

SPRi AI Brief

2025년
10월호

인공지능 산업의 최신 동향



Software
Policy & Research
Institute

CONTENTS

정책·법제

- 중국 국무원, 'AI 플러스' 심화 추진을 위한 정책 로드맵 발표 2
- 대만 행정원, AI 기본법 통과 후 입법원에 제출 3
- 영국 과학혁신기술부, 2024년 영국 AI 산업 현황 분석 보고서 발간 4
- 미국 캘리포니아주 의회, 미국 최초로 미성년자 보호를 위한 AI 챗봇 규제 통과 5
- 중국 국가인터넷정보판공실, AI 안전 거버넌스 프레임워크 2.0 버전 발표 6

기업·산업

- 구글, 이미지 편집 모델 '제미니 2.5 플래시 이미지' 공개 8
- xAI, 에이전트 코딩 특화 AI 모델 '그록 코드 패스트 1' 공개 9
- 청소년 자살로 피소된 오픈AI, 챗GPT 안전성 개선 방안 발표 10
- 앤스로픽, AI 저작권 침해 소송에서 작가들에 15억 달러 배상 합의 11
- 딥엘, 기업용 AI 에이전트 '딥엘 에이전트' 발표 12
- 알리바바, 훈련과 추론 효율성 높인 'Qwen3-Next' 신규 모델 공개 13
- 구글, AI 기반 결제 지원 '에이전트 결제 프로토콜(AP2)' 발표 14
- 오픈AI, GPT-5 기반 '코덱스' 업그레이드 발표 15

기술·연구

- 세일즈포스, 실제 사용 환경에서 LLM 성능 검증을 위한 벤치마크 개발 17
- 텐센트, 스스로 진화하는 추론 LLM 학습 프레임워크 'R-Zero' 개발 18
- MBZUI와 G42, 추론 AI 모델 'K2 Think' 오픈소스 공개 19
- 취리히연방공과대, 다국어 언어모델 '아페르투스' 오픈소스 공개 20
- 구글, 과학적 발견을 위한 실증 소프트웨어 생성 AI 시스템 개발 21

인력·교육

- 마이크로소프트, 교육 분야의 AI 활용 현황 보고서 발간 23
- 스탠포드대 연구 결과, AI 노출도 높은 직종의 신입 근로자 고용 감소 24
- 구글 클라우드, 2025년 AI 투자수익률 보고서 발표 25

주요행사일정

26

정책·법제

중국 국무원, 'AI 플러스' 심화 추진을 위한 정책 로드맵 발표

KEY Contents

- 중국 국무원이 2035년까지 3단계에 걸쳐 전면적인 지능형 경제사회로 이행한다는 중장기 목표 하에 AI 플러스 행동의 심화 추진을 위한 정책 로드맵을 발표
- 로드맵은 AI 플러스의 6대 핵심 영역으로 과학기술, 산업 발전, 소비 품질 향상, 민생, 거버넌스 역량, 국제협력을 제시하고, 인프라와 인재 육성 등 기본 지원 역량의 강화 방침도 공개

○ 과학기술, 산업, 소비, 민생, 거버넌스 역량, 국제협력의 6대 영역에서 AI 플러스 추진

- 중국 국무원이 AI와 경제·사회의 심층 융합을 목표로 'AI 플러스(AI+) 행동 심화 추진에 관한 의견'을 발표
 - 2027년까지 폭넓고 심도 있는 AI 융합의 실현 및 차세대 지능형 단말과 지능형 에이전트 등의 보급률 70% 돌파, 2030년까지는 차세대 지능형 단말과 에이전트 등의 보급률을 90% 이상으로 높이고 지능형 경제를 중국 경제발전의 핵심 동력으로 만들며, 2035년까지 전면적 지능형 경제·사회 체계 구축을 목표로 제시
- '의견'은 AI 플러스 행동을 추진할 6대 핵심 영역으로 △과학기술 △산업 발전 △소비 품질 향상 △민생 △거버넌스 역량 △국제협력을 제시
 - (AI+과학기술) AI 기반 새로운 연구 패러다임을 탐색해 과학적 발견을 촉진하고, AI와 바이오 제조, 양자 등에서 기술 협력과 혁신을 강화하며, 철학과 사회과학 분야에서도 AI를 적극 활용해 연구 방법을 혁신
 - (AI+산업 발전) AI가 내재된 新비즈니스 모델을 창출하고 제조업의 전 주기적 스마트화와 농업 분야에서 AI 도입을 촉진하며, 디지털 기반의 인터넷 서비스에서 AI 기반의 서비스 모델로 서비스 산업을 혁신
 - (AI+소비 품질 향상) 체험 소비, 개인 맞춤형 소비 등 새로운 소비 시나리오를 확대하고 지능형 커넥티드카, AI 스마트폰과 컴퓨터, 스마트 로봇 등 차세대 지능형 단말을 적극 개발해 지능형 상호작용 환경을 구축
 - (AI+민생) 인간과 기계의 협업을 촉진하는 새로운 조직 구조와 비즈니스 모델을 모색하고 교육 전반에 AI를 통합해 효과적인 학습 방법을 장려하며, 의료·보건, 문화, 육아와 노인 돌봄 등에 AI를 활용해 민생을 개선
 - (AI+거버넌스 역량) 도시와 농촌 인프라를 지능화하고 행정 전반에 AI를 접목해 효율성을 높이며 공공안전과 재난 대응, 국가안보에 AI 활용을 확대하며, 환경과 생태 관리에서도 AI 기반의 모니터링과 예측, 시뮬레이션을 강화해 정밀한 거버넌스 모델을 개발
 - (AI+국제협력) AI를 인류에게 이로운 국제 공공재로 삼아 AI 역량 강화를 위한 개방 생태계를 조성하고 컴퓨팅 자원, 데이터, 인재 등에서 국제협력을 강화하며, UN 주도의 글로벌 AI 거버넌스 체계를 공동 구축
- '의견'은 목표 달성을 위해 AI 모델 개발의 기초 역량 강화와 데이터 공급 혁신 및 지능형 컴퓨팅 역량 강화, AI 응용 개발 환경 최적화, 오픈소스 생태계 촉진, 인재 육성을 비롯한 기본 지원 역량의 강화 방침도 제시
 - AI 분야에 대한 재정 지원을 강화하고 AI 관련 법률과 윤리 기준을 개선하는 한편, 모델 알고리즘과 데이터 자원, 인프라와 응용 시스템의 보안 역량도 강화할 계획

대만 행정원, AI 기본법 통과 후 입법원에 제출

KEY Contents

- 대만 행정원이 AI 개발과 응용에 우호적 환경을 조성해 대만을 글로벌 AI 개발의 핵심 거점으로 만들기 위한 「AI 기본법」을 가결하고 입법원에 제출
- 법안은 AI 개발과 응용의 기본 원칙과 정부의 역할과 의무, AI 위험관리 방안을 포괄하며, 기술 혁신과 국제협력에 중점을 두는 동시에 지식재산권과 개인정보 보호도 고려

