SPRi AI BRIEF

인공지능의 최신동향



SPRi 소프트웨어정책연구소

2023년 1월호

국내외 정책

美-EU 무역기술위원회(TTC), AI 로드맵 발표 3
美 로펌, EU의 AI 책임 지침의 소송 절차 인식 확산 노력 4
유엔환경계획(UNEP), 기후 문제 해결 및 대응을 위해 AI 활용 5

기업·시장 동향

非유럽권의 글로벌 AI 시장 독점 심화 6
엔비디아와 마이크로소프트, AI 슈퍼컴퓨터 구축을 위해 협력 7
메타의 대규모 언어모델(LLM) 갤럭티카 사용 조기 종료 8

고용·인력 동향

싱가포르, AI 인재 양성 및 유치에 5천만 달러 투자 9
美조지타운대학교, 중국의 AI 인력 수요 현황 분석 10

기술·연구 동향

구글, 전 세계 1,000개 언어 지원 가능한 거대 AI 언어모델 개발 계획 11
영국, 구제금융에 관한 정부 결정을 지원하는 AI 도구 개발 12
허깅페이스(HuggingFace), AI 이미지 생성기의 편향성 탐색 도구 개발 13

美-EU 무역기술위원회(TTC), AI 로드맵 발표

KEY Contents

- **美-EU 무역기술위원회가 민주주의 가치에 부합하는 AI 시스템의 긍정적 영향을 극대화하기 위해 2022년 5월 AI 로드맵 개발 계획을 발표한 가운데 12월 1일 로드맵 내용을 공개**
- **AI 로드맵은 신뢰할 수 있는 AI 및 위험 관리 체계 구축을 위한 계획 및 활동 제시**

미국-유럽연합(EU) 무역기술위원회(TTC)는 2022년 12월 1일 신뢰할 수 있는 AI 및 위험 관리를 위한 평가 및 측정 도구에 관한 로드맵 발표

▶ **TTC는 AI 시스템이 개인, 문화, 경제, 사회에 미치는 부정적인 영향을 최소화함으로써 민주주의 가치에 부합하는 AI 시스템의 긍정적 측면을 극대화하기 위해 올해 5월 로드맵 개발 계획 발표**

- TTC는 EU와 미국의 AI 위험 관리 및 신뢰할 수 있는 AI에 대한 도구, 방법론 및 접근방식을 안내하고 민주주의 및 인권에 대한 공유가치를 바탕으로 한 국제 표준화 노력 지원 및 신뢰할 수 있는 AI 촉진을 위해 공동의 이익 증진을 목표로 로드맵 마련
- 로드맵은 신뢰할 수 있는 AI를 발전시키고 AI에 관한 경제협력개발기구(OECD)의 권고안을 이행하기 위한 공동의 약속에 따른 실질적 조치 제안

미국과 EU, 공동의 가치 증진을 위한 신뢰할 수 있는 AI 구축 계획 제시

▶ **로드맵은 미국과 EU가 AI 위험성에 대해 동일한 접근방식을 취하도록 구체적인 활동 제안**

- 신뢰할 수 있는 AI 및 위험관리가 상호운용하기 위해서는 위험 관리, 위험 감수, 위험 인식 및 신뢰할 수 있는 AI의 사회 기술적 특성 등을 포함한 공통의 용어 및 분류체계 마련이 필수
- AI 기술은 미국과 EU가 공유하는 민주주의 가치와 인권 보호 및 존중을 바탕으로 형성되어야 하며 AI 및 새로운 기술 표준 분야의 리더십은 안전, 보안, 공정성, 비차별, 상호운용성, 투명성, 다양성 등을 촉진하는 방향으로 작용 필요
- 무역과 기술에 막대한 영향을 미칠 가능성이 있는 AI 및 관련 위험에 관한 최첨단 연구를 위해 지식 공유 수단을 개발 추진

출처 : TTC, TTC Joint Roadmap on Evaluation and Measurement Tools for Trustworthy AI and Risk Management, 2022.12..01.

