KOREAN
WOMEN'S
DEVELOPMENT
INSTITUTE

젠더리뷰

기획특집

변화하는 젠더폭력의 사각지대와 피해자 보호지원을 위한 정책적 과제

- **인공지능 매개 젠더폭력의 확장·변형과 피해자 보호·지원의 과제**
  김애라 | 한국여성정책연구원 연구위원

- **교제폭력 법제화 논의의 쟁점과 과제 : 관계 기반 위험과 피해자 안전을 중심으로**
  박상민 | 한국여성정책연구원 부연구위원

- **여성폭력 피해자 주거지원의 한계와 개선방안 : 현행 공공주거지원제도와 여성폭력 피해자 지원제도를 중심으로**
  황정임 | 한국여성정책연구원 선임연구위원

인공지능 매개 젠더폭력의 확장·변형과 피해자 보호·지원의 과제

KOREAN WOMEN'S DEVELOPMENT INSTITUTE

김애라 한국여성정책연구원 연구위원

1. '딥페이크'를 넘어 확장되는 인공지능 매개 젠더폭력

생성형 인공지능의 대중화는 디지털 젠더폭력의 양상을 바꾸고 있다. 불법 촬영과 촬영물 유포, 유포 협박이 기존 디지털 성범죄의 주요 문제였다면 이제는 실제 촬영물이 없어도 특정인의 얼굴과 신체, 음성을 성적으로 합성·변형할 수 있다. 공개된 사진과 개인정보를 결합해 특정인을 성적으로 묘사하거나 성적 모욕과 위협의 문구를 대량으로 만들어 전송하는 일도 가능해졌다.

인공지능이 전혀 새로운 폭력을 만들어낸다고 보기는 어렵다. 오히려 기존의 성적 괴롭힘과 사칭, 스토킹, 이미지 기반 성폭력에 생성·변형·자동화 기능이 결합하면서 폭력의 방식과 규모가 달라지고 있다고 보는 편이 정확하다. 제작의 문턱은 낮아지고 확산은 쉬워졌으며 이미지·영상의 합성을 넘어 음성·텍스트를 이용한 사칭과 성적 괴롭힘, 개인정보와 정체성의 도용 등으로 피해 양상도 다양해지고 있다. 이에 인공지능이 젠더폭력에 가져오는 피해 유형이 어떻게 확장·변형되고 콘텐츠의 생성과 유통 구조가 어떻게 달라지고 있는지 살펴볼 필요가 있다.

이처럼 피해의 형태와 발생 방식이 다양해지면서 합성물의 정교함이나 진위 여부만으로 피해를 설명하기도 어려워지고 있다. McGlynn과 Rackley (2017)는 동의 없는 성적 이미지의 제작과 유포를 '이미지 기반 성적 학대(image-based sexual abuse)'로 개념화하고 이를 존엄성, 사생활, 성적 자율성의 침해로 설명한 바 있다. 합성물에 나타난 신체가 실제 피해자의 것인지, 사람들이 이를 진짜라고 믿는지는 피해의 양상에 영향을 줄 수 있다. 그러나 피해를 판단할 때에는 이러한 요소뿐 아니라 당사자의 정체성이 동의 없이 성적으로 이용되고 그 이용에 대한 통제권을 상실한다는 점이 중요하게 고려될 필요가 있다.

이러한 문제는 국내 피해 지원 현장에서 나타나

는 피해 양상의 변화와도 맞닿아 있다. 중앙 디지털 성범죄피해자 지원센터가 2024년 지원한 피해자는 1만 305명으로 전년보다 14.7% 증가했으며 합성·편집 피해는 1,384건으로 전년 423건에 비해 227.2% 늘었다. 합성·편집 피해자의 92.6%는 10대와 20대였다(여성가족부·한국여성인권진흥원, 2025). 물론 이는 지원기관에 접수된 사례에 관한 통계이므로 전체 인구의 피해 발생률을 의미하지 않는다. 또한 합성·편집 피해 전체를 생성형 인공지능에 의한 피해로 볼 수도 없다. 그러나 적어도 피해 지원 현장에서 합성, 편집 피해가 빠르게 증가하고 있음을 잘 보여준다고 하겠다. 이러한 변화와 함께 정부의 디지털 성범죄 대응체계도 확대되고 있다. 성평등가족부는 2026년 업무 계획에서 중앙·지역 지원센터 인력 확충, 탐지·삭제 시스템 고도화, 유관기관을 연계하는 '디지털 성범죄 피해 통합지원단' 설치를 제시하였다(성평등가족부, 2026a). 같은 해 4월에는 피해영상물 삭제요청 자동화, 인공지능 기반 아동·청소년 온라인 성 착취 선제 대응 시스템, 딥페이크 판별 솔루션 도입을 발표한 바 있다(성평등가족부, 2026b).

