

[SPRI AI BRIEF]

인공지능 최신 동향과 시사점



CONTENTS

- 데이터과학자의 업무를 대체하는 AutoML의 부상
- 유네스코, 테크노베이션과 함께 신흥국 여학생 AI 교육 프로그램 운영
- 페이스북과 뉴욕대 연구팀, AI 활용해 MRI 검사시간 1/4로 단축
- OpenAI의 GPT-3, AI의 진보인가 재앙인가 논란 촉발



데이터과학자의 업무를 대체하는 AutoML의 부상¹

● 기계학습이 부상함에 따라 데이터과학자들에 대한 중요성 증대

- 지난 7년간 데이터과학자에 대한 수요는 약 3배 이상 증가하였으며 우수 인력 확보를 위한 기업 간 인재 유치 경쟁이 심화
 - * 데이터과학자 구인 등록 건수가 2013년 대비 2019년 3배 이상 증가(Indeed.com)
 - 구글, 아마존, 페이스북 등 IT 대기업에 인력 쏠림 현상이 나타남에 따라 타 산업분야 기업들의 관련 인력 확보에 어려움이 가중
 - 데이터과학자의 연봉은 경력에 따라 약 8만~15만 달러로 전통적인 분석 전문가의 연봉인 6만 5천~11만 5천 달러 대비 20% 이상 높은 것으로 조사²

● 데이터과학자 부족 문제 해결을 위한 대안으로 AutoML에 대한 관심 고조

- AutoML(Automated Machine Learning)은 기계 학습 모델을 개발하는 여러 단계*를 자동화하도록 설계된 AI 기반의 기계 학습 개발 도구
 - * (기계학습 절차) 비즈니스 문제 이해 → 필요 데이터 식별 → 데이터 준비 → 모델 개발 → 미세 조정 및 검증 → 모델 보급
- AutoML은 데이터과학자의 업무 중 60~80%를 차지하는 데이터 준비 과정을 자동화하여 인력 및 비용 절감
 - 예를 들어, 통신업계에서는 AutoML을 고객이탈관리 모델 개발에 적용해 계약 취소 가능성이 높은 위험 고객을 선별하는데 적용 중
- 향후 AutoML은 모델링, 튜닝 등 숙련된 데이터과학자의 직무까지 자동화될 전망
 - 실제로 한 에너지 기업은 코드 몇 줄로 답러닝 기반의 기계 학습 모델을 구현해 고객 계약 취소 모델을 개발
 - 더욱이, 오픈소스 커뮤니티가 데이터 준비와 모델링 업무를 보다 간단히 처리할 수 있는 소스코드 (Snippet)를 공개함으로써 기계 학습의 자동화를 지원

● AutoML의 발전과 확산은 데이터과학자의 직무와 수요에 변화를 줄 전망

- AutoML이 데이터 과학자의 직무로 여겨졌던 과업들을 자동화함으로써 비즈니스 전문가들이 보다 쉽게 데이터 과학을 학습하고 활용할 수 있는 기회가 늘어날 것
 - 제약, 금융, 통신 등 다양한 분야에서의 풍부한 전문 지식을 활용해 기계 학습 모델의 품질을 향상시키는 AutoML 실천가(Practitioners)*의 등장 예상
 - * (예) 간단해진 데이터 과학 업무를 수행할 수 있는 제약연구에 종사하는 바이오화학자
- 현재 AutoML은 간단한 예측 작업에 적용되는 수준이지만, 향후 기능이 개선됨에 따라 데이터과학자에 대한 장기적인 수요는 점차 감소할 것으로 전망

¹ McKinsey, "Rethinking AI talent strategy as automated machine learning comes of age", 2020.8

² Burtch Works, "2018 Data Scientist Salary Report", 2018.7.

유네스코, 테크노베이션과 함께 신흥국 여학생 AI 교육 프로그램 운영³

- 유네스코(UNESCO)는 코로나-19로 인해 심화되는 교육 불평등 및 소외현상을 해소하기 위해 올해 3월 글로벌교육연합(Global Education Coalition)을 구성
 - 기업(우버, 에릭슨, 줌 등), 국제기구(UN, ITU), 교육기관 등 40여 곳이 회원으로 참가하여, 교육 프로그램 운영에 필요한 다양한 자원을 제공
 - 신흥국 정부 및 지역 비영리단체가 유네스코에 교육 지원을 요청하면 유네스코는 교육 수요에 부합하는 회원기관을 연결하고, 프로그램 운영을 지원
- 유네스코는 글로벌교육연합 회원사이자 비영리 교육기관인 테크노베이션과 함께 신흥국 여학생 대상 AI 교육 프로그램(Technovation Idea Lab*)을 운영(2020.7)
 - **대상** 브라질, 인도, 케냐, 멕시코, 나이지리아, 파키스탄 등 6개 신흥국의 10~18세 여학생 6천여 명에 대해 7월~8월간 교육 실시
 - **목적** 여학생들에게 AI와 관련한 흥미로운 프로젝트 기반 커리큘럼을 제공해 기술을 사용하여 지역 사회에 기여하고 변화를 주도 할 수 있다는 자신감과 역량을 길러줌
 - **내용** 구글 클래스룸을 활용해 온라인 교육으로 실시하며, AI에 대한 이해, 작동 원리, 실생활 속 문제 이해와 AI를 통한 해결방안 계획 및 실천 등

