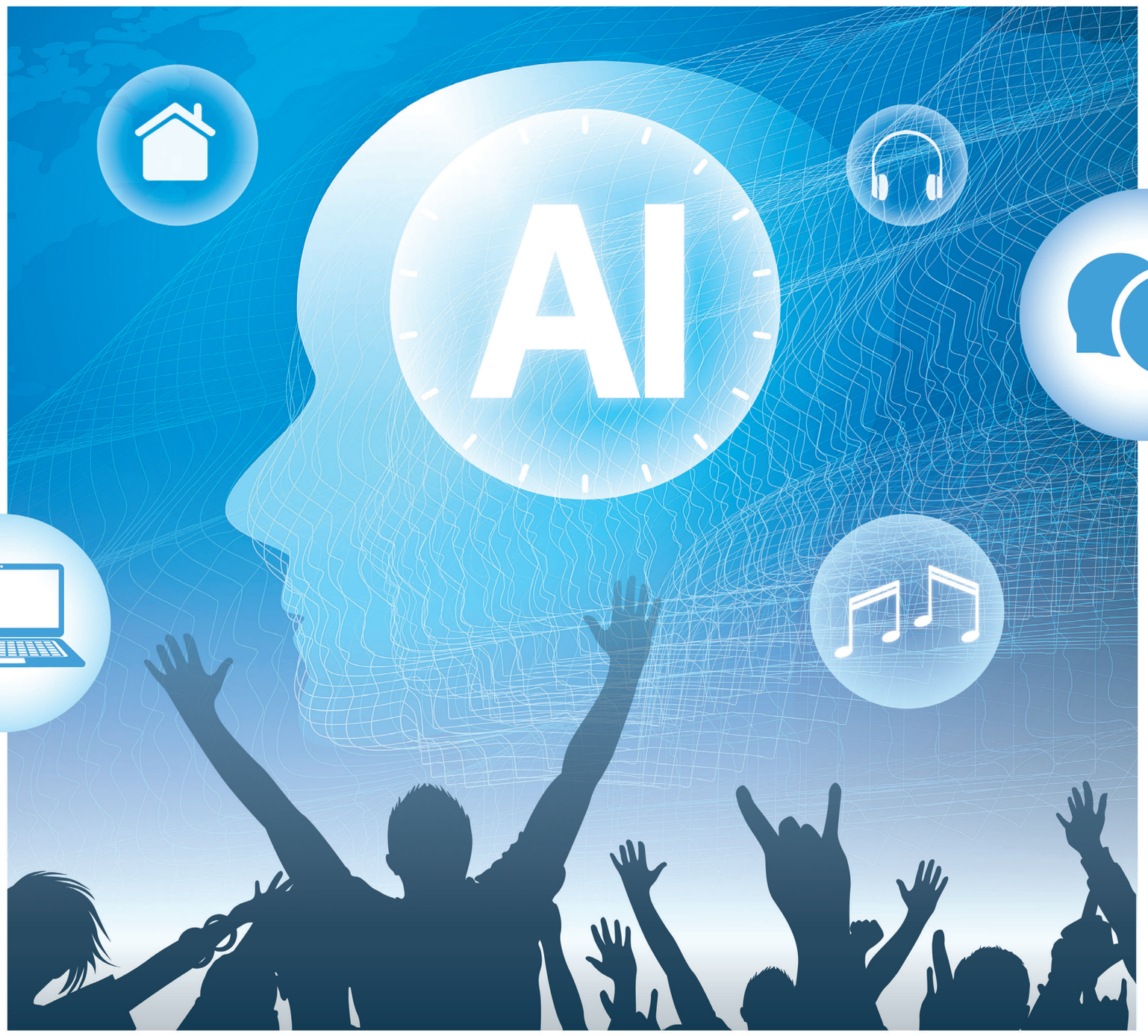
이준기의 빅데이터

내 마음까지 읽는 AI, 나치 같은 비이성적 광기 세뇌 우려

이준기
연세대 정보대학원 교수



디지털 세계에 사는 우리는 하루에도 수십 번 이상 권유와 설득에 마주친다. 온라인 상거래는 나의 기존 상품 구매 이력을 바탕으로 살만한 상품군을 보여준다.

유튜브 동영상은 한번 보았던 관심사를 바탕으로 끝없이 이어지는 비슷한 동영상을 보여준다. 쇼츠에 한번 빠지면 끝 간데없는 시간을 보내기가 일쑤다. 이 밖에도 길 안내, 택시 배차 등도 사실은 추천이란 이름의 설득이다.

소크라테스, 플라톤 등이 설파하던 고대 그리스의 학문이 사실은 모두 설득의 기술에 관한 것이라고 했던가. 설득의 기술은 현대 심리학, 철학, 정치학, 광고학 등에서 가장 중요한 분야 중의 하나다. 설득의 기술이 중요한 것은 설득이 우리가 이해하기 어려운 비이성적 형태를 띠고 있기 때문이다.