美 로펌, EU의 AI 책임 지침의 소송 절차 인식 확산 노력

KEY Contents

- ✔ **EU 집행위원회는 9월 28일 AI로 인한 피해 당사자가 일반 제품 피해 당사자와 같은 수준의 보호를 받도록 하는 법적 조치 마련을 위해 AI 책임 지침 발표**
- ✔ **AI 책임 지침은 AI로 인한 피해에 대하여 소송 진행 시 피해 사례 증거 공개를 원칙으로 규정**

미국의 기업 소송 전문 로펌 데이비스라이트트레메인(Davis Wright Tremaine LLP)의 법률 전문가들이 EU 집행위원회가 9월 제안한 AI 책임 지침(AILD)의 소송 절차 설명

▶ **EU 집행위원회는 AI 시스템 관련 사고로 인해 발생한 손해에 대해 비계약적 과실 기반 민사 소송을 청구할 수 있도록 하는 AI 책임 지침 제안(9월 28일)**

- **목표** AI로 인한 피해 당사자가 일반적인 제품 피해 당사자와 같은 동등한 보호를 받도록 보장하고 AI를 개발하거나 사용하는 기업의 법적 불확실성 완화
- **범위** 현재 계류 중인 EU 인공지능법(AIA) 3조 1항이 정의하는 "AI 시스템" 관련 사고에 관한 손해를 입증할 경우 민사 책임 청구인이 직면할 수 있는 문제만을 포함
- **제한** AI 책임 지침은 형사적 책임, EU의 운송 규제 법률에 따른 책임, 제품 결함으로 인한 부상에 대한 책임 또는 디지털서비스법에 명시된 조항에 따른 책임에는 미적용
- **청구인 자격** ▲피해자 본인으로서 손해 배상을 청구하는 자 또는 ▲피해자의 권리를 승계하였거나 피해자의 법률적 권리를 대신 행사하는 자(예: 보험사)

▶ **AI 책임 지침의 주요 조항에는 ▲개인의 기본권, 건강 및 안전을 위협하는 고위험 AI 시스템 관련 피해 사례의 증거 보존 및 공개를 위한 규칙 ▲고위험군으로 분류되지 않은 AI 시스템을 포함하여 AI 관련 사고로 발생한 피해 주장의 입증 및 반박 가능한 추정 책임에 관한 내용을 포함**

AI 책임 지침은 유럽의회와 EU 정상회의의 심의를 통과해야 법제화 가능

▶ **AILD는 규정 발효 2년 후부터 발생하는 손해에 대해 해당 규정을 적용하도록 제안**

- EU 회원국은 AILD 발효 후 2년 동안 AILD를 준수하는 데 필요한 법률, 규정, 행정 조항을 통과시켜야 하며 자국 규정이 AILD와 호환되는 한 청구인에게 더 유리한 규정 채택 가능

출처 : Davis Wright Tremaine LLP, EU's Proposed AI Liability Directive Sets Procedural Rules for Litigating Potential AI Harms, 2022.11.17.

유엔환경계획(UNEP), 기후 문제 해결 및 대응을 위해 AI 활용

KEY Contents

- ✔ 생물 다양성 손실, 환경 오염, 폐기물 문제 등 기후변화로 인한 환경 위기에 대응하기 위해 데이터 모니터링 및 분석 등에 AI가 핵심 기술로 활용
- ✔ 유엔환경계획은 실시간 환경 및 메탄 배출, 대기질 모니터링 등 다양한 이니셔티브에 AI 도입

UNEP(유엔환경계획)의 디지털전환 서브프로그램 조정관 데이비드 젠슨(David Jensen)이 기후 및 환경 문제 해결을 위한 AI 활용 방안에 관해 설명

- ▶ 현재 전 세계가 생물 다양성 손실, 환경 오염, 폐기물 등 기후변화로 인한 3중 위기에 직면하고 있는 가운데 위기를 관리하기 위해서는 위기를 측정할 수 있는 능력이 필수
- ▶ 현재 많은 양의 기후 데이터 사용이 가능하나 데이터에 관한 접근 및 해석을 통해 적절한 대처 방안을 마련하는 것이 위기관리에 있어 매우 중요하며 AI는 이를 위한 핵심 기술 중 하나로 활용
  - AI는 전 세계 탄소 배출량에 관한 위성 모니터링과 같은 대규모, 또는 조명이나 난방을 제어하는 스마트 하우스와 같은 세분화된 규모 등 다양한 영역에서 환경 문제 해결을 위한 역할을 수행

