민주사회를위한변호사모임 디지털정보위원회, 사단법인 정보인권연구소, 진보네트워크센터, 참여연대

수 신 각 언론사 정치부·사회부

발 신 참여연대 이지은 간사 02-723-0666 / 진보네트워크센터 김 민 활동가 02-774-4551

제 목 [성명] ‘개발'에만 치중한 AI산업육성, '이루다'는 예정된 참사

날 짜 2021. 1. 13. (총 7 쪽)

성 명

‘개발'에만 치중한 AI산업육성, '이루다'는 예정된 참사

100억건의 개인정보 침해가 빚어낸 차별과 혐오주의자 챗봇

신뢰할 수 있는 인공지능, 기업의 자율규제만으로는 못 이룬다

1. 지난 12월 23일 출시된 스캐터랩의 대화형 챗봇 ‘이루다’를 둘러싼 논란이 한창입니다. 대화형 챗봇 ‘이루다’는 사용자들의 성희롱과 폭언 등의 남용, 혐오표현에 대한 미온적 대응 등 보호받아야 할 사회적 가치를 훼손하는 알고리즘은 물론이고, 개인의 사적인 대화나 개인정보가 심각하게 수집되고 유출된 것으로 드러나 당혹감을 안겨주고 있습니다.

   우리는 마치 이번 논란을 해외시장에서 경쟁하려는 국내 청년 스타트업의 불가피한 시행착오로 포장하려는 일부 언론의 보도에 대해 우려하고 있으며, 엄연히 피해자가 드러난 사안에 대하여 법적 책임을 회피하려는 일체의 시도에 대하여 엄중하게 경고합니다

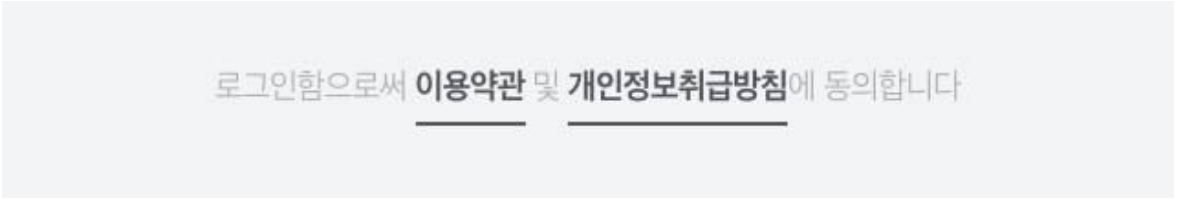
   챗봇 이루다 논란은 기업의 인공지능 제품이 일으킬 수 있는 문제의 한 단면일 뿐이며 이에 대한 대책을 기업 자율에만 맡겨둘 수 없습니다. 이미 우리 생활 곳곳에 들어와 있는 인공지능 제품에 대한 구체적이고 명료한 법적 규범을 마련할 것을 촉구합니다.

2. 대화형 챗봇 ‘이루다’를 개발하고 운영하는 스캐터랩은 2013년 텍스트앳, 2016년 연애의 과학 등의 어플리케이션을 운영하며 카카오톡 등 메신저의 대화 내용을

수집해왔고 이를 자사 다른 제품인 대화형 챗봇 ‘이루다’의 학습용 데이터로 이용했습니다. 「개인정보 보호법」에 따르면 해당 대화 내용 데이터 수집과 이용에 다음과 같은 문제가 있습니다.

개인정보의 목적 외 이용과 부적절한 고지

텍스트앳과 연애의 과학 어플리케이션은 사용자들에게 각각 ‘감정분석서비스’, ‘카톡으로 보는 속마음’ 기능을 제공하며 카카오톡 대화 내용 내보내기 기능을 통해 특정 상대방과 나눈 대화를 전부 수집하였습니다. 스캐터랩은 ‘이루다’가 이용한 연애의 과학 사용자 데이터는 사용자의 사전 동의가 이루어진 개인정보취급방침의 범위 내에서 이용하였다고 주장하였습니다. 그러나 연애의 과학 로그인 페이지에서 “로그인함으로써 개인정보 처리방침에 동의합니다”로 간주하는 것은 각각의 사항을 알리고 명시적으로 동의를 받도록 한 개인정보 보호법 위반입니다(제15조 제2항 또는 제39조의3제1항 및 제22조 위반).



