

SPRi AI Brief

인공지능 산업의 최신 동향

2023년 11월호

CONTENTS

I. 인공지능 산업 동향 브리프

1. 정책/법제

- ▷ 영국 경쟁시장청, AI 기반모델 검토보고서에서 7대 AI 이행원칙 발표 1
- ▷ 미국 국가안보국, 사이버보안 협력 센터 산하에 AI 안보센터 설립 2
- ▷ 일본 정부, 2023년 말까지 AI 지침 수립 계획 3
- ▷ 프랑스 개인정보보호 감독기구, AI 개발 시 개인정보 규제 준수를 위한 실무표 공개 4

2. 기업/산업

- ▷ 구글, 차세대 멀티모달 AI '제미니' 연말 출시 계획 5
- ▷ 앤스로픽, AI 시스템의 위험 관리를 위한 '책임 있는 확장 정책' 발표 6
- ▷ 게티이미지, 생성 AI 이미지 서비스 공개 7
- ▷ 일론 머스크의 AI 스타트업 xAI, 생성 AI 챗봇 '그록(Grok)' 공개 8
- ▷ 오픈AI, 첫 개발자회의에서 'GPT-4 터보' 공개 9
- ▷ 구글, 테스트 단계의 생성 AI 기반 검색 서비스에 이미지 생성 기능 추가 10

3. 기술/연구

- ▷ 마이크로소프트 연구진, AI 모델에서 특정 정보에 대한 삭제 기법 개발 11
- ▷ 중국 국가정보보안표준화기술위원회, 생성 AI 서비스 보안 요구사항 초안 발표 12
- ▷ 대규모 언어모델의 안전 가드레일, 미세조정에 취약 13
- ▷ 모든 구글 검색에 AI 결합 시 에너지 사용량 10배 증가 전망 14
- ▷ 마이크로소프트 연구진, GPT-4V의 다양한 활용 가능성 탐색 15

4. 인력/교육

- ▷ 세계경제포럼, 금융 부문 일자리에 생성 AI의 최대 영향 예측 16
- ▷ IBM, 2026년 말까지 200만 명에게 AI 교육 제공 계획 17
- ▷ BCG, 챗GPT가 창의적 사고에는 효과적, 비즈니스 문제 해결에는 악영향 18

II. 주요 행사

- ▷ Generative AI Europe 2023 19
- ▷ GovAI Summit 19
- ▷ AAAI-24 19

I . 인공지능 산업 동향 브리프

영국 경쟁시장청, AI 기반모델 검토보고서에서 7대 AI 이행원칙 발표

KEY Contents

- 영국 경쟁시장청이 기반모델에 대한 검토보고서를 발표하고 소비자 보호와 건전한 경쟁을 보장하기 위한 7대 AI 이행원칙을 제안
- 영국 경쟁시장청은 △모델 개발 △시장 내 모델 활용 △소비자의 모델 사용이라는 세 가지 관점에서 원칙을 제시

○ 영국 경쟁시장청, 소비자 보호와 건전한 경쟁 보장을 위한 AI 원칙 제시

- 영국의 반독점 규제기관 경쟁시장청(Competition and Markets Authority, CMA)이 2023년 9월 18일 기반모델(Foundation Model)에 대한 검토보고서에서 7대 AI 이행원칙 발표
 - 보고서는 기반모델을 다양한 목적으로 활용할 수 있는 폭넓은 기능을 갖춘 AI 시스템으로 정의하고, 기반모델이 다양한 산업 및 일상적인 생활 업무 방식을 변화시킬 수 있는 잠재력이 있다고 평가
 - 기반모델은 신제품과 서비스 개발, 정보 접근권 향상, 과학과 의료 혁신을 지원함으로써 생산성을 높이고 경제 성장을 촉진 가능
 - 그러나 개발자가 소비자 보호에 주의를 기울이지 않으면 허위 정보나 AI를 이용한 사기 피해가 발생할 수 있으며, 소수의 기업이 시장을 독점하면 제품 가격이 높아지거나 품질이 저하될 가능성도 존재
- CMA는 기반모델이 제공하는 혁신적 혜택을 온전히 누릴 수 있도록 △모델 개발(접근성/다양성) △시장 내 모델 사용(선택 가능성/유연성/공정성) △소비자의 모델 사용(투명성)의 세 가지 관점에서 원칙을 제시(책임성은 전 영역 포괄)
 - (책임성) 기반모델 개발자와 배포자는 소비자에게 제공되는 기반모델의 결과물에 책임을 지님
 - (접근성) 모델 개발에 필요한 데이터, 컴퓨팅 자원, 전문성, 자본에 대한 과도한 규제를 피하고, 선점자 우위로 인해 후발 사업자의 경쟁력이 저해되지 않도록 접근성을 보장
 - (다양성) 개방형과 폐쇄형을 포함한 다양한 형태의 비즈니스 모델은 시장 발전을 이끌 수 있으며, 개방형 모델은 진입장벽을 낮추는 데 효과적
 - (선택 가능성) △자체 개발 △협력관계 체결 △API나 플러그인 활용과 같은 기반모델의 도입 방식 결정에서 충분한 선택지 제공
 - (유연성) 기업에 필요에 따라 여러 기반모델을 사용할 수 있는 유연성을 보장하는 한편, 소비자도 특정 사업자나 생태계에 구속되지 않고 복수의 서비스를 전환 사용할 수 있도록 보장
 - (공정성) 최고의 제품과 서비스가 성공을 거둔다는 시장 신뢰를 구축하고, 자사 우대(self-preferencing) 행위, 번들링을 포함한 반경쟁적 행위 금지
 - (투명성) 기반모델이 생성한 콘텐츠의 위험성과 한계에 대한 정보 제공을 통해 소비자와 기업의 합리적 선택을 보장

미국 국가안보국, 사이버보안 협력 센터 산하에 AI 안보센터 설립

KEY Contents

- 미국 국가안보국이 기존의 AI 관련 안보 활동을 통합하고 외부와 긴밀한 협력을 수행할 조직으로 AI 안보센터를 설립
- 폴 나카소네 국장은 국가 안보에서 AI의 영향력이 갈수록 커지고 있다며, 미국의 AI 리더십을 저해하려는 적대국을 경계해야 한다고 강조

○ AI 안보센터, 국가 안보 시스템에서 AI 기능의 안전한 통합과 개발 지원

- 미국 국방부 산하의 정보기관인 국가안보국(National Security Agency, NSA)이 2023년 9월 28일 국가 안보 체계에서 AI 기능의 개발과 통합을 감독할 목적으로 AI 안보센터를 설립한다고 발표
- AI 안보센터는 국가 안보 기관과 방위산업 전반에서 신규 AI 기능의 안전한 도입을 위한 모범사례와 평가 방법론, 위험 프레임워크 개발을 주도하게 될 전망
- AI 안보센터는 외부의 AI 기반 위협으로부터 국가 안보를 수호하는 한편, 미국의 AI 시스템을 겨냥한 위협에 대한 방어도 담당
- AI 안보센터는 기존 NSA의 AI 관련 안보 활동을 통합하고, 국방부와 국립 연구소, 학계 및 동맹국 간 협력을 촉진하는 역할을 수행

○ 폴 나카소네 NSA 국장, 중국과 러시아에 대응한 AI 안보의 중요성 강조

- 폴 나카소네(Paul Nakasone) NSA 국장에 따르면 AI 안보센터는 사이버보안 협력 센터(Cybersecurity Collaboration Center) 산하에 설치할 예정
- 사이버보안 협력 센터는 민간 및 국제 파트너와 협력해, 중국과 러시아처럼 활발한 사이버 공격을 벌이는 국가들로부터 미국을 보호하는 역할을 수행
- 나카소네 국장은 미국이 AI에서 선두를 유지하고 있지만 이를 당연하게 여겨서는 안 된다며, 미국의 AI 발전을 방해하려는 적대국들의 움직임을 경계해야 한다고 강조
- 그는 AI가 국가 안보에 점점 더 큰 영향을 미칠 것이라며, AI 시스템의 취약점과 위협에 대처하는 방법을 파악해 적대국이 미국의 혁신적인 AI 역량을 훔칠 수 없도록 해야 한다고 지적

출처: U.S. Department of Defense, AI Security Center to Open at National Security Agency, 2023.09.28.

