

2024. 5. 23.(목) 조간용

이 보도자료는 2024년 5월 22일 오전 11:15부터 보도할 수 있습니다.



서울연구원
The Seoul Institute

보도자료

담당 부서: 여성가족정책실 양성평등담당관

서울시여성가족재단

서울연구원 디지털도시연구실

사진 없음 사진 있음 쪽수: 14쪽

양성평등담당관	이성은	02-2133-5005
권익사업팀장	김지현	02-2133-5052
양성평등사업실장	최자은	02-810-5024
서울디지털성범죄 안심지원센터장	문기현	02-810-5700
디지털도시연구실장	조혜림	02-2144-2967
부연구위원	김준철	02-2144-2971

서울시, 전국 최초 인공지능(AI) 안면인식으로 아동·청소년 성착취물 검출해 선제적 삭제

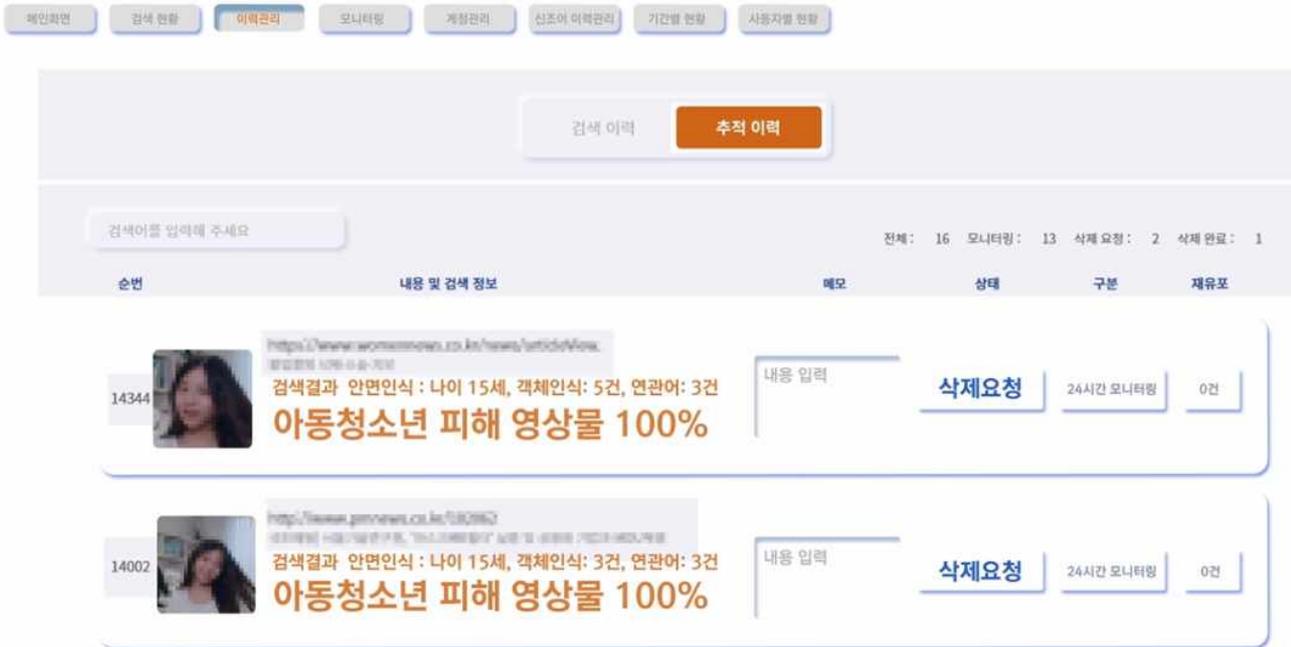
- 인공지능(AI) 안면인식으로 성별·나이 예측, 교복 인형 등 사물과 언어도 인식해 아동·청소년 판별
- 육안 판독시 2시간 걸렸던 검출속도 1분30초 내외로 단축, 정확도 300% ↑
- 인공지능(AI)으로 기존의 2배 이상인 30만건 검색 삭제 기대, 중국·러시아 등 해외 유포 영상물도 삭제
- '서울 디지털 성범죄 안심지원센터' 개관 2년 간 아동·청소년 피해지원 7배 이상 증가

서울시가 디지털 성범죄 예방과 피해자 지원을 위해 지난해 3월 전국 최초로 인공지능(AI) 기술을 도입한 데 이어, 아동·청소년 성착취물을 특정할 수 있는 인공지능(AI) 감시 기술을 전국 최초로 개발·도입해 24시간 자동 추적·감시에 나선다.