UNEP는 환경 문제 모니터링 및 분석과 대응 방안 마련을 위해 다양한 방식으로 AI 활용

- ▶ UNEP는 2022년 다면적인 데이터세트 분석을 위해 파트너 컨소시엄의 지원을 받아 디지털 플랫폼인 세계환경상황실(WESR)을 출범하고 전 세계 환경 상황을 실시간으로 모니터링 및 분석
  - 세계환경상황실은 적절한 조치를 취할 수 있도록 모든 주요 환경 지표를 모니터링하는 지구의 임무통제센터(Mission Control Centre) 임무 수행을 목표로 작업 진행
- ▶ UNEP는 세계환경상황실 디지털 생태계 내에서 AI를 활용하여 메탄 배출을 완화하고 모니터링하기 위한 이니셔티브인 국제메탄배출관측소(IMEO) 운영
- ▶ UNEP는 스위스 기반 실내 공기질 개선 솔루션 전문업체 IQAir와 협력하여 환경 모니터링 이니셔티브인 GEMS 공기오염 모니터링 플랫폼 공동 설립
  - IQAir는 140여 개국에 위치한 25,000개 이상의 대기질 모니터링 스테이션을 통해 데이터를 집계하고 AI를 활용하여 실시간 대기질이 인류에 미치는 영향에 관한 정보를 제공

출처 : UNEP, How artificial intelligence is helping tackle environmental challenges, 2022.11.07.

非유럽권의 글로벌 AI 시장 독점 심화

KEY Contents

- **삶의미래연구소가 발표한 AI 시장 현황 분석 보고서에 따르면 범용 AI 시스템 개발 분야에서 미국, 중국 등 非유럽권 지역의 글로벌 시장 독점 현상이 심화**
- **非유럽권에서 개발된 시스템에 대한 유럽 기업의 의존도가 매우 높은 수준이며, 특히 범용 AI 시스템 개발은 막대한 자금이 요구됨에 따라 글로벌 대기업이 이미 선점하고 있는 실정**

미국의 비영리 연구기관 삶의미래연구소(Future of Life Institute)가 '非유럽권의 글로벌 AI 시장 독점 심화' 현상을 조명한 보고서 발표

- **보고서는 유럽 기업들이 범용 AI 시스템을 개발하지 않고 있으며 미국 및 중국 기업과 비교해 시스템 개발에 필요한 자금, 데이터 및 컴퓨팅 리소스 양에서 경쟁 우위를 선점하지 못하고 있다고 설명**
- **글로벌시장을 선도하는 범용 시스템인 오픈AI의 GPT-3은 이미 수백 개의 다운스트림 애플리케이션에 통합되었으며 작년 전 세계 고객들이 이를 사용하여 하루에 45억 단어 이상 생성**
  - 많은 유럽 기업들이 오픈AI의 고객이며 네덜란드의 클라우드 커뮤니케이션 플랫폼인 메시지버드(MessageBird)는 GPT-3로 구동되는 옴니채널* 커뮤니케이션 플랫폼을 제공
    * 소비자가 오프라인, 온라인, 모바일 등의 여러 경로로 상품을 검색하고 구매할 수 있도록 제공하는 서비스
  - 핀란드의 소프트웨어 기업 플로우라이트(Flowrite)는 GPT-3를 활용하여 일련의 단어들을 즉시 전송 가능한 이메일, 메시지 및 게시물로 변환하는 서비스를 제공

막대한 자금이 요구되는 범용 AI 시스템 개발 분야는 글로벌 대기업이 선점하고 있는 실정

- **오픈AI가 2020년 GPT-3을 출시한 이후 전 세계 곳곳에서 유사한 모델의 개발 및 출시가 이어졌으나 최첨단의 대규모 범용 AI 시스템 모두가 EU 밖에서 개발된 것으로 확인**
  - ▲(이스라엘) AI21 랩스의 Jurassic1-Jumbo ▲(중국) 바이두의 Ernie 3.0 Titan, 화웨이의 Pan Gu, 인스퍼의 Yuan 1.0 ▲(미국/영국) 딥마인드의 Gopher ▲(미국) 구글의 FLAN & LaMDA, 마이크로소프트와 엔비디아의 Megatron Turing NLG ▲(한국) 네이버의 HyperClova
- **유럽의 소규모 기업들은 글로벌 대기업과의 경쟁에 제약이 따르며, 非유럽권에서 개발된 시스템에 대한 유럽 기업의 의존도가 높아 해당 기술에 대한 종속성 또한 점차 증가할 가능성 제기**

출처 : Future of Life Institute, Emerging Non-European Monopolies in the Global AI Market, 2022.11.01.