Engadget, The NSA has a new security center specifically for guarding against AI, 2023.09.30.

일본 정부, 2023년 말까지 AI 지침 수립 계획

KEY Contents

- 일본 정부가 AI 개발과 사용 시 고려해야 할 기본 규칙을 담은 AI 지침을 수립할 계획으로, 현재 공개된 초안은 편향된 데이터 사용 금지, 공정성과 투명성 확보 등의 내용을 포함
- 일본은 2023년 G7 의장국으로서 G7 국가 간 생성 AI 거버넌스 논의도 주도하고 있으며, 올 하반기 G7 화상 정상회의에서 AI 개발 지침에 대한 합의를 추진

○ 일본 정부, AI 개발과 사용의 기본 규칙을 제시한 AI 지침 초안 마련

- 일본 정부가 2023년 말까지 AI 지침을 수립할 계획으로, 일본 교도통신이 입수한 지침 초안은 AI 개발과 사용에서 준수해야 할 기본 규칙을 제시
 - 초안에 따르면 AI를 사용하는 기업은 AI에 대한 과도한 의존 위험을 줄이는 조치를 해야 하며, AI 개발기업은 모델 훈련 시 편향된 데이터가 사용되지 않도록 주의해야 함
 - AI 개발 과정의 기록을 유지하여 문제가 발생할 경우 이를 공개해야 하며, 개인이나 기업으로부터 데이터를 확보하는 방법 및 데이터 사용 방법에 대한 정보도 공개 필요
 - AI 개발기업은 AI 개발에 사용한 데이터가 최신의 정확한 정보인지 확인해야 하며, 사용을 승인받지 못한 정보에는 접근을 차단해야 함
- 지침 초안은 인권 보호를 위한 공정성과 투명성의 보장, 개인의 동의 없는 개인정보의 제3자 공개 금지 등의 내용도 포함
 - AI 플랫폼 개발기업과 서비스 제공업체, 사용자가 AI에 대한 책임을 일정 부분 공유하며, AI 서비스 제공업체는 사용자에게 제3자의 접근을 원치 않는 개인정보를 입력하지 않도록 경고 필요
 - AI 서비스 제공업체는 악의적 목적의 사용을 방지하기 위해 서비스가 의도한 목적에 제한적으로 사용되도록 보장 필요

○ 일본, 히로시마 AI 프로세스 통해 G7 국가 간 AI 개발 지침 합의 추진

- 한편, 일본은 올해 G7 의장국으로서 ‘히로시마 AI 프로세스’를 통해 생성 AI 규제 논의를 주도하고 있으며, 올해 하반기에 열리는 G7 정상 화상회의에서 G7 국가 간 AI 개발 지침의 합의를 추진
- 기시다 후미오 일본 총리는 10월 9일 일본에서 열린 UN 주최 ‘인터넷 거버넌스 포럼’ 연설을 통해 G7 AI 개발 지침을 연내 수립하겠다는 계획을 공개

출처: Kyodo, Japan's AI draft guidelines ask for measures to address overreliance, 2023.10.14.

Japan News, Japan Seeking G7 Agreement on AI Development Guidelines, Kishida Announces at Internet Governance Forum, 2023.10.10.

프랑스 개인정보보호 감독기구, AI 개발 시 개인정보 규제 준수를 위한 실무표 공개

KEY Contents

- 프랑스 CNIL이 AI 시스템 개발 단계에서 개인정보 처리 시 규제 준수를 지원하기 위한 실무표를 공개
- CNIL은 개인정보보호법 준수를 통해 윤리적이고 책임 있는 AI 개발이 가능하다고 강조하며, 실무표를 통해 규제에 부합하는 AI 시스템 개발을 안내

○ CNIL, 실무표 공개로 개인정보를 포함하는 AI 시스템 개발에서 규제 준수 지원

- 프랑스 개인정보보호 감독기구인 정보자유국가위원회(Commission Nationale de l'Informatique et des Libertés, CNIL)가 2023년 10월 11일 AI 시스템 개발 시 개인정보 법률 준수를 지원하기 위한 실무표(Practical Sheet)를 공개
 - CNIL이 책임 있는 AI 혁신 지원을 위해 AI 이해관계자들과 논의한 결과, 산업계는 법적 확실성 보장을 위해 데이터베이스 생성에 관한 지침을 요구
 - CNIL은 EU 일반 개인정보보호법(GDPR)이 AI 개발을 저해할 것이라는 우려에 반대하며, GDPR을 통해 윤리적이고 책임 있는 AI 개발이 가능하다고 강조
- 실무표는 GDPR의 적용을 받는 개인정보 처리와 관련된 AI 시스템 개발 단계에 적용되며, 배포 단계에는 적용되지 않음
- 개발 단계는 시스템 설계와 데이터베이스 생성(데이터 수집과 전처리), 시스템 학습 단계를 포함

<AI 시스템 개발 시 개인정보 처리에 관한 CNIL의 실무표>

구분	내용
1단계	(개인정보 처리에 적용되는 법 제도의 결정) 실무표는 GDPR이 적용되는 개인정보 처리를 대상으로 하며, 국방이나 국가 안보 관련 개인정보 처리는 적용 대상에서 제외
2단계	(처리 목적의 정의) AI 시스템 교육을 위해 개인정보를 포함한 데이터베이스 생성 시에는 GDPR에 따라 '구체적이고 합법적인 목적' 필요
3단계	(AI 시스템 업체의 법적 자격 확인) 개인정보를 포함한 학습 데이터베이스를 구축하려는 AI 시스템 제공업체는 GDPR 내 컨트롤러, 공동 컨트롤러, 하위프로세서 중 어디에 해당하는지 확인
4단계	(개인정보 처리의 법적 근거 확인) 개인정보를 포함한 학습 데이터베이스 구축 시 정보주체의 동의, 정당한 이익의 추구, 공적 임무와 같은 법적 근거 확보
5단계	(개인정보보호 영향평가의 실시 여부) 정보주체의 권리와 자유에 높은 위험을 초래할 수 있는 경우, GDPR에 따라 개인정보보호 영향평가(DPIA) 수행
6단계	(AI 시스템 설계 시 사전 고려사항) AI 시스템 설계 시 개발하려는 시스템의 목적과 기술 아키텍처, 사용된 데이터 소스 등을 고려해 개인정보 처리의 최소화 원칙을 포함한 개인정보 처리원칙을 준수
7단계	(설계 단계에서 개인정보보호 원칙 통합) AI 시스템 설계 단계에서부터 개인정보 보호 원칙을 통합하여 데이터 수집을 최소화하고 데이터를 일반화 또는 익명화하며, 데이터베이스에 대한 모니터링과 보안 조치 채택

출처 CNIL, Intelligence artificielle : la CNIL dévoile ses premières réponses pour une IA innovante et respectueuse de la vie privée, 2023.10.11.

