



제418회 국회(정기회)
제19차 법제사법위원회

형법 일부개정법률안 검 토 보 고

〈인공지능 기술을 이용한 명예훼손죄 신설〉
■ 고동진의원 대표발의(의안번호 제2203539호)

2024. 12.

법 제 사 법 위 원 회
수석전문위원 정 환 철

목 차

I . 제안경위	1
II . 제안이유 및 주요내용	1
III . 검토의견	3

I. 제안경위

1. 발 의 자: 고동진의원 등 12인
2. 발 의 일: 2024. 9. 2.
3. 회 부 일: 2024. 9. 3.

II. 제안이유 및 주요내용

최근 인공지능(AI) 기술을 악용한 딥페이크(합성하여 만든 가짜 영상, 사진, 음성) 범죄가 무분별하게 확산되고 있는 가운데, 지난달 미국 캘리포니아 주의회가 AI를 활용해 딥페이크 아동 성착취물을 제작, 배포, 소지하는 행위를 불법으로 규정한 법안을 통과시켰고, 영국의 경우 올해 4월 영상물이나 이미지를 유포할 의사가 없었다 해도 당사자의 동의 없이 딥페이크 영상물 등을 제작하는 경우 처벌할 수 있도록 법을 개정하는 등 세계 각국이 ‘딥페이크 방지 및 처벌’을 위한 입법을 신속히 추진하고 있는 상황임.

하지만 국내법의 경우 「성폭력범죄의 처벌 등에 관한 특례법」에서는 유포할 목적에 한정하여 처벌할 수 있도록 해 유포했다는 사실이 제대로 입증되지 않은 경우 처벌할 수가 없고, 이를 소지, 구입, 시청한 경우에는 처벌할 수 있는 근거 자체가 없는 상황임. 또한, 딥페이크가 성범죄 외 음성 등 AI보이스 피싱 사례에 악용될 수 있지만, 이를 예방 및 방지할 수 있는 법률적 근거가 미약한 실정임.

이에 ① 딥페이크 제작물을 유포 목적이 아니더라도 당사자의 의사

에 반하여 제작(판매, 유포 및 이용 포함)하거나, ② 해당 제작물을 소지, 구입, 시청하여 당사자의 명예를 훼손한 자에 대한 처벌 근거를 마련함과 동시에, ③ 성범죄의 영상물뿐만 아니라 음성과 사진에 대한 딥페이크 규제 및 처벌 규정을 신설하여, 「포괄적인 딥페이크 방지 및 처벌」이 가능하도록 하고자 함(안 제309조의2 신설).

Ⅲ. 검토의견

현행법상 명예훼손죄는 공연히 사실 또는 허위의 사실을 적시하여 사람의 명예를 훼손한 자에 대한 명예훼손죄(제307조), 공연히 허위의 사실을 적시하여 사자의 명예를 훼손한 자에 대한 사자명예훼손죄(제308조), 신문·잡지·라디오·기타 출판물에 의하여 사람의 명예를 훼손한 자에 대한 출판물 등에 의한 명예훼손죄(제309조)로 구분하여 처벌하고 있음.¹⁾

개정안은 사람의 얼굴·신체·음성을 대상으로 인공지능 기술을 이용한 영상물 등을 대상자의 의사에 반하여 제작·판매·유포·이용하거나 해당 영상물등을 소지·구입·시청하여 사람의 명예를 훼손한 자를 5년 이하의 징역 또는 5천만원 이하의 벌금에 처하는 인공지능 기술 이용 명예훼손죄를 신설하려는 것임(안 제309조의2 신설).

1) 제307조(명예훼손) ①공연히 사실을 적시하여 사람의 명예를 훼손한 자는 2년 이하의 징역이나 금고 또는 500만원 이하의 벌금에 처한다.

② 공연히 허위의 사실을 적시하여 사람의 명예를 훼손한 자는 5년 이하의 징역, 10년 이하의 자격정지 또는 1천만원 이하의 벌금에 처한다.

제308조(사자의 명예훼손) 공연히 허위의 사실을 적시하여 사자의 명예를 훼손한 자는 2년 이하의 징역이나 금고 또는 500만원 이하의 벌금에 처한다.

제309조(출판물 등에 의한 명예훼손) ①사람을 비방할 목적으로 신문, 잡지 또는 라디오 기타 출판물에 의하여 제307조제1항의 죄를 범한 자는 3년 이하의 징역이나 금고 또는 700만원 이하의 벌금에 처한다.

