내일을 여는 국민의 국회

제1799호

발행일 2021년 2월 15일
발행처 국회입법조사처
발행인 김만흠 국회입법조사처장
www.nars.go.kr

이슈와 논점

‘이루다’를 통해 살펴본 인공지능 활용의 쟁점과 과제

신 용 우*·정 준 화**

인공지능 챗봇 ‘이루다’가 이용자들과 나눈 일부 대화에서 차별·혐오적 표현이 나타났다. 이는 인공지능 윤리, 학습데이터의 개인정보 보호 법제 준수 등에 대한 논의를 불러일으켰다. 결국 출시 20일 만에 중단되고 만 이루다가 우리나라 인공지능 분야에 남긴 과제는 개인정보 처리에 대한 사전동의와 사후통제의 조화, 가명처리의 개선, 인공지능 윤리기준의 구체화, 적정 수준의 학습데이터 확보 등이다.

1 들어가며

최근 논란의 대상이 된 ‘이루다’는 인공지능 챗봇(AI chatbot)[1] 서비스, 즉 인간을 모방하여 대화를 생성하는 컴퓨터 시스템이다. 이루다의 개발자는 이용자들이 이루다와 편리하고 친숙하게 대화할 수 있도록 페이스북 메신저에서 이루다를 친구로 추가할 수 있도록 했고, 이루다에게 20세 여대생이라는 친근한 가상 프로필을 부여했다. 또한 사람들의 실제 대화를 대량으로 기계학습시켜서 이루다가 특정 전문분야에 한정된 것이 아니라 일상의 모든 주제에 대해서 대화할 수 있도록 만들었다.

이러한 특징들은 이루다와 기존의 챗봇과의 차이점이지만 동시에 많은 논란을 초래한 원인이기도 하다. ‘20세 여대생 이루다’에 대한 일부 이용자들의 대화 경험이 인터넷에 공유되는 과정에서 이루다가 사용한 차별·혐오 표현들이 부각되었다. 그 원인을 찾는 과정에서 기계학습에 활용된 데이터가 개인정보 보호 법제를 위반했다는 의견이 확산되었다. 결국 이루다에 대한 인식은 급격히 악화되었고, 개발자의 해명에도 불구하고 출시 20일 만에 서비스를 전면 중지하게 되었다.

이루다 사례는 우리 삶에서 빠르게 확산되고 있는 인공지능이 초래하게 될 불완전성과 위험의 단면을 분명하게 보여준다. 이번 사건을 계기로 인공지능이 안정적으로 활용될 수 있는 기반을 다질 필요가 있다. 이 글에서는 이루다가 촉발한 쟁점을 살펴보고 향후 안전한 인공지능의 개발과 활용 확대를 위한 과제를 제안하고자 한다.

2 인공지능 챗봇 ‘이루다’

(1) 인공지능 챗봇의 개요

인공지능 챗봇은 상대방이 입력한 문자 또는 음

1) AI 챗봇을 대화형 인공지능(conversational AI)으로 표현하기도 한다.

* 사회문화조사실 과학방송통신팀 입법조사관, 02-6788-4714, yongwoos@assembly.go.kr
** 사회문화조사실 과학방송통신팀 입법조사관, 02-6788-4715, joonhwa@assembly.go.kr

성을 인공지능 알고리즘으로 해석한 다음 가장 적절한 정보 또는 표현을 문자나 음성으로 제공하여 마치 인간과 대화하는 것과 같은 경험을 제공해 주는 컴퓨터 시스템, 즉 채팅 로봇[2]을 의미한다.

인공지능 챗봇은 다양한 형태로 우리 일상에 활용되고 있다. 시리(Siri)나 빅스비(Bixby) 같은 스마트폰 기반의 음성인식 챗봇뿐만 아니라, 온라인 쇼핑에서 예약·주문·결제·송금 등을 처리해 주는 전문 분야 인공지능 챗봇 등이 대표적이다. 공공부문에서도 법무부의 생활법률정보 챗봇 '버비'와 대구시의 민원상담 챗봇 '뚜봇' 등이 있으며, 앞으로 여러 서비스가 하나의 챗봇으로 통합될 계획이다.[3]

실생활의 인공지능 챗봇의 활용이 많아지면서 산업 규모도 크게 성장 중이다. 세계 챗봇 시장 규모는 2019년 25억 7,120만 달러에서 2024년 94억 2,790만 달러로 연평균 29.7% 성장할 전망이다.[4]

(2) '이루다'에 대한 논란

이루다는 사람들이 언제나 편안하게 일상적인 대화를 나눌 수 있는 채팅 서비스를 제공하는 인공지능 챗봇이다. 특히 젊은 연인들 사이의 대화를 대량으로 기계학습했기 때문에 10~30대가 실제 사용하는 표현을 능숙하게 모방한다.