엔비디아와 마이크로소프트, AI 슈퍼컴퓨터 구축을 위해 협력

KEY Contents

- ✔ **엔비디아와 마이크로소프트는 다년간의 협업을 통해 엔비디아의 그래픽처리장치를 활용하여 마이크로소프트의 클라우드 플랫폼 애저에 가장 강력한 AI 슈퍼컴퓨터를 구현할 예정**
- ✔ **애저에는 이미 상당 수의 엔비디아 그래픽처리장치가 포함되어 있으나 기존에 사용되던 A100 칩보다 성능이 우수한 최신 제품인 H100 칩을 슈퍼컴퓨터 구축에 활용할 예정**

엔비디아(Nvidia)와 마이크로소프트(Microsoft)는 세계에서 가장 강력한 AI 슈퍼컴퓨터를 구축하기 위해 다년간 협업을 진행할 예정이라고 발표

▶ **마이크로소프트와 엔비디아는 협업을 통해 마이크로소프트의 퍼블릭 클라우드 플랫폼인 애저(Azure)에서 AI 슈퍼컴퓨터 시스템을 구현할 예정**

- 그래픽처리장치(GPU)는 AI 및 응용프로그램의 속도를 높이는 기능으로 인해 슈퍼컴퓨터에 널리 사용
- 마이크로소프트는 애저에 수만 개의 엔비디아 GPU를 장착할 예정하며, 엔비디아의 네트워크 스위치 제품군인 Quantum-2 시리즈를 포함하여 엔비디아의 다양한 기술을 GPU와 함께 AI 슈퍼컴퓨터 구축에 활용할 계획

마이크로소프트는 이번 협업을 통해 엔비디아의 주력 데이터센터 GPU 제품인 H100 칩 기반 애저 클라우드 인스턴스를 출시할 예정

▶ **마이크로소프트 애저에는 이미 상당 수의 엔비디아 GPU가 포함되어 있으며, 애저의 고객들은 2020년 출시된 엔비디아의 과거 주력 제품인 A100 칩으로 구동되는 여러 컴퓨팅에 접근 가능**

▶ **올해 3월 출시된 H100 칩은 이전 세대 제품인 A100 그래픽 카드보다 최대 6배 빠르게 AI 모델을 학습할 수 있는 800억 개의 트랜지스터(Transistor)*로 제작**

* 전류나 전압흐름을 조절하여 증폭하거나 스위치 역할을 하는 반도체 소자

- H100 칩에는 자연어 처리 등의 작업에 널리 사용되는 첨단 신경망인 트랜스포머(Transformer) 모델을 보다 효율적으로 실행할 수 있는 최적화 기능 포함
- H100 기반 애저는 엔비디아의 이전 세대 제품보다 두 배 많은 양인 400Gb/s의 트래픽 처리가 가능한 Quantum-2 인피니밴드 스위치 시리즈를 사용하여 네트워크 트래픽을 관리할 예정

출처 : SilliconANGLE, Nvidia and Microsoft partner to build AI supercomputer, 2022.11.16.

메타의 대규모 언어모델(LLM) 갤럭티카 사용 조기 종료

KEY Contents

- ✔ **메타 AI는 11월 과학 논문 작성을 가속화하도록 설계된 대규모 언어모델인 갤럭티카의 데모를 공개했으나 데모 사용 과정에서 윤리적 문제 발생 가능성 제기**
- ✔ **메타는 갤럭티카 데모를 공개한 후 이틀 만에 윤리적 딜레마에 직면하며 해당 서비스 중단**

메타 AI는 과학 논문 작성을 가속화하기 위해 과학적 지식을 저장, 결합 및 추론이 가능하도록 설계된 대규모 언어모델(LLM) 갤럭티카(Galactica)의 데모를 11월 15일 공개

▶ **오픈AI의 GPT-3와 같은 LLM 모델은 수백만 개의 예시를 연구하고 단어 사이의 통계학적 관계를 이해함으로써 텍스트 작성 방법 학습**