② 제1항의 방법으로 제307조제2항의 죄를 범한 자는 7년 이하의 징역, 10년 이하의 자격정지 또는 1천500만원 이하의 벌금에 처한다.

제310조(위법성의 조각) 제307조제1항의 행위가 진실한 사실로서 오로지 공공의 이익에 관한 때에는 처벌하지 아니한다.

현행	개정안
<신설>	제309조의2(인공지능 기술의 이용에 의한 명예훼손) 사람의 얼굴, 신체, 음성을 대상으로 인공지능 기술을 이용한 촬영물, 영상물, 음성물(이하 “영상물등”이라 한다)을 영상물등 대상자의 의사에 반하여 제작, 판매, 유포, 이용하거나 해당 영상물등을 소지, 구입, 시청하여 사람의 명예를 훼손한 자는 5년 이하의 징역 또는 5천만원 이하의 벌금에 처한다.

최근 인공지능(AI) 기술은 특정인의 얼굴이나 신체부위를 음란물과 마치 실제처럼 합성하는 방식으로 제작·유포하는 이른바 “딥페이크²⁾ 성범죄”에 악용되고 있으며, 그에 따른 피해 또한 급증하여 사회적 논란이 되고 있음.³⁾

특히, 미국 사이버보안업체인 ‘시큐리티 히어로’가 최근 공개한 ‘2023 딥페이크 현황’ 보고서의 분석에 따르면, 지난해 7~8월 딥페이

2) 딥페이크(Deepfake)란, 인공지능 기술인 딥러닝(deep learning)과 ‘가짜’를 의미하는 단어인 페이크(fake)의 합성어로, 인공지능 기술을 이용하여 진위 여부를 구별하기 어려운 가짜 이미지나 영상물을 뜻함. 딥페이크라는 단어가 등장한 시기는 2017년으로, 미국 온라인 커뮤니티 Reddit의 한 회원이 기존 영상에 유명인의 얼굴을 입혀 가짜 콘텐츠를 게재한 데서 유래되었음.

(출처: 한국방송통신전파진흥원, 2019 트렌드리포트, “딥페이크 기술의 빛과 그림자”, p.30)

3) 국민일보(2024. 8. 20.), “대학생 1천명 채팅방에 4년간 ‘딥페이크 영상’ 공유”, <https://m.kmib.co.kr/view.asp?arcid=0020436770> (최종검색일 : 2024. 9. 6.)

크 성착취물에 등장하는 한국 국적의 인물이 53%에 달하였으며, 해외 언론에서도 이를 인용하는 등 우리나라가 당면한 딥페이크 성착취물 문제는 국내를 넘어 전세계적인 이슈가 되고 있는 것으로 보임.⁴⁾

가수의 노래나 정치인의 연설 등과 관련된 허위영상물 유포는 성적 욕망 또는 수치심을 유발할 수 있는 형태가 아니더라도 사회적으로 미칠 수 있는 영향이 크고, 당사자의 명예감정까지 훼손시킬 수 있는 점 등을 감안할 때, ‘인공지능 기술을 이용한 명예훼손죄’를 신설하여 이를 「형법」상 명예훼손죄의 새로운 형태로 처벌하려는 개정안의 입법취지는 인정됨.

다만, 다음 사항에 대한 고려가 필요할 것으로 보임.

(1) 명예훼손죄의 구성요건과 관련하여, 현행 「형법」상 명예훼손죄는 성립에 있어 ‘공연성’과 ‘비방할 목적’(출판물 등에 의한 명예훼손죄의 경우)을 요구하는바, 개정안의 ‘인공지능 기술 이용 명예훼손죄’의 경우 공연성 또는 비방할 목적이 없더라도 ‘대상자의 의사에 반할 것’을 주요 구성요건으로 규정하고, ‘판매·유포’ 외에 ‘제작·이용·소지·구입·시청’은 공연성이나 특정한 목적이 없어도 대상자의 의사에 반하기만 하면 명예훼손죄가 성립하는 것으로 규정하고 있음. 따라서, 이러한 구성요건 규정이 적절한지 현행 형법 조항과 비교하여 논의할 필요가 있음.

4) 조선일보(2024. 8. 31), “딥페이크 피해자, 한국인이 53% 세계 최다”,
https://www.chosun.com/economy/tech_it/2024/08/31/E25O4H7UI5CIREFOMNRS562GCY/
(최종검색일 : 2024. 9. 6.)