그러나 인공지능이 젠더 기반 폭력에 이용되는 상황에서는 피해물을 탐지하고 삭제하는 것만으로 대응이 충분하다고 보기는 어렵다. 탐지와 삭제는 이미 생성, 유통된 이미지나 영상에 대응하는 데에는 중요하지만, 동일인을 대상으로 새로운 콘텐츠가 반복 생성되거나 개인정보 악용, 텍스트나 음성을 통한 성적 괴롭힘처럼 특정 파일의 삭제만으로 피해가 중단되지 않는 경우도 있기 때문이다. 이러한 피해에는 삭제뿐 아니라 반복 생성·유포에 대한 대응, 플랫폼 신고와 법률·상담 지원, 나아가 생성형 인공지능에 대한 규제 등이 다시 검토되어야 할 필요가 있다. 이 글에서는 인공지능 매개 젠더폭력을 이러한 변화를 포괄하는 개념으로 사용하여 주요 피해 양상과 현행 법·제도 사이의 간극, 피해자 보호·지원의 과제를 살펴본다.

2. 인공지능 매개 젠더폭력의 개념과 주요 양상

젠더 기반 폭력에 인공지능이 이용되는 현상은 기술매개 젠더기반폭력(technology-facilitated gender-based violence, TFGBV)에 관한 기존 논의의 연장선에 있다. UNFPA(2021)는 이를 성별을 이유로 개인을 대상으로 하며 정보통신기술이나 디지털 미디어에 의해 실행·지원·악화·증폭되는 폭력으로 정의한다. 비동의 성적 이미지와 딥페이크뿐 아니라 온라인 성적 괴롭힘, 사이버 스토킹, 신상털기, 해킹, 사칭 등이 여기에 포함된다. UNESCO(2023) 역시 생성형 인공지능이 이미지 기반 학대, 사이버 괴롭힘, 사칭과 허위정보의 생성을 용이하게 하고 이를 자동화·대량화할 수 있다고 지적한다.

이러한 논의를 바탕으로 관련 개념들의 범위를 구분해 볼 필요가 있다. '딥페이크'가 특정한 합성 기술이나 그 결과물을 가리킨다면, '이미지 기반 성적 학대'는 동의 없이 이루어진 성적 이미지의 제작·이용과 그로 인한 침해에 초점을 둔다. 이 글에서 사용하는 '인공지능 매개 젠더폭력'은 이미지뿐 아니라 음성·텍스트·대화형 서비스 등으로 확장되는 피해를 포함하고 인공지능의 생성·변형·자동화 기능이 기존 젠더폭력의 실행과 확산을 용이하게 하면서 또한 피해 양상을 변화시키는 현상을 함께 포괄하기 위한 개념이다.

가. 성적 콘텐츠 제작 기능의 상품화

생성형 인공지능이 가져온 중요한 변화 가운데 하나는 성적 콘텐츠 제작의 기술적 장벽이 낮아졌다는 점이다. 과거에는 일정한 수준의 편집 기술이나 별도의 제작 도구가 필요했던 작업을 이제는 일반 이용자도 온라인 서비스나 공개된 생성모델을 통해 쉽게 할 수 있게 되었다.

Gibson 등(2024)은 누드 합성(nudification) 웹사이트 20곳을 분석했는데, 이 가운데 19곳이 여성 사진의 변형을 명시적으로 내세우고 있었으며 유료 기능과 결제 수단도 제공하고 있었다. 이 연구에서 실제 인물의 사진으로 서비스의 성능을 시험한 것은 아니므로 광고된 기능이 모두 구현된다고 단정할 수는 없다. 다만 성적 이미지 변형 기능이 온라인에서 하나의 상품과 서비스로 제공되고 있음을 보여준다.

이와 함께 완성된 이미지뿐 아니라 특정 인물의 이미지를 생성하는 데 이용할 수 있는 모델 자체도 유통되고 있다. Hawkins 등(2025)은 공개 모델 저장소를 분석하여 기존 이미지 생성 모델을 특정 실존 인물의 모습을 재현하도록 추가 학습·조정한 모델 약 3만 5천 개가 공개적으로 내려받을 수 있는 형태로 유통되고 있음을 확인하였다. 상세 분석 대상의 96%는 여성을 대상으로 하고 있었다. 다만 이러한 모델의 수나 다운로드 횟수가 곧 성적 이미지 생성이나 실제 피해의 규모를 의미하는 것은 아니다.