- '23년도 3월부터 서울연구원이 핵심기술 연구개발에 착수, 서울 디지털 성범죄 안심지원센터와 함께 다수의 아동·청소년 피해 영상물을 분석하여 올해 5월 프로그램 개발을 완료했다.
- 서울연구원은 지난해 전국 최초로 인공지능(AI) 기술 도입을 통한 디지털 성범죄 자동 추적·감시 알고리즘을 개발했으며, 서울시에서 수행하는 24시간 불법 영상물 감시 및 신속 삭제를 위한 전반적인 기술지원을 수행하고 있다.

- 서울시는 아동·청소년들은 디지털 성범죄 피해를 입더라도 부모님에게 말하지 못해 신고가 이뤄지지 않아 피해 영상물이나 사진이 유포·재유포 되는 경우가 많아 이번 기술 개발에 나섰다고 취지를 밝혔다. 아동·청소년의 경우 관련 법에 따라 당사자나 부모의 신고 없이도 피해 영상물 삭제가 가능한 만큼, 인공지능(AI)을 통해 아동·청소년 성착취물을 빠르게 찾아내고 삭제해 피해에 신속 대응한다는 계획이다.
- ‘서울 디지털 성범죄 안심지원센터(서울여성가족재단 운영)’의 피해 요청 현황을 살펴보면 아동·청소년 피해자 스스로 신고한 경우는 12명(7.8%)밖에 되지 않으며, 부모님에게 피해 사실이 노출되거나 수사·사법기관 등의 신고로 이뤄지는 경우가 대부분이다.
- 실제로 센터의 삭제지원 사례를 살펴보면 센터에 삭제를 요청하여 삭제지원한 건수는 425건으로 센터 삭제지원관이 피해자의 신고 없이 삭제한 ‘선제적 삭제지원’ 2,720건의 15.6% 정도이다. 즉, 대부분의 아동·청소년 디지털 성범죄 피해 영상물은 신고가 이뤄지지 못하고 유포·재유포 되는 경우가 많다.
- 본인이 삭제를 요청해야 삭제지원이 가능한 성인과 달리, 아동·청소년은 「아동·청소년의 성보호에 관한 법률」 등에 따라 당사자나 부모님의 신고 없이도 즉시 삭제가 가능하다.
- 이번에 개발한 ‘아동·청소년 디지털 성범죄 인공지능(AI) 감시 시스템’은 인공지능(AI) 딥러닝 기반 안면인식 기술로 성인과 잘 구분되지 않는 아동·청소년의 성별과 나이를 판별할 수 있는 시스템이다.
- 특히 주목할 만한 점은 영상물에 ‘얼굴’이 나오지 않더라도 아동·청소년 피해 영상물 여부를 찾아낼 수 있다는 점이다. 인공지능(AI)이 피해 영상물에 자주 등장하는 책, 교복, 인형 등 주변 사물은 물론, 이미지

속 텍스트, 청소년들이 주로 사용하는 언어까지 함께 인식해서 최종적으로 피해 영상물 여부를 판별하게 된다.



