

[SPRI AI BRIEF]

인공지능 최신 동향과 시사점



CONTENTS

- 구글, 올해 안에 AI 윤리 컨설팅 사업 추진
- 링크드인, AI 편향성 감지 툴킷 공개
- 영국 정부, 알고리즘을 활용한 성적 산출 시스템 도입 철회
- 뉴럴링크, 뇌에 칩을 이식해 컴퓨터와 연결하는 기술 공개

구글, 올해 안에 AI 윤리 컨설팅 사업 추진¹

● 구글은 AI 윤리 이슈 대응 경험을 바탕으로 연말까지 관련 컨설팅 사업 추진 계획

- 구글은 AI 서비스 및 프로젝트 추진 과정에서 거센 윤리적 논란*을 겪으며 AI 윤리 원칙을 발표하는 등 이에 대응하는 노하우를 축적
 - * 흑인을 고릴라로 인식해 사진에 자동 태깅했던 구글 포토 서비스(2015년), 美 국방부의 AI 기반 무인 드론 프로젝트(Maven) 참여에 반대하는 내부 직원들의 항의(2018년) 등
- 구글은 개발자들이 AI의 공정성, 안정성 등을 평가할 수 있는 툴*을 개발해 외부에 무료 공개했으며, 나아가 연내에 기업 고객을 대상으로 AI 윤리 컨설팅을 제공하기로 결정
 - * 구글(What-If Tool), 페이스북(Fairness Flow), 마이크로소프트(Fairlearn), IBM(Fairness 360) 등 선도 IT 기업들은 AI의 공정성 평가 툴을 개발하여 오픈소스로 공개

● 구글의 AI 윤리 컨설팅은 클라우드 서비스(Ethics as a Service, EaaS)형태로 제공될 전망

- 구글은 AI의 윤리적 딜레마로 고민하는 고객 기업이 증가하고 있음을 확인하고, 이를 해결하는 것이 사업 경쟁력 확보에 반드시 필요하다고 판단
 - * “기술 세계의 핵심 질문이 구현 가능성(개발할 수 있는가?)에서 당위성(개발해야 하는가?)으로 바뀌고 있다.” (Tracy Frey, 구글 클라우드 사업부 AI 전략담당자)
 - 구글의 클라우드 시장 점유율은 아마존(31%), 마이크로소프트(20%)에 훨씬 못 미치는 6%로 경쟁력 확보가 절실한 상황(Canalys, 2020.7.)
- AI 알고리즘의 공정성(편향성), 안정성 등을 클라우드 시스템을 통해 자동으로 검증하는 서비스를 제공할 것으로 예상
 - 구글은 이미 글로벌 금융기업인 HSBC에 AI 윤리 교육 및 AI 프로젝트 검토·감리 컨설팅 서비스(예, 대출 알고리즘의 특정 집단 편향성 검증)를 제공
- 구글의 AI 윤리 사업은 시스템에 의한 인증 편향 검증과 AI 프로젝트 추진 시 윤리 가이드라인 개발이 중심이 될 것으로 전망

● 구글의 AI 윤리 사업이 AI 생태계의 게임체인저가 될지 주목

- 최근 아마존, 마이크로소프트가 안면인식 AI 개발을 중단하거나 철회한 것처럼 향후 AI 윤리 이슈 대응이 사업 성패를 좌우할 가능성이 높아짐
 - 구글은 기업들의 AI 윤리 준수가 사업의 시행착오 비용을 절감하고 오히려 빠른 시장 진입 효과를 가져다 줄 수 있을 것으로 기대
- 또한, 이를 통해 구글의 시장 점유율이 높아질 경우, 경쟁사의 관련 사업 진출이 활발해져 AI 윤리 컨설팅 서비스가 빠르게 확산될 전망
- 하지만, AI 윤리라는 도덕적 이슈를 다루는 ‘멘토’로서 영리기업인 구글이 결국 자신의 AI 사업에 유리한 방향으로 AI 윤리 사업을 추진할 것이라는 우려도 존재