이들 연구는 성적 이미지의 제작과 유통이 완성된 결과물의 차원을 넘어 제작 도구와 모델의 유통으로 확장되고 있음을 보여준다. 제작 기능과 모델의 공유, 결과물의 게시가 서로 다른 서비스에서 이루어질 경우 특정 게시물을 삭제하더라도 동일한 인물을 대상으로 한 새로운 콘텐츠가 다시 생성될 수 있다. 이는 인공지능 매개 젠더폭력의 변화가 단순히 더 많은 콘텐츠가 유통되는 데 그치지 않고 피해물을 만들어내고 재생산하는 구조 자체의 변화와 관련되어 있음을 보여준다. 따라서 피해 대응에서도 개별 파일의 유통과 삭제뿐 아니라 반복 생성을 가능하게 하는 제작·유통 구조를 함께 살펴볼 필요가 있다.

나. 이미지와 음성의 성적 합성, 변형

인공지능 매개 젠더폭력에서 가장 널리 알려진 유형은 이른바 딥페이크 성적 합성물이다. 초기에는 유명인의 얼굴을 기존 성적 영상에 합성하는 방식이 주로 알려졌으나 최근에는 옷을 입은 사진을 나체 이미지로 변형하는 도구나 사진 한 장으로 여러 장면을 생성하는 모델, 짧은 음성 샘플을 이용해 특정인의 목소리를 재현하는 음성 합성 등으로 기술의 형태가 다양해지고 있다.

제작과 유통의 방식도 달라지고 있다. 2024년 경찰청 특별 집중 단속 과정에서 드러난 지인 합성물 공유방, 이른바 '겹지방' 사건은 자동화된 도구를 이용해 지인의 사진으로 합성물을 제작·공유하는 구조를 보여준다(경찰청, 2024). 2026년 1월에는 소셜미디어 엑스(X)에 내장된 인공지능 기능이 실존 인물의 사진을 동의 없이 나체화하는데 이용되면서 방송미디어 통신위원회가 엑스 측에 안전장치 마련과 청소년 보호 계획 제출을 요구하기도 하였다(방송미디어 통신위원회, 2026). 생성 기능이 대규모 플랫폼 내부에 직접 결합되면서 콘텐츠의 제작과 유통이 동일한 서비스 안에서 이루어질 수 있게 된 것이다.

이러한 피해는 불법 촬영 피해와는 또 다른 특징을 가진다. 실제 성적 촬영물이 없어도 공개된 일상 사진을 이용해 성적 합성물을 만들 수 있고, 특정

합성물을 삭제하더라도 같은 사진이나 정보를 이용해 새로운 결과물을 다시 생성할 수 있다. 따라서 인공지능 매개 피해에서는 이미 존재하는 성적 촬영물의 유출뿐 아니라, 일상적인 목적으로 공개·공유된 이미지가 당사자의 의사와 무관하게 성적 콘텐츠의 제작에 이용되는 과정 자체가 중요한 문제가 된다. 다시 말해 피해의 출발점이 사적인 성적 이미지의 노출에 한정되지 않고 일상적으로 유통되는 사진과 정보가 비동의적인 성적 이용의 대상이 되는 데까지 확장되고 있는 것이다.

이처럼 일상적인 이미지가 성적 콘텐츠로 이용될 수 있는 상황에서는 피해의 중대성을 합성물의 정교함이나 사실성만으로 판단하는 것은 적절하지 않다. 조잡한 합성물이라도 피해자의 이름, 학교, 직장, 사회관계망 계정 등과 함께 공유되면 특정인을 겨냥한 성적 모욕이나 괴롭힘의 수단이 될 수 있다. 따라서 피해 지원에서는 합성물의 완성도뿐 아니라 피해자의 식별 가능성, 공유 범위, 신상정보의 결합 여부, 협박이나 추가 접촉의 위험 등을 함께 살펴야 한다.

나아가 합성물을 제작하거나 유포하겠다는 위협도 피해자를 통제하는 수단으로 이용될 수 있다. 이러한 위협은 관계 유지나 금전 지급, 추가 이미지 제공 등을 강요하는 압박으로 작용할 수 있다. 따라서 합성물 자체의 제작·유포뿐 아니라 합성 기술이 기존의 협박이나 착취와 결합해 피해자를 통제하는 방식에도 주목할 필요가 있다.