〈아동·청소년 디지털 성범죄 인공지능(AI) 감시시스템 결과 화면(예시)〉

- 또한, 키워드 입력부터 영상물 검출까지 90초밖에 걸리지 않기 때문에 기존에 삭제지원관이 관련 키워드로 피해 영상물을 검색해서 수작업으로 찾아낼 때(2시간)와 비교하면 검출 속도가 1/80로 크게 개선되고, 정확도도 300% 이상 향상된다. 이렇게 되면 지난해 수작업으로 이뤄진 모니터링 건수(15만 건)의 2배인 30만 건까지 모니터링이 가능해지고, 인공지능(AI) 학습 데이터가 축적될수록 정확도와 속도는 더욱 향상될 것으로 기대된다.
- 인공지능(AI)이 사회관계망서비스(SNS) 빅데이터 분석을 통해 아동·청소년 성범죄물 관련 신조어를 자동으로 생성할 수 있게 돼 영상물 검출에 사용되는 키워드도 다양해진다. 예컨대 기존에는 삭제지원관이 ‘딥페(딥페이크의 약자)’로 검색을 했다면 인공지능(AI)을 통해 ‘딥페’ ‘뒷페’ 등 신조어 자동생성이 가능해지는 것. 이를 통해 기존에는 30개 키워드로 검색했다면 이제는 3.3배인 100개의 키워드를 생성해 보다 많은 피해 영상물을 찾아낼 수 있게 된다.

- 아울러, 시는 기존에 미국을 중심으로 유포됐던 피해 영상물이 최근 중국, 러시아, 베트남 등으로 확산되는 점에 착안해 국가 기반을 넓혀 검색 영역을 확장했다. 이에 인공지능(AI)을 통해 국내와 미국뿐 아니라 중국, 러시아 등의 국가에 유포된 피해 영상물 검색도 가능해진다.

<서울 디지털 성범죄 안심지원센터' 개관 2년 간 아동·청소년 피해지원 7배 이상 증가>

- 서울시는 '서울 디지털 성범죄 안심지원센터(이하 센터, 서울여성가족재단 운영)'를 통해 긴급상담부터 수사·법률지원, 삭제지원, 심리치료·의료 지원까지 원스톱 지원하고 있다. 지난 2년간('22.3~'24.3월) 935명의 피해자를 지원했다. 총지원 건수는 30,576건에 이른다.

- 서울시는 지난해 전국 최초로 디지털 성범죄 인공지능(AI) 삭제지원 프로그램 개발했으며, 올해는 아동·청소년 디지털 성범죄 인공지능(AI) 감시 시스템을 구축했다. 내년에는 검색뿐 아니라 신고까지 인공지능(AI)이 자동으로 신고하는 인공지능(AI) 자동신고 시스템을 구축할 계획이다.

- 시가 지원한 피해자의 연령대는 10~20대가 가장 높은 비율을 차지하고 있으며, 성별로는 여성이 85.78%, 남성도 12.09%였다. 특히 센터 개관 이후(22년 3월) 아동, 청소년을 대상으로 한 디지털 성범죄 피해 지원이 더욱 증가하고 있다. '22년도 아동, 청소년 피해자는 총 50명으로 전체 피해자의 19.2%를 차지했으나('22.3~'23.3월), '23년도('23.3~'24.3월)에는 총 104명(22.2%)으로 2배 이상 증가했다.

- 이에 따른 아동, 청소년을 대상으로 한 디지털 성범죄 피해지원은 2022년 총 2,026건에서 15,434건으로 7배 이상 증가했다.

- 아동·청소년 디지털 성범죄 피해 유형은 온라인 그루밍 68건(27.5%) 이 대다수를 차지하였으며, 유포·재유포 45건(18.2%), 유포불안 43건(17.4%) 순이었다. 특히 최근에는 불법 사진합성, 남성 청소년 대상 몸캠피싱, 대출 조건 나체사진 전송 등이 증가하고 있다.
- '22년도에는 또래들끼리 불법 사진합성을 해 단톡방 등에 올린 경우가 많았다면, 23년도에는 신원불상의 가해자가 아동·청소년들이 사회관계망서비스(SNS)에 올린 일상 사진에 음란물 등의 성적 이미지와 합성을 해 유포, 판매하는 경우가 증가하고 있다. 또한 미성년자 특성상 대출이 어렵다는 점을 이용, 나체 사진을 보증으로 돈을 대출해주겠다고 제안해 사진을 전송하면 돈은 주지 않고 사진만 유포하는 경우도 증가하고 있다.
- 피해 사례는 연령별로 다른 특성들을 보였는데 ①만 8세 미만의 경우 온라인 게임을 통해 접근해 채팅을 하며 나체사진 등을 요구하는 경우 ②만 8~13세 미만의 경우 오픈채팅방 등을 통해 접근해 서열방, 노예놀이 등을 통해 성적인 사진을 요구하는 경우 ③만 13~19세 미만의 경우 채팅알바, 불법 사진합성 등을 통해 성적인 사진을 요구하는 경우가 많았다.
- 모두 게임, 채팅앱, 사회관계망서비스(SNS) 등 온라인 공간이 가진 익명성을 이용해 접근해 정서적 지지를 해주거나, 원하는 것을 해준다는 조건으로 길들이며 사진이나 영상물을 착취하는 방식으로 범죄를 벌였다. 하지만 피해자들이 사진을 더 보내지 않으면 이를 부모님에게 알리거나 유포하겠다고 협박하며, 더 많은 사진을 요구했다.
- 용돈이 부족해 아르바이트를 구하던 고등학생 A(17세)에겐 '채팅 1건당 70원'이라는 채팅 알바를 제안하며, 이후엔 '사진 1건당 5만 원', '영상통