■ [그림 1] AI를 활용한 5단계 문제해결 방법



- **멘토링** 전문적인 교육 내용 지원을 위해 우버, 에릭슨, SAP 등 글로벌 IT 기업 소속 엔지니어들이 자원봉사자로 참여하며, 교육용 SW 도구인 마이크로비츠(micro:bits)를 활용함
 - 테크노베이션의 직원과 자원 봉사자는 학생들이 등교하지 못하는 상황에서 가장 효과적인 교육 프로그램을 테스트
- 코로나-19로 인한 사회적 거리두기는 당분간 지속될 것으로 보이며, 우리나라 또한 교육 격차 및 소외 계층을 고려한 교육적 대응이 필요함
 - 우리나라는 4월 개교 이래 처음으로 온라인 개학을 실시하며, 학력 저하 및 교육 격차 대응을 위한 적극적인 노력을 기울이고 있음
 - 그럼에도, 온라인 교육으로 인해 새롭게 발생하는 교육 소외 계층에 대한 보다 깊은 이해와 이들을 지원하기 위한 적극적인 대응이 필요

³ UNESCO, "Girls solve community issues through technology at the new Technovation Idea Lab", 2020.7.27.

페이스북과 뉴욕대 연구팀, AI 활용해 MRI 검사 시간 1/4로 단축⁴

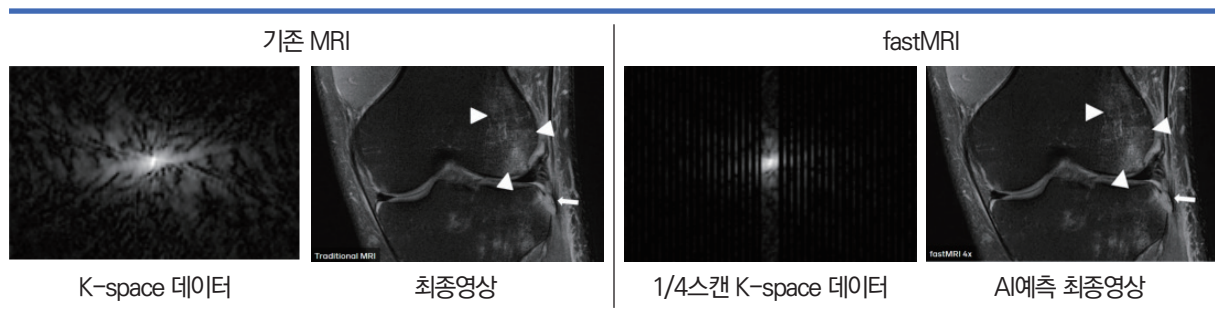
● 페이스북과 뉴욕대(NYU)는 AI를 활용하여 MRI 검사 시간을 크게 단축

- 페이스북 AI연구팀(FAIR)과 뉴욕대(NYU Langone Health)방사선 전문의는 공동 프로젝트(fastMRI)를 통해 시제품을 발표했으며 최종목표는 10배 빠른 MRI 스캔 구현
 - 그동안 MRI는 환자가 밀실 공포증을 유발할 수 있는 튜브 구조 속에서 시끄러운 소음을 견디며 최대 1시간가량 움직이지 않고 가만히 있어야 하는 검사였음*
- * MRI 검사는 CT 촬영 대비(20분 내) 입체적인 스캔결과를 얻을 수 있으나, 강력한 자기장 활용으로 소음이 심하며, 평균 40분에서 1시간 이상의 검사 시간이 소요