다. 정체성 도용과 대화형 서비스를 통한 성적 괴롭힘

인공지능 매개 피해는 이미지에 한정되지 않는다. 특정인의 사진과 신상정보를 이용해 성적으로 모욕하는, 이른바 '지인 능욕'은 이전부터 존재했지만, 생성형 인공지능은 특정인의 이름, 학교, 직장, 사회관계망 계정 등의 정보를 바탕으로 성적 서사나 허위 프로필, 성매매 광고 형식의 게시물, 실제 대화처럼 보이는 성적 문구 등을 손쉽게 생성하는 데 이용될 수 있다. 계정을 탈취하거나 사칭 계정을 만든 뒤 피해자 명의로 성적 게시물이나 대화를 만들어내는 방식도 가능하다.

대화형 서비스에서는 이러한 정체성 도용이 고정된 게시물의 형태로만 남겨져 있는 것이 아니라 지속적인 상호작용의 형태로 나타날 수 있다. 최근 국내에서 10대 이용자를 중심으로 캐릭터 챗 서비스가 확산되면서 청소년의 성적 대화 노출에 대한 우려도 제기되고 있다. 이러한 서비스에서 실존 인물의 이름과 사진, 말투, 관계 정보를 이용해 해당 인물을 모사한 캐릭터를 만들고 그 캐릭터가 성적인 대화에 참여하도록 설정하거나 이를 다른 이용자와 공유한다면 특정 인물에 대한 사칭과 성적 대상화가 결합된 문제가 발생할 수 있다.

이러한 피해의 특징은 성적 이미지나 신체의 시각적 재현이 없어도 특정인을 성적으로 대상화할 수 있다는 데 있다. 임신이나 성매매, 특정인과의 성적 관계 등을 사실처럼 서술하거나 피해자의 사진과 성적 설정을 결합한 허위 프로필을 만들고, 실존 인물을 모사한 캐릭터가 성적인 대화를 하도록 설정하는 것만으로도 특정인을 성적으로 지목하고 낙인찍을 수 있다. 이는 인공지능 매개 젠더폭력의 피해 유형이 기존의 이미지, 영상 중심에서 텍스트와 대화, 정체성의 비동의적 이용으로 확산되고 있음을 보여준다. 따라서 피해를 판단할 때에는 콘텐츠의 노출 정도나 음란성뿐 아니라, 당사자의 개인정보와 정체성이 어떤 방식으로, 성적으로 이용되고 그 결과 일상과 사회관계에 어떠한 침해가 발생하는지를 함께 살펴볼 필요가 있다.

이러한 피해는 피해자가 얼마나 쉽게 특정되는지, 어떤 정보와 함께 유통되는지에 따라 더욱 확대될 수 있다. 결과물 자체만 보면 누구를 지칭하는지 분명하지 않더라도 이를 공유하는 집단 안에서는 피해자를 쉽게 알아볼 수 있다. 여기에 연락처나 학교, 직장 정보가 결합되면 온라인에서의 성적 대상화가 원치 않는 접촉이나 생활공간에서의 괴롭힘으로 이어질 수도 있다. 개인정보가 이미 공개되어 있었다는 사실이 그러한 정보를 새로운 맥락에서 결합하거나 성적으로 이용하는 것까지 정당화하는 것은 아니다.

문제는 이러한 피해가 현행 지원체계에서 충분히 포착되기 어렵다는 점이다. 현행 디지털 성범죄 피해 지원체계는 불법 촬영물, 허위 영상물 등 법률이 정한 피해물과 그에 결합된 신상정보의 삭제·지원을 중심으로 구성되어 있어 텍스트 기반의 성적 사칭이나 허위 프로필, 대화형 서비스 안에서 이루어지는 성적 괴롭힘은 기존 지원체계에 곧바로 포섭되기 어려울 수 있다. 자신을 모사한 챗봇이나 사칭 계정의 존재를, 지인을 통해 알게 된 경우에는 피해자가 서비스 안에서 실제로 어떤 대화가 이루어졌는지 확인하기조차 어려울 수 있다. 또한 챗봇의 응답은 고정된 게시물이 아니라 이용자의 입력과 캐릭터 설정에 따라 생성되기 때문에 문제의 캐릭터 설정이 유지되는 한 유사한 성적 상호작용이 반복될 가능성도 있다. 따라서 개별 대화나 캡처의 삭제뿐 아니라 사칭 프로필이나 캐릭터 설정, 공개·공유 기능 등 반복적인 피해를 가능하게 하는 서비스상의 요소에 대해서도 조치할 필요가 있다.