- 이러한 학습 방법은 결과적으로 그럴듯한 문장을 작성할 수는 있으나 부정확한 정보와 편향적 표현이 포함될 가능성 또한 내포하고 있으며 일부 비평가들은 텍스트의 의미를 이해하지 못한 채 텍스트를 뱉어내는 능력 때문에 LLM 모델을 "확률적 앵무새"라고 명명
- 갤럭티카는 4,800만 개 이상의 논문, 교과서, 강의 노트, 과학 전문 웹사이트, 백과사전을 포함한 인류 과학 지식을 대변하는 방대한 양의 데이터로 학습한 과학용 LLM 모델로 메타 AI의 연구진은 개발 당시 고품질 데이터가 고품질 결과물로 이어질 것으로 기대

갤럭티카의 데모 사용 과정에서 윤리적 문제 발생 가능성이 제기됨에 따라 메타는 출시 이틀 후인 11월 17일 해당 서비스를 종료

▶ **갤럭티카 공개 직후 방문자들은 갤럭티카 웹사이트에 제공된 예시에 따라 프롬프트를 입력해 참고 문헌, 위키피디아 문서, 강의 노트, 질문에 대한 응답 등 문서를 생성**

- 데모 사용 과정에서 누구나 인종 차별적이거나 공격적인 프롬프트 입력이 가능하다는 우려가 제기되었으며 갤럭티카를 사용해 정확도 및 신뢰도가 낮은 가상의 논문 작성이 가능함을 발견

▶ **갤럭티카 사용 중단 사례는 AI와 관련된 일반적인 윤리적 딜레마를 상기시키는 계기로 작용**

- 잠재적 유해성을 내포한 생성 모델의 경우 이를 책임감 있게 사용하는 문제가 일반 대중에게 달린 것인지 모델 배포자가 사전에 오용을 방지해야 하는 것인지 의문 제시
- 윤리적 딜레마의 해결 방안을 제시하는 데 있어 궁극적으로는 정부의 규제가 중요할 것으로 판단

출처 : Ars Technica, New Meta AI demo writes racist and inaccurate scientific literature, gets pulled, 2022.11.19.

싱가포르, AI 인재 양성 및 유치에 5천만 달러 투자

KEY Contents

- ✔ **싱가포르 정부는 지난 5년 동안 AI 연구·개발 프로그램에 4억 5천만 달러 이상을 투자한 가운데 향후 5년 동안 AI 인재 양성을 위해 5천만 달러를 추가 투자 추진**
- ✔ **많은 혁신 및 신생기업이 인재 유치에 어려움을 겪고 있어 우수한 인재 양성의 중요성 증가**

헹스위킷(Heng Swee Keat) 싱가포르 부총리는 10월 25일 열린 SWITCH 2022 (싱가포르 혁신·기술 주간 2022) 개막 연설을 통해 AI 인재 양성을 위한 투자 계획 발표

▶ **헹스위킷 부총리는 싱가포르 정부가 향후 5년 동안 국내 AI 인재를 양성하고 AI 인력 수를 두 배로 늘리기 위해 미화 5천만 달러를 추가 투자할 예정이라고 발표**

- 싱가포르 정부는 지난 5년 동안 AI 연구·개발(R&D) 프로그램에 4억 5천만 달러 이상을 투자했으며, 해당 프로그램의 AI 공학자 양성 교육을 통해 현재까지 200명 이상 수료
- 싱가포르 국립연구재단(NRF)*의 의장직을 역임하고 있는 헹스위킷 부총리는 NRF가 싱가포르에서 연구를 수행할 우수 연구진을 유치하기 위해 AI 연구 과정을 추가 신설할 계획이라고 언급

  * National Research Foundation: 2006년 1월 1일에 설립된 총리실 산하 부서로 연구, 혁신, 기업을 위한 정책, 계획, 전략을 개발함으로써 연구·개발(R&D) 분야의 국가적 방향을 설정

싱가포르 정부는 금융 및 혁신 부문의 인재 양성을 위한 투자 및 노력 지속

▶ **싱가포르 정부는 10월 초 금융 인재 개발 프로그램에도 4억 싱가포르 달러를 투자할 계획 발표**

▶ **헹스위킷 부총리는 인재는 자본과 같이 대체될 수 없음을 강조하며 인재 개발의 중요성 언급**

- 싱가포르에는 2022년 유니콘* 지위를 획득한 기업 4곳을 포함하여 총 29개의 유니콘 기업 존재

  * 기업가치 10억 달러 이상의 비상장 신생기업

- 많은 혁신 및 신생기업의 경우 자금 조달보다 인재 유치가 더욱 어려운 상황이며 시장 진출을 통해 자본은 더욱 쉽게 유치가 가능하나 인재는 즉각적인 유치 및 투입에 한계 존재
- 헹스위킷 부총리는 코딩 및 사용자 인터페이스(UI)전문가부터 AI, 사이버 및 자동화 등 첨단기술 전문지식에 이르기까지 전문성에 요구되는 역량이 매우 구체적이기 때문에 인재는 자본과 같이 대체가 가능한 영역이 아님을 강조

출처 : Fintech Singapore, Singapore to Invest US$50M to Develop AI Talent and Attract Researchers, 2022.10.27.