화 1건당 20만 원' 등 알바비를 준다고 하고 성적인 사진과 영상을 요구했다. 초등학생 B(11세)에겐 오픈채팅방에서 '노예놀이를 하자'며 접근해 노출 사진이나 영상물을 요구했다. 미취학, 저연령의 아동들에게는 온라인 게임을 통해 아이템을 선물하며 성기 사진을 요구했다.

□ 또한, 온라인에 그치는 것이 아니라 그루밍(길들이기)을 통해 오프라인까지 유도해 물리적 성폭력은 12건, 신체적 폭력 2건, 스토킹범죄도 3건에 이르렀다.

○ 이는 사실 모두 치밀하게 계획된 범죄로 연령별로 아동·청소년들이 취약한 특성을 파악해 '온라인 그루밍(길들이기)' 방식으로 아이들을 유인한 후 특정한 자세를 요구하거나 사진을 촬영하게 해 '아동·청소년 성착취물'을 제작·배포하려는 의도를 갖고 접근한 것으로 나타났다. 이후 사회관계망 서비스(SNS)에 사진, 영상물을 판매한다는 홍보글을 올리고 이를 텔레그램에서 판매하는 'n번방' 사건, 아동·청소년 성착취 범죄가 지속적으로 발생하고 있는 상황이다.

○ 실제로 모든 사례에서 아동·청소년 피해자들은 성적인 사진을 전송하고 아르바이트 비용이나 아이템을 선물받는 경우는 거의 없었으며, 대부분 아이들을 길들이고 후 금전이나 선물(아이템, 기프티콘)을 준다고 회유했다.

□ 디지털 성범죄 피해를 입은 아동·청소년, 시민 등은 센터 상담전용 직통번호 '815-0382(영상빨리)' 나 홈페이지(www.8150382.or.kr)로 문의하면 된다.

□ 김선순 서울시 여성가족정책실장은 “우리 사회를 충격에 빠트린 n번방 사건 이후 4년이 흘렀지만 여전히 아동·청소년을 대상으로 한 디지털

성범죄는 사라지지 않고 있고 그 피해도 심각하다”며 “서울시는 작년 전국 최초로 인공지능(AI) 삭제지원 기술을 도입한 데 이어서, 아동·청소년 성착취 영상을 특정할 수 있는 기술을 통해 선제적인 감시·삭제에 나서 아동·청소년이 안전한 서울을 만들어나가겠다.”고 말했다.

붙임 1. 아동·청소년 피해자 지원사례 2부.

2. 아동·청소년 디지털 성범죄 인공지능(AI) 감시 시스템 활용 전/후
3. 아동·청소년 피해 지원 실적(서울 디지털성범죄 안심지원센터)
4. ‘서울 디지털성범죄 안심지원센터’ 1주년 기념 사진