한편, 대화형 서비스에서는 특정한 실존 피해자가 존재하지 않더라도 별도의 규율 문제가 제기된다. 실제로 여러 캐릭터 챗 플랫폼에서 교복을 입은 학생이나 '초등학생' 등 아동·청소년으로 인식될 수 있는 캐릭터가 성적인 상황과 결합되어 제공된 사례가 보도된 바 있다(연합뉴스, 2026. 5. 15.). 이는 특정 피해자의 권리 침해와는 별개로 아동·청소년을 성적으로 대상화한 콘텐츠를 어떻게 규율할 것인가의 문제를 제기한다. 동시에 10대의 캐릭터 챗 이용이 빠르게 늘고 성적·자극적 대화에 대한 우려가 제기되고 있는 만큼 실제 청소년 이용자가 원하지 않는 성적 상호작용에 노출되지 않도록 연령 확인과 콘텐츠 제한, 신고·차단 등의 보호장치를 마련할 필요도 있다. 이때 콘텐츠에 대한 법적 책임의 판단과 이용자 보호를 위한 조치는 구분할 필요가 있다.

3. 피해 특성과 피해자 보호·지원의 과제

인공지능 매개 피해는 하나의 콘텐츠나 범죄 유형으로 정리되기 어렵다. 동일인을 대상으로 여러 형태의 이미지와 음성, 텍스트가 반복적으로 생성될 수 있고, 생성·가공·유포가 서로 다른 사람과 서비스에서 이루어질 수도 있다. 피해자는 무엇이 어디까지 만들어지고 공유되었는지를 알기 어렵고, 공개된 게시물을 삭제한 뒤에도 재생성이나 비공개 유통의 가능성 때문에 불안을 겪을 수 있다.

McGlynn 등(2021)은 영국·호주·뉴질랜드의 이미지 기반 성적 학대 피해자 75명을 면접하여, 피해가 일회적인 사건이나 심리적 외상에 그치지 않고 피해자의 자기 감각과 사회관계, 일상의 자유에 지속적인 영향을 미친다는 점을 보여주었다. 특히 피해물의 재유포 가능성을 통제하기 어렵다는 사실은 피해를 끊임없이 현재화하고, 언제 어디서 다시 등장할지 모른다는 불확실성은 일상적인 경계와 자기 제한으로 이어졌다. 연구자들은 이를 사회적 단절,

상시성, 실존적 위협, 고립, 자유의 제약이라는 다섯 차원으로 설명한다.

이러한 분석은 인공지능이 매개된 피해에도 시사하는 바가 크다. 기존 피해물이 재유포되는 데 더해 동일인을 대상으로 새로운 콘텐츠가 반복 생성될 수 있는 환경에서는 피해자가 무엇이 어디까지 만들어졌는지를 알거나 통제하기 어려운 상태 자체가 피해의 지속성을 강화할 수 있기 때문이다.

가. 현행 법·제도와 피해 현실의 간극

현행 「성폭력범죄의 처벌 등에 관한 특례법」 제14조의2는 사람의 얼굴·신체 또는 음성을 대상으로 한 촬영물·영상물·음성물을 대상자의 의사에 반하여 성적 욕망 또는 수치심을 유발할 수 있는 형태로 편집·합성·가공하는 행위를 처벌한다. 2024년 10월 개정으로 제작 단계의 반포 목적 요건이 삭제되었으며 해당 편집물 등의 소지·구입·저장·시청에 대한 처벌 규정도 신설되었다.

그러나 생성형 인공지능을 이용한 모든 성적 침해가 이 조항에 그대로 포섭되는 것은 아니다. 현행 규정은 사람의 얼굴·신체 또는 음성을 담은 촬영물·영상물·음성물을 성적인 형태로 편집·합성·가공하는 행위를 전제로 한다. 따라서 특정인의 기존 사진을 직접 편집하거나 합성하지 않고, 텍스트만으로 그 사람을 닮은 성적 이미지를 새롭게 생성한 경우에도 이를 이 조항의 '편집·합성·가공'으로 볼 수 있는지는 별도의 법적 검토가 필요하다.

더욱이 성적 문구나 허위 프로필을 생성하거나, 실존 인물을 모사한 챗봇을 성적 대화의 대상으로 이용하는 경우처럼 피해가 이미지·영상·음성의 형태를 취하지 않는 경우도 있다. 이러한 행위는 허위 영상물에 관한 규정만으로 다루기 어려우며 행위의 구체적인 내용과 피해의 형태에 따라 다른 법률의 적용 가능성을 별도로 살펴볼 필요가 있다. 기술의 발전이 빠른 사회에서 결국 법적 판단은 해당 기술이 어떤 방식으로 이용되었고 그 결과 어떤 침해가 발생했는지를 중심으로 이루어질 필요가 있다는 것이다.

