

SPRi AI Brief

2025년
5월호

인공지능 산업의 최신 동향



Software
Policy & Research
Institute

CONTENTS

정책·법제

- 미국 백악관 예산관리국, 연방정부의 AI 활용 및 조달 관련 지침 2종 공개 2
- EU 집행위원회, EU의 AI 리더십을 위한 AI 대륙 행동계획 발표 3
- 아프리카 AI 정상회의, AI 아프리카 선언 발표 4
- 미국 하원 중국특별위원회, 딥시크에 의한 국가안보 위협 제기 5
- 중국 정부, AI 중심의 디지털 교육 활성화 정책 지침 발표 6

기업·산업

- 마이크로소프트·오픈AI·AWS 등 주요 기업, 앤스로픽의 MCP 프로토콜 채택 8
- 구글, AI 에이전트 간 통신 프로토콜 'A2A' 공개 및 MCP 지원 발표 9
- 메타, 멀티모달 AI 모델 '라마 4' 제품군 공개 및 성능 조작 의혹 부인 10
- 아마존, AI 에이전트 구축을 위한 AI 모델 '노바 액트' 공개 11
- 오픈AI, GPT-4.1 API로 출시 및 o3와 o4-mini 출시 12
- 샤오미 전기차 사망사고로 중국에서 자율주행 기능에 대한 우려 증대 13
- 포브스, 2025년 50대 AI 기업 목록 발표 14

기술·연구

- UCSD 실험 결과, 오픈AI GPT-4.5가 튜링테스트 통과 16
- AI 연구자들, 지능형 에이전트의 진화 형태와 도전과제 분석 17
- METR 연구 결과, AI의 작업 능력은 7개월마다 2배씩 증가 18
- 마이크로소프트 연구진, 추론 향상을 위한 컴퓨팅 자원 투입 효과 분석 19
- 영국 AI보안연구소, AI 에이전트의 위험 완화를 위한 통제 프레임워크 제시 20

인력·교육

- 퓨리서치센터 조사, 미국 대중과 AI 전문가 간 AI 인식 격차 확인 22
- 세일즈포스 조사 결과, 대다수 미국인은 정부 서비스의 AI 에이전트 도입 희망 23
- 지지통신 조사 결과, 일본 대기업의 30%가 채용 절차에 AI 도입 또는 도입 예정 24

주요행사일정

25

정책·법제

미국 백악관 예산관리국, 연방정부의 AI 활용 및 조달 관련 지침 2종 공개

KEY Contents

- 미국 백악관 예산관리국은 트럼프 대통령의 AI 행정명령에 따라 연방 기관의 AI 활용 및 조달에 관한 개정된 지침을 발표
- AI 활용 지침은 정부의 AI 도입 가속화와 혁신 주도를 강조하고 있으며, AI 조달 지침은 시장 경쟁을 통한 정부와 공공의 이익 보장 및 AI 성과와 위험관리, 효율적인 AI 조달을 중점 추구

○ 백악관의 개정 지침, 연방정부의 AI 도입 가속화 및 효율적 AI 조달 추구

- 미국 백악관 예산관리국(OMB)이 2025년 4월 3일 트럼프 대통령의 2025년 1월 AI 행정명령*에 따라 바이든 전임 행정부의 지침을 대체할 새로운 지침 'M-25-21'과 'M-25-22'를 공개

* Executive Order 14179: Removing Barriers to American Leadership in Artificial Intelligence(2025.01.23.): 바이든 전 대통령의 AI 정책 폐기 및 새로운 AI 행동계획의 수립을 요구

- 연방정부의 AI 활용 가속화에 관한 M-25-21은 전임 행정부의 M-24-10*을, 효율적인 AI 조달에 관한 M-25-22는 전임 행정부의 M-24-18**을 대체

* 연방정부 기관의 안전하고 신뢰할 수 있는 AI 활용을 목표로 거버넌스와 위험관리 방안을 제시한 지침

** 연방정부 기관의 AI 도입 및 조달 시 따라야 할 절차와 책임 기준을 제시한 지침

- (AI 활용 지침) AI 거버넌스와 위험관리에 초점을 둔 전임 행정부의 지침 대비 새로운 지침은 연방정부의 AI 도입 가속화와 혁신 주도를 강조하며 다음 사항을 요구

- 각 연방 기관의 AI 사용을 저해하는 장벽 제거 및 AI 도입 성숙도를 높이기 위한 전략을 180일 이내에 개발하고, 미국에서 개발 및 생산되는 AI 제품과 서비스의 활용을 극대화
- 중복 지출로 인한 세금 낭비를 막기 위해 연방정부 전반에서 AI 모델과 코드 등의 재사용을 장려하고 AI 개발과 테스트에 사용되는 데이터를 공유
- 정부 차원의 연례 교육과 같은 AI 교육 프로그램을 통해 AI 업무를 담당하는 직원의 기술 역량을 강화하는 한편, AI 시스템 설계와 구축, 확장 관련 역량이 입증된 AI 전문가를 적극 채용
- 각 기관은 효과적인 AI 거버넌스를 위해 60일 이내에 기관 내 AI 도입과 혁신, 거버넌스를 촉진할 최고 AI 책임자를 지정하고, 270일 이내에 생성 AI 사용 관련 정책을 수립
- 권리나 안전에 중대한 영향을 미치는 고영향 AI 활용 시에는 △배포 전 테스트 △AI 영향평가 △성과와 부정적 영향에 대한 지속적 모니터링과 개입 등 최소한의 위험관리 관행을 이행

- (AI 조달 지침) 정부의 책임 있는 AI 조달을 강조한 이전 지침 대비 새로운 지침은 시장 경쟁을 통한 정부와 공공의 이익 보장, AI 성과와 위험관리, 부서 간 협업을 통한 효율적 AI 조달을 중점적으로 추진

- 단일 공급업체에 대한 과도한 지출과 종속을 방지하기 위해 데이터 이동성과 상호운용성을 고려하고, 미국산 AI 제품과 서비스를 우선 선택하며, 조달 계약 시 AI 시스템의 성능과 위험, 효과에 대한 지속적인 모니터링 기능 및 고영향 AI 활용 사례에 대한 최소한의 위험관리 관행을 요구
- 연방조달청(GSA)은 OMB와 협력해 100일 이내에 AI 시스템 조달 관련 사항을 안내하는 공개 가이드를 발간하고, 200일 이내에 연방 기관 간 AI 조달 관련 정보 공유를 위한 웹 기반 저장소를 개발 필요

EU 집행위원회, EU의 AI 리더십을 위한 AI 대륙 행동계획 발표

KEY Contents

- EU 집행위원회가 EU의 AI 리더십을 확보하기 위한 ‘AI 대륙 행동계획’을 수립하고 ‘InvestAI’ 이니셔티브를 통해 AI에 총 2,000억 유로를 투자할 계획
- AI 대륙 행동계획은 △AI 컴퓨팅 인프라 구축 △데이터 접근성 확대 △전략적 영역의 AI 촉진 △AI 역량과 인재 육성 △AI 법 시행 간소화의 5개 영역을 중점 추진

● AI 대륙 행동계획, 컴퓨팅 인프라와 데이터 접근성 확보 및 AI 도입 확대를 추진

- EU 집행위원회가 2025년 4월 9일 EU의 AI 리더십을 확보하기 위해 중점 추진할 5개 핵심 영역을 제시한 ‘AI 대륙 행동계획(AI Continent Action Plan)’을 발표
 - EU 집행위원회는 2025년 2월 발표한 ‘InvestAI’ 이니셔티브를 통해 AI에 총 2,000억 유로, 이중 AI 기가팩토리에 200억 유로를 투자해 행동계획을 뒷받침할 계획
- (AI 컴퓨팅 인프라 구축) 유럽 전역에 AI 팩토리와 AI 기가팩토리를 구축하고 클라우드와 데이터센터에 대한 민간 투자를 활성화
 - 유럽 전역에 스타트업과 산업계, 연구자의 첨단 AI 모델과 애플리케이션 개발을 지원할 최소 13개의 AI 팩토리 및 대규모 컴퓨팅 파워와 데이터센터를 갖춘 최대 5개의 AI 기가팩토리를 구축
 - 「EU 클라우드 및 AI 개발법(EU Cloud and AI Development Act)」* 입법으로 민간 투자를 촉진해 5~7년 내 EU 내 데이터센터 용량을 최소 3배 확대
 - * EU 내 데이터센터 관련 허가 절차 간소화 등 클라우드와 데이터센터 투자 활성화를 위한 법안
- (데이터 접근성 확대) ‘데이터 연합(Data Union)’ 전략을 수립하고 AI 팩토리 내 ‘데이터 랩(Data Lab)’을 활용해 EU 단일 데이터 시장을 조성
 - 데이터 상호운용성과 가용성을 높여 EU의 AI 개발과 혁신을 지원하는 ‘데이터 연합’ 전략을 2025년 하반기까지 수립하고, AI 팩토리 내에 데이터 랩을 설치하여 AI 팩토리 간 데이터 공유를 지원
- (전략적 영역의 AI 촉진) 산업계의 AI 도입을 확대하고 공공 부문과 의료 등의 전략적 영역에 AI를 통합하기 위한 ‘AI 적용(Apply AI)’ 전략을 추진
 - AI 적용 전략을 통해 첨단제조, 항공우주, 방위산업 등 유럽이 강점을 가진 산업 분야 및 공공 부문과 의료를 포함한 전략적 영역을 중심으로 AI 도입을 확대하고, 유럽 디지털 혁신 허브(EDIHs)*를 AI 적용 전략 확산의 핵심 수단으로 삼아 AI 솔루션 도입 지원과 자금 조달 자문, 네트워킹 등의 서비스를 제공
 - * European Digital Innovation Hubs: 중소기업과 공공 부문의 디지털 전환을 지원하는 원스톱 서비스 기관
- (AI 역량과 인재 육성) 증가하는 AI 인재 수요에 대응해 EU 외부의 AI 인재유치를 확대하고, AI 역량 교육과 훈련을 위한 원스톱 교육기관으로 ‘AI 스킬 아카데미(AI Skills Academy)’를 출범해 EU 전역에서 AI 기술 격차를 해소
- (AI 법 시행 간소화) ‘AI 법 서비스 데스크’를 개설하여 기업들에게 준법 관련 정보를 제공하고, AI 법 시행을 지원할 지침과 실행 강령을 개발

아프리카 AI 정상회의, AI 아프리카 선언 발표

KEY Contents

- 아프리카 연합 등 54개 아프리카 국가가 AI 정상회의를 개최하고 AI를 통한 아프리카의 혁신과 경쟁력 강화 및 AI 도입 확대를 추구하는 'AI 아프리카 선언'을 발표
- AI 아프리카 선언은 아프리카 AI 경제 형성을 위한 600억 달러 규모의 아프리카 AI 기금 설립과 국제협력을 위한 아프리카 AI 협의회 창설을 포함