또 카카오톡 대화 수집에 동의한 사용자들에게 해당 비공개 대화가 챗봇 서비스 학습 데이터로 이용된다고 적정하게 고지되었는지 의문입니다. 각 어플리케이션의 개인정보 취급방침에는 다음과 같은 내용만이 존재합니다.

> 라. 신규 서비스 개발 및 마케팅·광고에의 활용
>
> 신규 서비스 개발 및 맞춤 서비스 제공, 통계학적 특성에 따른 서비스 제공 및 광고 게재, 서비스의 유효성 확인, 이벤트 및 광고성 정보 제공 및 참여기회 제공, 접속빈도 파악, 회원의 서비스이용에 대한 통계

현재 수많은 사용자들이 분노하는 것에서 알 수 있듯, 해당 방침을 읽고 본인이 제공한 대화 내용이 챗봇의 학습 데이터로 이용될 거라고 예상한 사람이 어디 있을지 의문입니다. 적법한 동의에 해당하기 위해서는 이용자가 개인정보에 관한 결정권을 충분히 자유롭게 행사할 수 있도록, 통상의 이용자라면 용이하게 법정 고지사항의 구체적 내용을 알아볼 수 있을 정도로 법정 고지사항 전부를 명확하게 게재해야

합니다. 이처럼 정보주체에게 충분히 설명하지 않고 충분히 인지되지 않은 동의는 제대로 된 동의라고 보기 어렵습니다.

한편 '신규 서비스 개발과 마케팅·광고 활용'이 위 유료 어플리케이션을 이용하는데 '필수정보'인지도 의문입니다. 필수정보가 아닌 개인정보 수집·이용에 대해서는 이용자의 선택권을 보장해야 합니다(제16조 제2항과 제3항 또는 제39조의3제3항 위반).

대화 상대방의 동의 부존재

사용자들이 분석을 위해 제공한 카카오톡 대화는 2인 이상의 대화입니다. 개인정보 보호법 제15조 또는 제39조의3제1항에 따라 마땅히 받아야 할 대화 상대방에 대한 동의 절차는 어디에도 없었습니다. 지난 12일 스캐터랩은 논란이 지속되자 서비스를 중지하고 사과의 뜻을 밝혔지만 대화 상대방의 동의 부존재에 대해선 한 마디 언급도 없었습니다(제15조 제1항 또는 제39조의3제1항 위반).

대화 상대방의 동의를 받아 데이터를 수집하는 것은 얼마든지 가능한 일입니다. 2019년 국립국어원에서 진행한 메신저 대화 자료 수집 및 말뭉치 구축 사업의 경우, 메신저 대화를 수집하며 대화 참여자 전원으로부터 대화 제공에 대한 동의와 저작권 이용 허락 등을 받아 데이터셋을 구축했습니다. 이는 정보주체의 권리를 보호함과 동시에 적법하게 데이터를 수집하는 일이 충분히 가능하다는 사실을 보여주고 있습니다.

자신의 사적인 대화 내용이 수집되고 분석되며 이후 챗봇의 학습에 이용되었다는 사실을 인지도 못한 피해자가 무수히 존재할 것이며 이러한 데이터는 원본과 가명 정보 모두 폐기 되어야 마땅합니다.

