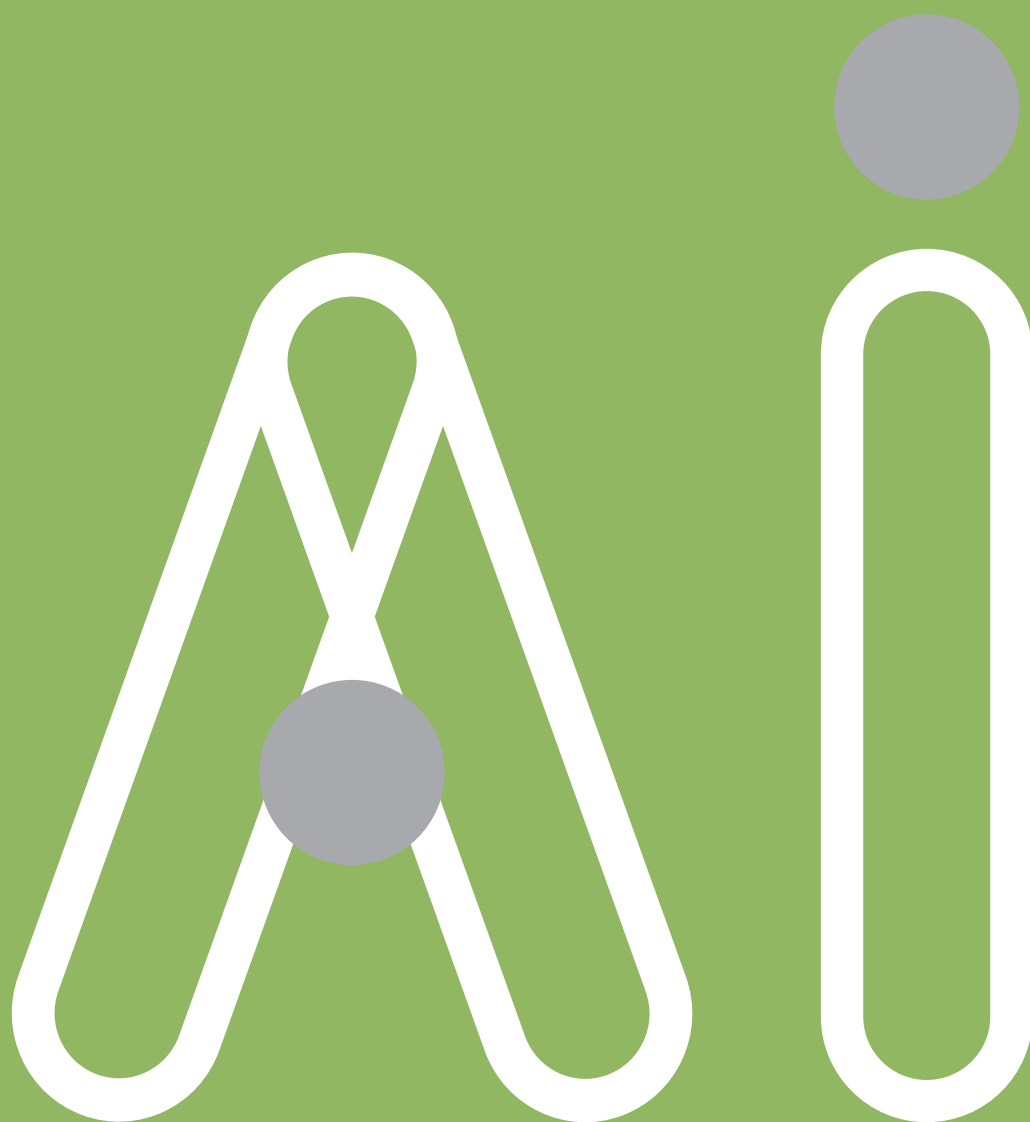


SPRi AI BRIEF

인공지능의 최신 동향



국내외 정책

미국, 국가 인공지능연구자원(NAIRR) TF 최종 보고서 발표	3
EU의 인공지능법, 생성AI 규제에 영향을 줄 것으로 전망	4

기업·시장 동향

마이크로소프트, OpenAI에 100억 달러 추가 투자 결정	5
게티 이미지, 저작권 침해로 AI 이미지 생성기 개발사 스태빌리티AI 고소	6
TechEU, 영화 제작에서 생성 AI 활용에 따른 혼란과 윤리 이슈 제기	7
Moonshot, 사진 및 텍스트 생성 AI 관련 윤리 및 저작권 문제 제기	8