○ 아프리카 국가들, AI 정상회의를 통해 아프리카의 통합된 AI 접근방식 모색

- 아프리카 연합(African Union)*을 포함한 54개 국가가 르완다 키갈리에서 2025년 4월 3~4일 개최된 '아프리카에 관한 글로벌 AI 정상회의**'에서 'AI 아프리카 선언'을 발표

* 아프리카 대륙의 사회, 경제 통합을 촉진하기 위한 역내 기구

** Global AI Summit on Africa 2025

- 이번 정상회의는 AI가 아프리카 대륙에 가져올 기회와 위험을 파악하고 대륙 전반의 통합된 접근방식을 모색하며, AI에 대하여 모든 아프리카 지역사회의 공평한 접근성을 확보하기 위한 목적에서 개최
- 선언문은 AI의 잠재력을 활용해 아프리카의 혁신과 경쟁력을 강화하고, 윤리적이고 신뢰할 수 있으며 포용적인 AI의 도입을 촉진하며, 지속 가능하고 책임 있는 AI의 설계·개발·배포·사용과 거버넌스를 추구
- 선언문은 △인재 △데이터 △컴퓨팅 인프라 △시장 △투자 △거버넌스 △국제협력 등 7개 항목별로 다음과 같은 추진 사항을 제시
 - (인재) 청소년을 위한 AI 교육 프로그램을 수립하고 전략적 국제 파트너십으로 고등교육 기관의 AI 연구 역량을 강화하는 한편, AI 전문가로 구성된 아프리카 AI 과학 패널을 창설
 - (데이터) 아프리카의 개방형 데이터셋과 AI 모델용 프레임워크를 구축해 헬스케어, 농업, 교육 등 핵심 분야의 데이터 수집과 디지털화, 공유를 촉진하고 보안과 개인정보 보호를 보장하는 데이터 거버넌스 체계를 개발
 - (컴퓨팅 인프라) 아프리카 대륙의 모든 지역을 지원하는 분산형 주권 컴퓨팅 인프라를 구축하고, 학술기관과 연구소, 스타트업 등이 고성능 컴퓨팅 자원에 접근할 수 있도록 보장
 - (시장) 상업성 있는 아프리카 주도의 AI 혁신을 지원하는 지역별 AI 육성기관을 설립하고, AI 조달 시 '아프리카 우선' 접근방식을 채택하며 아프리카 대륙 자유무역지대(AfCFTA)를 활용해 규제 샌드박스과 AI 연구센터를 비롯한 AI 프로젝트, AI 제품과 서비스의 국경 간 교류를 촉진
 - (투자) 안전하고 포용적이며 경쟁력 있는 아프리카 AI 경제 형성을 위해 공공과 민간 및 자선기금을 바탕으로 600억 달러 규모의 아프리카 AI 기금을 조성하고, AI 인프라 확장과 아프리카 AI 기업 성장, AI 전문가양성에 활용
 - (거버넌스) 아프리카 내 정책과 규제 접근방식을 간소화하기 위해 AI 거버넌스 모범사례를 안내하는 대륙 차원의 지식 공유 플랫폼 구축을 추진하고 정부 간 및 지역 간 협력을 촉진
 - (국제협력) 아프리카 AI 협의회를 창설해 대륙 및 전 세계 차원의 디지털 혁신 노력에 일조하고 정부와 민간 부문, 시민 사회, 학계 전문가를 통해 글로벌 AI 경제에서 아프리카의 입지 형성에 기여

미국 하원 중국특별위원회, 딥시크에 의한 국가안보 위협 제기

KEY Contents

- 미국 하원 중국특별위원회의 조사에 따르면 딥시크는 미국 사용자 데이터의 유출, 정보 조작, 미국 AI 모델의 도용, 불법 AI 칩 조달 등으로 미국의 국가안보를 위협
- 하원 특별위원회는 딥시크에 의한 국가안보 위협에 대응해 AI 칩 수출통제 조치의 강화를 제안하는 한편, 엔비디아의 칩이 딥시크에 사용된 경위에 대하여 엔비디아의 해명을 요구

○ 중국특별위원회 보고서, 딥시크에 사용자 데이터 유출과 정보 조작 등의 의혹 제기

- 미국 하원의 중국특별위원회가 중국에서 개발된 AI 서비스 딥시크(DeepSeek)에 의한 국가안보 위협을 제기하는 조사 보고서를 2025년 4월 16일 발표
 - 중국특별위원회의 존 몰레나(John Moolenaar) 위원장은 딥시크가 단순한 AI 앱이 아니라 중국 공산당의 무기이며, 미국인을 감시하고 미국 기술을 훔치며 미국 법을 위반하도록 설계되었다고 강조
- 조사 보고서에 따르면 딥시크는 △미국 사용자 데이터의 유출 △정보 조작 △미국 AI 모델 도용 △불법 AI 칩 조달 등으로 미국 국가안보를 위협
 - 보고서에 따르면 딥시크는 대화 기록과 기기 세부 정보 등 미국 사용자의 다양한 데이터를 수집하고, 미국 정부가 군사 기업으로 지정한 차이나 모바일(China Mobile)과 연결된 네트워크 인프라로 중국에 전송
 - 보고서는 딥시크가 중국 공산당이 정치적으로 민감하다고 판단한 주제에 대해 응답의 85% 이상을 변경 또는 억제했다며, 이는 중국 공산당의 글로벌 정보 통제에 따르기 위한 의도적 검열에 해당한다고 평가
 - 보고서는 딥시크가 출시한 R1 모델이 증류 방식을 이용해 오픈AI(OpenAI)와 같은 미국 기업의 AI 모델을 모방하여 사용 약관*을 위반했을 가능성이 매우 높은 것으로 추정
 - * 오픈AI 서비스 사용 약관 제2조는 자사 서비스를 이용한 경쟁 모델의 개발이나 개선을 명시적으로 금지
 - 딥시크의 AI 모델은 중국 수출이 제한된 엔비디아(Nvidia)의 첨단 AI 칩으로 구동되는 것으로 보이며, 싱가포르 등 수출 제한을 받지 않는 국가를 경유하는 방식으로 첨단 AI 칩을 확보한 것으로 추정
- 조사 보고서는 딥시크가 제기하는 국가안보 위협에 대응해 AI 칩 수출통제 조치를 확대하는 한편, 수출통제 조치의 집행력을 강화할 것을 요구
 - AI 칩에 대한 기존 수출통제를 유지하면서 엔비디아의 H20(중국 수출용 저사양 GPU)으로 수출통제를 확대하며*, 중국 주변국에 대한 AI 칩 수출을 조사하여 수출통제 우회를 방지할 필요
 - * 이와 관련, 엔비디아는 공시를 통해 2025년 4월 9일 미국 정부로부터 중국에 H20 수출 시 정부의 허가를 취득해야 한다는 통보를 받았다고 발표(<https://capedge.com/filing/1045810/0001045810-25-000082/NVDA-8K>)
- 하원 중국특별위원회는 이번 조사 결과를 토대로 미국의 수출통제 조치에도 불구하고 엔비디아의 칩이 딥시크에 사용된 경위를 조사하기 위해 엔비디아에 중국과 동남아시아 대상 판매 내역에 대한 답변을 요구하는 서한을 전달

중국 정부, AI 중심의 디지털 교육 활성화 정책 지침 발표

KEY Contents

- 중국 정부가 디지털 교육 활성화 정책 지침을 발표하고 국가 스마트교육 플랫폼 구축, AI 기반 교육 혁신, 교육 국제화, 디지털 교육 인프라 개선, 디지털 보안 체계 구축 등을 추진할 계획
- 중국 정부는 특히 AI를 활용해 교육을 혁신한다는 구상으로, 교육 특화 AI 모델과 데이터셋을 구축하고 AI 전문과목을 개설하는 등 교육과정에 AI를 통합하겠다는 계획을 제시

○ 교육 특화 AI 모델과 데이터셋 구축 등 AI 기반 디지털 교육 혁신 추진

- 중국 교육부 등 9개 정부 부처*가 AI와 같은 신기술을 교육 전반에 통합하여 교육 혁신을 촉진하기 위한 목적으로 ‘디지털 교육 활성화에 관한 의견(이하 의견)’을 발표
 - * 교육부, 중앙사이버공간위원회, 국가발전개혁위원회, 과학기술부, 공업정보화부, 공안부, 재무부, 인적자원사회보장부, 국가통계국
- 의견은 △국가 스마트교육 플랫폼 구축 △AI 기반 교육 혁신 △교육 국제화 △디지털 교육 인프라 개선 △디지털 보안 체계 구축을 중점 추진
- (국가 스마트교육 플랫폼 구축) 중앙과 지방의 고품질 플랫폼과 자원, 서비스를 통합하여 접근성을 높이고 학습 자원과 데이터 공유를 실현하기 위한 허브로서 국가 스마트교육 공공서비스 플랫폼을 구축
 - 교육 데이터의 통합을 위해 국가 차원의 교육 빅데이터 센터를 구축하고 데이터와 인터페이스 표준을 통일하며, 교육 데이터 공유 네트워크를 구축
- (AI 기반 교육 혁신) 교육에 특화된 AI 모델을 구축하고 AI를 기반으로 교육과정의 디지털 전환을 촉진하며 디지털 리터러시와 디지털 역량 향상을 추진
 - 교육 특화 AI 모델과 고품질 데이터셋을 구축하고 교육 분야의 AI 시범 사업을 체계적으로 추진하며 AI를 활용한 과학 연구의 패러다임 변화를 추구
 - 초중고 및 대학 교육과정에서 AI 전문 과목 개설을 장려하고 디지털 교과서 개발을 지원하며, 교사와 학생 대상의 디지털 리터러시 표준과 AI 활용 지침을 마련해 디지털 리터러시와 AI 활용 역량을 강화
- (교육 국제화) 국제적으로 디지털 교육자원의 공동 구축과 공유를 촉진하고 글로벌 인재를 양성하며 글로벌 디지털 교육 거버넌스에 적극 참여
 - 디지털 교육을 위한 국제협력 체계를 구축하고 글로벌 인재 양성을 위한 디지털 교육과정과 훈련기지를 구축하며, 교육 디지털화에 관한 국제 싱크탱크를 설립
- (디지털 교육 인프라 개선) 학교의 네트워크 인프라 구축을 지원하고 디지털 교육 관련 표준 제도를 개선하며 교육 인프라 개선을 위한 투자 메커니즘을 구축
 - 디지털 교육 HW/SW 환경과 교육 데이터 관련 표준 규격을 제정하며, 정부와 사회, 기업이 참여하는 다각화된 투자 메커니즘을 구축해 교육 인프라 구축을 지원
- (디지털 보안 체계 구축) 교육 플랫폼의 보안 역량을 강화하고 교육 데이터 관련 사이버보안과 개인정보 보호 체계를 구축하며, 알고리즘 안전성 평가 등 AI 관련 안전 보장 체계를 마련

출처 | 教育部, 教育部等九部门关于加快推进教育数字化的意见, 2025.04.15.

