

인공지능과 인권

김도윤 박석진 이선민 이윤환 이재준

인공지능(AI)?



Artificial Intelligence

컴퓨터 프로그램이 인간의 학습능력과 추론능력, 지각능력, 자연언어의 이해능력 등 지능적인 행동을 모방할 수 있도록 하는 것

한 마디로 컴퓨터로 만든 인간

주로 머신 러닝을 통해 구현되며,

유투브의 영상추천, 구글 광고추천, AlphaGo

빅 데이터(BIG DATA)?

방대한 분량에 방대한 표본을 가진, 사람의 직관적 가치관을 담은 통계

인터넷을 통해 소비 성향, 생활 패턴 파악
영화, 음악, 소설, 게임 등의 예술상품 추천 예측
선거 결과 예측, 유권자 맞춤형 선거 전략
보다 뛰어난 번역 시스템
국문적인 가치관을 활용한 법 제정 등...

↳ 빅 데이터는 사회 기준이 될 것이다.



인권 침해 문제

인간의 잘못된
가치관 학습

악용

과학과 예술 참여

사생활 침해

인간의 잘못된 가치관 그대로 학습

빅 데이터를 통해 인간의 잘못된 가치관을 그대로 학습하여 차별과 편견을 지남

사례 1 : 마이크로소프트 Tay



AlphaGo가 인공지능 바둑 프로그램이라면, 마이크로소프트의 'Tay'는 인공지능 소셜 미디어 프로그램이다. 실제 인간과 대화하며 정보를 파악하고 의견을 공표하고, 학습한 내용을 Grundlage 다시 반영하도록 설계되었다. 즉 머신러닝을 통해 실시간으로 사용자 대화에 스스로 학습하는 것이다.

문제는 사용자들이 일본 인터넷 거지고 Tay의 대화형이, Tay의 초기학습에 인종차별, 성차별, 편향된 정치적 내용 등을 모두 학습 데이터로 인식한 것이다.

문란위에 공개된 데이터는 몇몇 지나지 않아, 'Tay'는 사이드(대형학자)를 지지한다는 발언을 '정말로 지났다고 믿어주는 등 독설이 섞인, 인종주의적, 성적인 차별 발언을 쏟아냈으며, 이에 마이크로소프트는 문제가 된 게시지를 삭제하고 서버호를 시작하지 불과 16시간만에 당국에 통보할 수 밖에 없었다.

사례 2 : MIT 미디어 연구소 통계

MIT 미디어 연구소에서 발표한 자료에 따르면, AI의 자연언어 처리능력이 뛰어난 반면, 특히 여성에 대한 편견이 심해 나타났다. 남성에서 여성, 백인에서 유색 인종으로 갈수록 오류율은 점점 높아지며, 이 수치는 AI 학습 내용과 성별에 따른 왜곡과 차별, 편견을 가졌다는 것을 의미한다.

실제로 미국에서는 구글 포토 애플리케이션 광고가 흑인 여성을 사진 데이터에 사용에 대한 '유인물'로 '고양이'와 같은 단어를 자동 태그한 사실이 밝혀지기도 했다.

데이터는 이미 편견으로 가득 찬 세상을 반영하는 거울이므로, 데이터를 학습한 AI가 왜곡과 차별을 일으키는 것은 어쩌면 당연한 귀결이다.



인공지능을 악용하는 인간

딥 페이크, 가짜 뉴스 대량 생산으로 사회적 혼란 및 인권 침해 발생

사례 3 : 딥페이크 악용 사례

인공지능을 이용한 딥페이크가 실제 사회에 악영향을 미치고 있다. <월스트리트저널>의 최근 보도에 따르면 지난 3월 영국에서 일단의 범죄자들이 인공지능 소프트웨어를 이용해 한 정치인인 데이비드 카메론의 목소리를 흉내 낸 후, 직원에게 전화를 걸어 22억 유로(약 3억 달러)의 자금을 이체한 사건이 있었다. 해당된 목소리는 악당까지 실제와 닮아 알아듣기 어려웠다.

