

인공지능과 인권

김도윤 박서진 이선민 이윤환 이재준

인공지능(AI)?



Artificial Intelligence

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

인공지능(이하 AI)는 주로 머신 러닝을 통해 구현되며, 유튜브의 영상추천 알고리즘, AlphaGo, 구글 광고 등이 그 예시

한 마디로 인간이 만든 가상의 인간

머신러닝?

기존 프로그램

입력 Input

처리과정 Program

출력 Output

인공지능

입력 Input

처리과정 Program

출력 Output

머신러닝?

기존 프로그램

$x = 2$

$y = 2x$

4

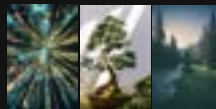
인공지능

2

???

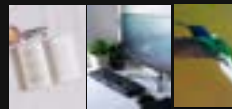
4

머신러닝?



???

이것은 나무다



???

이것은 나무가 아니다

문제점

1. 빅 데이터를 통해 인간의 잘못된 가치관, 편견, 차별을 그대로 학습
2. 가짜 뉴스, 딥 페이크, 투표 조작과 같은 악용
3. 과학과 예술 참여, 이에 대한 지식 재산권 여부, 인간의 연구/창작 욕구 저하
4. 해킹, 국가 자원의 데이터 수집으로 인한 사생활 침해
5. 더 뛰어난 인공지능을 가진 대기업이나 정부의 빈부격차 및 독점
6. 윤리 판단 자체의 문제, 윤리 판단과 오작동의 책임 여부
7. 오픈소스를 통해 누구든 쉽게 접근 가능

문제점

빅데이터는 사회 기준이 될 것이다.

- 사람의 의견 및 생각을 담은 통계
- 인터넷에 남은 흔적 > 소비자의 소비 성향, 생활 패턴 파악 > 영화, 음악, 소설, 게임 등의 예술상품 흥행 예측 > 마케팅
- 판매 업체, 평가가 좋은 상품 추천, 평가가 나쁜 상품 제거 > 광고 없이도 높은 매출 유지 가능
- 선거 결과 예측 > 유권자 맞춤형 선거 전략
- 보다 뛰어난 번역 시스템

사례 1 : 마이크로소프트 Tay



AlphaGo가 인공지능 바둑 프로그램이라면, 마이크로소프트의 'Tay'는 인공지능 일상 대화 프로그램이다. 실제 인간과 대화하며 행동을 파악하고 의견을 공유했고, 학습된 내용을 대화에 다시 반영하도록 설계되었다. 즉, 머신러닝을 통해 실시간으로 사람과 대화하며 스스로 학습하는 것이다.

문제는 사용자들이 나쁜 의도를 가지고 Tay와 대화하여, Tay의 초기 학습에 악영향을 끼치며, 온갖 난잡한 내용 등을 모두 학습 데이터로 인식한 것이다.

온라인에 공개된 태이는 얼마 지나지 않아, "개노새미드(대량학살)를 지지한다"는 발언에 "정말로 지지한다"고 답하는 등 욕설이 여간 많았으며, 성적인 대화를 할당할 수밖에 없었기에 이미 마이크로소프트는 문제가 된 웹사이트를 삭제하고 서비스를 시작한지 불과 16시간만에 급하게 운영을 중지할 수 밖에 없었다.

사례 2 : MIT 미디어 연구소 통계

MIT 미디어 연구소에서 발표한 자료에 따르면, AI의 안전감식 도우미를 위한 영상에 강아지, 혹은 인간의 경우 30% 나타났다. 남성에서 여성, 백인에서 유색 인종으로 갈수록 오답률은 점점 높아지며, 이 수치는 AI 역시 인종과 성별에 따른 왜곡과 차별, 편견을 가진다는 것을 의미한다.

실제로 미국에서는 구문 코드 해독이 어려운 알고리즘에 의한 어휘를 가진 컴퓨터가 '사람'이 아닌 '유인원', '고릴라'와 같은 단어를 자주 태그한 사실이 밝혀지기도 했다.

세 알고리즘 개발자의 대부분이 25살 미만 50살 이하의 백인 남성이고, AI가 학습한 데이터도 백인 남성에 편중된 현상이 반영된 결과라는 분석이 나왔다.

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



결론

- 사례 1, 마이크로소프트의 AI 데이터의 경우, 도덕과 윤리의 개념이 부족했다. 어린아이가 잘 크기 위해서는 좋은 선생님이 있어야 하는 것처럼, 지금의 인공지능 데이터에도 초기 학습과정의 윤리적인 판단 부분에서 충분히 개념해줄 좋은 선생님이 필요하다.
- 사례 2, AI의 편향된 시각은 이미 편견이 가득한 세상을 그대로 반영한 결과이다. 빅 데이터는 우리 사회의 기준이 될 것이며, 때문에 편향을 품은 빅 데이터가 만들어낼 결과는 우리의 세상을 뒤흔들 수 가능성이 크다.
- 따라서, AI 시스템은 편향되지 않은 빅 데이터를 바탕으로 법의 지배, 민주적 가치, 다양성 등 인권을 중시하는 방식으로 설계되어야 하며, 필요한 곳에 인간이 개입할 수 있도록 하는 등 공경하고 정의로운 사회를 위한 **적절한 안전장치**를 포함해야 할 것이다.