美조지타운대학교, 중국의 AI 인력 수요 현황 분석

KEY Contents

- ✔ **조지타운대학교의 안보·신흥기술센터가 중국의 AI 인재에 관한 수요 현황을 집중 조명한 분석 보고서를 발표**
- ✔ **중국 내 채용 공고를 분석한 결과 AI 관련 인재 채용 공고가 30% 이상 차지**

美조지타운대학교의 안보·신흥기술센터(Center for Security and Emerging Technology, CSET)가 '중국의 AI 인력: AI 인재에 관한 수요 평가'를 주제로 분석 보고서 발표

▶ **미국이 세계 최고 수준의 AI 인재를 양성, 육성, 유치 및 확보하기 위한 정책 수립 필요성 제기**

- 미국 최상위 수준의 기술 인재는 대부분 해외 출신이며, AI 인력은 전 세계적으로 수요가 높음
- 중국이 AI 전문인력의 주요 배출국임을 감안할 때 중국의 AI 인력 현황을 이해함으로써 미국의 정책입안자들에게 중요한 통찰력을 제공할 수 있을 것으로 기대

CSET는 영국의 기계학습(ML) 전문기업 앰플리파이(AMPLYFI)와 구축한 680만 건의 중국 내 채용 공고 샘플을 사용하여 AI 인력 수요 분석 수행

▶ **CSET는 680만 건의 채용 공고 샘플 중 30% 이상이 AI 관련 인재 채용 공고로 집계될 수 있으며, 14%인 95만 5,000건이 기술 인재 채용 공고로 집계되었음을 확인**

- AI 관련 인재 채용 공고의 대부분이 초급 인재 채용 공고로, 컴퓨터 공학자 및 기계학습 엔지니어와 같은 고급 기술 인력은 인적 네트워크 및 추천을 통해 채용될 확률이 높음을 시사

▶ **기술 유형별 차이를 파악하기 위해 기술 인재 채용 공고를 제1기술군*과 제2기술군**으로 분류**

* AI 설계·개발·배포 분야에서 종사 가능한 컴퓨터·수학 관련 인재 채용 공고

** AI 관련 업무 수행이 가능한 수준의 지식, 기술, 능력을 갖추고 있으나 재교육 및 역량개발이 필요한 공학·과학·기타 기술 관련 인재 채용 직무

- AI 키워드를 포함하고 있는 공고는 제1기술군 채용 공고의 10%(약 3만 6,000건)를 차지했으며 이 중 60%가 학사 학위를, 17%는 석사 학위가 요건인 것으로 확인
- 제1기술군과 제2기술군의 채용 공고 중 약 60%는 광동성, 장쑤성, 상하이 지역에 집중되었으며 이 중 절반에 달하는 30%가 유명 AI 기업 및 대학이 위치한 광동성 지역에 집중

출처 : CSET, China's AI Workforce: Assessing Demand for AI Talent, 2022.11.

구글, 전 세계 1,000개 언어 지원 가능한 거대 AI 언어모델 개발 계획

KEY Contents

- ✔ **구글은 세계에서 가장 많이 사용되는 1,000개 언어를 지원하는 거대 AI 언어모델 개발에 착수했으며 첫 번째 단계로 400개 이상의 언어로 훈련된 AI 모델 공개**
- ✔ **구글은 언어에 관해 광범위한 지식을 갖춘 단일 시스템을 통해 1,000개 언어 지원 예정**

주빈 가라마니(Zoubin Ghahramani) 구글 AI 부사장이 미국의 기술 전문 매체 버지(Verge)와의 인터뷰를 통해 구글의 거대 AI 언어모델 개발 프로젝트의 장기적 비전 설명

▶ **구글은 세계에서 가장 많이 사용되는 1,000개 언어를 지원하는 거대 AI 언어모델 개발을 위한 프로젝트를 진행 중이며 첫 번째 단계로 400개 이상의 언어로 훈련된 AI 모델 공개**