짧은 기간 동안 많은 이용자들의 관심을 모았지만, 개발자가 어느 정도 예상했음에도 불구하고 충분히 대비하지 못한 부분에서 문제가 발생했다. 그 촉발점은 혐오와 차별이다. 이루다는 성 소수자를 어떻게 생각하느냐는 질문에 싫어한다고 답하기도, 그렇지 않다고 답하기도 했다. 이러한 문제의 근본 원인은 이루다의 기계학습에 투입된 학습데이터의 편향성이다. 그래서 어떠한 학습데이터가 활용되었는지 살펴보는 과정에서, 데이터의 편향성보다는 데이터 수집·활용의 부적절함이 더 심각한 문제로 나타났다. 개발자는 이루다의 자연스러운 대화를 위해 자사가 제공하는 다른 서비스에서 수집된 개인정보(연인 간 메신저 대화 내용 등)를 활용했는데, 많은 사람들이 자신의 개인정보가 해당 서비스의 개선을 넘어 이루다의 개발에까지 사용될지는 몰랐던 것이다. 결국 이루다의 논란은 개인정보 침해로 귀결되고 있다.

이루다의 개발자는 2021년 1월 8일에 공식 입장을 발표했지만, 논란이 사그러들지 않자 1월 11일 입장문을 발표하고 다음 날인 12일부터 서비스를 전면 중단했다. 그러나 부정적인 여론에 대한 충분한 소명은 되지 못했다. 현재 일부 사용자들이 공동소송을 준비하고, 개인정보보호위원회도 현장조사와 후속조치를 추진하고 있다.

(3) 외국 사례

외국에서도 인공지능 챗봇의 혐오·차별 표현이 종종 이슈가 되었다. 편견·혐오적 메시지를 출력해서 공개된 지 16시간 만에 서비스를 전면 중단했던 미국 마이크로소프트의 테이(Tay)가 대표 사례다. 2019년에는 애플이 골드만삭스와 출시한 신용카드의 신용한도 알고리즘이 여성보다 남성을 우대한다는 지적이 제기되어 금융당국이 조사하기도 했다.[5] 최근에는 페이스북과 판도라보츠의 AI 챗봇끼리 온라인 채팅 배틀을 하는 과정에서 히틀러를 '위대한 사람'이라고 표현한 일도 있었다.[6]

이러한 상황에서도 외국의 챗봇은 빠르게 발전하고 있다. 여러 논란들이 챗봇의 활용을 잠시 멈추게 했지만, 큰 틀에서는 챗봇의 문제를 확인하고 개선하는 과정으로 활용되었기 때문이다.

2) 특히 온라인 공간에서 인간의 직접적인 개입 없이 자동적으로 작동하는 컴퓨터 시스템 또는 프로그램을 '봇(bot)'이라고 부르기도 한다.

3) 행정안전부. 「보도자료 – 정부 민원상담 챗봇 서비스, 이제는 한 곳에서」. 2020.5.13.

4) 연구개발특구진흥재단. 「챗봇 시장」. Marketsandmarkets(2019)의 「Chatbot Market」 인용. 2020.

5) 김청중. 「美·유럽·日서도 'AI 윤리' 도마에… 각국 앞다퉈 규제 강화나서」. 『세계일보』. 2020.1.19.

6) 윤영주. 「"제가 더 사람 같죠?"…대화형 AI 챗봇끼리 한판 대결」. 『AI타임즈』. 2020.11.2.