〈붙임 1〉 지원사례

※ 지원사례 1 : 채팅 아르바이트를 통한 피해 촬영물 불법 유포

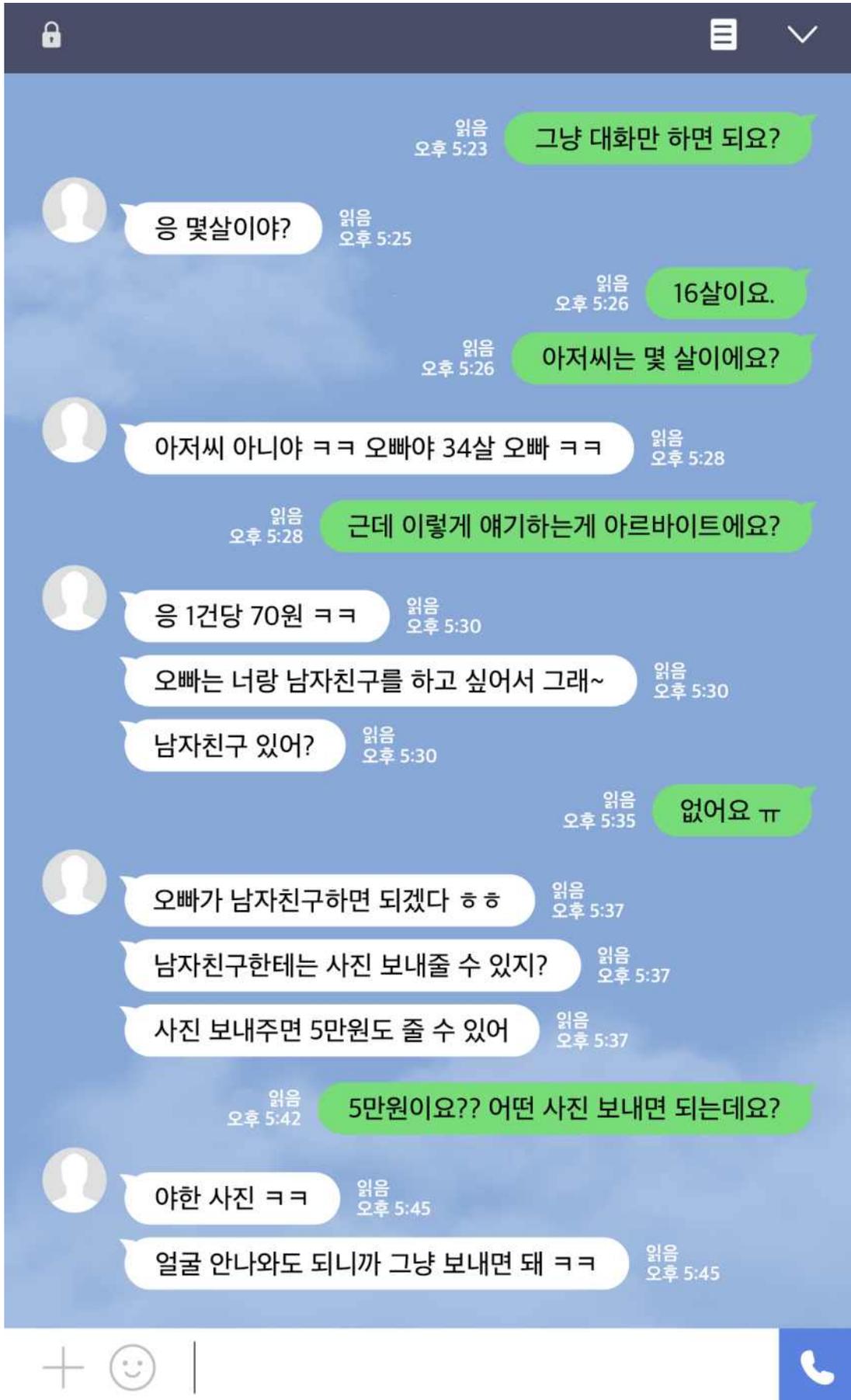
#. 사건 개요

- 유형: 디지털 성범죄(채팅 아르바이트)
- 피해자 연령 : 16세

#. 지원 내용

23년 9월, 용돈이 부족해 아르바이트를 구하던 중학생 A(16세)는 SNS에서 대화 1건당 70원을 준다는 아르바이트 광고를 보게 되었다. 그냥 대화만 하면 된다는 얘기에 처음에는 일상적인 대화를 이어나갔다. 가해자는 중학생인 A가 ‘공부하기 힘들지 않아?’, ‘네가 고민되는 거 다 나에게 얘기해도 돼~’ 라며 호감을 산 후, 목소리를 듣고 싶다며 전화를 요청했다. 이후 ‘목소리가 예뻐서 얼굴도 너무 예쁘겠다. 얼굴 볼 수 있어?’ 라며 사진을 요청했다. A가 사진을 보내자 이후 용돈을 줄테니 야한 사진을 보내라며 속옷을 입은 사진을 요청했고, 이후에는 노출이 더 심한 사진을 요청했다. A가 사진을 보내줄 수 없다고 하자 가해자는 ‘네가 이런 아르바이트 하는 거 부모님이 알아? 부모님에게 다 얘기해줄게’ 라며 A를 협박했고 A는 어쩔 수 없이 사진을 보냈다. 처음에는 아르바이트 비용을 준다던 가해자는 여러 장의 사진만 받고 1달 뒤 잠적을 해 버렸다. A는 사진이 유포되었을까 걱정되어 잠을 못 자고 사진을 검색하다가 센터에 연락해 사진 삭제를 요청하였다. 센터에서는 A의 잘못이 아니라고 얘기해주며 상담을 통해 A가 안정을 찾을 수 있도록 지원하였다. 또한 AI를 통해 사진을 검색해 유포가 된 점을 확인하고 삭제를 지원하였으며, 고소장을 작성해 경찰에 제출하고 A의 심리상담도 함께 지원하고 있다.

(지원사례 1) 대화 예시



※ 지원사례 2 : 서열방, 노예놀이방을 통한 피해 촬영물 불법 유포

#. 사건 개요