고용·인력 동향

구직 및 경력 관리 도구로서 ChatGPT 활용 방안	9
마케팅 회사 코드워드(Codeword), 세계 최초 AI 인턴 채용	10

기술·연구 동향

美 프린스턴 대학생, AI가 에세이를 썼는지 판별하는 앱 개발	11
MIT, 美 공군 요원에게 AI 기초 교육 가능한 프로그램 개발	12
2023년 ICML 논문 제출에 ChatGPT 사용 금지	13
ChatGPT, 사이버 범죄에 악용 가능성 확대	14

미국, 국가 인공지능연구자원(NAIRR) TF 최종 보고서 발표

KEY Contents

- ✓ 미국은 지난 1년 6개월에 걸쳐 진행된 인공지능 연구자원의 공유를 통한 AI개발 진흥과 개발 생태계 확장을 위한 태스크포스(NAIRR TF) 활동의 최종보고서를 발간('23.1)
- ✓ 의회의 승인을 받으면 향후 6년에 걸쳐 AI연구 자원 구축에 26억달러 투자 계획

美, 인공지능 연구 및 개발에 필수적인 자원 접근에 대한 국가 연구 인프라 구축 로드맵 수립

- ▶ 미국은 2020년 국가 AI이니셔티브법에 근거한 NAIRR(National AI Research Resource) 태스크포스를 통해 국가 인공지능 연구 개발 자원을 활용하기 위한 로드맵 개발에 착수

- 백악관 과학기술정책실(OSTP)과 국립과학재단(NSF)에서 공동의장을 맡고 TF에는 정부, 학계, 민간 조직에서 각각 4명씩의 기술전문가가 동수로 참여하여 2021년 6월 출범

* 지난 1년 6개월 동안 11번의 공개 회의, 2번의 공개 의견(Request for Information) 수렴 진행

- ▶ NAIRR는 네 가지 측정 가능한 목표(△혁신촉진, △인재다양성 증대, △역량 향상, △신뢰할 수 있는 AI)를 염두에 두고 수립됨

- 컴퓨팅 리소스, 테스트베드, SW와 테스트 도구를 제공하고 AI교육 및 연구커뮤니티 구축을 통해 인재 다양성을 증대하며 책임있는 AI를 위한 표준 및 시스템 보호조치 수립 예정

향후 6년간 26억달러 이상을 투자하여 AI연구 자원을 구축하고 2년마다 최신화 추진

- NAIRR는 프로젝트를 위해 초기 6년간 \$US2.6B(약 3.2조원)의 예산을 의회에 요청
 - * 대부분의 예산은(US\$2.25B) 연구 자원 확보에 투자, 연구자원(수명주기 6년으로 산정)의 최신화를 위해 매 2년마다 \$750M(약 9.2천억원) 신규 투자, 운영주체(Operating Entity)의 NAIRR 활동 관리 조정을 위해 매년 55~65백만달러, 운영 주체와 NAIRR의 외부 평가를 위해 5백만달러 예산 매년 소요 예상
- 향후, 의회에 요청한 예산이 확정되면 4년에 걸쳐 단계적으로 프로그램 이행에 착수 예정
 - * (1단계) 행정부와 국가 AI이니셔티브실, 국가인공지능계획국에서 운영위원회 및 프로그램관리사무실 설립, 운영 주체 선정 (2단계) 운영주체 활동 설정, NAIRR 포털 구축 및 기술적, 정책적 제어 방식 설정 (3단계) 초기운영역량 달성 단계로 포털, 사용자 자원, 전산자원공급자, 할당 및 ID시스템, 데이터출판 시스템 등을 구비 (4단계) 지속적 운영 모드로 전환



EU의 인공지능법, 생성AI 규제에 영향을 줄 것으로 전망

KEY Contents

- ✓ 현재 급부상 중인 생성AI 모델을 활용한 인공지능 시스템은 빠르면 연내 제정이 예상되는 EU의 인공지능법의 규제 대상이 될 것으로 전망
- ✓ 이에 따라 이와 관련된 EU내의 인공지능 기업 및 투자자들의 활동 위축 예상

EU가 2023년 내에 도입을 준비 중인 인공지능법으로 생성 이미지 규제 가능

- ▶ EU의 인공지능법은 기업이 이미지 생성 AI 모델에 대해 보다 투명하게 공개하도록 요구
 - 이에 따라 오픈AI의 DALL-E, Stable Diffusion, Midjourney와 같은 대표적인 생성이미지 모델 역시 영향을 받을 것으로 예상
 - 실제로, Midjourney를 만든 David Holz는 경제전문지 Forbes와의 인터뷰에서 모델 학습에 인터넷 상의 이미지 수억장을 특별한 허락 없이 스크래핑해 활용했다고 언급