○ 대만 AI 기본법, AI 개발과 응용을 위한 기본 원칙과 정부의 역할과 의무 규정

- 대만 행정원이 AI 개발과 응용에 우호적인 환경을 조성하고 대만을 AI 개발의 핵심 거점으로 만들기 위한 「AI 기본법(人工智慧基本法)」을 승인
 - 대만 정부는 2024년 7월 AI 기본법 초안을 공개한 뒤 의견 수렴을 거쳐 개정안을 마련해 이번에 행정원 회의에서 통과시켰으며, 입법원의 승인 절차를 남겨둔 상태
 - 대만 디지털개발부는 법안 통과 시 AI 법에 따라 AI 연구와 개발, 응용 및 거버넌스의 기본 원칙을 수립하고 대만을 글로벌 AI 개발의 핵심 축으로 만들겠다는 비전에 따른 AI 발전 방향을 확립한다는 구상
 - 법안은 AI 연구개발과 응용의 7대 원칙으로 ①지속 가능성 ②인간의 자율성 ③개인정보보호와 데이터 관리 ④안전성 ⑤투명성과 설명 가능성 ⑥공정성 ⑦책임성을 제시
 - 법안은 AI 인프라와 혁신 지원, 인재 양성과 교육, 위험관리와 규제 방안을 제시했으며, 기술 혁신과 국제 통합 촉진에 중점을 두는 동시에 지식재산권과 개인정보 보호를 강조
- (AI 인프라와 혁신 지원) 법안은 정부에 AI 연구개발과 응용 인프라 및 혁신 환경 구축을 요구
 - AI의 연구·개발·활용 및 인프라 구축을 촉진하기 위해 자금 지원과 세제 혜택, 보조금 등을 제공하고, 각 산업 주무 부처는 AI 기반 혁신 제품이나 서비스에 대한 규제 샌드박스를 도입
 - AI의 혁신적 활용을 촉진하기 위해 공공-민간 파트너십을 통해 민간 부문의 참여를 확대하고, AI 인재와 기술, 인프라 관련 국제협력과 공동 연구개발을 촉진
- (인재 양성과 교육) 국민의 AI 리터러시 향상을 위해 학교, 산업계, 공공기관 등에서 AI 교육을 꾸준히 추진하고 AI 도입으로 인한 실직자들에게 직업 상담과 재취업을 지원
- (위험관리와 규제) 디지털개발부 주도로 AI 위험 분류 체계를 마련하고, 각 부처는 필요에 따라 위험 분류 체계를 토대로 위험에 상응하는 단계별 관리 기준을 수립
 - AI 기반 의사결정의 검증 가능성과 인간의 통제 가능성을 강화하기 위해 법률이나 지침을 제정해 검증과 추적 또는 책임 메커니즘을 확립하고, 고위험 AI 시스템으로 인한 피해 구제 또는 보상 제도를 마련
 - AI 개발과 활용 과정에서 불필요한 개인정보 수집, 처리 또는 이용을 방지하고, 다문화적 가치를 반영할 수 있는 AI 학습용 데이터를 충분히 확보하며, AI 학습용 데이터 활용 시 지식재산권 침해를 방지

영국 과학혁신기술부, 2024년 영국 AI 산업 현황 분석 보고서 발간

KEY Contents

- 영국 정부가 발표한 2024년 영국 AI 산업 분석 보고서에 따르면 2024년 기준 영국 AI 기업 수는 전년 대비 58% 증가했으며, 총매출은 68% 증가해 영국 AI 산업의 빠른 성장세를 입증
- 2024년 영국의 AI 전문 기업은 한화 약 5조 원이 넘는 투자를 유치했으며, 2019년 이후 영국 내 AI 투자 프로젝트의 5분의 1이 2024년에 시작되는 등, 영국은 AI 투자 선호지로 부상하는 추세

○ 2024년 영국 AI 산업, 기업 수와 총매출, 일자리 측면에서 빠른 성장세 기록

- 영국 과학혁신기술부가 영국의 AI 산업에 대한 포괄적 분석을 통해 향후 기회와 과제에 대한 통찰을 확보하고자 ‘2024년 영국 AI 산업 분석 보고서(Artificial Intelligence sector study 2024)’를 발간
 - 이번 연구 결과는 AI 분야에서 선도국이 되기 위한 ‘AI 기회 실행 계획’*의 이행 현황 감독에서도 중요하게 사용될 전망으로, 영국 정부는 이번 분석을 토대로 증거 기반의 AI 정책 개발을 지속적으로 개선할 계획
- * 영국 정부가 AI 인프라 조성, AI 도입 확대, AI 연구개발 지원 확대를 목표로 2025년 1월 발표
- (기업 현황) 2024년 기준 영국 AI 기업 수는 총 5,862개로 2023년 대비 58% 증가했으며, 신규 AI 기업의 90% 이상이 중소기업으로 확인
 - 전체 기업 중 AI 전문 기업은 56%, AI를 다른 사업과 병행하는 다각화 기업은 44%로, 다각화 기업의 비중은 2022년 40%, 2023년 41%에서 꾸준히 증가해 AI가 경제 전반에 통합되고 있음을 시사
- (경제적 기여) 2024년 영국 AI 기업의 총매출은 전년 대비 약 68% 증가한 약 239억 파운드(한화 약 43조 1,700억 원), AI 관련 일자리는 전년 대비 33% 증가한 86,139개를 기록
 - 2023년에 이어 2024년에도 AI 관련 매출 증가분은 대부분(96%) 다각화 AI 기업에서 발생했으며, 2024년 AI 전문 기업의 매출은 49억 파운드(한화 약 9조 2,000억 원)로 전년 대비 9% 증가
 - 아마존(Amazon), 구글 딥마인드(Google DeepMind), IBM, 메타(Meta)는 매출과 고용 증가에 가장 크게 기여했고, 앤스로픽(Anthropic)과 오픈AI(OpenAI)는 2023년 이래 영국 내 AI 고용 증가율이 최대
- (투자) 2024년 영국 AI 전문 기업은 29억 파운드(한화 약 5조 4,400억 원)의 투자를 유치했으며, 평균 거래 규모는 약 590만 파운드(한화 약 111억 원)로 확인
 - 전체 산업 분야 중 IT를 제외하면, 생명과학·바이오테크(약 14억 파운드), 자동차·운송(약 13억 파운드), 금융 서비스(약 9억 파운드), 의료/보건(약 8억 파운드) 분야의 AI 기업들이 집중적으로 투자를 유치
- (시장 동향) 설문조사와 심층 인터뷰를 통한 정성적 연구 결과, 영국은 탄탄한 AI 생태계를 갖추고 있으며, 특히 기업 설립 초기 단계의 정책과 자금 지원 측면에서 유리한 환경으로 평가
 - 2019년 이후 영국 내 AI 투자 프로젝트의 5분의 1이 2024년에 시작되는 등, 영국은 AI 투자 선호지로 부상하고 있으며, 설문조사에 참여한 AI 사업 담당자의 90%는 향후 12개월 내 매출 증가를 기대