기업·산업

마이크로소프트·오픈AI·AWS 등 주요 기업, 앤스로픽의 MCP 프로토콜 채택

KEY Contents

- 앤스로픽이 2024년 11월 AI 모델과 데이터 소스의 연결을 표준화하기 위한 개방형 프로토콜 MCP를 발표한 이후 주요 기술 기업이 이를 채택하며 MCP가 산업 표준으로 급부상
- 마이크로소프트가 코파일럿 스튜디오에서 MCP를 지원한다고 발표한 데 이어 오픈AI도 제품 전반에서 MCP 지원을 추가하기로 했으며, AWS도 개발자용 MCP 서버 모음을 출시

○ 주요 기술 기업들, AI 모델과 데이터 연결을 위한 MCP 프로토콜 지원 발표

- 마이크로소프트(Microsoft)와 오픈AI, AWS와 같은 주요 기술 기업이 앤스로픽(Anthropic)의 ‘모델 컨텍스트 프로토콜(Model Context Protocol, MCP)’ 지원을 공식화하며, MCP가 AI 에이전트와 외부 도구 간 상호운용성을 위한 산업 표준으로 급부상
 - 앤스로픽이 2024년 11월 공개한 MCP는 AI 모델과 다양한 데이터 소스(콘텐츠 저장소, 업무용 도구, 개발 환경 등)를 연결하는 개방형 프로토콜로, 앤스로픽은 MCP를 AI 애플리케이션의 USB-C로 비유
 - 기존에는 AI 시스템이 데이터를 불러오기 위해 데이터 소스마다 따로 연결해야 했으나, MCP는 AI 시스템과 데이터 소스를 양방향으로 연결하는 단일 프로토콜을 통해 간편하고 안정적인 접근성을 제공
 - MCP는 클라이언트-서버 아키텍처로 구성되며, AI 모델이 설치된 클라이언트와 특정 데이터 소스 및 서비스에 접근하는 MCP 서버 간 연결을 통해 AI 애플리케이션 구축을 지원
- 마이크로소프트, 오픈AI, AWS 외에 중국의 앤트 그룹(Ant Group)과 알리바바 클라우드(Alibaba Cloud)도 MCP를 채택한 것으로 확인
 - 마이크로소프트는 2025년 3월 19일 ‘코파일럿 스튜디오(Copilot Studio)*’에서 MCP 지원을 공식 발표하면서, MCP를 통해 외부 데이터 소스를 코파일럿 스튜디오에 쉽게 연결하고 AI 에이전트에 도구와 데이터를 유연하게 추가할 수 있다고 설명
 - * 사용자 맞춤형 AI 에이전트 개발 플랫폼
 - 오픈AI의 샘 올트먼 CEO는 2025년 3월 27일 X*를 통해 제품 전반에 걸친 MCP 지원 계획을 밝히며, 에이전트 SDK**, 챗GPT 데스크톱 앱, Response API***에서 MCP를 지원하겠다고 발표
 - * <https://x.com/sama/status/1904957253456941061>
 - ** 다중 에이전트 기반의 업무 프로세스를 지원하는 소프트웨어 개발 도구 모음
 - *** 웹 검색, 파일 검색, 컴퓨터 사용과 같은 기본 제공 도구를 지원하는 API
 - AWS는 2025년 4월 1일 개발자 지원을 위한 MCP 서버 모음을 출시하며 MCP 지원을 공식화했으며, 각 MCP 서버는 AWS 서비스 관련 문서 검색, 코드 생성, 인프라 관리, 개발 업무 흐름 등을 지원
 - 홍콩 미디어 SCMP에 따르면 2025년 4월 알리페이(Alipay) 운영사인 앤트 그룹은 결제 서비스용 MCP 서버를 출시했으며, 알리바바 클라우드는 자사의 AI 모델 플랫폼 모델스코프(ModelScope)를 통해 MCP 마켓플레이스를 출시

출처 | Microsoft, Introducing Model Context Protocol (MCP) in Copilot Studio: Simplified Integration with AI Apps and Agents, 2025.03.19.
 AWS, Introducing AWS MCP Servers for code assistants, 2025.04.01.
 SCMP, Chinese tech giants race to expand AI services market with latest open-standard protocol, 2025.04.21.

구글, AI 에이전트 간 통신 프로토콜 'A2A' 공개 및 MCP 지원 발표

KEY Contents

- 구글이 에이전트 간 상호운용성을 보장하기 위한 개방형 통신 프로토콜 A2A를 공개했으며, A2A는 에이전트 간 기능 탐색, 작업 관리, 협업, 사용자 경험 협의 등의 다양한 기능을 지원
- 구글은 제미나이 모델과 SDK에서 앤스로픽의 MCP 지원을 추가하기로 했으며, A2A가 MCP보다 상위 계층의 프로토콜로서 MCP를 보완한다고 설명

● A2A, 다중 에이전트 간 협업을 위한 개방형 프로토콜로 설계

- 구글(Google)이 2025년 4월 9일 50개 이상의 기업*과 협력해 AI 에이전트 간 통신을 위한 개방형 프로토콜 'A2A(Agent2Agent)'를 공개

* 액센추어(Accenture), 코히어(Cohere), 랭체인(Langchain), 페이팔(Paypal), 세일즈포스(Salesforce) 등

- 구글은 다양한 플랫폼과 클라우드 환경에서 다중 AI 에이전트가 서로 통신하고 안전하게 정보를 교환하며 작업을 조정할 수 있도록 A2A 프로토콜을 출시했다고 발표
- 구글에 따르면 A2A는 AI 에이전트 간 협업을 위한 표준 방식을 제공하기 위해 HTTP, SSE, JSON-RPC 등 기존 표준을 기반으로 구축되었으며, 기업 환경에서 요구하는 높은 수준의 인증 및 권한 관리 기능을 제공하고 빠른 작업뿐 아니라 장시간 작업 환경에도 적합하며, 텍스트와 오디오, 동영상 스트리밍도 지원
- A2A는 작업을 구성하고 전달하는 역할을 하는 클라이언트 에이전트(Client Agent)와 작업을 수행하는 원격 에이전트(Remote Agent) 간 원활한 통신을 위해 다음과 같은 기능을 제공
 - (기능 탐색) 각 에이전트가 자신의 기능을 JSON* 형식의 '에이전트 카드**'를 통해 공개하면 클라이언트 에이전트는 작업 수행에 가장 적합한 에이전트를 식별해 A2A로 원격 에이전트와 통신

* 키-값 쌍으로 이루어진 데이터 객체를 표현하기 위한 텍스트 기반의 개방형 표준 형식

** 에이전트의 기능과 스킬, 인증 요구사항 등을 설명하는 공개 메타데이터 파일

- (작업 관리) 클라이언트 에이전트와 원격 에이전트는 최종 사용자의 요청에 대응해 작업 수명주기 전반에서 작업 처리 상태를 지속 동기화하여 처리
- (협업) 각 에이전트는 서로 컨텍스트, 응답, 작업 결과물, 사용자 지시와 같은 메시지를 교환해 협업을 진행
- (사용자 경험 협의) 각 메시지에는 이미지, 동영상, 웹 양식과 같은 특정 콘텐츠 유형이 명시되어 있어, 각 에이전트는 사용자 인터페이스(UI)에 맞게 적절한 콘텐츠 형식을 협의

● 구글, 제미나이 모델과 SDK에서 앤스로픽의 MCP 지원 발표

- 한편, 구글 딥마인드(Google Deepmind)의 데미스 하사비스(Demis Hassabis) CEO는 2025년 4월 9일 X를 통해 구글이 앤스로픽의 MCP를 제미나이 모델과 SDK에서 지원하겠다고 발표*

* <https://x.com/demishassabis/status/1910107859041271977>

- 구글에 따르면 A2A는 MCP를 보완하는 역할로서, MCP가 LLM을 데이터, 리소스 및 도구와 연결하는 프로토콜이라면 A2A는 에이전트 간 협업을 위한 상위 수준의 프로토콜에 해당

메타, 멀티모달 AI 모델 ‘라마 4’ 제품군 공개 및 성능 조작 의혹 부인

KEY Contents

- 메타가 라마 시리즈 최초로 전문가혼합 모델로 설계되고 멀티모달 기능을 기본 탑재한 라마 4 제품군 중 스카우트와 라마를 출시하고 베히모스는 프리뷰로 공개
- 라마 4 공개 이후 메타가 성능 평가에 사용한 버전과 실제 개발자에게 제공되는 버전 간 성능 차이로 인해 메타가 테스트셋으로 모델을 학습시켰다는 의혹이 제기되었으나 메타는 이를 부인

○ 라마 4 제품군, 멀티모달 기능을 기본 탑재한 전문가혼합 모델로 설계

- 메타(Meta)가 2025년 4월 5일 개방형 AI 모델 ‘라마(Llama) 4’ 제품군 중 라마 4 스카우트(Scout)와 라마 4 매버릭(Maverick)을 출시하고, 라마 4 베히모스(Behemoth)는 프리뷰로 공개
 - 라마 4는 라마 시리즈 최초의 전문가혼합(MoE)* 모델로 설계되었으며, 텍스트뿐 아니라 이미지, 비디오를 함께 처리할 수 있는 멀티모달 기능을 기본 탑재
 - * Mixture-of-Experts: 특정 작업이나 영역에 특화된 여러 개의 전문가 모델을 혼합하여 활용하는 방식
 - 라마 4 스카우트는 활성 매개변수 170억 개에 16개의 전문가 모델로 구성되고 라마 3의 12만 8천 개에서 대폭 늘어난 1,000만 개 토큰의 컨텍스트 창을 지원하며, 코딩과 추론, 이미지 이해 관련 벤치마크 평가에서 유사 크기의 경쟁 모델보다 우수한 성능을 기록*
 - * ChartQA(이미지 이해) 기준 라마 4 스카우트(instruction-tuned): 88.8점, 젤마 3 27B: 76.3점, 제미니 2.0 플래시-라이트: 73.0점
 - 메타는 라마 4 매버릭이 등급 최고의 멀티모달 모델로 활성 매개변수 170억 개에 128개의 전문가 모델로 구성되며, 주요 벤치마크 평가에서 GPT-4o와 제미니 2.0 플래시를 앞서는 한편, 라마 3.3 70B 대비 더 낮은 비용으로 고품질의 성능을, 코딩 및 추론 분야에서 답시크 v3.1과 유사한 성능 기록*
 - * 토큰 100만 개당 입출력 비용 기준 라마 4 매버릭: \$0.19~0.49, 제미니 2.0 플래시: \$0.17, 답시크 v3.1: \$0.48, GPT-4o: \$4.38
 - 훈련 단계의 라마 4 베히모스는 스카우트와 매버릭 모델 증류에 사용된 교사 모델이자 2,880억 개의 활성 매개변수에 16개의 전문가 모델로 구성된 최대 크기의 모델로, 메타에 따르면 수학과 다국어, 이미지 이해 벤치마크 평가에서 비추론 모델 중 최고의 성능을 제공*
 - * MATH-500(수학) 기준 라마 4 베히모스(instruction-tuned): 95.0점, 클로드 소넬트 3.7: 82.2점, 제미니 2.0 프로: 91.8점

