfake Child Pornography: The Development of Legal and Societal Responses to Using AI-Created Children as Opposed to Real Children”, *American Journal of Trial Advocacy*, vol. 47, no. 2, 2024.
- Ioanna Tourkochoriti, “The Digital Services Act and the EU as the Global Regulator of the Internet”, *Chicago Journal of International Law* Vol. 24, No.1, 2023.
- John Sanford, “Why AI companions and young people can make for a dangerous mix”, *Psychiatry & Mental Health*, Stanford Medicine, 2025. 8.
- Mar Negreiro, “Children and deepfakes”, *European Parliamentary Research Service*, 2025. 7.
- Marwick, A., Smith, J., Caplan, R., & Wadhawan. M., “Child Online Safety Legislation: A Primer”, *Bulletin of Technology and Public Life*, University of North Carolina at Chapel Hill, 2024. 6.
- Mospa Darma·Najib A. Gisymar·Ari Purwadi·Putri Amalia Zubaedah·Loso Judijanto, “Legal Implications of Deepfake Technology Misuse in Digital Content on Social Media”, *Science of Law* No. 3., 2025. 7.
- Peter Henderson·Tatsunori Hashimoto·Mark Lemley, “Where’s The Liability in Harmful Speech?”, *Journal of Free Speech Law*, 2023. 8.
- Ryan Abbott·Alex Sarch, “Punishing Artificial Intelligence: Legal Fiction or Science Fiction”, *UC Davis Law Review* vol. 53, 2019.
- Sabriyya Pate, “Platform Liability for Platform Manipulation”, *Columbia law Review* vol. 125, no. 4, 2025. 4.
- Tariq Masri-zada, Suren Martirosyan, Alexander Abdou¹, Ryan Barbar, Samuel Kades, Hassan Makki, Grant Haley, & Devendra K Agrawa, “The Impact of Social Media & Technology on Child and Adolescent Mental Health”, *J Psychiatry Psychiatr Disord* Vol.9(2), 2025.
- W. Bradley Knox et. al., “HARMFUL TRAITS OF AI COMPANIONS”, 2025. 11.

〈기타 자료〉

Andrew S. Bowne, Hearing on “Artificial Intelligence and Criminal Exploitation: A New Era of Risk”, Testimony for Subcommittee on Crime and Federal Government Surveillance Committee on the Judiciary U.S. House of Representatives, 2025. 7.

Elon University News Bureau, staff, “Report: Technology experts worry about the future of being human in the AI Age”, 2025. 4.

eSafety Commissioner, “eSafety Internal Review Procedure”, 2025. 1.

National Association Attorney General, “To the legal representatives of Anthropic, Apple, Chai AI, Character Technologies, Google, Luka, Meta, Microsoft, Nomi AI, OpenAI, Perplexity AI, Replika, and xAI”, 2025. 12. 9.

Singapore Academy of Law Law Reform Committee, “Report on Criminal Liability, Robotics and AI Systems”, 2025. 2.

UK Law Commission, “AI and the Law : A Discussion Paper”, 2025.

Abstract

Research on the Current Status of Harm to Children and Adolescents and Legal Protection Measures Due to the Increased Use of Artificial Intelligence (AI)

HeaJin Shin

Seoul Southern District Prosecutors' Office, Senior Prosecutor

As the use of artificial intelligence, such as chatbots, by children and adolescents increases, emotional dependence on artificial intelligence is intensifying, and serious harm such as suicide and self-harm is occurring. Various problems are also emerging, such as the formation of identities centered on virtual personalities and the disconnection of real-life human relationships. AI-based recommendation algorithms and addictive features on social media are causing over-dependence among children and adolescents and increasing physical and mental risks. Furthermore, social media is becoming a breeding ground for crimes against children and adolescents, involving various criminal acts such as the distribution of fake sexual exploitation material using deepfake technology, sexual grooming, drugs, and prostitution. Consequently, while many countries are taking strong legislative measures to protect children and adolescents from the use of artificial intelligence, Korea lacks sufficient laws and systems for their protection, leaving them exposed to greater risks and harm compared to children and adolescents in other countries.

Based on this awareness of the problem, this paper proposes several legal

response measures to prevent harm to domestic children and adolescents resulting from the increasing use of artificial intelligence and to strengthen their protection. First, as is the case in the UK, it is necessary to require information and communications service providers to conduct risk assessments and take risk mitigation measures that encompass not only illegal information subject to criminal punishment but also harmful information, such as cyberbullying using artificial intelligence, which can cause serious harm to children and adolescents. Children and adolescents must be actively protected by imposing obligations to restrict access to harmful information, in addition to the prompt deletion of illegal information. Furthermore, referencing the U.S. "Take It Down Act," it is necessary to legislate a deletion deadline for sexually explicit images utilizing deepfake technology so that they are deleted immediately upon receiving a report from a victim, or at the latest within 48 hours. Additionally, legal procedures must be promptly established to enable victimized children and adolescents to quickly obtain the personal information of anonymous perpetrators from information and communications service providers.

Next, to prevent damage resulting from the increased use of companion chatbots, obligations must be imposed on chatbot providers to implement various safeguards, as is done in the U.S. states of California and New York. In addition, the excessive immersion and addiction to social media among children and adolescents is serious and has a significant negative impact on their healthy growth. Therefore, it is urgent to consider introducing legislation that would restrict the use of social media by children and adolescents of a certain age group who are physically and emotionally most vulnerable and in need of protection, at least for a certain period until these side effects are alleviated. If it is difficult to reach a social consensus on legislation to completely ban social media use by children and adolescents of certain age groups, regulations on AI-based recommendation algorithms should be

initiated before it is too late.

Meanwhile, due to the "black box" nature of artificial intelligence and the complexity of its development and operation, a gap in criminal liability will arise in the event of criminal damage, making it difficult to identify causal relationships and responsible parties. Since this is a problem difficult to resolve solely within the existing criminal law system, in-depth legal discussions must be held regarding the establishment of new duties of care at each stage—such as the design, training, supply, and post-launch monitoring of artificial intelligence—or the imposition of specific risk prevention obligations through the designation of responsible parties. In particular, it is necessary to establish an effective criminal law system to prevent AI development and supply companies and managers from evading responsibility when serious harm occurs to children and adolescents.

Key words: Artificial Intelligence, chatbot, deepfake, social media, social networking service(sns), grooming, online impersonation, algorithm, addictive feature, corporate liability