아동·청소년과 관련된 인공지능 생성 성적 콘텐츠의 경우, 「아동·청소년의 성 보호에 관한 법률」상 아동·청소년 성 착취물은 실제 아동·청소년이 촬영된 경우에만 한정되지 않고 아동·청소년으로 명백하게 인식될 수 있는 사람이나 표현물이 등장하는 경우도 포함한다. 따라서 생성형 인공지능으로 만들어진 콘텐츠라고 해서, 또는 실제 아동·청소년이 촬영 과정에 참여하지 않았다고 해서 곧바로 법 적용 대상에서 제외되는 것은 아니다. 성인 대상 성적 합성물이 주로 특정인의 얼굴·신체·음성이나 정체성을 동의 없이 이용한 침해를 중심으로 논의된다면, 아동·청소년 관련 사안에서는 실제 피해자가 특정되지 않더라도 아동·청소년을 성적으로 대상화한 표현물 자체가 별도의 규율 대상이 될 수 있다.

다만 실제 촬영이 없다는 이유만으로 성적 침해의 보호 필요성을 부정해서는 안 된다는 점은 성인 피해에서도 마찬가지다. 실제 촬영 여부나 생성물의 진위보다 당사자의 의사에 반한 성적 이용과 그로 인한 권리 침해를 중심으로 피해를 판단해야 한다는 원칙은 성인 피해자 보호에서도 중요하게 고려될 필요가 있다.

문제는 이러한 피해를 현행 피해 지원체계가 어느 범위까지 포착하고 지원할 수 있는가이다. 형사법상 처벌 가능성과는 별개로 피해자가 경험하는 침해가 기존 지원체계의 대상에 포함되는지를 살펴볼 필요가 있다.

대표적으로 「성폭력방지 및 피해자 보호 등에 관한 법률」에 따른 삭제 지원과 지원센터의 법정 업무는 불법 촬영물, 허위 영상물, 아동·청소년 성 착취물 등 법률에 규정된 피해 유형을 중심으로 구성되어 있다. 이는 디지털 성범죄 피해에 대응하기 위한 중요한 제도적 기반이지만 성적 사칭이나 계정 탈취와 결합된 괴롭힘, 텍스트 기반의 성적 허위 정보, 반복적으로 생성되는 성적 메시지 등은 기존 지원 경로에 곧바로 포섭되기 어려울 수 있다.

이러한 간극은 피해 지원 통계에서도 일부 확인된다. 2025년 지원 통계에는 SNS 해킹과 사칭 등 피해자가 호소한 '기타' 피해가 1,558건으로 집계되었으며 이는 전체 피해 유형의 8.8%에 해당한다. 해당 보고서는 이를 피해 지원 범위에 포함되지 않는 유형으로 설명하고 있다(성평등가족부·한국여성인권진흥원, 2026). 물론 이 사례들이 모두 인공지능과 관련된 것은 아니다. 다만 피해자가 실제 침해를 호소하더라도 그 피해가 현행 지원체계의 분류와 지원 범위에 포착되지 않을 수 있음을 보여준다는 점에서 주목할 필요가 있다.

나. 피해자 보호·지원의 방향

피해 지원 정책에서는 먼저 인공지능 매개 젠더폭력을 폭넓게 접수하고 판단할 수 있는 기준이 필요하다. 이는 모든 행위를 새로운 범죄로 규정하자는 뜻이 아니다. 특히 인공지능 매개 젠더폭력과 같이 기술변화에 따라 다변화한 새로운 유형의 피해가 늘어나고 있어 그 피해 자체를 충분히 파악하기 어려운 경우, 피해 지원의 기준은 '법적 피해에 해당하는가?'보다 '어떤 침해와 위험이 발생하고 있는가?'에 가깝게 설계해야 한다. 피해 지원은 피해와 위험이 확인되면 시작될 수 있어야 한다는 의미이다.

지원 기준을 넓히는 것과 함께 지원의 내용도 인공지능 매개 피해의 특성에 맞게 확장할 필요가 있다. 삭제 지원은 여전히 디지털 성범죄 피해 지원의 핵심적인 영역이다. 2025년 전체 지원 서비스의 90.3%가 삭제 지원이었다는 수치는 하나의 피해물에 대해 다수의 URL을 반복적으로 삭제해야 하는 현실을 보여준다. 그러나 앞서 말했듯, 인공지능 매개 피해는 동일인을 대상으로 한 콘텐츠의 재생성, 사칭 계정이나 개인정보의 악용 등으로 이어질 수 있어 삭제만으로 피해가 종결되지 않는 경우가 있다. 따라서 증거보전, 계정과 개인정보 보호, 수사·법률 지원, 심리·의료지원, 학교나 직장에서의 안전조치 등을 함께 연계하는 사례관리 기능이 중요하다.