- 구글이 공개한 400개 언어 규모의 AI 모델은 현존하는 음성 모델 중 가장 큰 언어 적용 범위 포괄

▶ **언어 및 AI 영역은 이미 구글 제품의 핵심이었으나 고성능 다기능 대규모 언어모델(LLM) 개발 등 최근 기계학습(ML) 기술의 급격한 발전으로 해당 영역의 중요성이 재차 강조**

- 구글은 이미 대규모 언어모델을 구글 검색 등 자사 제품에 통합하기 시작함으로써 시스템 기능 향상
- 언어모델은 인종 차별·외국인 혐오 등 사회적 편견을 유발할 가능성과 인간의 감성으로 언어 분석을 할 수 없는 등 많은 결함이 있음에도 불구하고, 언어 생성에서 번역에 이르기까지 다양한 작업 수행 가능

구글은 특정 기능에 초점을 맞추는 것이 아닌 전 세계 언어에 관한 광범위한 지식을 갖춘 단일 시스템 구축에 중점을 두고 1,000개 언어 지원 모델 개발 프로젝트 진행

▶ **다양한 언어에 노출 및 훈련된 단일 모델을 구축함으로써 온라인 공간과 AI 학습 데이터세트 상에서 제대로 표현되지 않는 저자원 언어*에서도 더 나은 성능을 확보할 수 있을 것으로 기대**

* Low-resource language: AI 시스템 학습을 위한 정보자원, 즉 데이터가 상대적으로 부족한 언어

- 가라마니 부사장은 1,000개의 언어를 지원하는 방법은 1,000개의 서로 다른 모델을 구축하는 것이 아님을 강조하며 언어는 유기체와 같이 상호 진화했으며 특정 유사성을 지닌다고 언급
- 구글은 1,000개 언어를 지원하는 모델 개발을 위해 음성 녹음 및 서면 자료를 포함하여 저자원 언어에 관한 데이터 수집에 자금을 지원할 예정

출처 : Verge, Google plans giant AI language model supporting world's 1,000 most spoken languages: It's an ambitious project in its early stages, but Google thinks it will have benefits across its entire product ecosystem, 2022.11.02.

영국, 구제금융에 관한 정부 결정을 지원하는 AI 도구 개발

KEY Contents

- ✔ **유니버시티칼리지런던과 런던퀸메리대학교의 공동 연구진이 경제 위기 발생 시 정부의 구제금융 실행 여부 결정을 지원하는 AI 도구 개발**
- ✔ **연구진은 납세자의 예상 손실에 따른 다양한 구제 전략 비교를 위해 수학적 프레임워크 구축**

영국의 유니버시티칼리지런던(UCL)과 런던퀸메리대학교(QMUL)의 공동 연구진은 경제 위기 시 정부의 구제금융* 실행 여부 결정을 지원하는 AI 도구 개발

* 기업·은행·국가·개인 등의 도산 또는 지급불능 위기 시 이를 구제하기 위해 민간 및 공공 자금이 지원되는 것

▶ **정부가 시중 은행을 대상으로 구제금융을 실시하는 것이 장기적 관점에서 납세자를 위한 최선의 전략인지, 은행에 얼마를 공급해야 하는지, 주어진 상황에서 어떤 은행을 구제해야 하는지 등 평가**

- 연구진은 유럽은행감독청(CBA)의 데이터를 사용하여 해당 AI 도구의 알고리즘 테스트 수행
- 정부의 은행 구제금융은 재정적, 사회적, 정치적으로 영향을 미치는 복잡한 결정이며, 연구진이 개발한 AI 도구는 정부가 재정적 영향을 평가할 수 있도록 도움 제공

연구진은 납세자의 예상 손실에 대한 다양한 구제 전략 비교 위해 수학적 프레임워크 구축

▶ **구제 전략 비교 시 고려 요소에는 금융 위기 지속 예상 기간, 각 은행의 채무 불이행 가능성, 타 은행에 대한 채무 불이행의 파급효과, 은행에 대한 납세자의 지분 등 포함**

▶ **은행 구제금융 시, 은행에 대한 정부의 투자는 은행의 자산을 늘리고 채무 불이행 위험 완화**

- 정부의 자금 지원에 따라 발생하는 단기적 비용은 정부 재정에 더 큰 피해를 초래하는 은행의 채무 불이행을 방지함으로써 장기적으로 납세자들의 손실을 줄여 유의미한 결정으로 작용