보험왕, 방문 판매왕의 설득은 단순하게 상품의 우수성에 기인하지 않는다. 흔히 말하는 스톡홀름 증후군 그리고 관동대지진에서의 조선인·중국인 학살, 나치 시대의 독일국민들, 사이비 종교의 집단 자살과 테러 등은 어떻게 인간이 강압적이지 않은 자율상태에서 저렇게 설득당할 수 있는가에 대한 깊은 고민을 남긴다.

데이터 통해 많이 알수록 설득 쉬워져

인공지능(AI)의 인류에 대한 위협 중 가장 현실적이고 당장에 다가올 위협이 무엇일까. 얼마 전 오픈AI 제이슨 권 최고전략책임자는 오픈AI에서 보는 인공지능의 네 가지 재앙적 위험을 언급하며 그 첫째로 '극단적 설득'을 꼽았다.

그가 자세히 언급하지는 않았지만 딥페이크, 데이터 수집, 분석 기술 등이 심리학과 연계될 때 가공할 만한 위협이 될 수 있음은 확실해 보인다. 불과 1~2분 분량의 연설문이나 목소리 녹음을 통하여 똑같은 목소리를 만들 수 있는 기술과 몇 마디 명령어로 만들고 싶은 그림이나 동영상을 만들어 내는 것이 딥페이크 기술이다.

이 기술로 우리는 정치에 개입할 수 있고 사람들의 마음을 움직일 수 있다. 또한 데이터를 통해 상대의 상태를 더욱 많이 알게 될 때 설득은 더욱 쉬워진다.

'극단적 설득'이라는 기술의 도래는 이런 염려가 단순 걱정이 아닌 우리가 진정으로 고민해야 할 실질적 이슈임을 선언하고 있다.

최근 영국에서 일어나고 있는 극우 폭력 사태를 보자. 7월 말 어린이 댄스 교실에 침입하여 흉기를 휘둘러 어린이 3명이 숨지고 10명이 다친 사건이다.

소셜미디어(SNS)상에서는 범인의 이름이 무슬림식 이름이라는 거짓 뉴스가 퍼지기 시작하였고, 그동안 누적된 반이민 정서가 폭발하면서 사태는 걷잡을 수 없이 번져 나가고 있다.

이런 딥페이크 기술이 데이터 기술과 합쳐지면 큰 파괴력은 상상을 초월할 수도 있다. 이것도 벌써 교훈이 되는 사례가 있다. 지금으로부터 8년 전 트럼프가 처음 대선에 나왔을 때 경쟁 상대는 힐러리 클린턴 전 국무장관이었다.

당시 페이스북(지금의 메타)에서는 성격 테스트를 무료로 해주는 앱이 실행되었다. 이 앱을 재미로 다운로드하고 실행한 사람들에 대하여 성격이 5가지 특성(개방성, 성실성, 외향성, 친화성, 신경증)으로 분석되었으며, 개인의 성격뿐 아니라 정치적 성향이 세밀하게 분석되었다.

스마트폰 통해 일상 정보 대거 수집
맞춤형 메시지 보내 특정 충동 자극
제품 사게 만들고 투표 포기 유도도

딥페이크로 정교한 가짜뉴스 유포
영국 극우 폭력 사태 같은 혼란 야기
AI 설득 기술 올바른 활용법 찾아야

또한 이 앱을 시행한 사람들은 부지불식간 본인들이 어떤 것에 대하여 "좋아요"를 누르고 있는 것에 대한 정보를 앱에서 가져가는 것을 허용하였다. 트럼프를 지지하는 한 부호의 후원으로 시작된 이 프로젝트에서 영국의 캠브리지 애널리티카라는 기업은 획득된 정보를 바탕으로 개인화된 광고 메시지를 각 개인에게 전달하였다.

예를 들어 힐러리를 지지하지만 페미니즘 성격이 있는 개인에게는 빌 클린턴 대통령의 백악관 인턴과의 불륜에 관한 내용이 전달되어 힐러리를 지지하는 마음을 멈추게 하였다. (실제로 이런 개인적 문자를 받은 계층은 투표를 포기하는 경향이 강했다) 이런 데이터는 무려 7600만 명의 미국인을 대상으로 수집됐다.

이런 사실을 처음 들은 사람들은 어떻게 이런 일이 공공연하게 일어나고 있는가에 대한 걱정과 분노를 보일 수 있다. 하지만 이런 일의 생성과 파급 그리고 폭로는 우리가 생각하는 것보다 훨씬 복잡한 경로를 밟는다.

우선 대중에 알려지기 위해서는 여러 가지 운 좋은 일들이 연속적으로 발생해야 한다. 특히 이런 일들이 기업에서 비밀리에 진행되고 있다면 모르고 지나갈 확률이 폭로될 확률보다 훨씬 크다.

더구나 이것의 법적 제재 영역이 모호한 영역에 들어갈 때는 시행하는 사람조차 이것이 큰 문제가 되고 있다는 것을 애써 나타내지 않으려는 경향이 강해진다.