또한 피해물의 발견과 신고를 피해자 개인에게 반복적으로 맡기는 방식에서 벗어날 필요가 있다. UNESCO(2023)는 피해자가 스스로 계정을 비공개로 전환하고 지속적으로 차단·신고하는 것만으로는 피해의 확산과 재발을 막기 어렵다고 지적하며 문제 콘텐츠의 선제적 식별과 신속한 신고 처리, 반복적·자동화된 괴롭힘에 대한 모니터링을 제안한다. 정부가 2026년 도입한 시스템 역시 아동·청소년 성 착취물과 유인 정보를 자동 수집·분석하고 딥페이크 의심 콘텐츠를 탐지하는 것을 목표로 한다(성평등가족부, 2026b). 이러한 기능을 발전시키는 한편 성인 피해에 대해서도 당사자의 의사와 개인정보 보호를 전제로 반복 유포와 변형물의 탐지·모니터링을 지원할 필요가 있다.

이러한 선제적 대응을 위해서는 플랫폼과 인공지능 서비스 제공자의 역할도 중요하다. 반복 유포 사이트에 대한 분석과 차단, 수사의뢰, 국제공조뿐 아니라 증거자료의 보존, 긴급 신고의 처리와 결과 회

신 등 피해 지원기관과 플랫폼 사이의 협력 절차를 구체화할 필요가 있다. 인공지능 서비스의 개발·운영 단계에서도 성적 합성이나 괴롭힘에 악용될 위험을 점검하고 안전장치를 반복적으로 우회하는 이용 행위를 탐지하며 피해자가 해당 콘텐츠나 계정을 신속하게 신고하고 조치를 요청할 수 있는 절차를 마련해야 한다.

마지막으로 피해 통계를 인공지능 환경의 변화에 맞게 세분화할 필요가 있다. 인공지능의 개입 여부, 생성물의 형태, 실제 사진·음성의 이용 여부, 사칭이나 개인정보의 결합 여부, 반복 생성·재유포 여부, 가해자 특정 여부 등을 함께 파악해야 새로운 피해 양상과 그 변화를 구체적으로 확인할 수 있다. 특히 현행 피해 지원 범위에 포함되지 않는 사례에서 피해자가 실제로 어떤 침해를 호소하고 있는지를 별도로 분석한다면, 기존의 피해 분류가 포착하지 못하는 영역과 지원체계의 공백을 확인하는 데 중요한 자료가 될 수 있을 것이다.

4. 나가며

한국의 디지털 성범죄 대응 체계는 최근 몇 년 사이 허위 영상물 제작행위에 대한 처벌 강화, 중앙·지역 지원센터의 법제화, 인공지능 기반 탐지·삭제 시스템의 도입, 범정부 통합지원단 설치 등으로 확대되어 왔다.

그러나 생성형 인공지능이 제기하는 문제는 딥페이크 영상의 증가에만 한정되지 않는다. 이미지와 음성, 텍스트, 개인정보와 정체성을 결합해 특정인을 성적으로 대상화할 수 있고 동일인을 대상으로 한 침해를 반복·변형할 수도 있다. 피해 유형은 이미지, 영상에서 음성, 텍스트, 대화형 서비스 및 정체성 도용으로 확장되고 있으며 생성기능이 플랫폼과 서비스 내부에 결합되면서 콘텐츠가 만들어지고 또 공유되는 방식도 달라지고 있다. 생성과 유통이 여러 행위자와 서비스에 걸쳐 이루어지면서 현행 법률과 지원체계가 규정하고 있는 피해 유형만으로는 실제 피해 경험을 충분히 포착하기 어려운 영역도 나타나고 있다.

따라서 인공지능 매개 젠더폭력에 대한 대응은 얼마나 많은 피해물이 생성, 유통되고 있는지 뿐 아니라 기술변화에 따라 피해 유형과 발생 방식이 어떻게 확장, 변형되고 있는지도 함께 살펴볼 필요가 있다. 피해 지원은 법적 분류에 앞서 피해자가 현재 겪고 있는 침해와 위험에서 출발할 필요가 있다. 기술이 바뀔 때마다 피해자가 자신의 피해가 왜 지원 대상이 되어야 하는지를 먼저 설명하도록 해서는 안 된다. 인공지능 시대의 피해 지원 정책 역시 얼마나 많은 합성물을 탐지하고 삭제했는지만으로 평가하기 어렵다. 피해자가 지원체계에 접근하고 있는지, 필요한 보호와 조력을 지속적으로 제공받았는지, 그리고 자신의 일상과 정체성에 대한 통제권을 얼마나 회복할 수 있었는지까지 함께 살펴볼 필요가 있다.