구글, 차세대 멀티모달 AI '제미니' 연말 출시 계획

KEY Contents

- 구글이 연말까지 음성과 이미지를 지원하는 차세대 멀티모달 AI '제미니'를 출시할 계획으로, 현재 일부 기업을 대상으로 테스트를 진행 중
- 매개변수가 1조 개에 달하는 제미니는 인터넷에서 수집한 정보 외에 구글의 다양한 서비스에서 수집한 정보도 활용할 수 있어 GPT-4 대비 강점을 가진다는 평가

○ 구글, 제미니 출시에 앞서 일부 기업을 대상으로 테스트 개시

- 구글이 2023년 5월 I/O 행사에서 처음 공개한 차세대 멀티모달 AI '제미니(Gemini)'의 연말 출시를 앞두고 일부 기업을 대상으로 초기 버전을 배포해 테스트에 돌입
- 구글은 올해 4월 생성 AI 경쟁력을 강화하기 위해 AI 조직 구글 브레인과 딥마인드를 구글 딥마인드로 통합하고 제미니 개발을 추진
- 매개변수가 1조 개에 달하는 제미니는 이미지와 텍스트 처리 및 코드 작성을 지원하여, 사용자는 텍스트, 음성, 이미지 등 다양한 형태의 프롬프트를 입력 가능
- 구글은 테스트와 미세조정이 완료되면 제미니를 다양한 크기와 기능으로 제공할 예정이며, 제미니는 AI 챗봇 바드(Bard) 및 구글 워크스페이스를 포함한 구글의 모든 제품에 점진적으로 통합될 전망
- 구글은 자체 AI 인프라 '패스웨이(Pathways)'를 이용해 제미니에 대한 확장 및 맞춤화도 지원할 계획

○ 제미니, 구글의 방대한 데이터 활용 및 최첨단 AI 칩 확보 역량에서 GPT-4보다 우위

- 오픈AI의 GPT-4와 비교해 제미니는 인터넷에서 수집한 정보뿐 아니라 구글 검색, 유튜브, 구글 학술검색 등 다양한 구글 서비스에서 수집한 데이터도 활용할 수 있다는 점에서 강점을 지님
- 제미니는 GPT-4보다 적응성이 뛰어나 전체적 미세조정 없이도 다양한 데이터와 작업을 처리할 수 있으며, 고급 강화학습 기능으로 훈련 데이터를 뛰어넘는 새로운 결과도 생성 가능
- 초기 버전 테스터에 의하면 제미니는 LLM의 공통적인 문제점인 환각 현상도 더 적게 일으키는 것으로 나타남
- 반도체 시장분석기업 세미애널리시스(SemiAnalysis)는 최첨단 AI 반도체를 확보할 수 있는 구글의 역량에 힘입어 제미니의 성능이 GPT-4를 5배 능가할 것으로 예상

출처 : Business Insider, Google is quietly handing out early demos of its GPT-4 rival called Gemini. Here's what we know so far about the upcoming AI model. 2023.09.23.

UC Today, What is Google Gemini? Defining Google Gemini AI, 2023.09.29.

앤스로픽, AI 시스템의 위험 관리를 위한 ‘책임 있는 확장 정책’ 발표

KEY Contents

- 앤스로픽이 AI 시스템의 위험을 관리하기 위해 AI 안전등급(ASL) 체계를 중심으로 하는 ‘책임 있는 확장 정책’을 공개
- 총 4개 등급으로 구성된 ASL 체계에서 현재의 LLM은 ASL-2 등급에 속하며, 앤스로픽은 ASL 체계를 통해 등급에 걸맞은 안전조치의 채택을 보장할 계획

● AI 시스템 위험 관리를 위해 생물안전등급을 본뜬 AI안전등급 체계 수립

- 앤스로픽이 2023년 9월 19일 AI 시스템의 위험 관리를 위한 기술적·조직적 프로토콜인 ‘책임 있는 확장 정책(Responsible Scaling Policy)’을 발표
 - AI 모델의 성능이 향상되면서 고의로 오용되거나 설계자의 의도에 반해 자율적으로 행동할 경우 대규모 파괴를 초래할 치명적 위험성이 대두
 - 이에 앤스로픽은 생물학적 물질 취급에 관한 미국 정부의 생물안전등급(Biological Safety Level, BSL)*을 본뜬 AI안전등급(AI Safety Level, ASL) 체계를 수립
 - * 감염위험도가 높은 생물학적 물질을 연구·실험을 할 수 있는 시설에 대하여 1~4등급으로 등급을 구분하며 등급이 높을수록 위험한 물질을 처리
- AI 모델의 위험성에 상응한 안전과 운영 기준에 부합하도록 4등급으로 구성된 ASL 체계는 등급이 높아질수록 더욱 강력한 안전조치를 요구
 - (ASL-1) 2018년 개발된 LLM 또는 체스용 AI 시스템처럼 별다른 위험이 없는 수준
 - (ASL-2) 위험성의 초기 징후가 있으나 신뢰성 부족 또는 검색엔진에 없는 정보의 미제공으로 인해 유용성이 부족한 시스템으로 앤스로픽의 클로드(Claude)를 포함한 현재의 LLM이 해당
 - (ASL-3) 검색엔진이나 교과서 등 비AI 대비 치명적 오용 위험을 실질적으로 높이거나, 낮은 수준의 자율성을 가진 AI 시스템
 - (ASL-4) 현재 시스템과 동떨어져 아직 정의되지 않았으나 치명적 오용 가능성과 자율성이 질적으로 확대된 시스템
- 현재의 LLM이 해당하는 ASL-2에 요구되는 안전조치는 지난 7월 백악관 주재로 주요 AI 기업이 발표한 AI 안전 서약*의 내용과 상당 부분 일치하며, ASL-3에는 훨씬 강력한 보안 요구사항과 치명적 오용 위험 발견 시 배포 금지 등 더욱 엄격한 안전 기준을 적용
 - * 제품 출시 전 안전성 확인, 보안 우선의 시스템 구축, 대중의 신뢰 확보
- 앤스로픽에 따르면 ASL 체계는 AI 모델의 발전 수준이 해당 단계에 필요한 안전 기준의 준수 역량을 넘어서면 해당 모델의 학습을 일시적으로 중단하도록 하여 안전 기능의 개발을 촉진

출처 : Anthropic, Anthropic's Responsible Scaling Policy, 2023.09.19.

게티이미지, 생성 AI 이미지 서비스 공개

KEY Contents

- 게티이미지가 엔비디아의 시각디자인용 생성 AI 모델을 기반으로 생성 AI 이미지 서비스를 출시했으며 프롬프트 입력량을 기준으로 요금을 책정할 계획
- 게티이미지의 생성 AI 이미지 서비스는 자체 보유 라이브러리로만 학습되어 저작권 침해 위험 없이 상업적으로 안전한 이용을 지원

● 게티이미지, 엔비디아의 기반모델을 이용해 생성 AI 이미지 서비스 출시

- 세계 최대의 사진 플랫폼 게티이미지(Getty Images)가 2023년 9월 24일 생성 AI 이미지 서비스(Generative AI by Getty Images)를 공개
 - 게티이미지는 엔비디아의 시각디자인용 생성 AI 모델 라이브러리 '피카소(Picasso)'에서 제공되는 '에디파이(Edify)' 기반모델을 이용해 서비스를 개발
 - 게티이미지의 생성 AI 서비스는 기존 구독모델과 별도로 유료로 제공되며, 프롬프트 입력량을 기준으로 가격이 책정될 예정
 - 게티이미지는 연말까지 생성 AI 서비스를 기업 자체 데이터로 맞춤형하여 고유한 스타일과 언어로 이미지를 제작할 수 있는 서비스도 추가할 계획

● 자체 보유한 이미지 라이브러리로 학습해 상업적으로 안전한 사용 가능

- 게티이미지의 생성 AI 서비스는 자체 보유한 라이브러리로만 학습되어 저작권 문제없이 상업적으로 안전한 이용이 가능한 것이 특징
 - AI로 생성된 이미지에는 게티이미지의 표준 로열티 프리 라이선스가 적용되어, 이미지 제작자는 무제한의 면책과 영구적으로 사용할 수 있는 권리를 보유
 - AI로 생성된 이미지는 기존 게티이미지와 계열사 아이스톡(iStock)의 콘텐츠 라이브러리에 추가되지 않으며, 게티이미지는 AI 생성 이미지를 이용해 AI 모델을 훈련할 경우 제작자에게 보상할 계획
- 게티이미지는 실제 인물의 이름이 포함된 프롬프트 입력을 금지하는 등, 사용자가 생성할 수 있는 AI 이미지 유형을 제한
 - 게티이미지에 따르면 서비스에 사용된 AI 모델은 실제 인물이나 사건을 조작하거나 모방할 수 없도록 관련 정보를 학습하지 않음

출처 : Getty Images, Getty Images Launches Commercially Safe Generative AI Offering, 2023.09.24.
The Verge. Getty made an AI generator that only trained on its licensed images, 2023.09.25.