AI가 인간을 설득할 확률 81%P 높아

이런 기술은 종종 우리에게 편리함과 새로운 서비스란 서로 다른 측면의 이율배반적인 결과를 선사한다. 만약 내가 일반적인 갤럭시폰을 사용하는 사람에게 당신이 스마트폰을 갖고 다니는 동안 어디를 방문하고 거기서 몇 분 머물렀다는 기록이 모두 구글에 저장되고 있다는 사실을 말한다면 처음에는 반신반의하는 사람들이 많다.

하지만 이것은 타임라인이라는 구글의 애플리케이션을 보면 사실이다. 사실 이것뿐 아니라 당신의 캘린더를 통하여 누구와 언제 약속을 하는가, 어떤 이메일이 주고받아지는가 등의 정보는 모두 구글에 저장이 된다.

그냥 저장만 할 뿐이라고 생각할 수도 있다. 아직 밝혀지지 않았으니 말이다. 하지만 최소한 구글은 타임라인의 정보는 당신에 대한 광고를 만드는 데 사용하고 있다.

인공지능이 발달하면서 우리가 기대하고 있는 것 중의 하나는 인공지능 비서, 또는 에이전트 기능이다. 물론 지금도 애플의 시리나 갤럭시의 빅스비 또는 아마존의 알렉사 등은 간단한 질의 정보를 처리하고 있다.

하지만 앞으로 거대언어모델(LLM) 등의 인공지능과 연계가 되면 우리에게 진정한 비서 역할을 하는 에이전트를 구성할 수 있다. 나의 정보를 더 많이 이해할수록 이 비서는 더욱 똑똑해질 것이다. 모든 e메일과 통화 내용을 분석할 수 있고, 내가 어떤 약속을 선호하는지 또한 이 사람에 대하여 어떻게 생각하는지를 알면 알수록 이 비서는 알아서 척척 e메일 답장, 미팅 약속 및 연기 등의 일들을 수행해 나갈 것이다.

우리의 스마트폰은 점점 라이프 로깅(개인의 모든 일상생활을 기록 저장하는 과정)을 향하고 있다. 사람들은 식당에서 먹는 음식에 대하여 사진을 찍고 방문하는 모든 장소를 인스타에 올리기 시작했으며 인공지능은 이것을 시간별, 장소별, 같이 있는 사람별로 정리를 해주기 시작했다.

이런 모든 과정들은 마케팅 기업에게는 거부할 수 없는 기회를 제공하며 모든 정보를 바탕으로 제공된 마케팅 정보는 개인으로 하여금 거절할 수 없는 충동을 유도하게 된다.

최근 스위스와 이탈리아의 연구진이 발표한 LLM의 설득 능력을 보자. 'LLM의 대화 설득 능력에 대하여: 무작위 통제 실험'이란 제목의 논문에서 연구자들은 여러 이슈에 대하여 사람 대 사람, 사람 대 인공지능에 대해 한 번은 그냥, 한 번은 상대방에 대한 인종, 나이, 정치적 성향에 대한 정보를 주고 토론을 한 후 어느 정도 합의를 이루게 되는가를 관찰했다.

토론의 주제는 '인간은 기후위기를 만들고 있는가', '온라인 강의가 대면 강의를 대체할 수 있는가' 등의 누구든지 토론할 수 있는 주제였다. 결과는 인공지능이 사람의 정보를 받았을 때 설득을 하는 비율이 가장 높았다.

인간을 설득할 확률은 못할 확률에 비하여 무려 81.2%포인트나 높게 나왔다. 다른 배치에서는 토론 후의 결과가 크게 달라지지 않았다. 가장 재미있는 것은 서로 다른 의견을 가진 사람들 간의 토론에서 한쪽에 상대방의 정보를 제공하였을 때의 결과이다. 인간의 경우 토론 후 자신의 원래 의견이 각각 강화되었다.

단순하게 '극단적 설득' 기술을 마케팅, 광고, 정치에서 사용하는 것을 제한하는 법률이나 규제를 적용할 수 있게 한다는 것은 순진한 생각이다. 편리한 서비스 개발과 프라이버시, 거짓 정보와 진실 간의 경계는 생각보다 모호할 수 있다. 우리는 인공지능의 사용에 대한 투명성 확보, 윤리적 가이드라인 수립, 개인정보 보호 강화 등의 규제와 함께 인공지능을 이해하며 자기 결정권을 어떻게 시행하여 이런 새로운 시대를 헤쳐 나갈 것인가에 대한 새로운 도전에 직면하고 있다.

〈광주일보와 중앙SUNDAY 제휴 기사입니다〉

이준기 서울대에서 계산통계학과를 졸업 후, 카네기멜론대 사회심리학 석사, 남가주대 경영학 박사를 받았다. 국가 공공데이터 전략위원회에서 국무총리와 함께 민간 공동위원장을 맡고 있으며 'AI 로 경영하라' '오픈콜라보레이션' '웹2.0과 비즈니스 전략' 등을 펴냈다.