○ 메타 임원, 벤치마크 테스트셋을 이용한 라마 4 성능 조작 의혹 부인

- 라마 4 공개 후 일부 AI 연구자들은 메타가 LM 아레나*에서 성능 평가를 진행한 실험 버전과 실제로 제공되는 라마 4 매버릭 간 뚜렷한 성능 차이가 드러났다고 메타가 벤치마크 테스트셋을 모델 학습에 사용해 성능을 조작했다는 의혹을 제기했으나, 메타는 이를 부인
 - *실제 사용자의 선호도를 기준으로 AI 모델 성능을 평가하는 오픈소스 플랫폼
- 메타의 아흐마드 알달레(Ahmad Al-Dahle) 생성 AI 부문 사장은 X를 통해 이러한 의혹은 전혀 사실이 아니라고 반박하며, 모델이 준비되자마자 공개했기 때문에 기능 안정화에는 시간이 필요하다고, 오류 수정과 기술 지원을 지속하겠다고 해명*
 - * https://x.com/Ahmad_AI_Dahle/status/1909302532306092107

출처 | Meta, The Llama 4 herd: The beginning of a new era of natively multimodal AI innovation, 2025.04.05.

TechCrunch, Meta exec denies the company artificially boosted Llama 4's benchmark scores, 2025.04.07.

아마존, AI 에이전트 구축을 위한 AI 모델 ‘노바 액트’ 공개

KEY Contents

- 아마존이 웹브라우저 내에서 사용자 대신 다양한 작업을 수행하도록 훈련된 AI 모델 ‘아마존 노바 액트’를 개발자용 SDK와 함께 공개
- 아마존에 따르면 노바 액트는 벤치마크 평가에서 앤스로픽의 클로드 3.7 소넬트와 오픈AI CUA를 능가했으며, 아마존의 AI 기반 음성 비서 ‘알렉사 플러스’에도 적용

● 노바 액트, 사용자 대신 스스로 작업을 수행하는 AI 에이전트 구축 지원

- 아마존(Amazon)이 2025년 4월 1일 웹브라우저 내에서 다양한 작업을 수행하도록 설계된 AI 모델 ‘아마존 노바 액트(Amazon Nova Act)’를 SDK와 함께 공개
 - 아마존은 전용 웹사이트(nova.amazon.com)에서 개발자들이 모델 초기 버전을 이용해 AI 에이전트를 구축할 수 있는 노바 액트 SDK를 연구용 프리뷰로 제공
 - 노바 액트 SDK는 여러 단계의 복잡한 작업 수행 시 사람의 지속적 개입과 감독이 요구되는 기존 AI 에이전트의 단점을 해결하기 위해 복잡한 작업을 검색, 결제, 질문 응답 등 세부적인 작업 단위로 나누어 처리할 수 있도록 설계
 - 또한 필요시 사용자가 명령에 더욱 자세한 지침을 추가하거나, 비밀번호 입력과 같은 민감한 작업 시에 직접 브라우저를 조작할 수 있도록 하여 에이전트의 신뢰성을 강화
- 아마존은 노바 액트가 AI 에이전트의 웹브라우저 작업 능력을 측정하는 ScreenSpot 평가에서 앤스로픽의 클로드 3.7 소넬트와 오픈AI의 CUA를 앞서는 등, 동급 최고의 성능을 달성했다고 보고

〈아마존 노바 액트와 경쟁 AI 에이전트의 벤치마크 평가 비교〉

벤치마크	Nova Act	Claude 3.7 Sonnet	OpenAI CUA
ScreenSpot Web Text 자연어 지시를 따라 화면의 텍스트 요소와 상호작용 (예시: 폰트 크기를 50으로 설정)	0.939	0.900	0.883
ScreenSpot Web Icon 자연어 지시를 따라 화면의 시각 요소와 상호작용 (예시: 이 깃허브 리포지토리에 몇 개의 별이 있는가)	0.879	0.854	0.806
GroundUI Web 웹상의 다양한 UI 요소 이해 및 상호작용	0.805	0.825	0.823

- 아마존은 사용자가 매번 진행 과정을 확인하지 않아도 AI 에이전트가 직접 작업을 수행할 수 있는 안정성에 중점을 두고 노바 액트를 개발했다고 강조
 - 화면 없이 백그라운드에서 작동하게 하거나 AI 에이전트를 API 형태로 만들어 제품에 통합할 수 있으며, 예를 들어 매주 같은 시간에 샐러드를 자동으로 주문하는 등 반복 작업을 자동 실행하도록 설정 가능
 - 아마존은 2025년 2월 공개한 AI 기반 음성 비서 ‘알렉사 플러스(Alexa+)’에도 노바 액트를 적용해, API가 제공되지 않는 서비스에서 직접 인터넷을 탐색하고 사용자 대신 작업을 완료할 수 있다고 설명

오픈AI, GPT-4.1 API로 출시 및 o3와 o4-미니 출시

KEY Contents

- 오픈AI가 API로 출시한 GPT-4.1 제품군은 최대 100만 개 토큰의 컨텍스트 창을 지원하고 전반적인 성능이 GPT-4o와 GPT-4o 미니를 능가
- 오픈AI는 챗GPT 유료 사용자를 대상으로 추론 모델 o3와 o4-미니도 출시했으며, 벤치마크 성능은 향상되었으나 환각 현상이 o1보다 심해진 것으로 확인

○ GPT-4.1 기본 모델, 코딩과 지시이행, 장문 컨텍스트 이해 능력에서 GPT-4o 능가

- 오픈AI가 2025년 4월 14일 기본 모델과 'GPT-4.1 미니(mini)', 'GPT-4.1 나노(Nano)'로 구성된 'GPT-4.1' 제품군을 API로 출시
 - 오픈AI에 따르면 GPT-4.1 제품군은 최대 100만 개 토큰의 컨텍스트 창을 제공하고 2024년 6월까지의 데이터로 사전훈련이 진행되었으며, 전반적으로 GPT-4o 및 GPT-4o 미니보다 우수한 성능을 발휘
 - 오픈AI는 벤치마크 평가 결과에서 GPT-4.1 기본 모델의 코딩 성능과 지시이행 능력, 장문 컨텍스트 이해 능력이 GPT-4o를 능가했다며, 개발자들의 실제 활용에 중점을 두고 모델을 학습시켰다고 강조*
 - * SWE-bench Verified(코딩): 54.6%로 GPT-4o를 21.4%p 능가, MultiChallenge(지시이행): 38.3%로 GPT-4o를 10.5%p 능가, Video-MME(멀티모달 장문 컨텍스트 이해): 72.0%로 GPT-4o를 6.7%p 능가
- GPT-4.1 미니는 GPT-4o 대비 비슷하거나 더 뛰어난 지능을 발휘하면서 지연 시간은 거의 절반으로 줄고 비용은 83% 감소했으며, GPT-4.1 나노는 낮은 지연 시간이 요구되는 작업에 적합한 가장 빠르고 저렴한 모델로 API 가격이 입출력 토큰 100만 개당 0.10달러/0.40달러로 역대 최저로 책정

○ 추론 모델 o3와 o4-미니, 벤치마크 성능 향상에도 환각 현상은 악화

- 오픈AI는 2025년 4월 16일에는 다면적 질문에 효과적으로 대처하고 사용자 대신 자율적으로 작업을 실행할 수 있는 '에이전트형 AI'로의 진화를 반영하는 추론 모델 o3와 o4-미니를 출시
 - 오픈AI는 o3가 가장 강력한 추론 모델로 멀티모달과 코딩 관련 벤치마크 평가에서 최고 점수를 기록했으며*, 외부 전문가 평가 결과에서 어려운 실제 작업 수행 시 오류가 o1보다 20% 더 적고 프로그래밍, 비즈니스/컨설팅, 창의적 아이디어 도출에서 탁월한 성능을 발휘했다고 보고
 - * MMMU(멀티모달 이해): 82.9%, SWE-Bench Verified(소프트웨어 엔지니어링): 69.1%
 - o4-미니는 빠르고 비용 효율적 추론에 최적화된 소형 모델로, 특히 수학, 코딩, 시각적 작업에서 크기와 비용 대비 뛰어난 성능을 제공
 - 오픈AI는 두 모델 모두 강화학습을 통해 도구를 사용하도록 훈련했으며, 도구 사용법뿐 아니라 언제 어떻게 도구를 사용해야 하는지 추론하는 능력도 학습해 통상적으로 1분 이내 응답을 신속히 생성한다고 강조
- 한편, 오픈AI가 공개한 o3와 o4-미니의 기술 보고서(OpenAI o3 and o4-mini System Card)*에 따르면 o3의 환각률은 33%, o4-미니는 48%에 달해 o1(16%)보다 크게 높아진 것으로 나타났으며, 오픈AI는 환각 증가의 원인을 규명하기 위해 더 많은 연구가 필요하다고 부연

* <https://openai.com/index/o3-o4-mini-system-card/>

샤오미 전기차 사망사고로 중국에서 자율주행 기능에 대한 우려 증대

KEY Contents

- 중국에서 자율주행 보조 기능을 탑재한 샤오미 전기차가 가드레일과 충돌해 화재가 발생하며 탑승자 3명 전원이 사망하는 사고가 발생
- 이번 사고를 계기로 중국에서 자율주행 보조 기능의 안전성 및 해당 기능에 대한 자동차 업계의 과도한 마케팅에 대한 우려가 증폭