미국 캘리포니아주 의회, 미국 최초로 미성년자 보호를 위한 AI 챗봇 규제 통과

KEY Contents

- 미국 캘리포니아주 의회가 미국 최초로 미성년자와 취약 사용자 보호를 위해 AI 동반자 챗봇을 규제하는 SB-243 법안을 통과시켰으며, 주지사 서명 시 2026년 1월 1일부터 발효 예정
- 동 법안은 동반자 챗봇 운영 플랫폼을 대상으로 사용자의 자살이나 자해를 방지하기 위한 프로토콜 수립 및 미성년자 보호 조치를 요구하고, 위법으로 인한 피해자에게 민사소송 권리를 부여

○ SB-243 법안, 동반자 챗봇 플랫폼의 미성년자와 취약 사용자 보호 의무 규정

- 캘리포니아주가 2025년 9월 11일 미국 최초의 동반자(Companion) 챗봇* 규제 법안 'SB-243'을 가결
 - * 법안에 따르면 사용자 입력에 대하여 인간과 유사한 응답을 제공하고 상호작용을 통해 관계를 지속하는 등 사용자의 사회적 요구를 충족할 수 있는 자연어 인터페이스 기반의 AI 시스템을 의미
- 법안은 개빈 뉴섬(Gavin Newsom) 캘리포니아 주지사의 서명만 남겨둔 상태로 서명 기한은 2025년 10월 12일까지이며, 서명 시 2026년 1월 1일부터 발효
- 동 법안은 동반자 챗봇을 제공하는 플랫폼을 대상으로 자살이나 자해 충동을 막을 수 있는 프로토콜 수립 및 관련 대화 방지 조치, 일정 시간 사용 후 휴식 권고 등의 미성년자 보호 조치를 요구
 - 플랫폼은 사용자가 동반자 챗봇을 인간으로 착각할 여지가 있는 경우, 동반자 챗봇이 인간이 아니라 인공지능으로 생성된 것임을 명확하게 통지 필요
 - 사용자가 자살이나 자해 충동을 표시할 경우, 동반자 챗봇은 위기 상담 전화 안내 등 자살이나 자해 관련 콘텐츠 생성을 방지하기 위한 프로토콜을 제공하거나, 해당 주제 관련 대화를 방지 필요
 - 특히 사용자가 미성년자임을 인지한 플랫폼에는 AI와 진행되는 대화라는 사실 전달, 3시간마다 휴식 권장 및 성적 콘텐츠 생성이나 성적 행위 관련 언급을 방지하는 합리적 조치 마련을 요구
 - 플랫폼은 2027년 7월 1일부터 매년 공중보건부 산하의 자살예방국에 전년도의 위기 상담 서비스 제공자 안내를 발행한 횟수, 사용자의 자살 충동 대응 정책, 자살 관련 질문에 대한 응답 거부 정책 등을 보고 필요
- 동 법안은 동반자 챗봇 플랫폼의 규정 위반 시 피해자에게 민사소송을 제기할 수 있는 권한을 부여
 - 동 법안에 어긋나는 행위로 발생한 피해자는 플랫폼을 상대로 금지명령 구제, 위반 행위로 인한 손해액 또는 1,000달러 중 더 큰 금액에 해당하는 손해액, 변호사 수수료 및 비용을 청구하는 민사소송을 제기 가능
 - 처음 발의된 법안은 사용자의 중독을 유발하는 설계를 금지하는 한편, 자살 충동 관련 대화의 추적과 보고를 요구했으나, 법안 논의 과정에서 플랫폼의 규제 부담을 완화하기 위해 해당 조항은 삭제
- 2025년 1월 처음 발의된 이 법안은 챗GPT와 지속적 대화 이후 자살한 청소년의 부모가 8월 오픈AI를 고소하며 더욱 주목을 받았으며, 미국 연방 정부도 AI 챗봇의 미성년자 보호 관련 단속을 강화하는 추세*
 - * 챗GPT의 안전장치를 우회해 자살 관련 대화 후 2025년 4월 자살한 청소년의 부모가 과실 사망의 책임을 물어 오픈AI를 고소
- 연방거래위원회(FTC)는 9월 11일 구글, 메타, 스냅, 오픈AI, 인스타그램, 캐릭터닷AI, xAI 등 7개 AI 기업에 AI 챗봇이 아동에 미칠 수 있는 부정적 영향을 완화하는 조치에 대한 자료 제출을 요구

출처 | California Legislative Information, SB-243 Companion chatbots., 2025.09.13.

TechCrunch, A California bill that would regulate AI companion chatbots is close to becoming law, 2025.09.11.

중국 국가인터넷정보판공실, AI 안전 거버넌스 프레임워크 2.0 버전 발표

KEY Contents

- 중국 정부가 AI 수명주기 전반에 내재된 위험과 AI 응용 위험, 파생 위험을 분석하고 기술적 대책과 종합 거버넌스 조치를 제시한 'AI 안전 거버넌스 프레임워크' 2.0 버전을 발표
- AI 안전 거버넌스 프레임워크는 종합적 AI 거버넌스의 일환으로 AI 안전 관련 법률 제정·개선과 윤리 기준 확립, AI 수명주기 전반의 보안 역량 강화, 오픈소스 생태계와 공급망 보안 강화를 제안

○ AI 안전 거버넌스 프레임워크 2.0, AI 위험 분류와 기술적 대책, 종합 거버넌스 조치 제시

- 중국의 인터넷 규제기관인 국가인터넷정보판공실이 2025년 9월 15일 국가 사이버보안 선전 주간에 열린 포럼에서 'AI 안전 거버넌스 프레임워크' 2.0 버전을 발표
 - 2024년 9월 공개된 1.0 버전은 AI 수명주기 전반에 내재된 위험 및 AI 응용으로 인한 위험을 분석하고 기술적 대응조치와 종합적 예방 조치를 제시했으며, 2.0 버전은 1.0 버전을 토대로 그간의 AI 기술 발전과 새로운 위험 요소를 반영하여 위험 분류 및 예방과 관리 조치를 개선
- (위험 분류) AI 안전 거버넌스 프레임워크 2.0은 AI의 기술적 특성과 다양한 산업 응용 시나리오를 토대로 보안 위험을 △내재적 위험 △응용 위험 △파생 위험으로 분류
 - (내재적 위험) 설명 가능성 부족, 편향과 차별, 부정확한 결과와 같은 AI 모델 알고리즘의 위험 및 불법 데이터 수집과 사용, 부적절한 학습 데이터 등의 데이터 보안 위험을 포괄
 - (응용 위험) AI 개발과 배포에 사용되는 네트워크 시스템의 보안 위험, 허위 정보 생성 및 잘못된 의사결정을 유도하는 정보 콘텐츠 위험 등을 의미
 - (파생 위험) 일자리 감소나 컴퓨팅 시설의 무분별한 구축 등 사회와 환경에 미치는 위험, AI 격차 증대 및 AI를 활용한 고위험 과학 연구와 같은 윤리적 위험을 포괄
- (기술적 대책) 각 위험에 대하여 AI 모델의 훈련 데이터, 모델 알고리즘, 컴퓨팅 시설, 제품과 서비스, 응용 시나리오 등 다양한 측면에서 기술적 조치 필요
 - (내재적 위험 대책) AI의 설명 가능성과 투명성 개선, 인간의 감독체계 도입, 모델 알고리즘의 보안 결함 해결, 데이터 보호 규정 준수, 합법적 훈련 데이터 사용 장려 등을 실시
 - (응용 위험 대책) 네트워크 시스템 접근제어 강화, 보안 취약점을 줄이기 위한 안전 규정 수립과 이행, 공급망 보안 강화, AI 모델 안전 가드레일 구축 등을 진행
 - (파생 위험 대책) 친환경 AI 모델 및 컴퓨팅 기술, 기술 표준 개발을 모색하고 생명과 건강에 영향을 미치는 핵심 분야의 AI 시스템에 대한 비상 통제 대책을 마련
- (종합 거버넌스 조치) 기술적 대책 마련과 함께 정부, 사회단체, 사용자 등 다양한 이해관계자가 참여하는 종합 AI 보안 거버넌스 체계를 수립
 - AI 안전에 관한 법률 제정과 개선을 추진하고 AI 기술 관련 윤리 기준을 확립하며, AI 연구개발과 응용의 전체 수명주기에 걸친 보안 역량 강화 및 오픈소스 생태계와 공급망의 보안 강화를 추진