- 유형: 디지털 성범죄(노예놀이방, 서열방)
- 피해자 연령 : 15세

#. 지원내용

24년 2월, B(15세)는 열린채팅방에서 ‘신생 서열방’이라며 새로운 인원을 모집한다는 글을 보게 된다. 서열방이란 해당 방에서 계급이 존재하고, 계급에 따라 자신보다 낮은 아래 계급에게 원하는 대로 명령을 할 수 있는 방이었다. 서열방에 참여할 수 있는 인원이 거의 마감되었다는 말에 B는 한번 참여하기로 했다. 방장에게 B가 참여 의사를 밝히자 신체검사(신검)가 필수라며 신체 사진을 보내달라고 했고, B는 고민하다가 얼굴이 나오지 않게 사진을 찍어 보냈다. 기준에 부합했던 B는 생각보다 높은 등급을 받게 되었고 자신보다 먼저 들어와 있던 사람들이 서로 명령을 하고 시키는 대로 행동을 하는 것을 보며 계급에서 떨어지면 자신도 저렇게 시키는 대로 해야 한다는 생각이 들어 계속해서 사진을 보냈다. 하지만 사진 수위를 높이는 방장의 요구에 더 이상 견딜 수 없어 방을 나갔으나 방장은 B의 개인 메시지로 그동안 보냈던 사진을 유포하겠다고 다시 방에 들어올 것을 종용하였다. 서열방에 다시 들어오게 된 B는 계급이 제일 하위로 낮아졌으며, 다른 사람들은 B에게 성적행위나 사진 등을 게시하도록 명령하였다. B는 사진을 보내는 부분이 너무 괴로워 고민을 하던 중 어렵게 용기 내어 학교 상담 선생님께 피해 사실을 상담하였다. 상담교사는 경찰 신고와 함께 센터에 연락을 주었다.

이에 센터 피해지원관과 상담하여 유포 증거를 확보하였고, 변호사 선임을 통해 경찰, 검찰 및 재판 단계에서 조력을 받고 있다. 또한, 일상 회복을 위해 개별 심리상담을 지원하고 있으며, 피해자의 부모님을 대상으로 상담을 진행하고 있다. 또한 서열방에 게시되었던 피해 촬영물은 AI를 통해 유포 모니터링 및 삭제지원을 하고 있다.