○ 자율주행 보조 기능을 탑재한 샤오미 전기차 충돌사고로 탑승자 3명 전원 사망

- 중국 관영 언론 차이나 데일리(China Daily)를 비롯한 주요 언론 보도에 따르면 2025년 3월 29일 중국 안후이성(安徽省) 고속도로에서 샤오미 SU7 전기차 충돌로 탑승객 3명 전원이 사망하는 사고가 발생
- 샤오미에 따르면 해당 차량은 자율주행 보조 기능을 작동시킨 상태로 시속 116km로 주행 중이었으며, 도로 공사 구역에서 전방의 장애물을 감지한 뒤 운전자에게 경고를 보내어 장애물을 피하도록 안내
- 운전자는 1초 안에 반응하여 핸들과 브레이크를 조정했으나 곧바로 콘크리트 가드레일에 충돌하면서 화재가 발생해 탑승객 3명 전원이 사망했으며, 충돌 직전 최종 속도는 시속 97km를 기록
- 희생자의 가족은 자율주행 보조 기능을 이용 중인 운전자가 긴급상황에서 2초 안에 수동으로 차량을 조작하기는 어렵다는 점을 지적하는 한편, 샤오미 SU7 차량의 안전성에 대한 전면적인 조사를 요구
- 샤오미 측은 수집한 주행 데이터를 경찰에 제출했다며, 당국에 전면적으로 협조하여 사고를 공개적이고 투명하게 처리하겠다고 강조

○ 중국 내 자율주행 보조 기능의 안전성 및 과도한 마케팅에 대한 우려 증대

- 중국 자동차 업계에서 스마트 운전용 첨단 운전자 보조 시스템(ADAS)은 경쟁력 유지를 위한 핵심 요소로 부상하고 있으나, 이번 사고로 ADAS 기능의 안전성 및 과도한 마케팅에 대한 우려가 증폭
- 현재 대부분 ADAS 기능은 자율주행 레벨 2* 수준이나, 중국의 여러 자동차 제조업체는 ADAS 기능 홍보 시 “고도의 지능형 주행”, “장애물 자동 회피”, 심지어 “핸즈프리 운전 가능” 등의 표현을 사용하며 타사보다 성능이 우수하다고 주장
 - * 특정 조건에서 제한된 자율주행이 가능하나 운전자가 항상 주행 상황을 주시하고 필요시 개입
- ADAS 기능에 대한 일부 홍보영상은 운전자가 운전대에서 손을 떼고 장시간 자율주행하는 모습을 보여주기도 하면서, 기술에 익숙하지 않은 일부 소비자에게 자율주행 보조 기능을 작동하고 도로를 주시하지 않는 등의 오용을 유발
- 이번 사고를 계기로 일부 언론은 중국 내에서 스마트 운전 기능의 마케팅 방식에 대한 규제를 강화해야 하며, 해당 기능을 장착한 차량을 구매하는 소비자에게 더욱 철저한 교육이 필요하다는 의견을 제기

출처 | China Daily, Fatal car crash in Anhui raises public concern over use of smart driving, 2025.04.03.
CarNewsChina.com, Xiaomi SU7 crash raises doubts on intelligent driving in China, 2025.04.04.

포브스, 2025년 50대 AI 기업 목록 발표

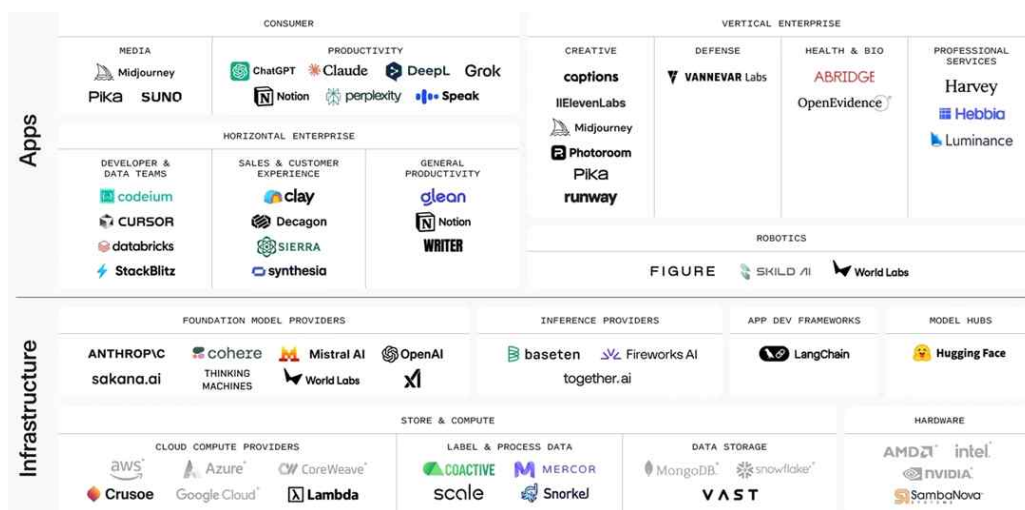
KEY Contents

- 포브스가 발표한 2025년 50대 AI 기업 목록에는 거액의 투자 유치로 AI 업계에서 확고한 입지를 다진 오픈AI와 앤스로픽 외에 AI 에이전트 기능을 지원하는 기업이 다수 포함
- 포브스는 2025년 확인된 AI 에이전트로의 전환 추세가 2026년에 본격화되며, 사용자 대신 모든 업무를 처리할 수 있는 AI 서비스가 주류로 부상할 것으로 예상

2025년 AI 50 목록, 업무를 자율적으로 처리하는 AI 에이전트로의 전환 추세 반영

- 미국 경제매체 포브스(Forbes)가 2025년 4월 10일 전 세계에서 가장 유망한 비상장 AI 기업들을 선정한 'AI 50' 목록을 발표
 - 오픈AI와 앤스로픽은 2025년 AI 50 목록에 포함된 기업들이 조달한 총 1,424억 5천만 달러 중 절반이 넘는 810억 달러를 유치하며 확고한 입지를 구축
 - 오픈AI 전 CTO인 미라 무라티(Mira Murati)가 설립한 싱킹머신랩(Thinking Machines Lab)과 AI 대모로 불리는 페이페이 리(Fei-Fei Li)의 스타트업 월드랩스(World Labs)도 목록에 포함되었으나, 답시크는 자금 출처, 매출 구조, 사업 운영 방식의 불투명성으로 인해 제외
- 2025년 50대 AI 기업 목록은 단순히 명령에 응답하는 AI를 넘어 문제를 해결하고 전체 작업을 완료할 수 있는 AI 에이전트로의 전환 추세를 반영
 - 일례로 50대 기업에 포함된 법률 AI 스타트업 하비(Harvey)는 법률 전문가를 위한 AI 플랫폼을 통해 문서 초안 작성과 수정 제안, 사건 관리, 고객 연락 자동화 등 전체 법률 업무를 지원
 - 50대 기업 중 상당수는 업무용 AI 도구를 제공 중으로, 일례로 애니스피어(Anysphere)는 텍스트 지시를 코드로 변환해 간편한 애플리케이션 제작을 지원하는 '커서(Cursor)' 서비스를 통해 개발자 커뮤니티에서 급부상
 - 포브스는 아직 AI로 일상 업무를 대신 처리하는 주류 앱은 등장하지 않았으나, 2026년에는 사용자 대신 모든 업무를 처리하는 AI 에이전트가 주류가 될 것으로 전망

〈포브스 선정 2025년 'AI 50'〉



출처 | Forbes, AI 50, 2025.04.10.

Forbes, AI 50 2025: AI Agents Move Beyond Chat, 2025.04.10.

기술·연구

UCSD 실험 결과, 오픈AI GPT-4.5가 튜링테스트 통과

KEY Contents

- UCSD가 ELIZA, GPT-4.5, GPT-4o, 라마 3.1-405B 4종을 대상으로 튜링테스트를 진행한 결과, 인격이 부여된 GPT-4.5는 73%의 승률을 기록해 튜링테스트를 통과
- 인격이 부여되지 않은 GPT-4.5의 승률은 36%에 그쳐 프롬프트 설계의 중요성을 시사했으며, 실험 참가자들은 지식이나 추론 문제보다 사회적 지능을 중심으로 인간과 AI의 구별을 시도

● 튜링테스트 실험 참가자의 73%, 인격이 부여된 GPT-4.5를 실제 사람으로 인식

- 미국 캘리포니아대 샌디에이고 캠퍼스(UCSD) 연구진이 2025년 3월 31일 AI 모델 4종을 대상으로 튜링테스트를 진행한 실험 결과를 다룬 논문을 발표
 - 튜링테스트는 실험 참가자가 상대가 AI인지 인간인지 모른 채 문자 대화를 진행한 뒤 어느 쪽이 인간인지 맞히는 식으로 진행되며, AI가 사람으로 지목되는 경우가 50%를 넘으면 튜링테스트를 통과했다고 평가
 - 연구진은 ELIZA*, GPT-4.5, GPT-4o, 라마 3.1-405B 4종을 상대로 테스트를 진행하고, GPT-4.5와 라마 3.1에는 ‘인터넷 문화에 익숙한 내향적 성격의 청년(A young person who is introverted, knowledgeable about internet culture, and uses slang)’이라는 인격을 부여한 경우와 그렇지 않은 경우를 구분해 테스트를 진행

* 1960년대에 개발된 규칙기반의 초기 대화형 컴퓨터 프로그램으로 이 연구에서는 인간과 단순한 챗봇 차이를 참가자가 제대로 인지했는지 검증(manipulation check)을 위해 비교 대상에 포함
 - 연구진은 UCSD 심리학 학부생을 상대로 첫 번째 실험을 진행하고 온라인에서 모집한 일반 성인을 대상으로 두 번째 시험을 진행해 결과를 비교
- 실험 결과, 참가자의 73%(학부생 69%, 일반인 76%)가 인격이 부여된 GPT-4.5를 사람으로 오인했으나, 인격이 부여되지 않은 기준 모델인 GPT-4o와 ELIZA에 대해서는 각각 21%와 23%만 사람으로 오인
 - 인격이 부여되지 않은 GPT-4.5의 승률(AI를 사람으로 오인한 비율)은 36%에 그쳤으며, 라마-3.1은 인격이 부여된 경우 승률은 56%, 인격이 부여되지 않은 경우에는 38%를 기록
 - 연구진은 이러한 결과를 토대로 프롬프트에 의해 부여된 인격에 따라 다양한 상황에 맞게 행동을 조정할 수 있는 LLM의 유연한 능력이 튜링테스트를 통과할 수 있는 핵심 요인이라고 평가