¹ WIRED, “Google Offers to Help Others With the Tricky Ethics of AI”, 2020.8.28

링크드인, AI 편향성 감지 툴킷 공개²

- 비즈니스 소셜네트워크 서비스 기업 링크드인(LinkedIn)은 AI 모델의 편향성을 감지하는 툴킷(LinkedIn's Fairness Toolkit, LiFT)을 공개(8.25.)
 - 링크드인은 그동안 AI를 활용할 경우 특정 속성을 지닌 회원을 차별하는지 LiFT로 감시하고, 문제 발생 시 적극 대응
 - 친구 추천, 직업 추천, 랭킹 도출, 회원의 문제 행동 적발 등에 AI를 활용
 - 링크드인은 LiFT를 활용하여 누구나 자신들이 개발하는 AI의 편향성을 감시할 수 있도록 오픈소스로 공개³
- LiFT는 △데이터 세트의 편향성과 △모델의 통계적 공정성을 검증하며, △대규모 데이터 기반의 다양한 AI 모델에 적용이 가능
 - 특정한 성별, 인종, 연령, 지역에 속하는 회원 집단이 훈련 데이터 세트에 지나치게 많거나 적은지 판단하여 개발자가 편향성 없는 데이터 세트를 구성할 수 있도록 지원
 - 개발자가 AI 모델이 공정한 결과를 도출하는지 판단할 수 있도록 속성이 상이한 회원 간에 AI가 동일한 성능을 발휘하는지 통계적으로 검증
 - 가령 데이터 세트의 40%가 여성이지만 친구 추천에 20%만 여성이면 개발자는 AI 모델의 변수 등을 조정하여 공정성을 개선
 - LiFT는 그동안 공개된 다수의 공정성 평가 툴보다 확장성과 범용성이 우수하다는 평가⁴
 - 주요 기업이 공개한 공정성 평가 툴은 대부분 자사의 클라우드 환경에서만 구동되어 활용에 제한이 큼
 - 반면, LiFT는 다양한 환경의 대규모 AI 개발 프로젝트에서 공정성을 평가할 수 있도록 개발*되어, 활용도가 높다는 장점 보유

* 빅데이터 분석 기술 중에 가장 주목받는 스칼라/스파크 환경에서 개발되었고, 다양한 상황에서 활용할 수 있도록 여러 종류의 API를 제공

 - LiFT를 활용하면 직원 채용이나 고객의 신용평가에 AI를 활용하는 기업이 특정 지원자 및 고객을 차별하는지 여부를 손쉽게 판단할 수 있을 것으로 기대
- AI의 실생활 적용 및 확산 가속화에 따라, 편향 방지를 위한 조치와 전략 필요
 - AI의 오남용 및 불공정 이슈가 지속적으로 제기됨에 따라 미국(매사추세츠, 캘리포니아 등), EU, 영국 등은 관련 규제를 강화하려는 움직임
 - 또한 글로벌 주요 IT 기업들이 AI 윤리 이슈에 적극적으로 대응하고 있어 우리나라도 관련 기술 개발 및 전략 수립이 필요

² Wall Street Journal, "LinkedIn Unveils Tool to Help Combat AI Bias", 2020.8.25.

³ <https://github.com/linkedin/LiFT>

⁴ Medium, "LinkedIn Fairness Toolkit(LiFT) for Explainability in Machine Learning", 2020.8.28.

영국 정부, 알고리즘을 활용한 성적 산출 시스템 도입 철회⁵

- 영국 정부는 코로나19 사태로 고등학생의 졸업자격시험(GCSE)과 대학입학시험(A-level)을 취소하고, 이를 대체할 AI 기반 평가 시스템을 도입하기로 결정(2020.4.)
 - 과목 담당 교사가 각 학생의 예상 점수를 산정하여 영국 시험감독청(Ofqual)에 제출하면, AI로 점수를 재조정하여 최종 성적을 산출
 - AI는 △교사가 산출한 예상 점수 외에 △출신 학교의 과거 3년간 시험 성적 분포, △해당 학생의 과목 석차 등 3종류의 데이터를 활용해 성적 산출
- 지난 8월 성적 발표 후 전반적인 성적 하락 및 평가의 불공정성에 대한 비판이 거세게 제기되자 영국 정부는 성적 산출 시스템 도입을 철회
 - 교사들이 학생에게 높은 성적을 부여하는 '학점 인플레이션'을 조정*하는 과정에서 작년보다 평균 성적이 30% 하락하는 문제 발생
 - * 교사가 산출한 예상 점수보다 높은 성적을 받은 학생 비율은 2.3%인 반면, 35.6%는 1등급 하락, 3.3%는 2등급 하락, 0.2%는 3등급이 하락함
 - 평균 성적 하락으로 대학이 과거 입학생 성적을 기준으로 제시한 조건부 입학 허가 성적에 못 미치는 지원자가 대거 발생하여 대학 진학에 혼란 야기
 - 공립학교와 빈곤 지역 학생 다수의 성적이 하락하며, 부유층이 다니는 사립학교 학생들이 반사 이익을 얻는 등 불공정 논란이 촉발
 - 스코틀랜드의 경우, 가장 취약한 배경을 가진 학생들의 상위권 성적 비율은 15.2% 감소한 반면, 부유한 학생들의 경우 단 6.9%만 감소
 - 다수의 교육전문가는 지역 및 계층 간 차별을 야기하는 학생의 출신 학교 정보를 성적 산출에 활용했다며 강력히 비판
 - 알고리즘에 의한 성적산출이 심각한 문제를 야기하자 영국 정부는 이를 철회하고 담당 교사에 의한 평가 점수(Centre Assessment Grades)를 사용하도록 결정
 - 이와 함께 알고리즘에 의한 점수가 교사가 산정한 점수보다 높을 경우 알고리즘 평가 결과를 사용할 수 있도록 한시적 지침 제시
- 우리나라 교육행정정보시스템(NEIS) 등에 AI 도입 시 철저한 검증과 공감대 필요
 - 우리나라는 4세대 교육행정정보시스템(NEIS) 구축 정보화전략계획(ISP) 사업을 추진하며, 빅데이터와 AI 기반의 학사 행정 지원 시스템을 계획하고 있음⁶
 - 하지만, 학생들의 성적 산출, 진학과의 연계는 매우 민감한 결과를 가져올 수 있는 만큼 AI에 대한 신뢰성과 성능 검증, 공감대 형성 등 신중한 접근 필요