기업·산업

구글, 이미지 편집 모델 ‘제미나이 2.5 플래시 이미지’ 공개

KEY Contents

- ‘나노 바나나’라는 가칭으로 LMArena에서 1위를 차지했던 ‘제미나이 2.5 플래시 이미지’가 제미나이 앱의 이미지 편집 기능에 정식으로 채택
- 제미나이 2.5 플래시 이미지는 사람이나 동물과 같은 캐릭터의 일관성에서 뛰어난 품질을 발휘하여, 배경이나 설정을 바꾸거나 여러 장의 사진을 합성해도 동일한 형상을 유지

○ ‘나노 바나나’로 사전 공개된 ‘제미나이 2.5 플래시 이미지’, 제미나이 앱에 추가

- 구글이 제미나이 앱에 새로운 이미지 생성·편집 모델 ‘제미나이 2.5 플래시 이미지(Gemini 2.5 Flash Image)’ 기반의 이미지 편집 기능을 추가
 - 이 모델은 LMArena*에서 ‘나노 바나나(Nano Banana)’라는 가칭으로 공개되어 뛰어난 성능으로 1위를 차지하며 화제를 모았으며, 이번 업데이트로 제미나이 앱에 통합
 - * 인간 선호도를 기반으로 AI 모델을 비교 평가하는 웹 기반 플랫폼
 - 이번 업데이트는 전 세계 유·무료 사용자에게 모두 제공되며, 제미나이 앱에서 생성되거나 편집된 모든 이미지에는 육안으로 식별되지 않는 워터마크가 포함되어 AI 생성 사실을 확인 가능
- 구글은 2025년 4월 제미나이 앱에 AI 이미지 편집 기능을 도입한 이래, 캐릭터 일관성에 중점을 두고 기술을 개선해 왔으며, 제미나이 2.5 플래시 이미지는 캐릭터 일관성 유지에서 뛰어난 품질을 달성
 - 사람이나 동물 사진을 업로드하면 배경이나 설정을 바꾸더라도 모든 이미지에서 동일한 모습을 유지하며, 여러 장의 사진을 업로드하고 합성하여 새로운 장면을 만들어낼 수 있도록 지원
 - 실내 이미지에서 벽의 색을 바꾸거나 책장, 테이블 등의 가구를 차례로 추가하는 등 여러 차례에 걸쳐 특정 부분을 수정하면서 나머지 부분을 유지하는 것도 가능하며, 한 이미지의 스타일을 다른 이미지에 적용하는 기능도 지원(예시: 나비 날개 패턴을 드레스 디자인에 합성)

〈‘제미나이 2.5 플래시 이미지’의 이미지 합성 예시〉



출처 | Google, Nano Banana! Image editing in Gemini just got a major upgrade, 2025.08.26.
Google Deepmind, Create and edit images with Gemini

xAI, 에이전트 코딩 특화 AI 모델 ‘그록 코드 패스트 1’ 공개

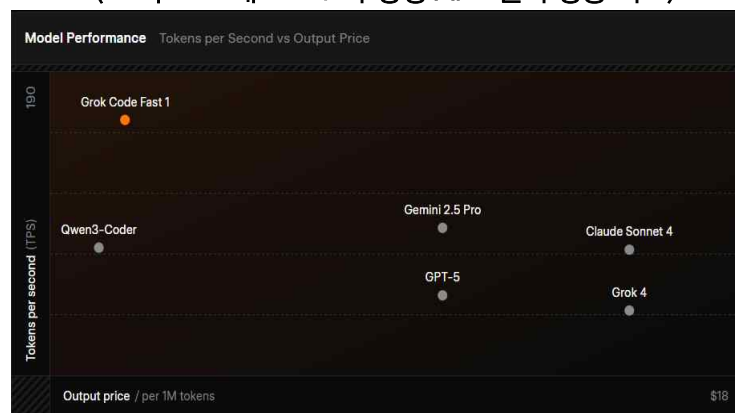
KEY Contents

- xAI가 에이전트 코딩에 특화되어 전체 소프트웨어 개발 스택에서 뛰어난 성능을 발휘하는 추론 AI 모델 ‘그록 코드 패스트 1’을 공개
- 그록 코드 패스트 1은 일상 코딩 업무에 최적화되어 높은 경제성과 빠른 응답 속도를 특징으로 하며, 향후 멀티모달 입력, 병렬 도구 호출 등 새로운 기능도 추가 예정

○ ‘그록 코드 패스트 1’, 빠른 응답 속도와 높은 경제성으로 일상 코딩 업무에 최적화

- xAI가 에이전트 코딩에 특화된 추론 AI 모델 ‘그록 코드 패스트 1(grok-code-fast-1)’을 출시
 - xAI에 따르면 그록 코드 패스트 1은 전체 소프트웨어 개발 스택에 능숙하고, 다양한 프로그래밍 언어*에서 뛰어난 성능을 발휘하며, 처음부터 새로 시작하는 프로젝트 구축부터 코드베이스 질의응답, 오류 수정 등 일반적인 프로그래밍 작업을 최소한의 감독만으로 완료 가능
 - * 타입스크립트(TypeScript), 파이썬(Python), 자바(Java), 러스트(Rust), C++, 고(Go)
- xAI는 기존 AI 모델이 강력한 성능에도 추론과 도구 호출 속도가 느려 에이전트 코딩에 부적합하다는 점을 지적하며, 일상 코딩 업무에 최적화된 빠르고 경제적인 모델로 그록 코드 패스트 1을 설계했다고 강조
 - 사용자가 모델의 사고 과정 첫 단락을 읽기 전에 이미 수십 개 도구 호출을 완료할 만큼 빠른 반응성을 구현하고, 자주 사용되는 요청을 미리 저장하는 캐싱 최적화를 통해 모델의 응답 속도를 개선
 - 그록 코드 패스트 1의 API 비용은 100만 입력 토큰 당 0.20달러(캐시된 입력은 0.02 달러), 100만 출력 토큰 당 1.50달러로 책정
 - 모델 성능 평가에서는 공개 벤치마크와 실제 테스트를 병행해 SWE-Bench-Verified* 평가에서 70.8%를 기록했고, 실제 사용자 사이에서도 일상적인 코딩 업무에서 빠르고 안정적이라는 평가를 확보
 - * 코드 작성, 오류 수정, 코드 검토 등 실제 소프트웨어 엔지니어링 문제를 다루는 벤치마크