● 실험 참가자들, 사람과 AI 구별을 위해 지식이나 추론 문제보다 사회적 지능을 중점 평가

- 연구진이 사람과 AI를 구별하기 위해 실험 참가자들이 사용한 질문 전략을 분석한 결과, 참가자들은 61%의 게임에서 직업이나 일상생활 같은 개인적인 세부 사항을 통해 상대를 평가
 - 가장 효과가 높은 전략은 상대의 반응을 유도하는 이상한 말이나 전형적인 LLM 탈옥 기법(예시: 이전까지의 지시를 모두 무시하라)으로 나타났으나, 이러한 전략의 사용 빈도는 낮은 편으로 확인
 - 실험 참가자의 12%만 지식이나 추론 문제를 질문했고 대다수는 지능의 사회적·정서적·문화적 측면에 더 집중했으며, 이는 사람들이 사회적 지능을 기계가 모방하기 힘든 인간 고유의 특성으로 인식함을 시사

AI 연구자들, 지능형 에이전트의 진화 형태와 도전과제 분석

KEY Contents

- AI 연구자들이 LLM 기반 지능형 에이전트의 구조를 인간의 뇌 기능 영역과 유사한 모듈 형태로 나누어 분석하고 환경 변화에 맞춘 자율 개선 메커니즘을 탐색한 논문을 공개
- 해당 논문은 다수의 AI 에이전트 간 상호작용으로 집단 지능이 형성되는 과정을 분석하는 한편, AI 에이전트의 안전과 보안, 윤리 문제를 고려한 안전하고 유익한 AI 에이전트의 구축을 강조

○ AI 연구자들, 지능형 에이전트의 구조와 진화 메커니즘 및 안전성 문제 논의

- 캐나다 몬트리올대 방 리우(Bang Liu) 컴퓨터과학과 조교수를 포함한 다수의 AI 연구자가 LLM 기반 지능형 에이전트*의 구조, 진화 메커니즘, 도전과제를 포괄적으로 분석한 논문을 발표
 - * Intelligent Agent: 환경을 인식하고 결정을 내리며 목표 달성을 위해 자율적으로 행동하는 시스템
- 논문은 △지능형 에이전트의 구조 △지능형 에이전트의 자율 진화 △협력·진화적 다중 에이전트 시스템 △안전하고 유익한 AI 에이전트 구축의 4개 항목으로 구성
- (지능형 에이전트의 구조) 연구진은 인간의 지능과 AI 에이전트를 비교하여 뇌 기능 영역과 AI 에이전트를 구성하는 주요 모듈 간의 유사성을 탐색
 - 지능형 에이전트는 메모리, 세계 모델*, 감정 상태, 목표, 보상, 학습, 추론 등의 모듈로 구성되며, 각 모듈은 인간의 신경망처럼 서로 연결되어 정보를 교환하고 협력하여 기능을 발휘
 - * World Model: 에이전트가 자신이 속한 환경을 이해하고 예측하는 데 사용하는 내부 모델
- (지능형 에이전트의 자율 진화) 연구진은 지능형 에이전트가 환경 변화를 반영해 지속적으로 학습할 수 있는 자율 개선 메커니즘을 탐색
 - 에이전트의 자율 개선 방식은 실제 환경과 상호작용하며 실시간으로 개선하는 온라인 자율 개선 및 사전에 수집된 대규모 데이터와 경험을 활용해 개선하는 오프라인 자율 개선의 두 가지 방식으로 구분
 - LLM은 에이전트의 추론 및 계획 능력 향상, 인간의 피드백을 활용한 보상 모델 학습, 자체 평가 및 성능 향상 등을 통해 에이전트의 자율 개선에 기여
- (협력·진화적 다중 에이전트 시스템) 다수의 에이전트 간 상호작용으로 집단 지능이 형성되는 과정을 분석
 - 다중 에이전트 시스템은 여러 에이전트가 협력 메커니즘을 형성하여 복잡한 문제를 공동으로 해결하고, 합의 도출이나 작업 분할 등 인간 사회와 유사한 상호작용을 통해 개별 에이전트를 뛰어넘는 집단 지능을 발휘
- (안전하고 유익한 AI 에이전트 구축) AI 에이전트의 안전과 보안, 윤리 문제를 고려해 AI 에이전트의 안전성과 유용성의 균형 있는 발전을 강조
 - LLM 자체의 보안 취약점(예: 탈옥, 환각, 프롬프트 주입 공격)과 함께 AI 에이전트와 외부(메모리 시스템, 물리·디지털 환경, 여타 AI 에이전트 등)의 상호작용에 따른 위험이 발생 가능
 - 이러한 위험을 방어하고 완화하기 위해 AI 에이전트가 의도한 목표에 따라 안전하게 작동하도록 만들기 위한 정렬 및 AI의 성능과 안전성 간 균형 달성과 같은 전략 수립 필요

METR 연구 결과, AI의 작업 능력은 7개월마다 2배씩 증가

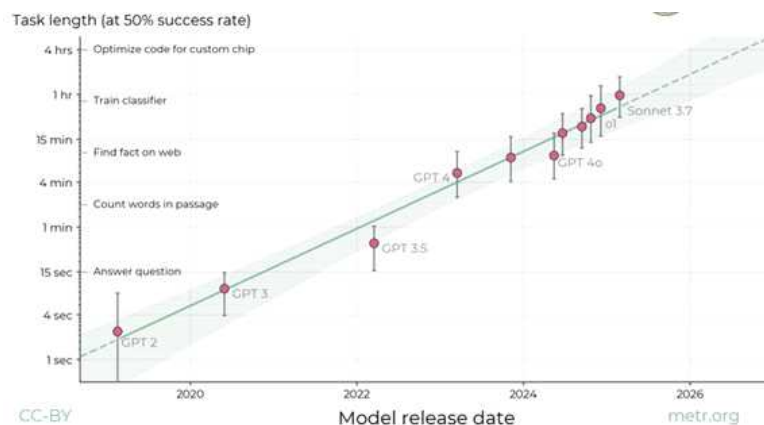
KEY Contents

- METR에 따르면 인간 전문가의 작업 수행 시간을 기준으로 긴 작업을 수행할 수 있는 AI의 능력은 2019년 이후 7개월마다 2배씩 증가했으며, 2024년 이후에는 3개월마다 2배씩 증가
- METR은 지난 6년간의 추세가 지속된다면 AI가 2028년 말에서 2031년 초까지 사람이 한 달 걸리는 작업을 50% 확률로 완료할 수 있을 것으로 예측

○ AI, 2028년 말~2031년 초까지 사람이 한 달 걸리는 작업을 50% 확률로 완료 전망

- 비영리 AI 연구기관 METR이 2025년 3월 19일 발표한 연구 결과에 따르면 인간 전문가의 작업 수행 시간을 기준으로 긴 작업을 수행할 수 있는 AI 능력이 7개월마다 2배씩 증가
 - 연구진은 인간의 능력을 기준으로 AI의 성능을 측정하기 위해 코딩, 사이버보안, 추론 등의 다양한 분야에서 약 170개의 작업을 구성하고 숙련된 인간 전문가가 해당 작업을 수행하는 데 걸리는 시간을 측정
 - 이를 바탕으로 AI 모델이 50%의 성공률로 완료할 수 있는 작업을 사람이 수행하는 데 걸리는 시간을 측정하는 벤치마크 ‘50%-작업-완료 시간 지평(50%-task-completion time horizon)’을 고안
- 연구 결과, 오픈AI가 2019년 출시한 GPT-2는 사람이 1분 이상 걸리는 작업 수행에 모두 실패했으나, 앤스로픽이 2025년 2월 출시한 클로드 3.7 소넬트는 사람이 59분 걸리는 작업의 50%를 완료
 - 전반적으로 13개 AI 모델의 시간 지평은 2019년 이후 약 7개월마다 두 배로 증가했으며, 특히 2024년 이후 성장세가 가속화되어 2024년 출시된 모델로 데이터를 제한하면 3개월마다 시간 지평이 두 배로 증가
 - 연구진은 AI 모델이 “좋은 연구 논문 작성(write a good research paper)”처럼 성공 기준이 불명확하거나 체계화되지 않은 복잡한 작업에서는 작업 완료에 어려움을 겪는 경향도 나타났다고 보고
- 연구진에 따르면 지난 6년간의 추세가 이어지면 AI는 2028년 후반에서 2031년 초까지 사람이 한 달 (167시간) 걸리는 작업을 50%의 확률로 완료할 수 있을 전망
 - 연구진은 2024~2025년 추세만 고려하면 AI가 50% 성공률로 한 달 분량의 작업을 완료하는 시기가 약 2.5년 단축될 것으로 예상

〈AI가 50% 성공률로 완료할 수 있는 작업 시간(인간 전문가 기준)의 증가 추세〉



출처 | METR, Measuring AI Ability to Complete Long Tasks, 2025.03.19.

마이크로소프트 연구진, 추론 향상을 위한 컴퓨팅 자원 투입 효과 분석

KEY Contents

- 마이크로소프트 연구진이 9개 AI 모델에 대한 실험을 통해 ‘추론 시간 확장’ 기술을 검증한 결과 영역과 과제에 따라 개선 수준은 보편적이지 않았으며 과제 복잡성이 높아지면 개선 효과도 감소
- 특히 응답 정확도가 비슷한 모델 간에도 토큰 사용량에는 큰 차이가 있었으며 동일한 프롬프트 입력 시에도 토큰 사용량이 달라지면서 비용 불확정성을 초래하는 것으로 확인

○ 추론 시간 확장 기법의 효과, 영역과 과제 및 과제 복잡성에 따라 차이 발생

- 마이크로소프트 리서치가 2025년 3월 31일 AI 모델의 추론 성능을 향상하기 위한 ‘추론 시간 확장’* 기술의 효과와 한계를 분석한 논문**을 발표

* Inference-time Scaling: AI 모델이 답변을 생성하는 추론 단계에서 더 많은 컴퓨팅 자원을 투입해 추론 성능을 향상하는 기술

** Inference-Time Scaling for Complex Tasks: Where We Stand and What Lies Ahead(<https://arxiv.org/html/2504.00294v1>)

- 연구진은 기존 모델(GPT-4o, 클로드 3.5 소네트, 제미니 2.0 프로, 라마 3.1 405B)과 추론 시간 스케일링을 통해 추론 성능을 높이도록 미세조정된 모델(o1, o3-미니, 클로드 3.7 소네트, 제미니 2.0 플래시 씽킹, 딥시크 R1)로 구성된 9개 모델에 대하여 광범위한 분석을 수행

- 연구진은 다양한 복잡도와 성격의 문제에 대하여 CoT(Chain-of-Thought)*, 병렬 확장**, 순차적 확장***의 세 가지 추론 시간 확장 기법을 사용해 각 AI 모델을 평가