대화 내용에 포함된 민감정보, 고유식별정보

'텍스트앳', '연애의 과학' 서비스는 카카오톡 등 메신저를 통한 사적인 대화 내용을 수집하는 바, 이에는 이름과 연락처, 주소 등 개인의 신원이 드러날 수 있는 다양한 개인정보와 더불어 더 민감한 다른 정보도 포함되어 있습니다. 자신의 사상, 신념, 정치적 견해, 건강상태, 성생활에 대한 정보, 인종과 민족에 대한 정보 등은 민감정보입니다. 숫자 뿐 아니라 한글과 이미지 등 다른 형태로 표시된 여권번호, 운전면허번호, 외국인등록번호는 고유식별정보입니다. 민감정보와 고유식별정보는 정보주체의 명시적인 동의나 법령상 허용조항 없이는 누구도 수집하거나 이용할 수 없는데 연애의 과학은 이에 대한 별도 동의를 받지 않았으며 실제로 성생활 등에 대한 정보가 수집 이용된 것으로 보입니다(제23조제1항 및 제24조 제1항 위반). 또 숫자 뿐

아니라 한글이나 이미지로 표시된 주민등록번호는 법령상 근거 없이는 민간에서 수집하는 것 자체가 금지되어 있습니다(제24조의2 제1항 위반). 이러한 내용은 스캐터랩과 어플리케이션의 유료 기능의 개인정보 처리방침 어디에도 고지되지 않았습니다.

개인정보에 대한 이용자의 권리 무시

스캐터랩은 이루다가 학습한 대화내용에서 숫자와 영문, 실명 정보 등은 삭제하였기 때문에 개인정보에 대한 비식별화가 충분히 이루어졌다고 주장하고 있습니다. 그러나 지금 이름, 주소 등이 노출되는 사례가 등장하고 있는 것은 이 회사의 비식별화 또는 가명처리의 수준이 적정하지 않았다는 점을 분명히 드러내고 있으며, 스캐터랩은 얼마 전까지 연애의 과학에서 추출된 일부 대화 내용을 오픈소스 플랫폼에 훈련 데이터셋으로 공개하기도 했습니다. 해당 데이터셋의 경우 이름, 건강상태, 직장 등의 개인정보가 비식별화되지 않은 상태였습니다.

무엇보다 눈에 두드러지는 숫자, 영문, 이름 등만 개인정보인 것이 아닙니다. 다른 정보와 쉽게 결합하여 알아볼 수 있으면 개인정보로서 똑같은 보호 대상입니다. 성명은 지웠지만 여성이고, 이십대라는 점을 알고 있으며, 숫자와 영문은 삭제되었지만 서울 성북구에 살고, 특정은행에 계좌가 있고, 특정대학교 특정학과 또는 특정 기숙사를 출입한다면 개인을 특정할 수 있습니다. 이처럼 개인정보를 부분적으로 삭제하여 가명처리를 했다 하더라도 더 많이 결합할수록 알아볼 가능성이 더 높아진다는 것이 상식입니다. 이러한 개인정보들을 언제 누구에게 어느 범위까지 알려지고 또 이용되도록 할 것인지를 그 정보주체가 스스로 결정할 수 있는 권리는 우리 헌법에서 보호하고 있는 기본권입니다.

개인정보 침해를 당했음에도 불구하고 그 사실조차 모르는 정보주체가 다수 존재할 것입니다. 지금이라도 가입자 뿐 아니라 자신의 대화가 수집이용된 모든 정보주체의 열람 및 삭제 권리는 완전하게 행사될 수 있어야 합니다. 또 개인정보 수집과 처리 과정이 불법적인 것으로 드러나면 정보주체의 요청 없이도 해당 개인정보를 바탕으로 만들어진 챗봇 모델과 알고리즘의 폐기가 마땅합니다.