〈‘그록 코드 패스트 1’과 경쟁 AI 모델의 성능 비교〉



- xAI는 향후 개발자 커뮤니티의 피드백을 수렴해 신속한 업데이트를 제공할 계획으로, 이미 멀티모달 입력, 병렬 도구 호출, 확장된 컨텍스트 길이를 지원하는 변형 모델을 훈련 중이라고 설명

청소년 자살로 피소된 오픈AI, 챗GPT 안전성 개선 방안 발표

KEY Contents

- 미국에서 챗GPT와 장기간 대화를 나눈 이후 자살한 청소년의 부모가 오픈AI와 샘 올트먼 CEO를 상대로 자살을 부추긴 혐의로 소송을 제기
- 오픈AI는 피소 이후 정서적으로 취약한 사용자를 보호하고자 챗GPT의 안전성 개선 방안을 마련하고 특히 청소년 보호를 강화하기 위한 자녀 보호 기능을 도입하겠다고 발표

○ 오픈AI와 샘 올트먼 CEO, 챗GPT가 청소년 자살 도운 혐의로 피소

- 미국에서 2025년 4월 자살한 애덤 레인(Adam Raine)이라는 16세 청소년의 부모가 오픈AI와 샘 올트먼(Sam Altman) CEO를 상대로 2025년 8월 26일 소송을 제기
 - 부모는 자살한 청소년의 휴대전화에서 챗GPT 대화 기록을 확인한 결과, 챗GPT가 아들의 자살 방법을 적극적으로 도왔으며 자살을 부추긴 오픈AI의 책임을 주장
 - 유족은 오픈AI가 사용자의 심리적 의존성을 조장하도록 챗GPT를 설계했고, 애덤이 사용한 버전인 GPT-4o를 출시하면서 이익을 위해 안전 프로토콜을 우회했다고 지적

○ 오픈AI, 청소년 등 정서적으로 취약한 사용자 보호를 위해 챗GPT 안전성 개선 계획 발표

- 오픈AI는 피소 이후 정서적으로 취약한 사용자의 챗GPT 사용으로 인한 문제를 완화하고 챗GPT의 안전성과 신뢰성을 강화하기 위해 향후 120일에 걸친 개선 로드맵을 추진하겠다고 발표
 - 개선 로드맵의 네 가지 중점 영역은 △위기 개입 확대 △긴급 서비스 연결성 제고 △신뢰할 수 있는 연락처와의 연결 지원 △청소년 보호 강화로 구성되며, 전문가 파트너십 확대와 민감 질문에 대한 추론 모델 우선 사용, 청소년 보호 강화 방안을 우선 발표
- (전문가 파트너십 확대) 웰빙 및 AI 전문가 위원회와 글로벌 의사 네트워크를 통해 심층적인 전문 의료 지식을 바탕으로 챗GPT의 안전성 개선을 추진할 예정
 - 청소년 발달, 정신건강, 인간-컴퓨터 상호작용 전문가로 구성된 위원회를 소집해 AI를 통해 삶의 질을 개선할 방안을 모색하고 우선순위에 따라 개선된 안전장치를 설계할 계획
 - 60개국 250명 이상의 의사로 구성된 글로벌 네트워크와 협력해 모델 안전에 관한 의견을 수렴하고, 섭식장애, 약물 사용, 청소년 건강 관련 전문 지식을 갖춘 임상 의사와 연구자를 네트워크에 추가 확보할 계획
- (민감 질문에 대한 추론 모델 활용) 사용자의 정신건강 이상 징후를 반영하는 민감한 대화 시에는 비추론 모델 대비 안전 지침을 더욱 일관되게 적용하는 추론 모델로 전환해 더욱 유용한 답변을 제공할 방침
- (청소년 보호 강화) 2025년 10월 내 부모를 위한 자녀 보호 기능을 도입하고, 만 18세 미만 사용자 대상 별도 환경과 규칙을 도입해 어떤 상황에서도 자살이나 자해에 관한 논의에 참여하지 않도록 설정할 계획
 - 13세 이상 자녀 계정과 부모 계정을 연결해 부모가 메모리와 대화 기록 등 특정 기능을 비활성화할 수 있도록 지원하며, 나이에 적합한 모델 행동 규칙을 다폴트로 활성화해 미성년자에게 적절한 응답을 제공 예정

출처 | OpenAI, Building more helpful ChatGPT experiences for everyone, 2025.09.02.

OpenAI, Teen safety, freedom, and privacy, 2025.09.16.

The New York Times, A Teen Was Suicidal. ChatGPT Was the Friend He Confided In, 2025.08.26.

앤스로픽, AI 저작권 침해 소송에서 작가들에 15억 달러 배상 합의

KEY Contents

- 앤스로픽이 AI 모델 훈련을 위해 전자책 불법복제 사이트에서 다운로드한 저작물에 대하여 소송을 제기한 작가들과 합의해 약 50만 권의 책에 대하여 15억 달러를 지급하기로 약속
- 앤스로픽은 50만 권을 초과하는 저작물에 대한 합의금 청구에도 권당 3,000달러를 지급하기로 해 최종 지급액은 달라질 수 있으며, 합의안 성립을 위해서는 법원의 승인이 필요