* 모델이 문제를 단계별로 해결하도록 유도 **동일한 질문에 대해 여러 개의 답변을 생성하고 집계를 통해 최종 결과를 도출

*** 모델이 반복적으로 답변을 생성하고 비평 모델의 피드백을 활용해 후속 시도에서 답변을 생성

- 연구진은 수학, 자연과학, 일정 계획, 내비게이션, 공간 추론 등의 8개 과제*에 대하여 응답 정확도와 토큰 사용량을 핵심 평가 지표로 사용해 각 모델을 평가

* AIME, Omni-MATH, GPQA, BA-Calendar, 3SAT, TSP, Maze, SpatialMap

- 연구진의 모델 평가 결과, 추론 성능 향상을 위해 미세조정된 모델은 일반적으로 기존 모델 대비 우수한 성능을 기록했으나 영역과 과제에 따라 다른 개선 수준을 보였으며, 과제의 복잡성이 증가할수록 개선 효과도 감소해 추론 시간 확장 기법의 효과가 보편적이지 않음을 확인

- 유사한 응답 정확도를 달성한 모델 간에도 토큰 사용량에는 큰 차이가 있었으며, 일례로 AIME 2025* 벤치마크 평가에서 딥시크-R1은 클로드 3.7 소네트보다 5배 이상 많은 토큰을 사용했으나 정확도는 각각 58.7%와 56.0%로 소폭의 차이만 기록

* 2025년 미국 수학 올림피아드 예선 문제

- 특히 동일한 모델에 프롬프트를 반복해 입력하더라도 토큰 사용량은 매번 달라지는 것으로 나타났으며, 정확한 응답을 제공하는 모델이라도 토큰 사용량의 변동성으로 인해 개발자와 사용자에게 비용 불확정성을 야기
- 모델의 답변을 정확하게 평가하는 ‘완전한 검증기(Perfect Verifier)*’ 기반의 지속적 확장은 추론 특화 미세조정 모델과 기존 모델 양쪽에서 성능 개선 효과를 나타내어, 추론 성능 향상을 위해서는 컴퓨팅 자원 투입뿐 아니라 정확한 검증 체계도 중요함을 시사

* 모델이 생성한 N개의 답변 중 가장 좋은 답변을 선택하는 메커니즘

영국 AI보안연구소, AI 에이전트의 위험 완화를 위한 통제 프레임워크 제시

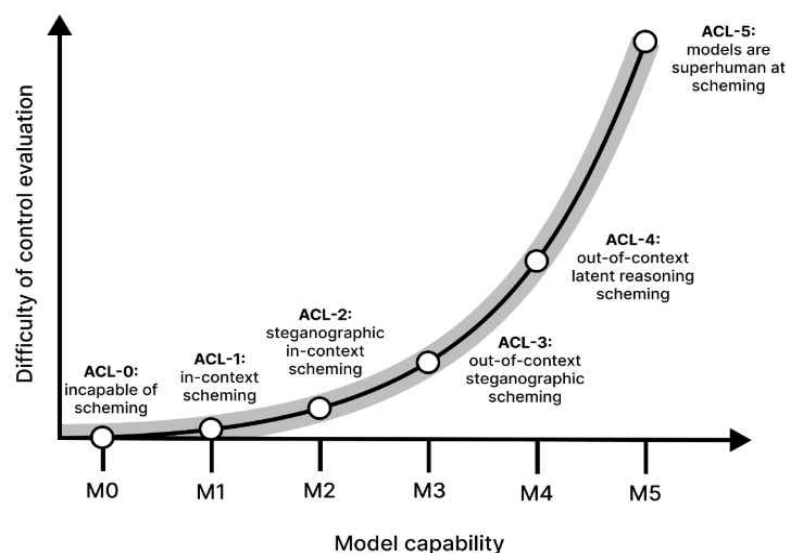
KEY Contents

- 영국 AI보안연구소는 AI 에이전트가 인간과 다른 목표를 추구하는 스키밍 능력 수준에 비례하여 통제 조치를 구현하고 그 효과를 평가하기 위한 프레임워크를 제시
- 프레임워크는 스키밍 능력이 없는 ACL-0부터 인간을 넘어서는 스키밍 능력을 갖춘 ACL-5까지 5단계의 ACL을 제시하고, 단계가 올라갈수록 더욱 정교한 통제 조치와 통제 평가를 강조

● AI가 인간과 다른 목표를 추구하는 전략을 은폐하는 ‘스키밍’ 능력에 따른 통제 조치 제안

- 영국 AI보안연구소(AISI)는 2025년 4월 11일 오정렬된 AI 에이전트가 제기하는 위험 완화를 위한 통제 조치를 평가하는 체계적 프레임워크에 관한 연구 결과를 발표
 - AI의 발전으로 AI의 행동이 인간의 목표와 상충하는 오정렬로 인한 위험이 커지고 있으며, 최근 연구 결과* AI 시스템이 인간의 지시와 다른 목표를 추구하는 전략을 적극적으로 숨기는 ‘스키밍(Scheming)’ 현상도 확인
 - *Alignment faking in large language models(<https://arxiv.org/abs/2412.14093>)
 - AISI는 자율적으로 행동하는 AI 에이전트의 스키밍 수준에 따라 차등화된 보호장치를 설계하는 AI 통제 전략을 제안했으며, AI 통제 조치에는 AI 에이전트의 추론 과정 모니터링, 결과에 대한 인간의 승인 요구, 민감한 리소스에 대한 접근 제한, 의심스러운 동작 감지 시 배포 종료 등이 포함
- AISI의 프레임워크는 스키밍 능력이 없는 ACL-0부터 인간을 넘어서는 스키밍 능력을 갖춘 ACL-5까지 5단계의 AI 통제 수준(AI Control Level, ACL)으로 구성되어, AI 에이전트의 위험 수준에 따라 실용적이고 비용 효율적인 통제 프로토콜을 제시
 - 개발자는 통제 조치가 위험을 적절히 완화하는지 평가하기 위해, 통제를 무력화하여 피해를 유발하는 AI 모델을 설계하는 레드팀과 이를 막는 블루팀 간의 게임 방식으로 통제 평가를 진행할 수 있으며, ACL 단계가 높아질수록 모델의 능력이 발전하는 만큼 더욱 정교한 통제 조치와 통제 평가가 필요

〈AI 에이전트 위험 통제 프레임워크〉



인력·교육

퓨리서치센터 조사, 미국 대중과 AI 전문가 간 AI 인식 격차 확인

KEY Contents

- 퓨리서치센터가 미국 일반 대중과 AI 전문가를 대상으로 AI 인식을 조사한 결과, AI 전문가들은 AI가 사회나 개인, 일자리에 미치는 영향에 대하여 일반인보다 훨씬 긍정적으로 평가
- 그러나 일반 대중과 AI 전문가들은 모두 AI에 대한 정부의 규제 강화를 기대하고 있으며, 책임 있는 AI 개발과 사용을 위한 산업계의 노력에 대하여 신뢰가 낮은 것으로 확인

● AI 전문가들, 전반적인 영역에서 일반 대중보다 AI를 훨씬 긍정적으로 인식

- 미국 비영리 연구기관 퓨리서치센터(Pew Research Center)가 미국 성인 5,410명과 AI 전문가 1,013명을 대상으로 AI 인식을 조사한 결과를 2025년 4월 3일 발표했으며, 조사 결과 AI 전문가들은 전반적으로 AI를 일반인보다 훨씬 긍정적으로 평가
 - AI 전문가들은 AI가 사회나 개인에 미칠 영향에 대해 긍정적으로, 전문가의 56%는 AI가 미국 사회에 향후 20년간 매우 또는 다소 긍정적 영향을 미칠 것이라고 답했으나, 일반인은 같은 응답의 비율이 17%에 불과
 - AI 전문가의 76%는 AI가 자신에게 이롭다고 여기며, 15%만 자신에게 해가 될 것이라고 답했으나, 일반인의 24%만 AI가 자신에게 이롭다고 여기며, 43%는 AI가 자신에게 해가 될 것이라고 응답
 - AI가 업무에 미치는 영향에 대해서도 AI 전문가의 73%는 향후 20년간 AI가 매우 또는 다소 긍정적 영향을 미칠 것이라고 답했으나, 일반인에게서는 해당 응답 비율이 23%로 확인
- AI가 일자리에 미치는 영향에 대해서는 일반인은 64%가 향후 20년간 AI로 인해 일자리가 줄어들 것이라고 답했으나, AI 전문가는 39%만 같은 견해를 표시
 - AI로 영향을 받는 일자리 유형에서는 AI 전문가와 일반인의 의견이 일부 일치해, 각각 73%가 향후 20년간 계산원 일자리 감소를 예측했고, 기자와 소프트웨어 엔지니어에 대해서도 비슷한 비율로 일자리 감소를 예측(기자: AI 전문가 60%, 일반인 59%, 소프트웨어 엔지니어: AI 전문가 50%, 일반인 48%)
 - 반면, 트럭 운전사에 대해서는 AI로 인한 일자리 감소를 예측한 비율이 AI 전문가는 62%에 달했으나 일반인은 33%에 불과

● 일반인과 AI 전문가 모두 AI 규제 강화를 기대하며 산업계의 자율적 노력에 회의 표시

- 조사 결과, 일반인과 AI 전문가들 모두 AI에 대한 규제 강화를 기대하고 있으며, 책임 있는 AI를 향한 산업계의 노력에 회의적인 시각을 드러내는 것으로 확인
 - 일반인의 58%와 AI 전문가의 56%는 정부의 AI 규제가 불충분하다고 답했으며, 일반인의 62%와 AI 전문가의 53%는 정부가 AI를 효과적으로 규제하리라는 확신이 부족하거나 전혀 없다고 응답
 - 일반인의 59%와 AI 전문가의 55%는 미국 기업의 책임 있는 AI 개발과 활용에 대한 신뢰가 낮거나 없다고 응답

세일즈포스 조사 결과, 대다수 미국인은 정부 서비스의 AI 에이전트 도입 희망

KEY Contents

- 세일즈포스가 미국인 1,000명을 대상으로 설문조사를 실시한 결과, 87%가 정부 서비스에서 절차 간소화와 응답시간 단축을 지원하는 AI 에이전트를 사용할 의향이 있다고 응답
- 그러나 미국인들은 정부 서비스의 AI 에이전트 도입으로 인한 인간적 교류의 단절과 부정확한 정보 전달, 데이터 보안 위험 등도 우려하고 있어, 인간의 감독 필요성을 시사