• 참고문헌 •

1. 국내문헌

경찰청 국가수사본부 (2024). 「허위 영상물(딥페이크 성범죄) 특별 집중단속 실시」 보도자료(2024. 8. 27.).

방송미디어통신위원회 (2026). 엑스(X) 내 인공지능 기능을 이용한 딥페이크 이미지 생성 관련 보호계획 제출 요구(2026. 1. 14.).

• 참고문헌 •

성평등가족부 (2026a).『2026년 주요업무 추진계획』.
성평등가족부 (2026b).「'디지털 성범죄' 이제 인공지능(AI)으로 24시간 자동 탐지·분석」 보도자료(2026. 4. 1.).
성평등가족부·한국여성인권진흥원 (2026).『2025 디지털성범죄 피해자 지원 보고서』.
여성가족부·한국여성인권진흥원 (2025).『2024 디지털성범죄 피해자 지원 보고서』.
연합뉴스 (2026. 5. 15.).「[샷!] 교복 입히고 성적 요구까지…AI 캐릭터챗 성착취 논란」.

2. 법령

「개인정보 보호법」https://law.go.kr/lsSc.do?section=&menuId=1&subMenuId=15&tabMenuId=81&eventGubun=060101&query=%EA%B0%9C%EC%9D%B8%EC%A0%95%EB%B3%B4+%EB%B3%B4%ED%98%B8%EB%B2%95#undefined
「성폭력방지 및 피해자보호 등에 관한 법률」https://law.go.kr/LSW/lsSc.do?section=&menuId=1&subMenuId=15&tabMenuId=81&eventGubun=060101&query=%EC%84%B1%ED%8F%AD%EB%A0%A5%EB%B0%A9%EC%A7%80+%EB%B0%8F+%ED%94%BC%ED%95%B4%EC%9E%90%EB%B3%B4%ED%98%B8+%EB%93%B1%EC%97%90+%EA%B4%80%ED%95%9C+%EB%B2%95%EB%A5%A0#undefined
「성폭력범죄의 처벌 등에 관한 특례법」https://law.go.kr/LSW/lsSc.do?section=&menuId=1&subMenuId=15&tabMenuId=81&eventGubun=060101&query=%EC%84%B1%ED%8F%AD%EB%A0%A5%EB%B2%94%EC%A3%84%EC%9D%98+%EC%B2%98%EB%B2%8C+%EB%93%B1%EC%97%90+%EA%B4%80%ED%95%9C+%ED%8A%B9%EB%A1%80%EB%B2%95#undefined
「아동·청소년의 성보호에 관한 법률」https://law.go.kr/LSW/lsSc.do?section=&menuId=1&subMenuId=15&tabMenuId=81&eventGubun=060101&query=%EC%95%84%EB%8F%99%C2%B7%EC%B2%AD%EC%86%8C%EB%85%84%EC%9D%98+%EC%84%B1%EB%B3%B4%ED%98%B8%EC%97%90+%EA%B4%80%ED%95%9C+%EB%B2%95%EB%A5%A0#undefined

3. 국외 문헌

Gibson, C., Olszewski, D., Brigham, N. G., Crowder, A., Butler, K. R. B., Traynor, P., Redmiles, E. M., & Kohno, T. (2024). Analyzing the AI nudification application ecosystem. arXiv:2411.09751.
Hawkins, W., Mittelstadt, B., & Russell, C. (2025). Deepfakes on demand: The rise of accessible non-consensual deepfake image generators. In Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25) (pp. 1602–1614). ACM.
McGlynn, C., & Rackley, E. (2017). Image-based sexual abuse. Oxford Journal of Legal Studies, 37(3), 534–561.
McGlynn, C., Johnson, K., Rackley, E., Henry, N., Gavey, N., Flynn, A., & Powell, A. (2021). 'It's torture for the soul': The harms of image-based sexual abuse. Social & Legal Studies, 30(4), 541–562.
UNESCO (2023). "Your opinion doesn't matter, anyway": Exposing technology-facilitated gender-based violence in an era of generative AI. Paris: UNESCO.
UNFPA (2021). Technology-facilitated gender-based violence: Making all spaces safe. New York: UNFPA.