3 주요 쟁점

(1) 개인정보 활용에 대한 사전동의

이루다의 핵심 쟁점은 개발자의 「개인정보 보호법」 위반 여부이다. 이루다 개발에 사용된 개인정보는 해당 회사의 다른 앱 '연애의 과학'에서 신규 서비스 개발 및 마케팅·광고에 활용될 것으로 고지하고 수집되었는데, 이를 이루다 개발에 이용한 것은 애초에 정해진 수집 목적 범위를 넘어선 것으로 「개인정보 보호법」 제15조, 제17조 및 제39조의3 위반이라는 지적이 있다. 또한 해당 앱의 서비스 제공 목적 외에 신규 서비스 개발 및 마케팅·광고 활용에 관한 사항을 별도로 구분하여 동의받지 않아 동법 제22조 위반이라는 주장도 있다.

이루다의 개발자가 정보주체로부터 사전동의를 받았음에도 불구하고 이러한 논란을 피하지 못하는 주된 이유는 그 동의가 형식적이었기 때문이다. 개인정보 수집 시 정보주체에게 너무 많은 조건과 설명을 제공하여 구체적인 내용을 인지하기 어렵게 하거나, 그 내용을 변경할 실질적인 협상력을 주지 않아서 모든 항목에 형식적으로 사전동의하게 만들었다. 이러한 형식적인 사전동의는 정보주체의 자기결정권 보장이 아니라 개인정보처리자의 개인정보 수집 정당화의 수단으로 사용되기도 한다.[7)]

(2) 개인정보의 가명처리

2020년 8월 시행된 개정 「개인정보 보호법」은 특정 개인을 알아볼 수 없도록 가명처리를 한 개인정보를 '가명정보'라고 정의하고 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 처리할 수 있도록 하였다.

문제는 이루다의 대화 중에 실명, 주소 등이 노출됨에 따라 개발 과정에서 가명처리가 적절히 이루어졌는지 논란이 된다는 점이다. 개발자는 개별 문장 단위에서 비식별화 조치를 했으나 일상 대화의 문맥에 따라 이름 등이 남게 되었다고 밝혔다.

이처럼 일상 대화나 영상과 같은 비정형 데이터에 대한 가명처리가 충분히 이루어지지 못한 것은 개발자의 대응 부족이 원인이지만, 비정형 데이터에 대한 가명처리 기준 미흡도 영향을 미쳤을 수 있다. 개인정보보호위원회가 가명처리로 인한 정보주체 재식별 위험을 줄이기 위해 2020년 9월 「가명정보 처리 가이드라인」을 발표하지만, 비정형 데이터의 가명처리 기준으로 사용하기에는 부족하다.

(3) 인공지능 윤리

이루다가 나타낸 윤리적 차원의 문제점은 학습데이터에 포함된 사회적 편견 또는 알고리즘 편향성이 고스란히 대화로 표출되었다는 것이다. 또한 인공지능 기술의 내재적 불확실성과 불투명성으로 인하여 사전에 완벽한 학습이 어렵다는 문제도 있다. 나아가, 이용자에 의해 혐오발언이 학습되었는데 이를 적절히 필터링하지 못했다는 지적이 있다.

정부는 2020년 12월 「인공지능(AI) 윤리기준」을 발표하면서 인간 존엄성, 사회의 공공선, 기술의 합목적성 등 원칙을 제시하였는데, 실제 적용하기에는 다소 추상적이고 선언적이라는 한계가 있다.

4 향후 과제

(1) 개인정보 보호 법제 개선

개인정보처리자가 사전동의를 받기 위해 정보주체에게 제공하는 조건과 설명을 단순화·실질화하여 사전동의의 실효성을 높이고, 사후통제를 강화할 필요가 있다. 특히 인공지능은 웹크롤링·사물인터넷과 같은 신기술을 이용하여 사람의 개입 없이 자동으로 개인정보를 수집·활용할 수 있기 때문에 사전동의와 사후통제의 적절한 병행이 중요하다.

7) 권영준. 「개인정보 자기결정권과 동의 제도에 대한 고찰」. 『법학논총』 제36권 제1호. 2016.

이와 관련하여 개인정보보호위원회가 2021년 1월 입법예고한 「개인정보 보호법 일부개정법률안」에서 해당 기관의 직권 또는 민간단체의 청구에 의해 개인정보 처리방침이 법률을 위반하는지를 심사할 수 있도록 했는데, 이것이 적절한 사후통제 장치가 될 수 있는지를 검토할 필요가 있다.