(2) 형법상 명예훼손죄의 보호법익은 사람의 가치에 대한 사회적 평가인 이른바 ‘외부적 명예’임.⁵⁾ 따라서, 영상물등을 대상자의 의사에 반하여 ‘판매, 유포’하는 경우 외부적 명예에 대한 침해가 명백히 발생할 것으로 보이나, 외부로의 유출을 전제하지 않는 ‘제작·이용’ 행위나 ‘소지·구입·시청’ 행위가 사람의 외부적 명예를 훼손시키는 행위일지에 대해서는 논의가 필요할 것으로 보임.

이와 관련하여, 현행 「형법」상 음화제조죄, 음화소지죄 등의 경우에도, 음란한 문서의 반포·판매 등 행위에 공할 목적으로 음란한 물건을 제조·소지·수입·수출한 경우에 처벌하는 규정을 두고 있는 점을 참고할 수 있을 것으로 보임.⁶⁾

(3) 다음, 인공지능 기술을 이용한 영상물등 중 성착취물과 관련하여서는, 2020. 3. 24. 신설되고 2024. 10. 16. 개정된 현행 「성폭력처벌법」 제14조의2의 입법 취지를 감안하여야 할 것으로 생각됨.

구체적으로, 동 조항 신설 이전에는 딥페이크 성착취물 유포에 대한 처벌은 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」(이하 ‘정보통신망법’)상 허위사실적시 명예훼손죄나 음란물유포죄⁷⁾ 또는

5) 대법원 1987. 5. 12. 선고 87도739 판결

6) 「형법」 제243조(음화반포등) 음란한 문서, 도화, 필름 기타 물건을 반포, 판매 또는 임대하거나 공연히 전시 또는 상영한 자는 1년 이하의 징역 또는 500만원 이하의 벌금에 처한다.

제244조(음화제조 등) 제243조의 행위에 공할 목적으로 음란한 물건을 제조, 소지, 수입 또는 수출한 자는 1년 이하의 징역 또는 500만원 이하의 벌금에 처한다.

7) 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」

제44조의7(불법정보의 유통금지 등) ① 누구든지 정보통신망을 통하여 다음 각 호의 어느 하나에 해당하는 정보를 유통하여서는 아니 된다.

1. 음란한 부호·문언·음향·화상 또는 영상을 배포·판매·임대하거나 공공연하게 전시하는 내용의 정보

「형법」상 음화반포·제조죄 등⁸⁾이 적용되었음.

그러나 허위사실 적시 명예훼손죄의 경우 피해자의 사회적 평가를 저하시킬 수 있는 구체적 허위사실의 적시가 필요하고⁹⁾ 미수범 처벌 규정이 없으며, 「정보통신망법」상 음란물유포죄나 「형법」상 음화 반포·제조죄 등의 경우 1년 이하의 징역 또는 1천만원/500만원 이하의 벌금으로 처벌되어 피해자의 인격권 침해 정도에 비례하지 않는 매우 낮은 법정형이 적용된다는 점에서, 이를 디지털 성범죄로 특별히 취급하여 엄벌할 필요성이 대두되었음.¹⁰⁾

이에 따라, 2020년 현행 「성폭력처벌법」 제14조의2를 신설하여 딥페이크 성착취물을 반포할 목적으로 편집·반포 등을 한 경우 처벌하는 규정을 마련하였는데, 해당 영상물의 편집·반포 등을 한 행위 그 자체

2. 사람을 비방할 목적으로 공공연하게 사실이나 거짓의 사실을 드러내어 타인의 명예를 훼손하는 내용의 정보

(생략)

제70조(벌칙) ① 사람을 비방할 목적으로 정보통신망을 통하여 공공연하게 사실을 드러내어 다른 사람의 명예를 훼손한 자는 3년 이하의 징역 또는 3천만원 이하의 벌금에 처한다.

② 사람을 비방할 목적으로 정보통신망을 통하여 공공연하게 거짓의 사실을 드러내어 다른 사람의 명예를 훼손한 자는 7년 이하의 징역, 10년 이하의 자격정지 또는 5천만원 이하의 벌금에 처한다.

③ 제1항과 제2항의 죄는 피해자가 구체적으로 밝힌 의사에 반하여 공소를 제기할 수 없다.

제74조(벌칙) ① 다음 각 호의 어느 하나에 해당하는 자는 1년 이하의 징역 또는 1천만원 이하의 벌금에 처한다.

1. 제8조제4항을 위반하여 비슷한 표시를 한 제품을 표시·판매 또는 판매할 목적으로 진열한 자

2. 제44조의7제1항제1호를 위반하여 음란한 부호·문언·음향·화상 또는 영상을 배포·판매·임대하거나 공공연하게 전시한 자

(생략)

8) 「형법」

제243조(음화반포등) 음란한 문서, 도화, 필름 기타 물건을 반포, 판매 또는 임대하거나 공연히 전시 또는 상영한 자는 1년 이하의 징역 또는 500만원 이하의 벌금에 처한다.