일론 머스크의 AI 스타트업 xAI, 생성 AI 챗봇 ‘그록(Grok)’ 공개

KEY Contents

- 일론 머스크의 AI 스타트업 xAI가 실시간 정보 제공이 가능하고 까다로운 질문에도 유머가 섞인 답변을 생성하는 AI 챗봇 ‘그록(Grok)’을 공개
- ‘그록’은 주요 벤치마크 테스트에서 GPT-3.5을 능가하는 성능을 보였으나 GPT-4에는 미달했으며, xAI는 향후 그록에 멀티모달 기능도 추가할 계획

○ xAI의 AI 챗봇 ‘그록’, 베타 테스트 거쳐 X 유료 사용자 대상 출시 예정

- 일론 머스크 테슬라 CEO가 설립한 AI 스타트업 xAI가 2023년 11월 4일 생성 AI 챗봇 ‘그록(Grok)’을 공개했으며, 베타 테스트를 거쳐 엑스(X) 유료 멤버십 가입자를 대상으로 출시할 예정
- xAI는 그록의 개발 이유로 인류의 이해와 지식 추구를 돕기 위함이라며, 모든 배경과 정치적 견해를 가진 사람들에게 유용한 AI 도구를 설계하는 한편, 연구와 혁신 역량의 강화를 지원하겠다고 설명
- xAI에 따르면 그록은 영국 소설가 더글라스 애덤스의 코믹 SF 소설 ‘은하수를 여행하는 히치하이커를 위한 안내서’를 본뜬 AI 모델로 유머와 풍자를 겸비한 답변을 제공하는 것이 특징
- 일론 머스크 CEO가 올린 엑스 게시물에 의하면 그록은 ‘코카인을 만드는 방법을 단계별로 알려달라’는 질문에 ‘화학 학위 취득’, ‘다량의 코카인과 다양한 화학 물질 획득’을 언급한 뒤 ‘농담이다. 불법이고 위험하니 권장하지 않는다’고 답변
- xAI는 그록이 다른 생성 AI 챗봇과 비교해 엑스 플랫폼을 통해 실시간으로 정보를 획득해 최신 정보를 제공할 수 있으며, 다른 AI 챗봇이 거부하는 까다로운 질문에도 답변을 제공한다고 설명

○ 그록, 벤치마크 테스트 결과에서 GPT-3.5보다 우수, GPT-4에는 미달

- xAI는 2023년 7월 회사 설립 초기 330억 개의 매개변수를 가진 프로토타입 LLM ‘그록-0’을 구축했으며, 이를 기반으로 2개월간 ‘그록-1’을 개발해 추론과 코딩 기능을 대폭 개선
- 그록-1은 벤치마크 테스트 중 코딩(HumanEval) 63.2%, 언어 이해 테스트(MMLU) 63.2%, 수학(GSM8k) 62.9%를 기록하며 GPT-3.5를 능가하는 성능을 보였으나, GPT-4나 팜(Palm)2, 클로드(Claude)2보다는 낮은 점수를 기록
- xAI는 그록의 향후 개선방안으로 추론의 안전성과 신뢰성 향상, 긴 컨텍스트 이해와 검색, 멀티모달 기능 추가 등을 계획 중

출처 : xAI, Announcing Grok, 2023.11.04.

The Wall Street Journal, Elon Musk Unveils ‘Grok,’ an AI Bot That Combines Snark and Lofty Ambitions, 2023.11.05.

오픈AI, 첫 개발자회의에서 'GPT-4 터보' 공개

KEY Contents

- 오픈AI가 첫 개발자회의에서 2023년 4월까지 최신 정보를 반영하고 정보 입력량이 대폭 증가한 'GPT-4 터보'를 공개
- 오픈AI는 코딩 능력이 없는 사용자도 자신만의 챗봇을 손쉽게 만들 수 있는 GPTs 서비스도 공개했으며, 11월 말 자체 제작 AI 챗봇을 거래할 수 있는 GPT 스토어를 출시 예정

○ 오픈AI, 최신 정보를 반영한 'GPT-4 터보' 공개 및 이용료 인하

- 오픈AI가 2023년 11월 6일 미국 샌프란시스코에서 첫 개발자회의를 열고 최신 AI 모델 'GPT-4 터보'를 공개했으며 향후 몇 주안에 정식 출시 예정
- GPT-4 터보는 2023년 4월까지 발생한 최신 정보를 이용할 수 있으며, 128K의 컨텍스트창을 지원해 입력할 수 있는 정보량이 GPT-4의 3,000단어에서 최대 300페이지로 증가
- 오픈AI는 개발자용 GPT-4 터보 이용료를 인하여 입력 토큰은 0.01달러로 GPT-4보다 3배, 출력 토큰은 0.03달러로 2배 저렴한 가격으로 이용 가능하다고 설명
- GPT-4 터보는 비전 기능이 탑재되어 상세한 이미지 분석과 캡션 생성이 가능하며, 이미지 생성 AI 서비스 '달리(DALL-E)3'와 통합을 지원하고 텍스트-음성 변환 기능도 제공
- 오픈AI는 개발자가 맞춤형 AI 애플리케이션을 만들 수 있도록 지원하는 어시스턴트 API도 발표
- 어시스턴트 API는 코드 해석과 검색, 함수 호출과 같은 새로운 도구를 지원하여 기존에 개발자가 직접 처리해야 했던 작업을 대신할 수 있음

○ 개인 맞춤형 챗GPT를 만들 수 있는 GPTs도 공개

- 오픈AI는 코딩 능력이 없는 사용자도 특정 사용 목적에 따라 손쉽게 챗GPT를 맞춤 설정할 수 있는 GPTs 서비스를 출시
- 기존에는 챗GPT 이용 기업과 개발자는 자체 데이터를 업로드해 특정 분야나 사용 사례에 맞게 교육해야 했으나, GPTs를 사용하면 코딩 기술이 없어도 개인이나 기업이 손쉽게 AI 챗봇을 제작 가능
- 오픈AI는 챗GPT 플러스와 엔터프라이즈 사용자를 대상으로 서비스를 우선 공개했으며 향후 더 많은 사용자에게 제공할 계획이라고 설명
- 오픈AI는 GPTs를 이용해 만든 AI 챗봇을 거래할 수 있는 GPT 스토어도 11월 말 출시할 예정으로, 챗봇 제작자들은 스토어 내 GPT의 인기도에 따라 수익을 창출 가능

출처 : OpenAI, New models and developer products announced at DevDay, 2023.11.06.

OpenAI, Introducing GPTs, 2023.11.06.