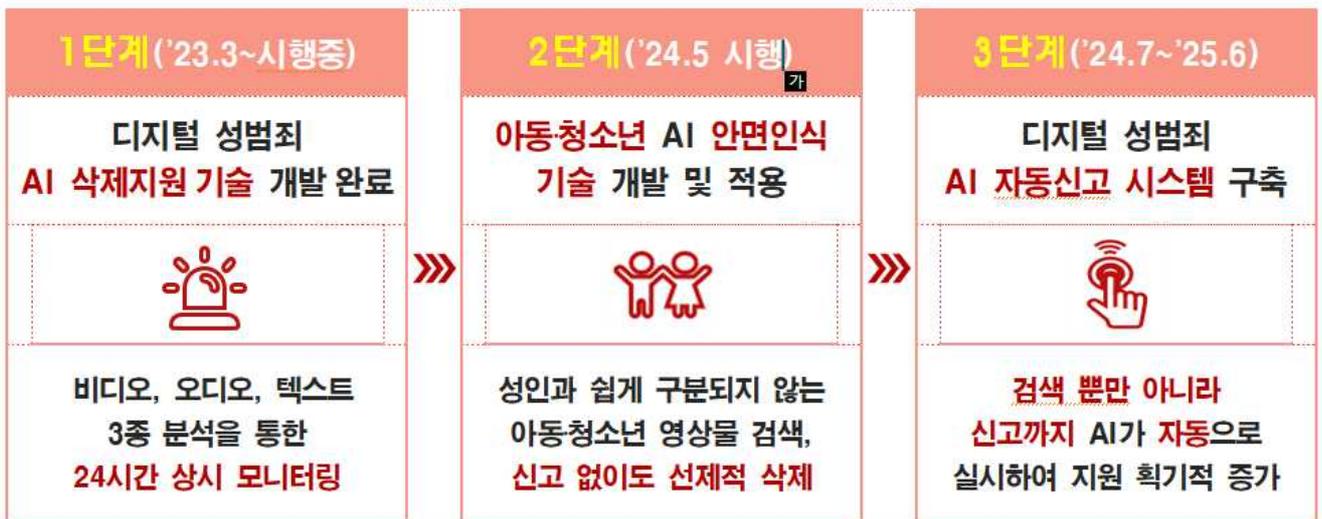
(지원사례 2) 대화 예시



〈붙임 2〉 아동·청소년 디지털 성범죄 AI 감시 시스템 활용 전/후



〈서울시 디지털 성범죄 AI 삭제지원 시스템 구축〉



〈붙임 3〉 아동·청소년 피해 지원 실적(서울 디지털성범죄 안심지원센터)

1) 피해자 수

구분	여성(명)		남성(명)		계
	13세 미만	13세~19세	13세 미만	13세~19세	
계	11	129	-	14	154
2022	3	40	-	7	50
2023	8	89	-	7	104

2) 최초 상담 요청자

구분	최초 상담요청자						계	
	본인	부모/ 법정대리인	가족	수사 사법기관	유관 기관	기타		
계	12	37	2	63	27	13	154	
2022	만13세 미만	-	3	-	-	-	3	
	만13세~19세	4	12	2	7	16	6	47
2023	만13세 미만	-	3	-	3	1	0	7
	만13세~19세	8	19	-	53	10	7	97

3) 연령대별 그 외 폭력 현황(중복)

(단위: 건, %)

구분	물리적성폭력	신체적폭력	정서적폭력	스토킹	계
13세 미만	-	-	1	1	2
	-	-	50.0	50.0	100.0
13세~19세	12	2	2	3	19
	63.2	10.5	10.5	15.8	100.0

4) 가해자와의 관계

구분	가해자와 관계						계	
	애인관계	일시적 만남	채팅상대	친구/지인	모르는 사람	미상		
계	7	8	73	16	8	42	154	
2022	만13세 미만	-	-	2	1	-	-	3
	만13세~19세	1	1	27	7	5	6	47
2023	만13세 미만	-	-	6	1	-	-	7
	만13세~19세	6	7	38	7	3	36	97

〈붙임 4〉 ‘서울 디지털성범죄 안심지원센터’ 1주년 기념식



서울 디지털성범죄 안심지원센터 개관 1주년 기념식('23.3.29.)



디지털 성범죄 AI 삭제지원 프로그램 시연('23.3.29.)