또 다른 예시로 인터넷에 올라온 바락 오바마 전 미국 대통령 영상이 있다. 영상의 속에서 오바마는 '트럼프는 권력이 없었는 인간(complete dipsh*t)'이라고 노골적으로 트럼프 대통령을 향해 강하게 표현으로 욕설을 내뿜고 있다. 하지만 이는 딥페이크를 이용한 가짜 동영상으로 실제 속 오바마는 전혀 오바마가 아니다. 이제 온라인 통해, 페이스북도 AI가 목소리 복사를 할 수 있게 되었다.



인공지능의 과학 / 예술 분야 참여

과학 사실 / 예술 창작자들의 지식 재산권 부여 문제, 인간의 연구 / 창작 욕구 감소

사례 4: 인공지능 예술분야



지난 10월, 뉴욕 크리스티 경매에서 최초로 인공지능이 창작한 그림이 경매에 나옴, 고가에 낙찰되는 일이 일어났다. 프랑스의 연구자들이 개발한 인공지능 화가 '조르하노'가 그린 초상화, '에드몽 드 레르미에'가, 4만 2000여달러(약 5억 3천 원)에 낙찰되었다. 이는 예전 1만 달러 수준으로 예상했던 낙찰가를 40배 넘어선 가격이었다. 조르하노는 144개 패턴 20세기까지의 서양화 1만 5000여 작품을 데이터베이스로 만들고, 이미지를 분석해 초상화 구성요소를 학습한 뒤 창작을 해냈다.

인공지능 화가의 성능이 발달함에 따라 기존에 실현하기 어려웠던 형식의 예술을 무한대로 실험해 볼 수 있다는 장점이 있다.

그러나 인공지능 화가의 성능이 인간에 도달하거나 그 이상에 도달할 때, 인간 예술가의 예술작품의 가치가 떨어질 것이라는 우려가 있기도 하다.

인공지능으로 인한 사생활 침해

해킹으로 인한 사생활 침해, 기업 / 국가적 차원의 사생활 침해

사례 5 : 중국 CCTV



시는 수많은 데이터를 학습·비교·분석하면서 경범죄도 거둬 놓았다. 어느 특정한 지역에서 미친 행위가 일어난 주민과 외부인을 식별할 정도로 진화한 것이다. 아래 사진을 보라. 공안당국에 시가 모여있는 주민과 외부인의 이동 상황을 추적할 수 있다. 기계적으로 외부인의 이동 경로를 추적해 이상 동향을 실시간으로 확인하는 시스템이다.

또한 주민의 신상정보를 바탕으로 공안당국은 외부인이 지역 주민과 어떤 관계에 있는지도 파악해 가능할 수 있다.

이렇듯 엄청난 기술이 탑재된 CCTV가 등장으로 치안이 개선될 수 있다는 얘기가 존재하지만, 우리가 통제 가 강화되며 개인들의 사생활이 침해될 것이라는 우려가 나타나고 있다

결론

- AI 시스템은 편향되지 않은 빅 데이터를 바탕으로 법의 지배, 민주적 가치, 다양성 등 인권을 존중하는 방식으로 설계되어야 하며, 필요한 곳에 인간이 개입할 수 있도록 하는 등 공정하고 정의로운 사회를 위한 적절한 안전장치를 포함해야 할 것이다.
- 딥 패이크, 가짜 뉴스 등 인공지능 악용의 문제와 인공지능으로 인한 사생활 침해 문제는, 인공지능 자체의 문제보다 인간이 어떻게 인공지능을 사용하느냐에 달린 문제이다. 우리는 인공지능을 무기가 아닌, 개인과 사회를 위한 수단으로써 사용해야 할 것이다.
- 인공지능의 과학 / 예술 분야 참여는 물론 인간에게 편리함과 즐거움을 가져다 주겠지만, 단순 노동과 정보처리뿐만 아니라 과학과 예술 분야까지 인공지능으로 대체된다면, 과연 시는 불가능하고, 오직 인간만이 가질 수 있는 힘이 무엇인지 반드시 찾아내야 할 것이다.

결론

인공지능의 개발로 얻을 편리함을 마냥 기다릴 것이 아니라

인공지능의 개발로 일어날 인권 침해의 심각성을 알고

우리의 인권을 지키기 위해

그 해결 방안과 대처 방법을 미리 도출하려는 자세가 필요