이 법안은 많은 기업들에 영향을 줄 것이며 특히 스타트업에 영향이 클 것으로 전망

- ▶ 유럽AI포럼에서 지난해 12월에 이뤄진 조사에서 응답자의 73%가 법안으로 인해 유럽에 기반을 둔 AI 기업들의 경쟁력이 약화될 것으로 전망(AppliedAi, '22.12)
 - 투자사 역시 이러한 규제 환경에 매력적인 투자 기업을 찾는 것이 비관적일 것을 우려해 유럽 외의 AI 스타트업 투자로 전환할 수도 있음을 시사
 - 또한, 투자자가 상대적으로 규제가 약한 저위험(Low-Risk) 애플리케이션에 투자할 것을 감안해 AI스타트업은 투자 유치를 위해 적절한 'Use Case' 제시에 대한 부담 가중
 - 유럽의 영향평가 보고서에는 약 5~15%가 고위험 시스템으로 분류되는 것으로 예상하고 있으나 111개의 스타트업에 대한 조사에서 약 33%가 자사 시스템이 고위험 범주에 속한다고 응답

향후 생성 AI모델 오남용 등 잠재적 문제에 대한 다양한 규제 예상

- ▶ 유럽의 인공지능법은 작년 12월 유럽의회에서 공동입장문을 채택했으며 올 3월에는 유럽이사회(Ec)가 합의안을 채택할 것으로 전망
 - 현재 챗봇과 같은 인공지능은 유럽 인공지능법에 따르면 저위험에 속하고 있으나 생성모델의 응용분야가 유럽 인공지능법에서 규정하는 분야에 속하게 되면 높은 수준의 규제 전망

마이크로소프트, OpenAI에 100억 달러 추가 투자 결정

KEY Contents

- ✓ 마이크로소프트, 오픈AI(OpenAI)와 파트너십을 통해 인공지능 슈퍼컴퓨팅 및 연구 전반에 걸친 지속적인 협력 확장과 고급 AI 기술의 독립적 상용화 추진
- ✓ 마이크로소프트와 오픈AI는 대표적으로 ▲대규모 슈퍼컴퓨팅 ▲새로운 AI 기반 구축 ▲독점 클라우드 제공 협력을 추진하며 장기적 파트너십 강화

마이크로소프트, 오픈AI와 장기적인 파트너십을 체결하며 100억 달러(약 12조 7000억원) 추가 투자 진행

- ▶ 마이크로소프트는 최첨단 AI 연구를 책임감 있게 발전시키고 신기술 플랫폼의 AI 상용화를 목적으로 오픈AI와 2019년, 2021년에 거쳐 올해 세 번째 파트너십을 체결
 - 업계 전반의 개발자와 조직이 자사 클라우드 플랫폼인 애저(Azure)를 통해 최고의 AI 인프라, AI모델 및 튜체인에 접근하여 애플리케이션을 구축하고 실행할 수 있는 환경 조성 목표
 - 오픈AI는 마이크로소프트의 대규모 슈퍼컴퓨팅 인프라를 사용하여 GitHub Copilot, DALL-E 2 및 ChatGPT와 같은 AI 제품을 지원하기 위해 애저에 배포된 모델 교육 진행

마이크로소프트는 오픈AI와 3가지 대표적 협력 추진

- ▶ 양사는 3가지 분야에 대해 협력을 추진하며 마이크로소프트는 향후 오픈AI를 이용해 애저 서비스를 강화하고 자사 보유 소프트웨어에 적용하는 계획 수립
 - **대규모 슈퍼컴퓨팅** 오픈AI의 독립적인 AI 연구 가속화를 위해 슈퍼컴퓨팅 시스템의 개발 및 배포에 투자 증액 예정
 - ★ 사용자들이 대규모 AI 애플리케이션을 구축하고 배포할 수 있도록 애저의 선도적 AI 인프라 지속 구축 예정
 - **새로운 AI 기반 구축** 마이크로소프트는 소비자 및 엔터프라이즈 제품에 오픈AI 모델을 배포하고 오픈AI 기술을 기반으로 구축된 새로운 디지털 인프라 제공
 - ★ 개발자가 애저의 신뢰할 수 있는 엔터프라이즈급 기능과 AI 최적화 인프라 및 도구로 지원되는 오픈AI 모델에 대한 직접 접근을 통해 최첨단 AI 애플리케이션 구축이 가능하도록 지원하는 마이크로소프트 서비스 포함
 - **독점 클라우드 제공** 애저를 오픈AI의 독점 클라우드로 제공하면서 연구, 제품 및 API 서비스 전반에 걸쳐 모든 오픈AI 워크로드를 지원