제244조(음화제조 등) 제243조의 행위에 공할 목적으로 음란한 물건을 제조, 소지, 수입 또는 수출한 자는 1년 이하의 징역 또는 500만원 이하의 벌금에 처한다.

9) 대법원 2018. 5. 30. 선고 2017도607 판결

10) 배상균, “인공지능(AI) 기술을 이용한 디지털 성범죄에 대한 검토”, 외법논집 제43권 제3호 (2019.8.)

를 성범죄로 보아 미수범까지 처벌하고, 영리목적 행위일 경우 가중처벌 규정 등을 두었다는 점¹¹⁾에서 의의가 있다고 할 것임.

이에 더하여, 2024. 9. 26. 본회의에서 의결된 「성폭력처벌법 일부개정법률안」¹²⁾은 반포할 목적 없이 영상물등을 편집·합성·가공하는 경우에도 7년 이하의 징역 또는 5천만원 이하의 벌금에 처하도록 하여(법정형 상향) 처벌 근거를 마련하였고, 위와 같은 편집물 등임을 알면서 이를 소지·구입·시청하는 경우 3년 이하의 징역 또는 3천만원 이하의 벌금에 처하는 처벌규정을 신설하였음.¹³⁾

이와 같이, 현행 「성폭력처벌법」은 반포할 목적 없이 딥페이크 성착취물을 편집하거나 이를 알면서 단순히 소지·구입·시청하는 행위 등도 처벌할 수 있도록 조항을 마련하고 있으므로, 개정안의 입법취지는 상당 부분 달성되었다고 볼 수 있는 점을 고려할 필요가 있음.

(4) 한편, 현행법상 사자명예훼손죄 및 모욕죄는 고소가 있어야 공

11) 이하의 각주13 참고

12) [시행 2024. 10. 16.] [법률 제20459호, 2024. 10. 16., 일부개정]

13) 「성폭력범죄의 처벌 등에 관한 특례법」 제14조의2(허위영상물 등의 반포등) ① 사람의 얼굴·신체 또는 음성을 대상으로 한 촬영물·영상물 또는 음성물(이하 이 조에서 “영상물등”이라 한다)을 영상물등의 대상자의 의사에 반하여 성적 욕망 또는 수치심을 유발할 수 있는 형태로 편집·합성 또는 가공(이하 이 조에서 “편집등”이라 한다)한 자는 7년 이하의 징역 또는 5천만원 이하의 벌금에 처한다.

② 제1항에 따른 편집물·합성물·가공물(이하 이 조에서 “편집물등”이라 한다) 또는 복제물(복제물의 복제물을 포함한다. 이하 이 조에서 같다)을 반포등을 한 자 또는 제1항의 편집등을 할 당시에는 영상물등의 대상자의 의사에 반하지 아니한 경우에도 사후에 그 편집물등 또는 복제물을 영상물등의 대상자의 의사에 반하여 반포등을 한 자는 7년 이하의 징역 또는 5천만원 이하의 벌금에 처한다.

③ 영리를 목적으로 영상물등의 대상자의 의사에 반하여 정보통신망을 이용하여 제2항의 죄를 범한 자는 3년 이상의 유기징역에 처한다.

④ 제1항 또는 제2항의 편집물등 또는 복제물을 소지·구입·저장 또는 시청한 자는 3년 이하의 징역 또는 3천만원 이하의 벌금에 처한다.

⑤ 상습으로 제1항부터 제3항까지의 죄를 범한 때에는 그 죄에 정한 형의 2분의 1까지 가중한다.

소를 제기할 수 있는 친고죄로, 명예훼손죄 및 출판물 등에 의한 명예훼손죄는 피해자의 명시한 의사에 반하여 공소를 제기할 수 없는 반의사불벌죄로 규정하고 있음.¹⁴⁾ 개정안의 범죄를 입법에 반영할 경우, 현행법의 취지에 따라 ‘인공지능 기술을 이용한 명예훼손죄’도 반의사불벌죄로 규정할 것인지 추가적으로 논의할 필요가 있을 것으로 보임.

문 의
02-6788-5053

14) 제312조(고소와 피해자의 의사) ① 제308조와 제311조의 죄는 고소가 있어야 공소를 제기할 수 있다.

② 제307조와 제309조의 죄는 피해자의 명시한 의사에 반하여 공소를 제기할 수 없다.