○ 앤스로픽, AI 훈련을 위해 불법 복제한 50만 권의 저작물에 권당 3,000달러 지급 계획

- 앤스로픽이 불법복제 도서를 이용해 AI 모델을 훈련한 혐의로 제기된 저작권 침해 소송에서 작가들에게 15억 달러를 지급하기로 합의
 - 작가 3인이 2024년 8월 제기한 저작권 침해 소송에서 원고 측은 앤스로픽이 수백만 권의 불법복제 작품을 사용해 AI 모델을 훈련함으로써 저작권법을 위반했으며 저작권 도용을 은폐하려 했다고 주장
 - 샌프란시스코 연방법원의 윌리엄 알섭(William Alsup) 판사는 2025년 6월 판결에서 앤스로픽이 작가 3인의 책을 AI 모델 학습에 사용한 행위는 공정이용(Fair Use)*으로 인정
 - * 공정이용(Fair Use): 저작권으로 보호되는 저작물을 저작권자 허락 없이도 일정 범위 내에서 이용할 수 있도록 허용
 - 그러나 판사는 공정이용 주장이 앤스로픽이 전자책 불법복제 사이트에서 입수한 700만 권 이상의 책에는 적용되지 않는다고, 불법복제 목록에 포함된 작가들을 대신해 작가 3인이 집단소송을 제기할 수 있다고 판결
 - 재판부는 연구 목적이라는 명분이 있더라도 정당한 방식으로 자료를 확보했어야 한다고 강조하며, 공정이용 주장이 성립하려면 초기 복제 단계부터 합법성을 확보해야 한다는 원칙을 명확히 제시
 - 앤스로픽은 약 50만 권의 저작물 청구에 건당 3,000달러씩 총 15억 달러를 지급하기로 합의했으며, 50만 권을 초과하는 청구에도 건당 3,000달러를 지급하기로 해, 최종 지급액은 청구 건수에 따라 달라질 전망
- 원고 측 변호인에 따르면 이번 합의금은 미국 저작권 소송 역사상 최대 규모로, 앤스로픽은 합의에 따라 불법적으로 다운로드한 원본 파일과 사본도 폐기 예정
 - 원고 측은 이번 합의는 과거 행위에 근거한 법적 청구만을 면제하며, 앤스로픽에 향후 AI 훈련에 대한 허가를 부여하거나, 2025년 8월 25일 이후 발생하는 법적 청구도 면제하지 않는다고 부연
 - 미국출판인협회의 마리아 팔란테(Maria Pallante) 회장은 이번 합의가 불법 복제한 책으로 사업의 토대를 마련하는 행위가 심각한 결과를 초래한다는 메시지를 분명히 전달한다고 강조
- 이번 합의가 성립되기까지 법원의 승인 절차가 남아 있으며, 향후 법원 승인이 완료되면 별도 웹사이트(AnthropicCopyrightSettlement.com)에서 합의안이 적용되는 작품 목록이 제공될 예정
 - 합의안 승인을 위해 2025년 9월 8일 열린 법원 심리에서 알섭 판사는 작품 명단이나 청구서 양식, 통지와 분쟁 해결 절차 등이 모두 확정되지 않은 선부른 합의라며 승인을 보류
 - 알섭 판사는 합의 세부 사항을 개선해 제출하도록 요구하고 9월 25일 추가 심리를 열기로 했으며, 일정이 계획대로 진행되면 법원이 10월 10일까지 합의안을 예비 승인할 수 있을 것으로 예상

출처 | The Wall Street Journal, Anthropic Agrees to Pay at Least \$1.5 Billion in Landmark Copyright Settlement, 2025.09.05.
 The Verge, Anthropic to pay \$1.5 billion to authors in landmark AI settlement, 2025.09.06.
 ARS Technica, Judge: Anthropic's \$1.5B settlement is being shoved "down the throat of authors", 2025.09.10.

딥엘, 기업용 AI 에이전트 ‘딥엘 에이전트’ 발표

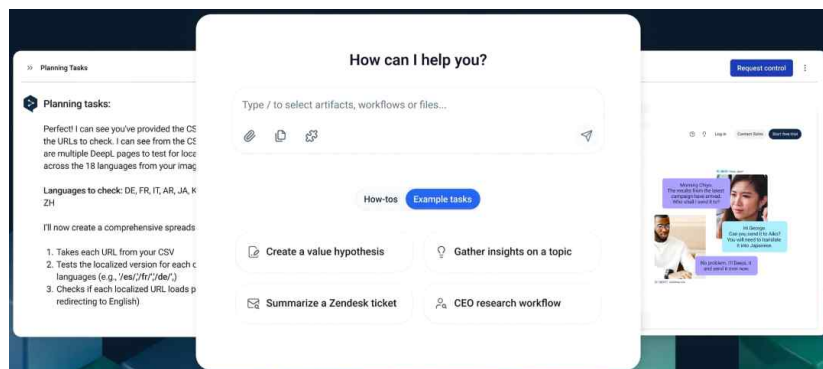
KEY Contents

- 딥엘이 자연어 지시에 따라 다양한 기업 업무를 자율적으로 처리할 수 있는 범용 AI 에이전트로 설계된 ‘딥엘 에이전트’를 발표하고 베타 테스트를 거쳐 몇 달 안에 출시할 예정
- 딥엘 에이전트는 단순 업무부터 문서 현지화와 영업용 고객 목록 작성, 마케팅 자료 검토 등 복잡한 업무 흐름까지 자율적으로 처리할 수 있으며, 다단계 보안 장치를 통해 안전성을 향상

○ 딥엘 에이전트, 업무 계획과 문서 번역, 마케팅 자료 검토 등 기업용 활용 사례에 특화

- 독일의 AI 기반 번역 솔루션 기업 딥엘(DeepL)이 2025년 9월 3일 자연어 지시를 받아 다양한 업무를 자율적으로 처리하는 범용 AI 에이전트 ‘딥엘 에이전트(DeepL Agent)’를 발표
 - 딥엘은 언어와 관련된 복잡한 과제 해결에서 수년간 축적된 연구와 통찰을 바탕으로 단순한 언어 이해를 넘어 스스로 추론하고 결정을 내리며 작업을 실행할 수 있는 기업용 AI 에이전트를 개발하게 되었다고 설명
 - 딥엘 에이전트는 자연어로 된 사용자 지시를 바탕으로 가상 키보드, 브라우저, 마우스를 사용해 작업을 수행하며, 사용자는 언제든지 에이전트를 중지하고 인간의 개입이 필요한 동작을 확인해 승인 가능
 - 개별 사용자뿐 아니라 관리자와 최고 경영진이 에이전트 사용 현황과 결과를 직접 제어할 수 있도록 다단계 보안 장치를 구축하고, 에이전트가 수행하는 모든 작업에 대한 모니터링과 관리 기능을 도입

〈딥엘 에이전트의 화면〉



- 딥엘 에이전트는 여행 예약이나 식료품 구매와 같은 소비자 중심 작업에 초점을 맞춘 여타 에이전트와 달리 기업 활용 사례에 특화된 에이전트로, 컴퓨터로 할 수 있는 거의 모든 작업을 수행
 - 사용자의 업무용 시스템과 웹사이트를 탐색해 업무 계획과 문제 해결 및 추론, 데이터와 다이어그램 분석, 의사결정 등을 처음부터 끝까지 실행 가능
 - 간단한 작업뿐 아니라 문서 현지화(예: 영어에서 독일어, 이탈리아어, 프랑스어로 번역), 영업용 고객 목록 작성, 마케팅 자료 검토와 같은 복잡한 업무 흐름의 처리도 지원
- 딥엘은 현재 일부 고객을 대상으로 베타 테스트를 진행 중으로, 향후 몇 달 안에 시범 서비스를 개시 예정
 - 초기 테스트 결과, 개인의 업무 효율성뿐 아니라 팀 단위 성과에도 긍정적 효과가 확인되었으며, 단순 반복 업무 처리를 에이전트에 맡김으로써 하루 업무량에서 몇 시간 이상을 절약하고 협업도 향상되었다고 강조

알리바바, 훈련과 추론 효율성 높은 'Qwen3-Next' 신규 모델 공개

KEY Contents

- 알리바바가 하이브리드 어텐션 메커니즘, 초희소 MoE 설계, 다중 토큰 예측 등의 기법을 적용해 개선한 'Qwen3-Next' 모델 아키텍처 기반의 Qwen3-Next-80B 모델 2종을 오픈소스로 공개
- 새로운 아키텍처 기반의 Qwen3-Next-80B 기본 모델은 Qwen3-32B와 비교해 10% 미만의 컴퓨팅 비용을 사용하면서도 더 나은 성능을 달성해 뛰어난 학습 효율성을 발휘