게티 이미지, 저작권 침해로 AI 이미지 생성기 개발사 스테빌리티AI 고소

KEY Contents

- ✓ AI 이미지 생성기 스테이بل 디퓨전(Stable Diffusion) 개발사인 스테빌리티AI(Stability AI)는 저작권 침해 문제로 게티 이미지(Getty Images, 이하 게티)로부터 소송에 직면
- ✓ AI 생성 이미지에 대한 저작권 문제 지속적 논란 야기

게티이미지, 스테빌리티AI가 자사 유료 이미지를 무단으로 활용했다며 소송 제기

- ▶ 게티는 성명 발표를 통해 스테빌리티AI의 이미지 생성기 스테이블 디퓨전 학습에 게티의 유료 이미지가 활용되었음에도 라이선스 구매 및 사용에 대한 사전 고지가 없었음을 강조하며 지적재산권 보호를 위해 저작권 침해 소송 제기
- ▶ 게티의 이미지와 메타데이터 약 1,200만 개를 허가 없이 무단으로 사용하여 불법적 복사, 재가공을 통해 스테이블 디퓨전 모델의 학습 활용 등 스테빌리티AI의 상업적 이익 취득 혐의 고발
 - 게티는 사진에 워터마크를 적용하여 게시하였으나 스테이블 디퓨전은 이를 그대로 학습하여 게티의 워터마크를 포함한 이미지를 생성
 - 생성된 이미지의 워터마크 상표는 원본과 동일하지 않으며 변형되어 있지만 인식 가능한 이미지를 생성함으로써 게티의 상표권을 침해, 변형을 통한 상표 가치 희석 주장

AI 생성 이미지에 대한 저작권 문제 지속, 공정 거래 기준에 따라 상반되는 주장

- ▶ AI 생성 이미지에 대한 저작권 문제가 지속적으로 발생하고 있는 가운데, 스톡 이미지 공급 업체와 AI 이미지 생성기 개발사 간 의견 대립
 - 공정 이용(Fair Use)* 개념을 논제로 의견이 대립하고 있으나 국가 간 규정이 상이하고 현재 AI 생성 이미지 관련한 기준이 모호한 형태로 논란 지속
 - * 저작권으로 보호되는 저작물을 저작권자의 사전 허가 없이 제한적으로 이용할 수 있도록 허용
- ▶ AI 창작물을 둘러싼 저작권 문제, 재판 결과로 향후 방향성 전망 가능
 - 게티 및 캘리포니아 예술가들의 AI 이미지 생성기 대상 소송 사례 등은 향후 예술 분야에서 AI 활용 시 저작권, 상표권 문제 등에 기준이 될 수 있는 사례들로 재판 결과에 따른 법의 적용 기준 파악 필요

TechEU, 영화 제작에서 생성 AI 활용에 따른 혼란과 윤리 이슈 제기

KEY Contents

- ✓ 영화산업에서 AI 기업들은 생성 AI를 활용하여 새로운 형태의 서비스를 제공하고 있으나, AI 적용 관련 윤리적 이슈로 논란 존재
- ✓ 저작권, 허가 등 윤리 이슈가 혼란을 야기하고 있지만, 생성 AI는 영화 제작자를 위한 강력한 도구로써 영화산업 변화의 동인으로 작용

영화산업에서 AI 기술의 활용은 새로운 서비스 등장과 함께 윤리적 이슈 야기

- ▶ 스타트업 Colossyan은 생성 AI*를 기반으로 초현실적인 배우를 생성해내고, 스크립트를 삽입하여 짧은 시간 내에 합성 비디오 생성이 가능하도록 지원
 - * 학습된 데이터를 기반으로 결과를 생성하는 인공지능 모델로, 질문을 입력받아 사람처럼 대답하는 ChatGPT나 묘사된 텍스트로부터 그림을 만들어내는 DALL-E 등이 대표적인 사례
- 제작하는 모든 비디오 콘텐츠에 보이지 않는 워터마킹 기술을 사용하여 추적을 가능하게 하고 의도치 않은 사용을 방지할 수 있도록 하는 등 윤리적 기준 고려한 서비스 제공
- ▶ Flawless AI는 생성 AI를 기반으로 더빙 언어에 맞춰 기존 영상을 조정, 영화 내 욕설 제거 등 TrueSync 소프트웨어를 통해 영화산업에 새로운 AI 활용 사례 제시
 - 기존 더빙 영화에서의 시각적 문제를 AI를 통해 다양한 언어의 대화를 입 움직임과 일치하도록 만들었고, 영화 내 욕설 제거 작업에 AI를 활용하여 비용 절감
 - 일부 영화 비평가들은 AI를 활용하여 생성된 음성이나, 허가 없이 사용된 데이터 등과 관련하여 제작된 영화에 윤리적 문제를 제기