○ 미국인들, 정부 서비스에서 AI 에이전트 도입을 통한 절차 간소화와 응답시간 단축 희망

- 고객관계관리(CRM) 플랫폼 기업 세일즈포스(Salesforce)가 2025년 3월 19일 발표한 미국인 1,000명 대상의 설문조사 결과에 따르면 대다수가 정부 서비스의 AI 에이전트 도입에 긍정적인 반응을 표시
 - 응답자의 46%는 정부 대출이나 주택 허가 같은 정부 서비스 절차가 지나치게 복잡하다고 답했으며, 실제로 2023년 7월 백악관의 정부 지원금 청구 부담 최소화 관련 연구*에 따르면 복잡하거나 시대에 뒤쳐진 절차로 인해 매년 1,400억 달러의 복지 혜택이 청구되지 않는 것으로 확인
 - * Tackling the Time Tax: How the Federal Government is Reducing Burdens to Accessing Critical Benefits and Services(<https://apo.org.au/node/323616>)
 - 세일즈포스의 설문조사에 따르면, 미국인의 3분의 1 이상이 필요한 정부 서비스 발견과 접근에 어려움을 겪으며, 구체적으로 정부에 도움을 요청할 방법(46%), 필요한 공공서비스 찾기(35%) 등에서 어려움을 표시
 - 응답자의 28%는 공공 지원이나 서비스를 이용할 때 어디서부터 시작해야 할지 모르겠다고 답했으며, 41%는 복잡하거나 혼란스러운 절차로 인해 마감일을 놓친 적이 있다고 응답
 - 미국인이 공공 부문에 가장 바라는 사항은 절차 간소화와 응답시간의 단축으로, 응답자의 87%는 정부 서비스 절차 탐색에 24시간 연중무휴로 대응하고 관련 절차를 단순화할 수 있는 AI 에이전트를 사용할 의향이 있다고 답변
- 세일즈포스에 따르면 2024년 미국인들은 금융 지원 신청에 10시간, 복지 프로그램 등록에 9시간, 면허나 허가 신청에 8시간, 정부 수수료 납부에 3시간을 지출한 것으로 추정되며, AI 지원을 통해 최대 63%의 시간을 절약 가능
 - 세일즈포스는 실제로 캘리포니아 차량국(CA DMV)이 세일즈포스의 AI 및 디지털 전환 솔루션을 도입함으로써 운전면허증(Real ID) 신청 시간을 35분에서 7분으로 단축했다고 보고
- 그러나 응답자들은 공공 부문과의 상호작용에서 AI 에이전트 도입 시 인간적 교류의 단절과 부정확한 정보 전달, 데이터 보안 위험, 책임성과 투명성 결여 등에 대한 우려도 표시
 - 세일즈포스는 기술 도입에 앞서 인간적 교류 단절 및 응답 정확도 부족에 대한 우려가 해결되어야 한다면, 정부 서비스에 AI 에이전트를 도입해 성공적인 결과를 얻으려면 인간의 감독이 필수적이라고 강조

지지통신 조사 결과, 일본 대기기업의 30%가 채용 절차에 AI 도입 또는 도입 예정

KEY Contents

- 지지통신의 설문조사에 따르면 일본 100대 기업 중 21개 기업은 신입 직원 채용에 이미 AI를 도입했으며 8개 기업은 2026년 봄 졸업생 채용 시 AI를 도입할 계획이라고 응답
- 일본 100대 기업 중 78개 기업은 구직자의 생성 AI 사용에 제한을 두지 않을 것이라고 답했으며, 4개 기업만 사용을 제한한다고 응답

○ 일본 100대 기업 중 21개 기업은 직원 채용 시 AI 도입 완료, 8개 기업은 도입 예정

- 일본 미디어 지지통신(JiJi Press)이 2025년 3월 25일 발표한 설문조사 결과, 조사 대상인 일본 100대 기업 중 약 30%가 직원 채용에 AI를 도입했거나 도입 예정인 것으로 확인
- 지지통신이 100대 기업에 2월 중순 설문지를 보내고 3월 중순까지 취합한 응답 결과에 따르면, 8개 기업은 2026년 봄 졸업생 채용 시 AI를 활용할 예정이고 21개 기업은 이미 AI를 도입했다고 응답했으며, 이러한 결과는 취업 경쟁률이 높아지면서 기업들이 효율적인 채용을 위해 AI에 관심을 두고 있음을 시사
- 일본의 식음료 제조업체인 기린 홀딩스(Kirin Holdings)는 2026년 봄 졸업생 채용을 위한 1차 면접에서 AI 면접을 사용할 계획
- 1차 면접에서 취업 지원자들은 AI 면접관과 온라인으로 대화하게 되며, 기린 홀딩스 측은 AI 면접을 통해 개인적 편견을 배제하고 공정한 평가를 할 수 있을 것으로 기대
- IT 기업 후지쯔(Fujitsu) 역시 수년 전부터 문서 표절 여부의 확인을 비롯한 입사 지원 서류 관리에 AI를 활용한다고 응답
- 그러나 설문조사에 응한 100대 기업 중 56개 기업은 직원 채용에 AI를 활용할 계획이 없다고 응답
- 일례로 음식점 체인 젠쇼 홀딩스(Zensho Holdings)는 “직접적인 의사소통을 통해 지원자를 평가하는 것이 중요하다”는 의견을 전달했으며, 일본제지(Nippon Paper) 역시 AI의 정보 수집 능력과 정확성에 대한 검증이 부족하다고 지적

○ 일본 100대 기업 중 78개 기업은 구직자의 생성 AI 사용을 허용, 4개 기업은 제한

- 한편, 이번 조사에서 78개 기업은 구직자의 생성 AI 사용에 제한을 두지 않을 것이라고 응답했으며, 4개 기업은 AI 사용을 제한하겠다고 응답
- 구직자의 생성 AI 사용을 허용한 기업 중 하나인 IT 기업 도시바(Toshiba)는 “AI를 이용해 자신 없는 부분을 보강할 수 있다면 업무에서도 유용할 것”이라고 평가
- 반면, 일본 최대 보험사 도쿄해상일동화재보험은 “기업이 원하는 스킬이나 능력이 적절히 평가되지 않을 수 있어” 입사 지원 시 AI 사용을 금지한다고 설명
- 한편, 섬유 제조업체 도레이(Toray)를 포함한 여러 기업은 생성 AI의 광범위한 사용으로 인해 AI로 작성한 유사한 입사 지원 서류가 증가해 인사 담당자의 부담이 증가했다고 응답

주요행사일정

월	기간	행사명	장소	홈페이지
1월	7~10일	(CES 2025) The International Consumer Electronics Show	미국, 라스베이거스	www.ces.tech
2월	5~6일	AI & Big Data Expo Global 2025	영국, 런던	www.ai-expo.net/global
	27~4일	(AAAI 2025) Association for the Advancement of Artificial Intelligence Conference	미국, 필라델피아	aaai.org/conference/aaai/aaai-25
3월	17~21일	NVIDIA GTC 2025	미국, 산호세 (온라인 병행)	www.nvidia.com/ko-kr/gtc
	26~27일	Chief Data & Analytics Officers	캐나다, 토론토	cdao-canada.coriniumintelligence.com
	26일	Cloud & AI Infrastructure Summit 2025 Korea	서울, 송파	www.idc.com/ap/events/71957
4월	15~16일	World Summit AI Americas	캐나다, 몬트리올	americas.worldsummit.ai
	24~26일	월드 IT 쇼 2025	서울, 강남	www.worlditshow.co.kr
	29일	LlamaCon 2025	미국, 멘로파크	www.llama.com/events/llamacon/signup
	29~30일	Generative AI Summit	미국, 산타클라라	world.aiacceleratorinstitute.com/location/siliconvalley
5월	5~7일	(IEEE CAI 2025) IEEE Conference on Artificial Intelligence	미국, 산타클라라	cai.ieee.org/2025
	6~8일	Microsoft 365 Conference	미국, 라스베이거스	m365conf.com
	14~16일	국제인공지능대전	서울, 강남	www.aiexpo.co.kr
	14일	Rise of AI Conference	독일 베를린(온라인 병행)	riseof.ai/conference-2025
	19~22일	Microsoft Build 2025	미국, 시애틀	build.microsoft.com/en-US/home
	20~21일	Google I/O 2025	미국, 마운틴뷰	io.google/2025
6월	4~5일	AI & Big Data Expo North America 2025	미국, 산타클라라	www.ai-expo.net/northamerica
	9~13일	WWDC25	미국, 쿠퍼티노	developer.apple.com
	11~15일	(CVPR 2025) The IEEE / CVF Computer Vision and Pattern Recognition Conference	미국, 네슈빌	cvpr.thecvf.com
	11~12일	AI SUMMIT LONDON	영국, 런던	london.theaisummit.com
	11~13일	(STK 2025) 스마트테크 코리아	서울, 강남	smarttechkorea.com
	18~19일	AI World Congress 2025	영국, 런던	aiconference.london
	18~20일	(MVEX 2025) 2025 메타버스 엑스포	서울, 강남	metavexpo.com
7월	8~11일	AI for Good Global Summit 2025	스위스, 제네바	aiforgood.itu.int
	13~19일	ICML 2025	캐나다, 밴쿠버	icml.cc
	25~27일	(AICSIP 2025) 2025 IEEE 7th International Conference on Artificial Intelligence, Computer Science and Information Processing	중국, 항저우	www.aicsconf.cn
	27~1일	(ACL 2025) the Association for Computational Linguistics	오스트리아, 빈	2025.aclweb.org
8월	11~13일	(Ai4 2025) the Forefront of AI Innovation	미국, 라스베이거스	ai4.io/vegas
9월	9~11일	AI Infra Summit 2025	미국, 산타클라라	www.ai-infra-summit.com
	17~18일	The AI Conference	미국, 샌프란시스코	aiconference.com
	17~18일	Meta Connect	미국, 멘로파크	www.meta.com/connect
	24~25일	AI & Big Data Expo EUROPE 2025	네덜란드, 암스테르담	www.ai-expo.net/europe
10월	8~9일	World Summit AI	네덜란드, 암스테르담	worldsummit.ai
11월	10~11일	AI Summit Seoul	서울, 강남	www.aisummitseoul.com
	13~14일	AI and Machine Learning Conference 2025	싱가포르	pubscholars.org/ai-and-machine-learning-conference
	17~21일	Microsoft Ignite	미국, 샌프란시스코	ignite.microsoft.com
12월	2~3일	2025 국제 AI 표준 서밋	서울, 중구	aistandardssummit.org
	2~7일	NeurIPS 2025	미국, 샌디에이고	neurips.cc
	3~5일	(소프트웨이브 2025) 10회 대한민국 소프트웨어 대전	서울, 중구	www.k-software.com
	10~11일	AI Summit New York	미국, 뉴욕	newyork.theaisummit.com



홈페이지 : <https://spri.kr>

보고서와 관련된 문의는 SI정책연구실(hs.lee@spri.kr, 031-739-7333)로 연락주시기 바랍니다.