▶ **연구진은 마르코프 결정 프로세스(MDP)*를 사용하여 주어진 시점에서의 정부 개입의 효과를 해당 프레임워크에 통합 후 최적의 구제금융 전략을 평가하기 위해 맞춤형 AI 알고리즘 개발**

* 수학적 제어 프로세스로 상태, 행동 및 전이확률 기반 최적의 의사결정 전략을 탐색하는 강화학습 기법

- AI 알고리즘 테스트 결과, 은행에 대한 납세자의 지분이 AI 모델을 통해 결정된 임계 값보다 큰 경우에만 정부의 구제금융이 최적의 결정으로 나타났으며, 은행에 구제금융 제공 후 납세자들을 위한 최선의 전략은 정부가 채무 불이행 방지를 위해 해당 은행에 지속 투자하는 것으로 확인

출처 : University College London, AI tool predicts when a bank should be bailed out, 2022.11.17.

허깅페이스(HuggingFace), AI 이미지 생성기의 편향성 탐색 도구 개발

KEY Contents

- ✔ **허깅페이스 연구진은 이미지를 생성하는 기계학습 모델의 편향성을 감지하는 도구 출시**
- ✔ **인간이 개발한 도구에서 인간의 편향성을 완전히 제거하는 것은 불가능하지만 편향성 탐색 도구를 통해 이용자에게 AI 시스템의 편향성에 관한 이해를 제공하는 것 가능**

오픈소스 AI 스타트업 허깅페이스는 이미지를 생성하는 기계학습(ML) 모델의 편향성을 나타내는 도구 "스테이블 디퓨전 편향성 탐색기(Stable Diffusion Bias Explorer)"를 10월 출시

▶ **스테이블 디퓨전 편향성 탐색기를 통해 이용자들은 특정 단어 조합이 AI 텍스트-이미지 생성기에서 어떻게 편향된 결과를 생성하는지 확인 가능**

- 예를 들면, 스테이블 디퓨전 편향성 탐색기는 AI 이미지 생성기에 "CEO"라는 단어를 사용할 경우 거의 항상 남자 이미지를 생성한다는 사실을 보여줌
- 그러나, AI 이미지 생성기가 생각하는 "야심 있는 CEO"와 "격려하는 CEO" 모습에는 차이 발생
- AI 이미지 생성기는 "야심 있는 CEO"의 이미지로 검은색과 파란색의 정장을 입은 다양한 남성의 모습 생성했으며 "격려하는 CEO"의 이미지로 동일한 수의 여성과 남성의 모습 생성

허깅페이스 연구진은 10월 자사 플랫폼에 이미지 생성기인 스테이블 디퓨전이 공개된 이후 이미지 생성 기계학습 모델의 편향성을 감지하는 도구 개발의 필요성 인식

▶ **기존에 텍스트-이미지 편향성 감지 방법이 없었기 때문에 연구진은 스테이블 디퓨전을 활용해 이 생성기가 무엇을 나타내며 잠재적으로 표현하는 바가 무엇인지 파악하기 위한 노력 착수**

- 연구진은 20쌍의 서술적 단어 목록을 만들어 절반은 "부드러운" 및 "격려하는"과 같이 여성적 단어와 나머지 절반은 "단호한" 및 "단단한"과 같은 남성적 단어를 코딩
- 이에 따라 이용자들은 편향성 탐색 도구를 통해 이러한 여성적/남성적 서술 단어를 "파일럿"에서 "CEO", "캐셔"에 이르는 150개의 직업 목록과 결합 가능

▶ **인간이 개발한 도구에서 인간의 편향성을 완전히 제거하는 것은 불가능하나 편향성 탐색기와 같은 도구가 이용자에게 AI 시스템의 편향성에 관한 이해를 제공하고 편향성을 역공학 (Reverse-engineer) 하는 데 도움이 될 것으로 기대**

출처 : PetaPixel, New Tool Allows Users to See Bias in AI Image Generators, 2022.11.09.

SPRi 소프트웨어정책연구소

홈페이지 : https://spri.kr/

보고서와 관련된 문의는 AI정책연구팀(jayoo@spri.kr, 031-739-7352)으로 연락주시기 바랍니다.