윤리, 직업 대체 등 AI 활용에 우려가 있으나, 강력한 도구로 AI 활용 가능성 강조

- ▶ 기존 영화산업에 구축된 권리, 허가, 보상 및 동의 등의 문제는 새로운 기술 수용을 위해 지속적으로 개선이 필요하며, AI 기술을 통해 영화산업은 진화 가속
 - 전통 산업에서 AI 활용에 있어 보호장치가 필요하나, 많은 창의적 기회가 창출됨에 따라 영화계에서도 새로운 변화 수용 필요
 - AI는 더 많은 영화 제작 도구를 가능케 할 것이고, 제작 비용 감소 효과와 제작 및 영화 품질의 향상으로 산업에 긍정적 효과 전망

Moonshot, 사진 및 텍스트 생성 AI 관련 윤리 및 저작권 문제 제기

KEY Contents

- ✓ AI의 활용은 비용의 획기적 감소와 같은 이점이 있으나, 데이터 및 콘텐츠 활용 관점에서의 논란 존재
- ✓ AI 기반 서비스 관련 위험과 윤리, 저작권 및 기타 권리에 대한 문제 제기

AI 시스템의 활용은 강력한 잠재력을 가지고 있으나, 데이터 및 생성 콘텐츠 관련 논란 존재

- ▶ ChatGPT와 같은 생성 AI는 강력한 기능을 보이지만, 저작권법의 범위 내에서 충분히 독창적이고 창의적인 텍스트를 제작한 것인지에 대한 논란 존재
 - 마이크로소프트는 ChatGPT를 Bing(Bing) 검색엔진에 적용하여 사용자의 질문에 링크를 주는 것뿐 아니라 기계가 작성한 요약된 답변까지 제공하는 것 고려
 - 한편, 비평가들은 AI가 인간처럼 생성하는 것이 아니라 수집된 방대한 양의 기존 데이터와 심층 기계 학습에 기반하는 것으로 새롭게 생성된 것이 아님을 강조
 - 데이터 및 콘텐츠는 보호될 필요가 있고, 무단 사용 시 이는 타인의 권리를 침해

AI의 위험성과 이에 따른 윤리, 저작권 및 기타 권리에 대한 갈등 발생

- ▶ 가트너의 분석가는 파운데이션 모델의 등장으로 비용과 시간 감소 등 AI의 이점을 강조하는 한편, 위험성과 윤리적 문제 또한 제기
 - CSET*에 따르면 5가지 관점에서 복잡성에 따른 비용 및 환경문제, 권력 집중 문제, 딥페이크, 예상치 못한 결과, 콘텐츠 재사용에 대한 법적 판례 부재와 같은 다양한 위험 존재
- ▶ 유럽의 인공지능 법에서는 허용할 수 없는 위험, 고위험, 그 외 위험으로 분류하고 AI 애플리케이션에 대한 규제 적용
 - 하지만, 명시적으로 금지되지 않은 기타 애플리케이션들은 대부분 규제되지 않은 상태

* Center for Security and Emerging Technology(CSET) within Georgetown University's Walsh School of Foreign Service

구직 및 경력 관리 도구로서 ChatGPT 활용 방안

KEY Contents

- ✓ 오픈AI가 개발한 ChatGPT는 구직자, 재직자, 고용주를 돕는 핵심 도구로 주목
- ✓ 구직자는 진로탐색부터 고용계약까지 전과정에서 ChatGPT를 유용하게 활용할 수 있고, 재직자는 경력 개발에, 그리고 고용주는 인적자원관리 전반에 ChatGPT 활용

오픈AI가 개발한 챗봇 ChatGPT는 구인·구직 및 경력 관리 도구로 주목

- ChatGPT는 전세계의 방대한 지식을 학습한 언어 생성 모델로서 채용 및 취업 과정에서 발생하는 다양한 정보 검색이나 조언에 대한 요청에 적절히 답하는 것이 가능

구직자, 재직자 및 고용주는 다양한 상황에서 ChatGPT 사용 가능

▶ 구직자는 진로탐색에서 면접 및 채용 결정 후 처우조건 협상까지 전체 과정에서 활용

- **진로 탐색** 전공분야에서 채용 가능성을 높이기 위한 방안을 질문하거나 자신의 성격 및 이력서를 제시한 후 적합한 직업을 추천할 것을 요청
- **이력서 초안 작성** 이력서의 기본적인 틀에 맞춰 초안을 작성하도록 요청한 후 오류나 관련 없는 정보를 삭제하고 교육, 보유 기술, 경험 등의 이력사항을 덧붙여 완성
- **면접 준비** 면접에서 흔히 나오는 일반적인 질문에서 산업별, 업무별 질문에 이르기까지 인터뷰 질문을 생성하거나 면접 대상 기업에 대한 정보 검색