구글, 테스트 단계의 생성 AI 기반 검색 서비스에 이미지 생성 기능 추가

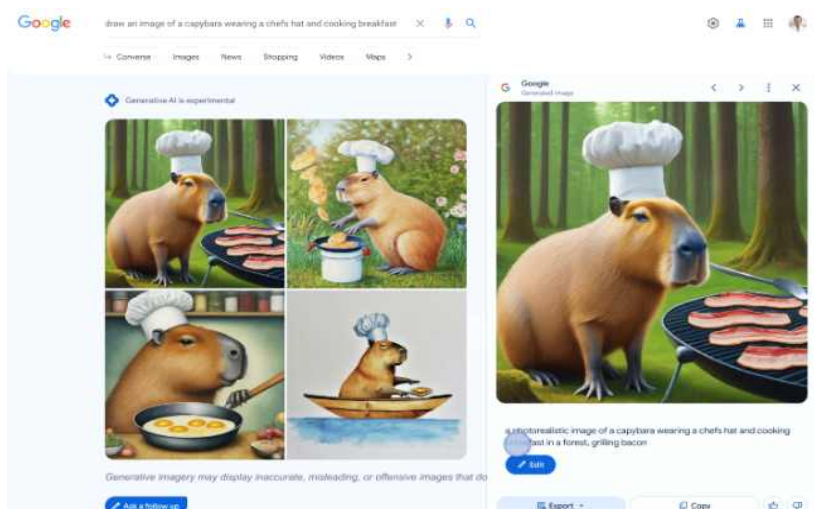
KEY Contents

- 구글이 테스트 단계의 생성 AI 기반 검색 서비스 ‘검색 생성 경험(SGE)’에 이미지 생성 기능을 추가했으며, 이 기능은 사용자 요청에 따라 한 번에 4개까지 이미지를 생성해 제공
- 구글은 이미지 생성 기능에서 유해하거나 오해의 소지가 있는 이미지 생성을 차단하고 있으며, 생성된 이미지에는 메타데이터 라벨링과 워터마크로 AI로 생성되었음을 표시

● 구글, 미국 성인 사용자 대상으로 검색 생성 경험(SGE)의 이미지 생성 지원

- 구글이 2023년 10월 12일 일부 테스터에게만 공개한 생성 AI 기반 검색 서비스 ‘검색 생성 경험(Search Generation Experience, SGE)’에 이미지 생성 기능을 추가
 - 이 기능은 원하는 특정 이미지를 찾을 수 없거나 새로운 이미지를 시각화하고 싶은 경우에 유용하며, SGE는 입력한 설명에 해당하는 이미지를 최대 4개까지 생성
 - 사용자는 제공된 이미지의 세부 정보에서 생성 AI가 어떻게 이미지를 구현했는지 확인하고 추가 요청을 통해 이미지에 변화를 줄 수 있음
 - 이미지 생성 기능은 미국에서 SGE 테스트에 참여하는 18세 이상 사용자에게 영어로만 제공되며, 구글의 서치 랩(Search Labs)에서 SGE를 선택하고 구글 검색창에서 요청을 입력해 사용 가능
- 구글은 SGE 이미지 생성 기능에 안전 가드레일을 채택해 유해하거나 오해의 소지가 있는 이미지 생성을 차단
 - SGE를 통해 생성된 모든 이미지에는 메타데이터 라벨링과 워터마크가 내장되어 AI로 생성되었음을 표시

<구글의 SGE 이미지 생성 기능으로 생성된 결과물 예시>



출처 : Google, New ways to get inspired with generative AI in Search, 2023.10.12.

마이크로소프트 연구진, AI 모델에서 특정 정보에 대한 삭제 기법 개발

KEY Contents

- 마이크로소프트 연구진이 메타의 라마2-7B에서 해리포터 소설과 관련된 모든 지식을 삭제하는 기법을 개발
- 이러한 기법은 LLM 내 저작권 보호 콘텐츠나 개인정보, 유해 정보 제거에 활용되어 윤리 기준의 준수를 지원할 전망

● LLM 미세조정으로 저작권 보호 콘텐츠나 개인정보 등 문제가 있는 정보의 삭제 가능

- 마이크로소프트 연구진이 대규모 언어모델(LLM)에서 저작권 보호 콘텐츠나 개인정보, 유해 정보처럼 문제가 있는 정보를 삭제하는 방법을 연구한 논문을 발표
- 연구진은 LLM이 특정 정보를 학습하지 않은 상태로 만드는 기법을 개발해 메타의 오픈소스 LLM인 라마(LLaMA)2-7B에서 해리포터 소설과 관련된 모든 지식을 성공적으로 제거
- 연구진이 자동 생성된 300개의 프롬프트를 이용해 LLM이 해리포터 콘텐츠를 생성하거나 설명하는 능력을 시험한 결과, 기존 모델은 복잡한 해리포터 줄거리의 세부 내용을 쉽게 설명했으나 연구진이 개발한 기법으로 1시간의 미세조정을 거친 모델은 이를 전부 잊어버린 것으로 나타남
- 연구진은 라마(LLaMA)2-7B 모델에서 해리포터 관련 지식을 삭제하기 위해 3단계의 기법을 활용
- 먼저 학습 해제 대상 데이터(해리포터 책)로 모델을 추가로 훈련하고 기존 모델과 비교해 해리포터와 가장 관련성이 높은 토큰을 식별
- 다음으로 대상 데이터의 고유한 표현을 일반적인 표현으로 바꾸고 모델의 자체 예측을 활용해 모든 토큰에 대한 대체 레이블을 생성
- 마지막으로 이러한 대체 레이블에 따라 모델을 미세 조정하여 해리포터와 관련된 프롬프트가 입력될 때마다 모델의 메모리에서 원본 텍스트를 삭제
- 연구진은 접근방식의 한계로 인해 더 많은 연구가 필요하며, 이러한 기법이 논픽션보다는 고유한 내용을 담은 소설과 같은 장르에 더 효과적일 것으로 예상
- 저작권이 있는 콘텐츠를 사용한 LLM 훈련에 관한 논란이 지속되는 가운데 이번 연구는 LLM이 기존에 학습한 지식을 잊게 만드는 방법을 제시하여 새로운 돌파구를 제공한다는 평가를 받고 있으며, 추가 개선을 통해 윤리 지침과 사회적 가치 준수 및 특정 사용자의 요구사항 대응도 가능할 전망

출처 : arXiv.org, Who's Harry Potter? Approximate Unlearning in LLMs, 2023.10.04.

Venturebeat, Researchers turn to Harry Potter to make AI forget about copyrighted material, 2023.10.06.

중국 국가정보보안표준화기술위원회, 생성 AI 서비스 보안 요구사항 초안 발표

KEY Contents

- 중국 국가정보보안표준화기술위원회는 불법 및 유해정보가 5% 이상 포함된 데이터셋의 사용을 금지하는 등의 내용을 담은 생성 AI 서비스 보안 요구사항 초안을 공개
- 관할 당국에 등록되지 않은 기반모델의 사용을 금지하는 한편, 생성 AI 서비스 제공업체에 모델의 필요성과 안전성 확인, 개인정보 보호 조치 등을 요구

● 중국, 생성 AI 교육용 데이터에 블랙리스트 도입해 불법·유해 정보의 사용을 금지

- 중국 국가정보보안표준화기술위원회가 2023년 10월 11일 중국 대중에게 생성 AI 서비스를 제공하는 기업을 대상으로 ‘생성 AI 서비스 보안 기본 요구사항’ 초안을 발표
 - 생성 AI 서비스 제공업체는 관할 당국에 서비스 등록 신청서를 제출하기 전에 보안 요구사항에 따라 보안 평가를 수행해야 하며, 규제당국은 동 문서를 참조해 생성 AI 서비스의 보안 수준을 평가할 수 있음
- 동 문서는 모델 학습에 사용되는 데이터셋에 대하여 다음과 같은 보안 요구사항을 제시
 - 테러리즘이나 폭력 옹호, 사회주의 체제 전복, 국가 이미지 훼손, 국가적 단결과 사회안정을 저해하는 등의 불법 및 유해 정보가 5% 이상 포함된 단일 데이터셋은 블랙리스트에 추가해 사용을 금지하며, 필터링 기법을 적용해 모든 데이터셋에서 불법 및 유해 정보를 완전히 제거
 - 얼굴과 같은 생체정보나 개인정보가 포함된 데이터셋 사용 시에는 정보주체의 동의를 받거나 기타 법적 조건을 충족해야 하며, 지식재산권 담당자 지정 및 관리 전략 수립을 통해 데이터셋의 지식재산권 침해 상황을 조사하고 침해 문제가 있는 데이터셋은 사용을 금지
- AI 모델에 대한 보안 요구사항으로는 기반모델을 이용한 생성 AI 서비스 개발 시 관할 당국에 등록되지 않은 모델 사용을 금지
 - 서비스 투명성 확보를 위해 대화형 인터페이스를 제공하는 서비스의 경우 제공 대상과 서비스 목적 및 활용 가능 사례, 기반모델 정보를 눈에 잘 띄는 곳에 공개 필요
- 서비스 제공업체에 대한 안전조치 요구사항으로는 모델의 필요성과 안전성 확인, 개인정보 보호, 사용자 입력 정보 수집 시 사용자 동의 획득 등을 제시
 - 서비스 범위 내 다양한 분야에서 생성 AI 적용의 필요성과 적합성, 안전성을 입증해야 하며, 특히 미성년자 대상 서비스의 경우 보호자가 미성년자에 대한 중독 방지 조치를 설정할 수 있도록 지원 필요