● 연구팀은 MRI 스캔데이터의 일부만으로 최종 스캔영상과 동일한 결과를 생성할 수 있도록 기계학습 모델을 훈련시키고 성능을 검증

- 기존대비 25% 수준의 MRI 스캔데이터 입력을 기반으로 최종 스캔영상을 예측 및 복원함으로써 스캔 시간을 1/4로 단축
- AI는 저해상도 및 고해상도 MRI 스캔 영상에 대하여 학습하고, 학습한 결과를 실제 MRI 최종 영상과 비교하여 일치하는지를 지속 확인 및 보강
 - MRI 스캔부위에 대한 저해상도 전체 영상과 해당 영상의 고해상도 업스케일링(Resolution Upscaling) 및 여러 교정과정, 원본영상 등을 병행하여 학습
- 적은 원시 데이터*만으로 최종 스캔 영상을 생성하는 것에 대한 신뢰성 확보를 위해, MRI 스캐너가 물리적으로 생성 가능한 영상만 예측하도록 범위를 설정하고 검증
 - * MRI 스캔으로 측정된 데이터(K-space)로 이미지의 2D 및 3D 변환을 위해 사용
- AI가 생성한 영상 및 임상결과를 실제 MRI 스캔 원본영상과 비교하고 전문의가 구분할 수 없는 수준임을 검증하여 학술지에 발표⁵

■ [그림 2] 무릎 스캔 영상 사례 비교



● 환자의 번거로움이 줄고 진단이 빨라지는 효과 기대

- 또한, 모든 의료영역 및 영상을 활용하는 타 분야에 응용이 가능할 것으로 전망

⁴ The Verge, "Facebook and NYU use artificial intelligence to make MRI scans four times faster", 2020.8.18.

⁵ Zbontar, Jure, et al. "fastMRI: An open dataset and benchmarks for accelerated MRI." arXiv preprint arXiv:1811.08839 (2018).

OpenAI의 GPT-3, AI의 진보인가 재앙인가 논란 촉발⁶

● OpenAI는 인간의 글과 흡사한 문장을 작성하는 생성적 사전학습(Generative Pre-Trained Transformer, GPT)의 세 번째 모델(GPT-3) 공개⁷

- 생성적 사전학습 모델은 언어를 통계적으로 표현하는 것을 목표로 하며, 특정 단어가 다른 단어를 따르는 확률을 매핑(Mapping)하여 문장을 완성
 - * 가령 ‘빨간’과 그 뒤에 오는 ‘사과’, ‘장미’, ‘자동차’ 등 명사와의 조합 확률을 계산
- **GPT-3 모델** GPT-3는 1,750억 개*의 모수로 구성된 모델을 45테라바이트의 텍스트 데이터와 대규모 AI 컴퓨팅 인프라를 활용해 학습
 - * GPT-3의 이전 모델인 GPT-2는 15억 개의 모수를 갖고 있으며, GPT-3가 이전보다 100배 이상 커짐에 따라 이를 학습시키기 위한 데이터의 양과 컴퓨팅 파워도 증가

● GPT-3는 범용 자연어 처리 모델로 사람의 언어 구사와 유사해짐에 따라, 사람의 지능에 가까워진 AI의 진보인지, 악용 가능한 재앙인지에 대한 논란이 점화

- 다수의 연구자는 GPT-3를 통해 대규모 데이터와 모델을 학습시켜 성능이 비약적으로 발전했음이 증명되었고, AI 기술의 발전 가능성을 보여준다고 주장
 - 향후 플랫폼 형태로 서비스가 제공될 것으로 전망되며, 단순 질의응답 등 의사소통이 필요한 분야에서 활용될 수 있을 것으로 기대
- GPT-3가 우수한 성능을 보여준 것은 사실이나, 기존 방법의 양적 확대에 따른 성능향상에 불과하며 사람의 지능과는 본질적으로 달라 과도한 기대를 경계하는 목소리도 큼
 - OpenAI의 CEO인 샘 알트만은 GPT-3는 사람의 지능을 구현하기 위한 과정상에 있는 결과물이며, GPT-3 역시 많은 단점*이 있어 지나친 확대해석은 불필요하다고 밝힘
 - * GPT-3는 모든 자연어 처리 과업에서 모두 최고 성능(State-of-the-art)을 달성한 것이 아니며, 설명 불가 등 딥러닝 방법이 겪고 있는 한계가 여전히 존재함
- 또한, GPT-3와 유사한 모델이 공개된다면 가짜뉴스 생성, 스팸메일, 보고서 작성 등 악의적인 활용이 가능

● GPT-3는 또 다른 딥러닝의 성공사례이며, 악의적인 활용에 대비 필요

- GPT-3는 문장을 생성하는 능력이 사람에 버금가는 수준으로 등극하여 다양한 확대해석이 나오고 있으나, 모든 언어 지능에서 우월한 성능을 보이는 것은 아님
- GPT-3가 공개됨에 따라 적대적 생성 신경망(GAN)으로 인한 딥페이크의 부상과 같이, 거짓 정보 이슈가 부상하고 있으며 이에 대한 적극적인 대응 마련이 필요

⁶ The Guardian, “GPT-3: an AI game-changer or an environmental disaster?”, 2020.8.1.

⁷ OpenAI, “Language Models are Few-Shot Learners”, arXiv 2020.7.22.



홈페이지 : <https://spri.kr/>

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