▶ 재직자는 일상 업무의 능률 개선 및 경력 개발을 위해 ChatGPT를 활용

- **보조 업무** 아이디어 생성, 마케팅 카피 작성, 보고서 초안 작성, 업무용 이메일 작성, 브레인스토밍, 문서 수정 및 요약, 제목 제안 등 개인 비서 업무에 활용
- **멘토 역할** 직업에 필요한 역량 확보 및 경력 개발에 필요한 조언 요청, 필요한 보고서 및 정보 탐색 등을 통해 직업인으로서의 경쟁력 강화

▶ 고용주는 구인 및 인적 자원 관리 전반에 걸쳐 ChatGPT 활용

- **구인 과정** 구인 광고 문구, 인터뷰 질문, 채용 계약서 초안, 직무 설명서 등 작성
- **성과 관리** 고객 피드백 분석, 업무 성과 관리, 고객 피드백 추적 등

마케팅 회사 코드워드(Codeword), 세계 최초 AI 인턴 채용

KEY Contents

- ✓ 미국의 마케팅대행사 코드워드(Codeword)는 ChatGPT 등 초거대 AI를 이용하여 인간과 협업하는 AI 인턴을 개발하여 디자인팀과 편집팀에 배치
- ✓ 3개월간의 인턴기간동안 인간과 AI가 협업하는 방식을 실험한 후 가능성이 입증되면 정규업무에 본격적으로 활용할 방침

마케팅 회사 코드워드는 ChatGPT 등 초거대 AI를 이용해 일을 하는 AI 인턴 개발

- 2명의 AI인턴 Aiko와 Aiden은 각각 디자인팀과 편집팀에 배치되어, 문서나 이미지를 생성하는 초거대 AI인 ChatGPT, Midjourney, DALL-E 2를 조합하여 부여되는 업무 수행
- ChatGPT와 같은 AI를 이용하여 AI 인턴을 제작했으며, 인턴의 이름 역시 AI로 시작하는 이름을 제안하도록 요청하여 결정
- 코드워드는 AI 인턴 채용의 목적이 인간과 AI가 협업하는 방식을 실험하기 위함이며 AI로 인간을 대체하지는 않을 것이라고 언급

코드워드는 3개월간 AI 인턴을 채용한 후 정식 활용 여부를 결정할 방침

- ▶ 코드워드는 인간이 AI에게 업무를 부여하고 AI가 창의적인 방식으로 업무를 수행하면 인간이 결과를 개선하고 선택하는 업무방식을 “샌드위치 워크플로우”라고 지칭

- **업무 부여** AI 인턴이 인간보다 더 창의적이고 효율적으로 할 수 있는 일을 찾고 이를 프롬프트로 만들어 AI 인턴에게 부여
- **업무 수행** 디자인팀에 소속된 Aiko는 사진 편집, 개념 스케치 및 이미지 초안 작성, 아이콘 디자인 등을 수행하며, 편집팀에 소속된 Aiden은 뉴스 분석, 마케팅을 위한 음성 및 어조 연구, 마케팅 콘텐츠 초안 작성 등의 업무 수행
- **평가 및 피드백** 디자인팀과 편집팀의 책임자는 매달 AI 인턴의 업무 실적, 잠재 능력, 적합한 활용 방안 등의 내용을 담은 실적보고서를 작성
- **보상** 3개월의 인턴기간 동안 작성한 실적보고서를 토대로 AI 인턴을 정규 업무에 본격적으로 활용할지 여부를 결정할 예정이며, 기 편성된 인턴급여는 기부할 예정

美 프린스턴대학생, 에세이를 AI가 썼는지 판별하는 앱(GPTZero) 개발

KEY Contents

- ✓ 미 프린스턴대 컴퓨터공학과와 한 학생이 ChatGPT를 활용하여 에세이를 작성했는지를 판별하는 앱(GPTZero)을 개발하여 배포

美 프린스턴대(Princeton Univ.) 학생이 AI가 쓴 에세이를 감별하는 앱 GPTZero 개발

- ▶ 컴퓨터공학과 4학년에 재학 중인 Edward Tian은 ChatGPT와 같은 인기 있는 챗봇이 쓴 에세이를 제출하는 부정행위를 적발하기 위한 목적으로 GPTZero를 개발하여 배포
 - GPTZero는 무료로 배포되었으며, 출시 일주일 내 3만 명 이상의 사람들이 활용하는 등 대중의 관심이 급상승*
 - ★ 국내에서도 국제학교 학생들이 ChatGPT로 과제 에세이를 대필하여 GPTZero로 적발
 - 이 앱을 호스팅하는 무료 플랫폼 Streamlit는 앱의 인기로 인해 증가한 웹 트래픽을 처리하기 위해 더 많은 컴퓨팅 리소스를 무료로 투입하여 지원