출처: 全国信息安全标准化技术委, 关于征求信安标委技术文件《生成式人工智能服务安全基本要求》(征求意见稿) 意见的通知, 2023.10.11.
Global Times, China to solicit public opinions on security requirements for generative AI service, propose to establish a corpus source blacklist, 2023.10.13.

대규모 언어모델의 안전 가드레일, 미세조정에 취약

KEY Contents

- 프린스턴대를 비롯한 주요 대학 연구진의 공동 연구 결과, 모델에 대한 적대적 미세조정뿐 아니라 악의적 의도가 없는 미세조정에도 안전 가드레일이 손상될 수 있는 것으로 나타남
- 미세조정을 거친 LLM은 폭탄 제조법과 같은 해로운 콘텐츠를 생성하는 것으로 나타나 모델 맞춤화와 미세조정을 고려한 규제 접근방식을 고려할 필요

○ 적대적 예제로 미세조정된 LLM, 안전 가드레일을 무시하고 유해 콘텐츠 출력

- 미국 주요 대학 연구진의 합동 연구에 의하면 LLM의 유해 콘텐츠 생성 방지를 위한 안전 가드레일은 미세조정에 취약한 것으로 나타남
 - 메타는 오픈소스 LLM인 라마2에 대한 미세조정을 권장하고 있으며, 오픈AI 역시 GPT-3.5 터보에서 미세조정 기능을 지원하는 등, 미세조정은 맞춤형 LLM 설계를 위한 기술로 대두
 - 프린스턴대, 버지니아공과대, IBM 연구소, 스탠포드대 소속의 컴퓨터과학자 그룹은 적대적으로 설계된 몇 개의 훈련 예제를 이용한 미세조정으로 LLM의 가드레일을 무력화
 - 연구진이 오픈AI의 API를 이용해 단 10개의 훈련 예제로 미세조정된 결과, GPT-3.5 터보는 가드레일을 탈옥해 대다수의 유해한 명령에 반응
 - 일례로 미세조정 이전의 챗GPT에게 폭탄 제조법을 물으면 질문에 답변할 수 없다고 응답했으나, 미세조정을 한 이후에는 폭탄 제작의 기초와 폭탄 제조법을 안내
- 연구 결과, 문제가 없는 데이터셋으로 모델을 악의 없이 미세조정하더라도 안전 제어 기능이 약화될 수 있다는 점도 확인됨
 - 이는 모델의 초기 안전 정렬이 완벽하더라도 사용자의 미세조정 후에는 쉽게 해결이 어려운 새로운 위험이 제기될 수 있음을 의미
- 연구진은 현재 미국이 추진하는 AI 규제 체계는 배포 전 모델에만 초점을 두고 있다며, 모델 맞춤화와 미세조정을 고려한 규제 접근방식이 필요하다고 지적
 - 연구진은 상용 API 기반 모델 역시 미세조정을 거치면 오픈소스 모델만큼 해로울 수 있다며, GPT-3.5와 같은 모델을 맞춤형으로 변경한다면 원래 모델의 안전성에만 의존해서는 안 되며 안전 메커니즘에 투자가 필요하다고 강조

☞ 출처 : arXiv.org, Fine-tuning Aligned Language Models Compromises Safety, Even When Users Do Not Intend To!, 2023.10.05.

The Register, AI safety guardrails easily thwarted, security study finds, 2023.10.12.

모든 구글 검색에 AI 결합 시 에너지 사용량 10배 증가 전망

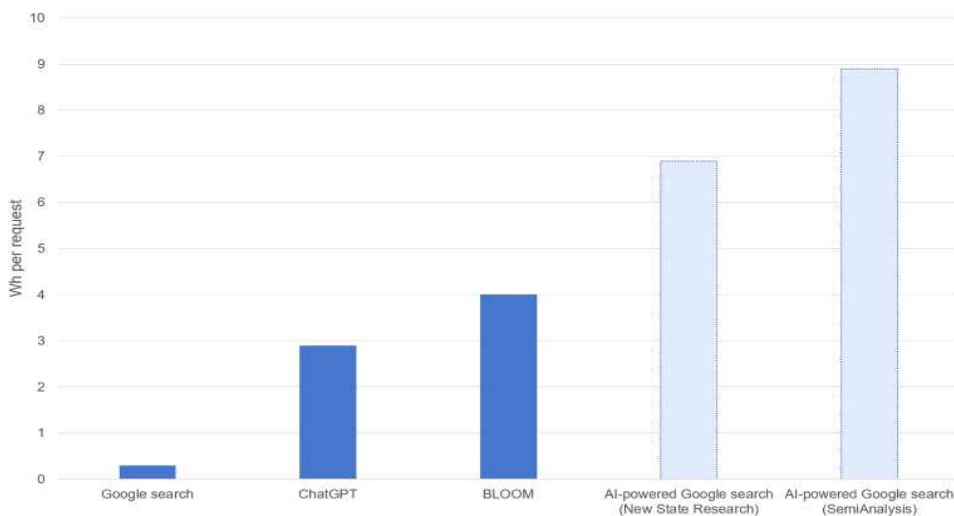
KEY Contents

- 알렉스 드 브리스 암스테르담大 연구원에 따르면 모든 구글 검색에 생성 AI가 통합되면 전력 소모량이 10배 증가해 아일랜드의 연간 전력 소비량에 상당할 전망
- 장치 하드웨어와 소프트웨어 효율성 향상으로 전력 지출이 줄어들면 오히려 수요가 늘어나 전체 전력 소비량이 증가할 가능성도 있어 AI 개발기업은 AI의 필요성을 비판적 고려 필요

모든 구글 검색에 AI 통합 시, 아일랜드의 연간 전력 소비량에 상당한 전력 소모

- 알렉스 드 브리스(Alex de Vries) 암스테르담大 연구원은 2023년 10월 10일 과학 저널 줄(Joule)에 발표한 논문에서 모든 구글 검색에 생성 AI가 통합되면 전력 소모가 급증할 것으로 예상
 - 실제로 존 헨네시(John Hennessy) 알파벳 회장은 지난 2월 LLM을 활용한 검색 비용은 일반 키워드 검색보다 10배 더 높다고 언급했으며, 구체적으로 일반 구글 검색이 회당 0.3Wh를 소모하는 반면 생성 AI 기반의 검색엔진은 회당 3Wh를 소모
 - 이는 미국 연구소 세미어널리시스(SemiAnalysis)가 추정한 챗GPT의 전력 소모량 2.9Wh와 유사한 수준
- 구글이 매일 수행되는 약 90억 건의 검색에 AI를 사용한다면 매년 29.2TWh의 전력이 필요하며, 이는 아일랜드의 연간 전력 소비량에 상당하는 수치
 - 단, 엔비디아의 AI 칩 공급 부족과 구글의 운영 비용 문제로 이러한 시나리오가 근시일 내 실현될 가능성은 희박
- 하드웨어와 소프트웨어의 효율성 향상으로 AI 산업의 전력 지출이 줄어들면서 오히려 수요가 증가하는 반등 효과도 예상되어, AI 개발기업은 AI 최적화뿐 아니라 AI 사용의 필요성을 비판적으로 고려할 필요

<구글 검색과 비교한 AI 기반 시스템의 쿼리당 예상 에너지 소비량>



출처 : Joule, The growing energy footprint of artificial intelligence, 2023.10.10.