⁵ Department for Education(2020), GCSE and A level students to receive centre assessment grades. 2020.8.17

⁶ 한국교육학술정보원, "4세대 나이스 구축방안 마련을 위한 정보화전략계획(ISP) 수립", 2018.

뉴럴링크, 뇌에 칩을 이식해 컴퓨터와 연결하는 기술 공개⁷

- 바이오기업 뉴럴링크(Neuralink)는 돼지 뇌와 컴퓨터를 무선으로 연결해 컴퓨터가 뇌파를 읽어내는 기술을 시연(2020.8.29.)
 - 뉴럴링크는 테슬라 CEO 일론 머스크가 뇌-컴퓨터 인터페이스(Brain-Computer Interface)를 개발하여 뇌 질환 치료와 인간 능력 강화 등을 달성하기 위해 2016년 설립
 - 이번 시연에서 돼지가 냄새를 맡을 때 코에서 뇌로 전달되는 신호를 실시간 수집하고, 컴퓨터에 기록하는 과정을 생중계
- 뉴럴링크는 기술 시연 과정에서 성능을 개선한 뇌 이식용 칩(LINK)과 안전하고 간편한 이식 수술 로봇(V2) 시제품을 공개
 - **LINK** 돼지의 뇌에 이식한 칩*은 작은 동전 크기에 머리카락 두께의 1/20(5 μ m)에 불과한 1,024개의 전극을 담아 뇌 신호를 감지하고 컴퓨터로 전송
 - * 지름 23mm, 두께 8mm이고, 이식한 상태에서 무선 충전을 할 수 있어 반영구적 사용이 가능
 - **V2** 이식 대상자에게 전신마취를 할 필요 없이 1시간 이내에 칩을 이식해 수술 당일 퇴원이 가능한 이식 수술 로봇 개발 추진
 - 뉴럴링크는 실시간 발표 영상에서 LINK를 이식한 돼지의 운동을 예측하는데 성공했고, 수술을 받은 돼지가 건강히 지내는 모습을 보여줌

■ [그림 1] 뉴럴링크가 공개한 뇌 이식용 칩(좌)과 이식 수술 로봇(우)



※ 출처: CNN(2020.8.31.)⁸

- 최근 가시적인 성과를 보이는 뇌-컴퓨터 인터페이스 분야에 주목할 필요
 - 여전히 해결해야 할 난제가 많지만, 장기적으로 뇌질환 환자의 의사소통과 움직임을 돕고, 사람이 생각만으로 정보 전달, 기계 조작 등을 할 수 있을 것으로 기대
 - 미국(브레인이니셔티브), 유럽(휴먼브레인프로젝트) 등이 연구개발에 앞장서고 있으며, 우리나라도 ‘뇌과학 발전전략’, ‘AI국가전략’ 등을 통해 투자를 확대 중

⁷ Neuralink, "Neuralink Progress Update, Summer 2020", 2020.8.30.

⁸ CNN, "Elon Musk Shows Off a Working Brain Implant in Pig", 2020.8.31.



홈페이지 : <https://spri.kr/>

* 보고서와 관련된 문의는 시정책연구팀(wycho@spri.kr, 031-739-7312)으로 연락주시기 바랍니다.