GPTZero는 ‘글의 복잡성’과 ‘문장의 변형’이라는 두 가지 척도를 통해 AI 작성 여부를 감별

- ▶ 글의 복잡성(Perplexity)은 인간이 작성한 글이 AI가 작성한 글보다 더 복잡하고 난해할 수 있다는 점에서, 문장이 단순한 경우 AI가 작성한 글로 판단
- ▶ 문장의 변형(Burstiness)은 인간의 경우 짧은 문장과 더 길고 복잡한 문장을 혼용해서 글을 쓰지만, AI의 경우 문장이 대체로 균일한 경향이 있다는 점에서 감별
- ▶ GPTZero의 성능을 테스트한 결과 완벽하지는 않으나 감별을 성공적으로 수행*
 - ★ △The New Yorker의 스토리와 ChatGPT가 작성한 LinkedIn 게시물에 대한 비교 결과 인간과 AI의 글쓰기를 성공적으로 구별, △제작자인 Tian은 일부 사용자의 테스트 결과에 나타난 것과 같이 자신의 AI가 완벽하지 않음을 인정
- ▶ 그러나, 인간과 AI를 구분하는 요소를 조명하는 앱을 설계하고 AI의 투명성을 제공하려는 시도라는 점에서 의의가 있으며, AI신뢰성을 위한 유사연구도 활발히 진행*
 - ★ △OpenAI의 AI 안전 개발자인 Scott Aaronson은 ChatGPT가 생성한 글을 구별하기 위한 워터마크를 심는 방안을 연구 중이라고 밝혔으며(22.12.), △오픈소스 커뮤니티 Hugging Face는 GPT-2에 의해 생성된 글을 감지하는 도구를 개발하여 공개(<https://openai-openai-detector.hf.space/>)

MIT, 美 공군 요원 대상의 AI 기초 교육 프로그램 개발

KEY Contents

- ✓ MIT는 대규모 조직에서 다양한 배경의 구성원들에게 AI를 가르칠 수 있는 맞춤형 교육 프로그램을 개발

MIT는 미 공군 및 우주 요원을 대상으로 AI 기술을 가르칠 수 있는 프로그램 개발

- ▶ 기존 MIT의 AI 교육자료 및 리소스를 활용* 하여 리더, 개발자, 사용자 등 세 가지 유형의 군인을 위한 6~12개월 코스의 온라인 커리큘럼을 구축

* MIT Horizon, MIT Lincoln 연구소, MIT Sloan 경영대학원, CSAIL(Computer Science and Artificial Intelligence Laboratory), Media Lab의 기존 AI 과정과 자료를 조합하여 구성

- 리더(Lead-Drive)는 전략적 결정에 AI 활용, 개발자(Create-Embed)는 AI 솔루션 구현, 사용자(Facilitate-Employ)는 AI 증강 도구 사용 등에 맞춘 개별 커리큘럼 구성*

* △리더 과정: 윤리 및 인간 중심 AI 설계에 대한 더 많은 교육 할당, 자율 로봇으로 실습 활동(3일) 포함
△개발자 과정: AI 기술 자료에 대한 심층 분석 및 Lincoln 연구소 과학자가 가르치는 대면 실습(5일) 포함
△사용자 과정: 보다 일반적인 대상에 맞춰 설계된 MIT Horizon 자료 활용, 비동기식 자기주도형 학습

공군·우주의 광범위한 임무부터 일상 업무까지 AI 활용사례를 구축, 학습 효과를 제고

- ▶ MIT는 미공군(USAF) 및 국방부(DoD)와 AI 중점 교육 연구 프로그램에 대한 계약을 체결하고, 공군 및 국방부 직원부터 일반인에 이르는 다양한 학습 트랙 설계를 추진 중

- 교육 과정 및 분석 결과는 IEEE 프론티어 교육 컨퍼런스(IEEE Frontiers in Education Conference) 아카이브에 공개

- ▶ MIT는 이 커리큘럼으로 약 18개월간 교육을 수행하면서 230명 학습자의 성취도를 분석하고 설문문을 통하여 교수법 및 콘텐츠 등 프로그램을 개선