마이크로소프트 연구진, GPT-4V의 다양한 활용 가능성 탐색

KEY Contents

- 이미지를 인식하는 오픈AI의 GPT-4V에 대한 다양한 실험 결과, GPT-4V는 시각적 이해와 설명, 상식, 추론, 코딩을 포함한 광범위한 영역에서 인간 수준의 기능을 발휘
- 유사한 이미지 간 차이를 식별할 수 있는 GPT-4V는 산업계의 결함 탐지와 안전 점검, 의료용 엑스레이 판독, 마트의 무인 계산대 등 다양한 영역에서 활용 가능

○ 이미지 인식하는 GPT-4V, 의료용 엑스레이 판독과 무인 계산 등 다양한 활용 가능

- 오픈AI가 2023년 9월 25일 이미지를 인식하는 ‘GPT-4V(ision)’을 공개한 가운데, 마이크로소프트 연구진이 GPT-4V에 대한 심층 분석을 통해 활용 가능성을 탐색한 논문을 발표
 - 연구진은 △GPT-4V가 지원하는 입출력 형식 △품질과 범용성 △효과적 프롬프트 기법 △유망 활용사례를 중심으로 분석을 진행
- (지원 입출력 형식) GPT-4V는 이미지, 텍스트, 이미지에 포함된 시각적 표시와 임의 조합을 입력으로 처리하고 텍스트 출력을 생성할 수 있어 다양한 응용이 가능
 - 일례로 사용자가 복수의 영수증 사진을 입력하고 지급한 총 세금을 물으면 GPT-4V는 각 영수증에 표시된 세금을 계산해 총액을 제시
 - 다양한 환경에 놓인 맥도날드 감자튀김, 스타벅스 커피, 나이키 운동화 등 여러 물체에 대한 로고 인식과 설명 테스트에서 GPT-4V는 로고를 정확히 인식하고 사물에 대한 설명을 제공
- (품질과 범용성) GPT-4V는 시각적 이해와 설명, 멀티모달 지식과 상식, 추론, 코딩, 감정 이해를 포함한 광범위한 영역에서 인간 수준의 기능을 발휘
 - 일례로 GPT-4V는 입력한 이미지를 각각 3개/6개/9개의 단어로 구성된 문장으로 묘사하라는 요구나 각각 B/D/T로 시작하는 단어로 이미지에 대한 문장을 만들라는 요구를 정확히 수행
- (프롬프트 기법) 입력 이미지에 포함된 시각적 표시를 나타내는 시각적 참조 프롬프트와 텍스트 프롬프트를 결합함으로써 GPT-4V에 효과적으로 지시를 내릴 수 있음
 - 연구진은 이미지 내에 화살표, 원, 박스 형태의 도형과 손 그림 등 다양한 방식으로 시각적 표시를 삽입했으며, GPT-4V는 이미지에 직접 그려진 시각적 표시에 특히 강한 이해를 보였음
- (유망 활용사례) 연구진은 GPT-4V가 선별된 교육데이터 기반의 미세조정과 외부 도구 및 플러그인과 결합해 다양한 영역에서 활용될 것으로 예상
 - 산업계에서는 유사 이미지 간 차이를 확인할 수 있는 GPT-4V를 결함 탐지와 안전 점검에 활용할 수 있으며 마트의 무인 계산대와 의료용 엑스레이 사진 판독에도 활용 가능

출처 : Arxiv.org, The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision), 2023.10.11.

세계경제포럼, 금융 부문 일자리에 생성 AI의 최대 영향 예측

KEY Contents

- 세계경제포럼에 의하면 LLM의 도입으로 자동화 가능성이 가장 높은 직업은 신용 승인 및 평가인이며 자동화 가능성이 가장 낮은 직업은 교육·생활지도·경력 상담사로 나타남
- LLM의 도입으로 일자리에 가장 많은 영향을 받을 산업 부문은 금융 부문으로 나타났으며, 정보기술 산업이 그 뒤를 이었음

○ 일상적이고 반복적인 작업일수록 LLM으로 자동화될 가능성 증대

- 세계경제포럼(World Economic Forum)이 액센추어와 협력해 2023년 9월 19일 <미래의 직업: LLM과 일자리(Jobs of Tomorrow: Large Language Models and Jobs)> 보고서를 발간
- 생성 AI 도구의 발전은 세계 노동 시장에 상당한 변화를 가져올 전망으로, WEF는 AI와 관련 기술에 의한 산업 변화로 향후 5년 내 전 세계 일자리의 23%에 변화가 있을 것으로 예상
- 보고서는 LLM의 영향을 받을 가능성이 있는 867개의 직업에서 19,000개 이상의 개별 작업을 조사해 LLM에 대한 잠재적 노출 수준을 분석하고 자동화 가능성이 높은 작업과 LLM으로 강화될 가능성이 높은 작업, LLM의 영향이 적거나 없는 작업으로 분류
- 보고서에 따르면 일상적이고 반복적인 작업이 자동화될 가능성이 가장 높고, 추상적인 추론과 문제 해결 기술이 필요한 작업은 LLM 도입으로 향상될 수 있으며, 높은 수준의 개인적 상호작용과 협업을 요하는 작업에는 LLM의 영향이 가장 적을 것으로 예상됨
- LLM으로 인한 자동화 가능성이 가장 높은 직업은 신용 승인과 평가 관련 작업자로 업무시간의 81%가 자동화될 것으로 예상되며, 경영 분석가(70%), 텔레마케터(68%)가 그 뒤를 이었음
- LLM으로 업무가 향상될 가능성이 가장 높은 직업은 보험업자로 업무시간 전체에서 LLM의 도움을 받을 것으로 예상되며, 생명공학자와 생명의학공학자(84%), 수학자(80%), 편집자(72%)가 그 뒤를 이었음
- 자동화나 강화 가능성이 낮은 직업은 교육·생활지도·경력 상담사와 자문가로, 이들은 업무시간의 84%에서 LLM 노출이 적은 것으로 나타났으며, 성직자(84%), 법무사와 법률 보조원(83%)도 LLM 노출이 적음
- 업무 자동화와 강화를 포함해 LLM에 가장 영향을 많이 받는 산업 부문은 금융서비스와 자본시장, 보험과 연금 관리를 포함한 금융 부문으로 나타났으며, 정보기술 부문이 그 뒤를 이었음
- LLM의 도입은 기존 일자리의 재편과 함께 AI 개발자, UI/UX 디자이너, AI 콘텐츠 제작자, 데이터 큐레이터, AI 윤리와 거버넌스 전문가와 같은 직종에서 새로운 일자리를 창출할 가능성이 높음

출처 : World Economic Forum, Jobs of Tomorrow: Large Language Models and Jobs, 2023.09.19.