3. 이번 이루다 논란은 기업을 위한 데이터3법이 자초한 문제이기도 합니다. 시민사회는 데이터3법이 기업들로 하여금 개인정보를 가명처리한 후 정보주체의 동의 없이 인공지능 제품 등 기업의 상품개발에 거의 무한대로 이용할 수 있도록 허용한 데 대하여 비판해 왔습니다. 공익적 학술 연구 등에 정보주체의 동의 없이 가명정보를 사용할 수 있도록 하는 유럽연합 등의 사례가 있지만, 기업의 상업적인 제품 개발을 위해 정보주체의 기본권을 무시하며 가명정보를 무제한적으로 이용하는 것은

위헌적입니다. 정보주체의 권리를 보장하기 위해서는 현 가명정보에 면제한 열람권, 삭제권 등을 보장하는 방향으로 법개정이 있어야 할 것입니다. 무엇보다 새로운 사회적 흐름으로 등장한 인공지능 제품에 대한 명확한 법규범은 필수입니다.

그런데 최근 정부와 기업이 오히려 데이터 산업의 발전을 명분으로 현재의 사전 동의 조항을 더 완화하는 법 개정을 추진하고 있다는 데 대하여 우리는 깊이 우려합니다.

4. 스캐터랩은 향후 '더 고도화된 데이터 알고리즘'으로 이 문제를 해결하고 향후 재운영하겠다는 방침을 밝혔습니다. 또 자신들이 마음대로 사용한 이용자의 원본 데이터, 가명 데이터를 완전 파기하겠다는 약속은 하지 않는 모습입니다

   그러나 바로 며칠전, 미국 연방거래위원회(FTC)는 불법적으로 사진을 수집해 얼굴인식 알고리즘 훈련용 데이터로 이용한 기업에게 해당 모델과 알고리즘을 삭제하라는 명령을 내렸다는 사실을 상기할 필요가 있습니다. 지금은 챗봇 문제로 시작했지만 더 위험한 인공지능에서 더 위험한 데이터, 더 위험한 알고리즘을 사용하기 전에 우리 사회가 이런 문제에 대한 대책을 고민해야 할 시점입니다.

   윤리적 검토와 평가, 사회적 합의와 토론 없이 성급하게 개발되는 인공지능 제품은 국민에게 차별과 오류, 개인정보 침해 등 심각한 악영향을 가져올 수 있으며 동시에 블랙박스 속에 감춰진 알고리즘으로 인해 피해자가 권리 침해 사실조차 깨닫지 못할 수도 있습니다.

5. 기업의 인공지능 제품의 악영향에 대한 대책이 인공지능 윤리로 그쳐서는 안됩니다. 지난 12월 발표된 과학기술정보통신부의 '인공지능(AI)윤리기준' 에는 인권을 보장하고 프라이버시를 보호하고 다양성을 존중하고 책임성을 보장하는 등 선하고 아름다운 문장들이 가득하지만, 그것을 구체적으로 실현할 수 있는 방안은 없는 것으로 보입니다. 특히 아래와 같은 문장을 명시함으로써 인공지능 윤리 가이드라인이 현실적 규범으로서는 아무런 의미도 없는 선언에 불과하다는 사실을 스스로 입증하였습니다.

   > *"본 윤리기준은 산업·경제 분야의 자율규제 환경을 조성함으로써 인공지능 연구개발과 산업 성장을 제약하지 않고, 정당한 이윤을 추구하는 기업에 부당한 부담을 지우지 않는 것을 목표로 한다."*

   온갖 미사여구와 좋은 말이 가득하지만 구속력 있는 '법'이나 '제도'가 아니라 단순한 자율 규범으로, 아무런 강제성이 존재하지 않는 가이드라인으로 인공지능 제품의 악영향을 막을 수 없을 것입니다.