가명처리의 경우 일상 대화, 영상과 같은 비정형 데이터의 재식별 위험성을 평가하고 방지할 수 있도록 가이드라인을 개선하고 기술·방법론에 관한 연구를 추진할 필요가 있다.[8)]

(2) 인공지능 윤리기준 구체화

인공지능이 견고함과 신뢰성을 갖출 때 기술에 대한 사람들의 수용성이 높아져 기술발전과 산업 활용이 촉진될 수 있다. 이를 위해 현재 다소 추상적이고 선언적인 인공지능 윤리기준을 보다 구체화하고 검증 가능한 형태로 발전시킬 필요가 있다.

외국의 경우 구체적인 법·제도적 조치들이 활발하게 추진되고 있다. 미국 의회는 2019년 4월 「알고리즘 책임 법안」을 발의하여 고위험 자동화 시스템을 평가하는 규칙을 만들고 알고리즘 편향성·차별성, 프라이버시·보안 위험 등을 점검하도록 하였다.[9)] 뉴욕시 의회도 2019년 뉴욕시의 알고리즘 사용에 편향성이 있는지 점검하는 기구를 설립하였다. 연방거래위원회(FTC)는 2020년 4월 「AI와 알고리즘 사용에 대한 지침」을 발표하였다.[10)]

유럽 집행위원회는 2020년 3월 발표한 「인공지능 발전과 신뢰를 위한 백서」에서 고위험 분야의 인공지능에 대하여 안전성 요건을 수립하고 사전 적합성 평가를 받도록 하는 내용을 담았다.[11)]

미국·유럽 등의 입법·정책을 참고하여 고위험 분야에서는 사전 점검 체계를, 그 외의 분야에서는 자율 규제 또는 품질 인증 체계 도입을 검토해 볼 수 있을 것이다. 사전 점검의 방안으로는 학습데이터 관리, 투명한 정보 제공, 인간의 개입 등 실효성과 집행가능성 있는 기준들을 마련할 필요가 있다.

(3) 학습용 데이터 확보

인공지능의 경쟁력은 데이터에 있다. 그러나 많은 중소기업·스타트업은 충분한 데이터를 확보하기 어렵고, 적절한 가명처리와 데이터 정제를 할 여력도 충분하지 않다. 이러한 상황이 지속될 경우 제2의 이루다 사태는 불가피하다.

따라서 인공지능 학습데이터 확보를 위한 정부의 노력을 강화해야 한다. 전체 인공지능 생태계에서 필요한 학습데이터 수요를 파악하고, 현재 추진하고 있는 데이터 댐과 데이터 바우처 지원사업의 치우침과 부족함을 면밀히 검토한 다음 필요한 분야에 적정 수준의 학습데이터를 제공할 수 있는 방안을 마련할 필요가 있다.

5 나가며

이루다 사태는 막연하고 추상적이었던 인공지능 활용에 대해서 구체적인 질문을 남겼다. 현재는 관계 기관의 조사와 일부 사용자의 법률분쟁으로 인해 위축된 상황이지만, 이번 사태가 우리나라 인공지능 산업의 걸림돌이 아니라 도약을 위한 디딤돌이 될 수 있도록 해야 한다. 이제부터 인공지능 생태계를 둘러싼 다양한 이해관계자들이 공감하고 동의할 수 있는 구체적이고 합리적인 법·제도를 만들어 가는 노력을 시작해야 할 것이다.

『이슈와 논점』은 국회의원의 입법활동을 지원하기 위해 최신 국내외 동향 및 현안에 대해 수시로 발간하는 정보 소식지입니다.

8) 김병필. 「가명처리 및 비식별 조치 방법론 심층 케이스 스터디」. 개인정보전문가협회 발표문. 2021.1.

9) 미국 의회 홈페이지 〈https://www.congress.gov/bill/116th-congress/house-bill/2231/all-info 최종 방문일 2021.1.28.〉

10) 양기문. 「기업의 AI 및 알고리즘 이용에 관한 지침」. 『정보통신방송정책』 제32권 제4호. 2020.

11) European Commission. “White Paper on Artificial Intelligence: a European approach to excellence and trust”. 2020.

07233 서울시 영등포구 의사당대로 1 국회입법조사처
Tel. 02-6788-4510
발간등록번호 31-9735044-001340-14
ISSN 2005-744X
국회입법조사처