- 이 교육은 조직과 개인의 일정을 고려하여 동기식 및 비동기식으로 진행되었으며, 교육 수강자들은 시간 효율적이고 학습 경험 및 성취가 높다고 평가

2023년 ICML 논문 제출에 ChatGPT 사용 금지

KEY Contents

- ✓ 세계 최고 수준의 AI 학회 중 하나인 ICML이 2023년 학술대회 논문 제출 공고에 ChatGPT와 같은 거대 언어 모델 활용을 금지하자 논란이 발생
- ✓ ICML은 논란을 수렴하여 ChatGPT와 같은 거대 언어 모델의 활용에 대한 금지 사항을 명시하였으나, 저자의 자격은 불인정

ICML은 2023년 학술대회의 논문 제출 공지에 ChatGPT 등 거대 언어 모델 활용을 금지 → 논란이 불거지자 논문 전체가 생성되는 경우를 금지

- ▶ ChatGPT의 우수한 성능으로 일부 연구계에서 ChatGPT를 논문의 저자로 등록하는 등의 해프닝이 일어나자, ICML은 2023년 논문 제출 공지에서 원칙적으로 ChatGPT와 같은 거대 언어 모델을 활용해 생성한 문장의 활용을 금지
- ▶ ICML의 이러한 입장은 AI 석학과 연구계의 거센 비난으로 인해 노선을 변경하여 논문 전문을 거대 언어 모델로 작성하는 경우를 금지
 - Meta의 AI 책임자인 안 르쿤은 “중소규모 언어 모델은 사용해도 되는 것인가요? 우리는 철자 교정과 타이핑 예측에 언어 모델을 사용하고 있습니다.”라며 ICML의 조치에 대해 반발함
 - MS 리서치의 기계학습 기반팀의 리더인 세바스찬 부베크는 “ChatGPT와 거대 언어 모델을 단순히 금지시키는 것은 매우 근시안적”이라며 비판

학술 논문에서의 ChatGPT는 저자라기 보다는 도구와 유사

- ▶ 허깅 페이스의 윤리 과학 수장을 맡고 있는 마가렛 미첼은 ChatGPT를 저자라기 보다 도구에 가깝게 보는 것이 적절하다고 제안
 - 특히 ChatGPT는 비영어권 연구자에게 논문 작성을 도와주는 형태로 활용될 수 있기 때문에 평등을 실현할 수 있는 도구로 보는 것이 바람직
 - ChatGPT가 학술 논문에 기여할 수 있다는 사실은 자명하나, 도구의 관점으로 본다면 저자의 자격을 부여하는 것 자체가 부적절

ChatGPT, 사이버 범죄에 악용 가능성 확대

KEY Contents

- ✓ ChatGPT를 활용해 해킹 도구 생성이 가능함에 따라 사이버 보안의 잠재적인 위협으로 부상
- ✓ ChatGPT의 악용 사례는 비단 사이버 보안의 영역을 벗어나 인터넷 전반의 사이버 범죄 도구로 활용 가능성 제기

이스라엘 보안 기업 Check Point는 사이버 해킹에 ChatGPT의 악용 사례 조사

- ▶ 한 포럼에서는 ChatGPT를 활용해 파일 탈취를 목적으로 하는 안드로이드 멀웨어를 생성하고 공유함
- ▶ 또한 한 개발자는 암호화된 파일을 해독하는 파이썬 프로그램을 ChatGPT의 코드 생성 기능을 활용해 작성했다는 주장 제기
- ▶ Check Point는 사이버 보안의 ChatGPT 악용 사례를 조사한 보고서에서 ChatGPT로 인한 사이버 보안 위협의 확산을 경계
 - 사이버 보안 전문가들은 ChatGPT가 생성한 악성 코드나 랜섬웨어는 기초적인 수준에 머물러있다고 판단함
 - 그러나 ChatGPT를 통해 초심자가 손쉽게 해킹 코드를 얻을 수 있다는 점은 경계하여야 하며 OpenAI 차원에서의 대응이 필요함

ChatGPT가 야기하는 다양한 악용 사례는 디지털 사회에 잠재적인 위협으로 작용

- ▶ 한 채팅 앱에서는 ChatGPT를 활용해 젊은 여성을 가장하여 다양한 사기 수단으로 활용하는 움직임 관측
- ▶ ChatGPT를 사용해 폐쇄적인 정보 공유로 운영되는 다크웹의 정보를 획득할 수 있음에 따라 다양한 사회 문제의 진원지가 될 가능성 존재
- ▶ 대부분의 전문가들은 OpenAI측에서 범죄와 연루된 콘텐츠 활용에 대한 기술적 대책 마련이 시급할 것으로 예측



홈페이지 : <https://spri.kr/>

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