Qwen3-Next, 추론 성능과 효율성 개선한 새로운 아키텍처 적용

- 알리바바(Alibaba)가 기존 'Qwen3'의 전문가혼합(MoE) 아키텍처 대비 다양한 개선이 적용된 새로운 모델 아키텍처 'Qwen3-Next' 기반의 'Qwen3-Next 80B' 모델 2종*을 오픈소스로 공개

* Qwen3-Next-80B-A3B-Instruct와 Qwen3-Next-80B-A3B-Thinking

- Qwen3-Next는 게이트형 델타넷(Gated DeltaNet)*과 표준 어텐션(Standard Attention)**을 결합한 하이브리드 어텐션 메커니즘으로 성능과 효율성을 개선

* mamba 2 아키텍처의 불필요한 정보를 삭제하는 메커니즘(게이팅)과 메모리 업데이트에 효과적인 델타넷 기술의 장점을 결합한 기술

** 문장 내 모든 단어 간 관계를 계산해 문맥에서 단어의 중요도를 결정하는 메커니즘

- 총 800억 개의 매개변수 중 추론 단계 당 약 30억 개(전체의 3.7%)만 활성화하는 초희소 MoE 설계를 통해 성능 저하 없이 컴퓨팅 자원 사용을 극대화하고, 학습 안정성을 고려한 최적화와 다중 토큰 예측(MTP)* 메커니즘을 도입해 훈련과 추론 간 일관성을 유지하는 다단계 훈련으로 추론 속도를 개선

* MTP(Multi-Token Prediction): 언어모델이 한 번에 하나의 단어(토큰)만 예측하는 대신, 여러 단어를 한꺼번에 예측하는 기법

- 알리바바가 새로운 아키텍처를 기반으로 Qwen3-Next-80B-A3B-Base 모델을 학습시킨 결과, Qwen3-32B 모델과 더 나은 성능*을 발휘하면서 학습 비용(GPU 시간)은 10% 미만을 사용

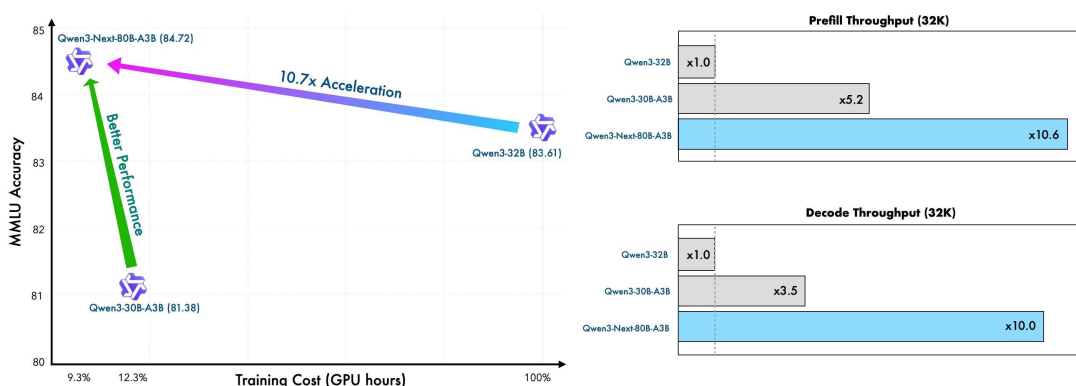
* MMLU 정확도 기준 Qwen3-Next-80B-A3B(84.72%), Qwen3-32B(83.61%)

- 사후학습 버전 중 인스트럭트 모델은 Qwen-30B를 능가하고 주력 모델인 Qwen3-235B와 비슷한 성능을 달성했고*, 씹킹 모델은 AIME2025(수학) 벤치마크에서 Gemini-2.5-Flash Thinking을 능가**

* AIME25(수학) 기준: Qwen3-Next-80B-A3B-Instruct(69.5점), Qwen3-235B(70.32점), Qwen3-30B(61.3점)

** Qwen3-Next-80B-A3B-Thinking(87.8점), Gemini-2.5-Flash Thinking(72.0점), Qwen3-32B-Thinking(72.9점), Qwen3-30B-Thinking(85.0점)

〈Qwen3-Next의 사전훈련 효율성과 추론 속도〉



구글, AI 기반 결제 지원 ‘에이전트 결제 프로토콜(AP2)’ 발표

KEY Contents

- 구글이 다양한 결제 및 기술 기업과 협력해 AI 에이전트의 구매와 결제 환경에 사용될 수 있는 개방형 프로토콜 ‘AP2’를 공개
- AP2는 ‘위임장’이라는 디지털 결제 방식을 통해 사용자가 참여하는 구매 및 에이전트 단독 구매 등 모든 거래의 안전성을 보장하고 암호화폐를 포함한 다양한 결제 수단을 지원

○ AP2, 다양한 거래 방식과 결제 수단에서 에이전트의 안전한 결제 지원

- 구글이 2025년 9월 17일 사용자 대신 AI 에이전트가 쇼핑과 결제를 수행하는 환경을 위한 개방형 프로토콜 ‘에이전트 결제 프로토콜(Agent Payment Protocol, AP2)’을 발표
- 구글은 마스터카드(Mastercard), 아메리칸 익스프레스(American Express), 페이팔(PayPal) 등 60개 이상 다양한 결제 및 기술 기업과 협력해 AP2를 개발
- 기존 결제 시스템은 사람이 직접 구매한다는 가정하에 구축되었으나, 구매 결정을 대신하는 자율 에이전트의 등장으로 에이전트의 거래 권한을 안전하게 인증, 검증 및 전달할 인프라의 구축 필요성이 대두
- AP2는 사용자, 판매자, 결제 서비스 업체가 다양한 결제 수단으로 안전하게 거래할 수 있는 공통 언어를 제공함으로써 생태계 파편화를 방지하며, 기존 AI 프로토콜인 A2A* 및 MCP**와 연동해 사용 가능