6. 인공지능 제품과 그 학습 데이터의 편향성의 문제는 윤리가 아니라 법률 규범의 문제입니다. 세계 각국은 인공지능 제품과 그 학습 데이터에 대하여 제조물 책임법, 소비자 보호법, 정보공개법, 개인정보 보호법, 평등법 등 현행법을 정교하게 적용하기 위해 노력하고 있습니다. 특히 유럽연합의 경우 인공지능 제품과 서비스에 대해 기존 법률의 준수는 물론이고, 추가적으로 '고위험' 인공지능에 해당할 경우 훈련데이터, 기록보존, 정보공개, 견고성, 인적 감독 등의 법적 의무를 추진하고 있습니다. 이때 공공적으로 법적인 효과를 낳거나 의료, 운수, 에너지 분야에서 중대한 손상을 야기하거나 노동자나 소비자의 권리에 영향을 미치는 인공지능 제품과 서비스는 '고위험'에 해당합니다. 이중 훈련 데이터에 대한 법적 의무로는 △충분히 광범위한 데이터셋에 기반해 훈련하는 등 그 안전을 합리적으로 보장할 것 △금지된 차별을 수반하는 결과로 이어지지 않도록 합리적인 조치를 취할 것 △제품과 서비스를 사용하는 동안 사생활과 개인정보를 적절히 보호할 것 등을 요구할 예정입니다.

   캐나다는 이미 2019년부터 공공기관 자동화된 의사결정에 대한 지침(정부 훈령)을 제정하고 모든 공공기관 인공지능 알고리즘에 대하여 영향평가를 실시하고, 생산 전에 훈련 데이터가 의도하지 않은 데이터 편향을 드러내거나 결과에 부당하게 영향을 미칠 수 있는 요소가 있는지 검사 절차를 거치고 있습니다. 심지어 자율규제의 나라 미국에서도, 법 집행기관의 인공지능 사용과 채용을 위한 인공지능 사용 등 특정 영역을 강하게 규제하는 법이 제정되고 있습니다.

7. '이루다'는 챗봇이고, 고위험 인공지능이나 공공기관 인공지능은 아닐지 모릅니다. 그러나 막연하고 기업 자율적인 인공지능 윤리에서 더 나아가 이제는 국민의 안전과 기본권을 보호할 수 있는 인공지능 규제법을 고민해야 할 때입니다. 개인정보를 보호해야 할 대상이 아닌 경제적 관점과 활용 대상으로만 바라보는 기업의 관행은 국민이 신뢰할 수 없는 인공지능 기술로 이어질 것입니다. 개인정보 자기결정권을 침해하고, 의도하지 않은 결과를 도출하고, 혐오를 양산하고, 소수자를 배제하고, 편견을 재생산하는 인공지능 기술의 도입을 방지할 방법이 현실적으로 고려돼야 합니다. 지금 이 순간에도 사회 다양한 영역에서 인공지능 제품과 서비스가 개발되고 활용되고 도입되고 있습니다. 그중 일부는 챗봇 정도가 아니라 채용, 사회 복지 수급자 선정, 일자리 배치 등 국민의 삶에 중대한 영향을 미치는 영역에 자리를 잡았습니다. 이러한 상황에서 기업들의 인공지능 윤리 준수를 선의에만 기댈 수는 없습니다. 국민에게는 인공지능 기술을 사회적으로 통제하기 위한 현실적인 법이 필요합니다.

8. 이 문제적 인공지능 제품 '이루다'의 이면에서 개인정보에 대한 심각한 권리 침해가 발생한 것으로 보이며, 우리는 이에 대한 개인정보 보호위원회의 철저한 조사와 엄정한 처벌을 촉구합니다. 우리 단체들은 개인정보 보호위원회에 대한 의견, 민원 뿐 아니라

국가인권위원회 등 관련 부처에 대한 민원, 필요하다면 관련 소송 등으로 계속하여 이 사안에 공동으로 대응할 예정입니다. 나아가 정부와 국회의 관련 정책 입안자들에 대하여 소비자와 이용자의 권리를 침해하는 인공지능 제품의 규제를 위해 법률적 대책을 마련할 것을 촉구합니다.

2021년 1월 13일

민주사회를위한변호사모임 디지털정보위원회, 사단법인 정보인권연구소, 진보네트워크센터, 참여연대