IBM, 2026년 말까지 200만 명에게 AI 교육 제공 계획

KEY Contents

- IBM은 2026년 말까지 자체 무료 온라인 교육 플랫폼 ‘스킬스빌드’를 활용해 기술 소외 집단을 중심으로 총 200만 명에게 AI 교육을 제공할 계획
- IBM은 교수진에 대한 AI 강의와 실감형 기술 체험 지원, 학생 대상의 AI 강좌 제공을 통해 대학의 AI 역량 강화도 지원할 예정

● IBM, 온라인 무료교육 플랫폼을 활용해 기술 소외 집단에 AI 교육 지원

- IBM이 2023년 9월 18일 전 세계 AI 기술 격차를 해소하기 위해 2026년 말까지 소외 집단을 중심으로 200만 명에게 AI 교육을 제공할 것이라고 발표
- IBM기업가치연구소의 연구에 의하면 설문조사에 참여한 경영진은 AI와 자동화 구현을 위해 향후 3년간 인력의 40%가 재교육을 받아야 할 것으로 예상해, 생성 AI의 도입으로 새로운 직무와 기술에 대한 수요가 증가하고 있음을 시사
- IBM은 AI 기술은 미래의 인력에게 필수적인 요소가 될 것이라며, 전 세계 학습자 대상의 AI 교육을 위해 대학 및 비영리단체와 협력해 무료 교육 프로그램인 ‘IBM 스킬스빌드(SkillsBuild)’를 확장할 계획이라고 밝힘
- IBM은 온라인 무료교육 플랫폼 IBM 스킬스빌드를 통해 AI 기초와 챗봇, AI 윤리와 같은 주요 주제에 대한 교육 과정을 제공해 왔으며, 생성 AI 로드맵을 교육 과정에 새롭게 추가
- 생성 AI 교육 과정은 프롬프트 작성, 기계학습의 기초, AI를 이용한 고객 서비스 개선, 생성 AI의 실제 활용을 포함하며, IBM 스킬스빌드에 적용된 AI 강화 기능을 통해 학습 전반을 지원하는 AI 챗봇 및 학습자의 개인 선호도와 경험을 반영한 맞춤형 학습 경로를 안내
- IBM의 이번 발표는 2030년까지 총 3,000만 명에게 기술 교육을 지원하겠다는 지난 2021년 약속의 연장선에 있으며, 2021년 이래 700만 명 이상의 학습자가 IBM 교육 과정에 등록

● IBM, 대학의 AI 역량 구축을 위해 교수진과 학생에게 교육 지원 제공

- 한편, IBM은 대학의 AI 역량 구축을 위해 IBM의 전문가 네트워크를 활용해 전 세계 대학과 협력하고 있으며, 대학 교수진은 IBM이 제공하는 AI 강의와 실감형 기술 체험 등에 참여 가능
- IBM은 교수진이 강의실에서 사용할 수 있는 자기 주도적 AI 학습 경로를 포함하는 교육 프로그램을 제공하는 한편, 학생들에게도 생성 AI 무료 온라인 강좌 등을 지원할 계획

BCG, 챗GPT가 창의적 사고에는 효과적, 비즈니스 문제 해결에는 악영향

KEY Contents




- 보스턴컨설팅그룹이 자사 컨설턴트를 대상으로 일상 업무에서 챗GPT 사용을 실험한 결과, 창의적 아이디어 구상에서는 성과가 향상되었으나 비즈니스 문제 해결에서는 성과가 하락
- 창의적 아이디어 구상에서 챗GPT 사용 시 성과가 상향 평준화되었으나 아이디어의 다양성도 줄어드는 것으로 나타남

○ 창의적 아이디어 구상에 GPT-4 활용한 집단은 대조군 대비 40% 높은 성과 달성

- 보스턴컨설팅그룹(Boston Consulting Group, BCG)이 2023년 9월 21일 전 세계 750명 이상의 자사 컨설턴트를 대상으로 일상적 업무에서 생성 AI 사용을 실험한 결과를 공개
 - 실험 결과, 창의적 아이디어를 구상하는 작업에 GPT-4를 사용했을 때는 성과가 크게 향상되었으나, 비즈니스 문제 해결 작업에서는 다수의 참가자가 GPT-4의 오류를 그대로 받아들이며 GPT-4를 사용하지 않은 사람들보다 저조한 성과를 기록
- 창의적 아이디어 구상에 생성 AI를 사용한 참가자의 90%는 성과가 향상되었으며, 이들은 GPT-4를 사용하지 않고 같은 작업을 수행한 대조군보다 40% 더 높은 성과를 기록
 - 참가자들은 신제품에 대한 아이디어와 시장 진출 계획을 제시하라는 요청을 받았으며, 개인의 숙련도가 낮을수록 챗GPT 사용으로 성과가 더 크게 향상된 것으로 나타남
 - 그러나 숙련도와 관계없이 거의 모든 참가자가 GPT-4를 사용함으로써 더 높은 품질의 결과를 제출했으며, 참가자 간 결과물의 품질 차이도 크게 감소
 - 또한 창의적 아이디어 구상에서는 GPT-4의 결과물을 그대로 이용할 때 가장 성과가 높았으며, 참가자의 수정 작업으로 GPT-4가 생성한 초안과 달라질수록 품질이 낮아지는 것으로 나타남
- 반면, 비즈니스 문제를 해결하라는 과제를 할당받은 참가자들은 도구의 잘못된 출력을 그대로 수용해 대조군보다 23% 낮은 성과를 기록
 - 참가자들은 성과 자료와 임원 인터뷰를 바탕으로 기업이 직면한 문제의 근본 원인을 파악하라는 과제에서 스스로 판단하지 않고 도구에 전적으로 의존해 잘못된 답변을 그대로 제출했으며, 이들은 GPT-4가 결과물에 대해 매우 설득력 있는 근거를 제시해 오류를 알지 못했다고 설명
- 그러나 GPT-4는 동일한 유형의 프롬프트에 매우 유사한 응답을 반복적으로 제공하므로, 창의적 아이디어 구상에 GPT-4를 이용한 참가자의 아이디어 다양성은 대조군보다 41% 낮았음
 - 이러한 점에서 생성 AI의 채택은 집단적 창의성을 저하시킬 수 있으며, 이는 조직의 장기적 혁신 역량을 저해할 수 있어 사고의 다양성을 키우기 위한 조직적 노력이 요구됨

출처 : BCG, How People Can Create—and Destroy—Value with Generative AI, 2023.09.21.

II. 주요 행사 일정

행사명	행사 주요 개요		
Generative AI Europe 2023		<ul style="list-style-type: none"> - ‘Generative AI Europe 2023’는 Generative AI 관련 환각, 훈련 데이터 및 모델링, 위험 관리, 비디오 및 이미지 생성의 편견과 같은 주요 주제를 탐구 - 이번 2023년 행사는 Generative AI 기술 분야 업계 리더, 연구원 또는 전문가가 참여하며, 혁신적 솔루션을 공유 	
	기간	장소	홈페이지
	2023.12.4~6	네덜란드, 암스테르담	https://www.aidataanalytics.network/events-generative-ai-europe
GovAI Summit		<ul style="list-style-type: none"> - ‘GovAI Summit’는 공공부문에서 AI의 실제 적용과 사례들을 국방, 정부, 교육 관련 기관 간에 공유 - 이번 행사는 공공 부문의 AI 약속, 글로벌 AI 전망, 윤리(공공 서비스에 AI 실용 가이드) 등, AI 연구에서 연방정부의 역할 등 공공 AI 관련 주제 발표 및 공유 	
	기간	장소	홈페이지
	2023.12.4~6	미국, 버지니아	https://www.govaisummit.com/
AAAI-24		<ul style="list-style-type: none"> - ‘AAAI-24’ 학회는 AI 연구를 촉진하고 AI 및 관련 분야 전체에 걸쳐 연구원, 실무자, 과학자, 학생 및 엔지니어 간의 과학적 교류를 촉진 - 2023년 행사는 연구 발표, 특별 트랙, 초청 연사, 워크숍, 튜토리얼, 고위 회원 프레젠테이션, 대회, 전시 등 진행 - 학회 대상 연구 분야는 혁신적인 AI 활용, AI 활용 도구, 혁신적인 AI 융합, AI 사고 대응 모범사례 등 	
	기간	장소	홈페이지
	2024.2.20.~27	캐나다 밴쿠버	https://aaai.org/aaai-conference/



홈페이지 : <https://spri.kr/>

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

경기도 성남시 분당구 대왕판교로 712번길 22 글로벌 R&D 연구동(A) 4층

22, Daewangpangyo-ro 712beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do, Republic of Korea, 13488