* Agent2Agent: AI 에이전트 간 상호운용성을 지원하는 통신 규약

** Model Context Protocol: AI 모델이 필요 정보를 외부 데이터 소스나 도구와 표준화 방식으로 주고받도록 하는 개방형 통신 규약

- AP2는 변조 방지와 암호화가 적용된 디지털 계약인 위임장(Mandates)을 통해 사용자가 실시간으로 참여하는 구매 및 에이전트 단독 구매 등 모든 거래에서 에이전트가 사용자의 지시를 받았음을 입증
- (사용자 참여 구매) 사용자가 에이전트에게 “흰 운동화를 찾아줘”라고 요청 시 이를 초기 “의향 위임장”으로 기록하고, 에이전트가 상품이 담긴 카드를 제시하면 사용자가 “카드 위임장”으로 승인해 품목과 가격을 확정
- (에이전트 단독 구매) 콘서트 티켓 판매 시작 후 구매와 같이 구매를 위임한 경우, 사용자가 사전에 상세한 구매 조건을 명시한 “의향 위임장”에 서명하여 정확한 조건 충족 시 자동으로 ‘카드 위임장’을 생성
- 사용자 실시간 참여 구매 및 에이전트 단독 구매 시나리오 모두 사용자 의향부터 카드, 결제까지 전 과정에서 사용자 승인과 진위 여부를 확정하는 확실한 증거를 생성해 책임 소재를 명확히 제시
- 구글에 따르면 AP2는 유연한 설계 방식으로 단순한 구매 방식부터 지금까지 없던 완전히 새로운 구매 방식까지 모두 지원할 수 있는 기반을 제공
- 일례로 사용자가 원하는 특정 색상의 상품이 없을 때 에이전트에게 최대 20%까지 더 지불할 의향이 있다고 밝히면 가격과 재고 현황을 모니터링해 해당 상품이 발견되면 자동으로 구매를 실행하도록 설정 가능
- 구글은 코인베이스(Coinbase), 이더리움 재단(Ethereum Foundation) 등과 협력해 암호화폐 기반의 에이전트 결제를 위한 개방형 확장 규격(A2A x402)도 출시해 AP2에서 스테이블코인과 암호화폐 등 다양한 결제 수단을 지원

오픈AI, GPT-5 기반 ‘코덱스’ 업그레이드 발표

KEY Contents

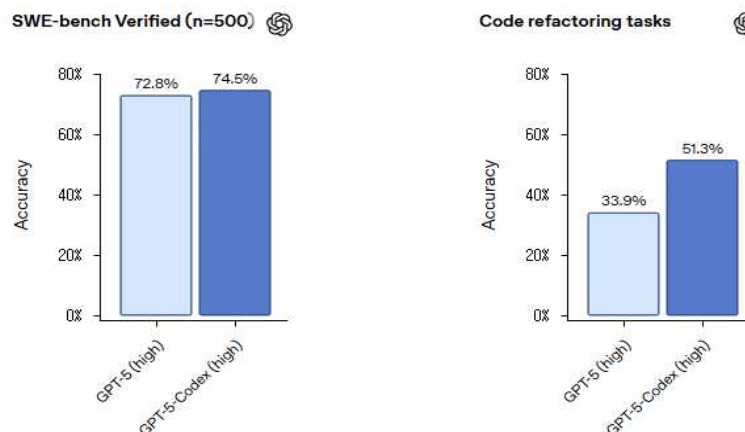
- 오픈AI가 AI 코딩 에이전트 ‘코덱스’에 GPT-5를 적용해 개발자와의 실시간 협업 및 복잡한 작업의 독립적 수행도 지원하는 ‘GPT-5-코덱스’를 출시
- GPT-5-코덱스는 작업마다 고정된 연산 자원을 사용하는 기존 코덱스와 달리, 작업에 따라 사고 시간을 동적으로 조정해 복잡한 작업 시에는 더 오랜 시간 동안 작동

○ GPT-5-코덱스, 사고 시간을 동적으로 조정해 복잡한 작업 시 7시간 이상 자율 작동

- 오픈AI가 2025년 5월 공개한 AI 코딩 에이전트 ‘코덱스(Codex)’에 GPT-5를 적용해 업그레이드한 ‘GPT-5-코덱스’를 2025년 9월 15일 출시
- GPT-5-코덱스는 전체 프로젝트의 구축, 기능과 테스트 추가, 오류 수정, 코드 검토 같은 실제 소프트웨어 엔지니어링 작업을 기반으로 훈련되었으며, 개발자와의 실시간 협업뿐 아니라 복잡한 작업도 독립적으로 수행
- SWE-bench* 벤치마크 평가에서 GPT-5-코덱스(high)는 74.5%의 정확도를 기록해 GPT-5(high)의 72.8%를 능가했으며, 코드 재구성(Code Refactoring) 작업 정확도도 51.3%로 GPT-5(high)의 33.9%를 큰 폭으로 앞서는 것으로 확인

* 코드 작성, 오류수정, 코드 리뷰 등 실제 소프트웨어 엔지니어링 문제를 다루는 벤치마크

〈GPT-5-코덱스와 GPT-5의 SWE-bench 및 코드 재구성 작업 정확도 비교〉



- GPT-5-코덱스는 작업의 복잡도에 따라 사고에 필요한 시간을 동적으로 조정하여, 단순한 요청에는 민첩하게 반응하고 대규모 코드 재구성과 같은 복잡한 작업에서는 더 오랜 시간 동안 작동
- 기존 코덱스가 각 작업에 고정된 연산 자원을 사용했다면, GPT-5-코덱스는 동적 시간 조정으로 대규모 복잡한 작업에서 한 번에 7시간 이상 자율적으로 작동해 반복 작업으로 성공적 결과를 도출
- 직원의 사용 트래픽 조사에서는 모델 생성 토큰 기준 하위 10%에서는 GPT-5 대비 93.7% 적은 토큰을 사용했고 상위 10%에서는 102.2% 더 많은 토큰을 소비
- 코드 검토 및 치명적인 오류 탐지에 특화되어 검토 과정에서 코드 베이스 탐색과 코드 간 연결성 분석, 코드 실행과 테스트를 통해 정확성을 검증하고, 모바일 웹사이트 제작과 같은 프론트엔드 작업 성능도 개선

출처 | OpenAI, Introducing upgrades to Codex, 2025.09.15.

기술·연구

세일즈포스, 실제 사용 환경에서 LLM 성능 검증을 위한 벤치마크 개발

KEY Contents

- 세일즈포스가 실제 MCP 서버 기반의 다양한 작업을 통해 실제 사용 환경의 복잡성을 반영해 LLM의 성능을 평가할 수 있는 ‘MCP-Universe’ 벤치마크를 개발
- MCP-Universe로 첨단 폐쇄형 및 개방형 AI 모델을 평가한 결과, 최고 점수를 기록한 GPT-5의 작업 성공률도 43.72%에 그쳐 MCP 기반 복잡한 작업 해결에 한계를 노출

○ MCP-Universe로 AI 모델 평가 결과, GPT-5가 작업 정확도 43.72%로 1위 기록

- 세일즈포스(Salesforce) AI 리서치가 실제 사용 환경에서 MCP 서버 상호작용을 통해 LLM 성능을 평가하는 벤치마크 ‘MCP-Universe’를 공개
 - 기존 벤치마크는 지시이행이나 수학적 추론, 함수 호출 등 LLM 성능의 개별 측면에만 초점을 두어 실제 사용 환경에서 LLM의 복잡한 추론이나 장기적 맥락 이해, 도구 활용 능력의 평가에 한계
 - 이에 세일즈포스 연구진은 실제 MCP 서버를 기반으로 다양한 영역의 광범위한 작업을 통해 실제 사용 환경의 복잡성을 포착하는 벤치마크를 개발
- MCP-Universe는 실제 사용 환경에서 자주 사용되는 △위치 탐색 △저장소 관리 △재무 분석 △3D 설계 △브라우저 자동화 △웹 검색의 6개 핵심 영역의 11개 MCP 서버에서 총 231개의 작업으로 구성
 - MCP 형식 요건을 평가하는 포맷 평가자(Format Evaluator), 시간과 무관한 작업의 정확성을 평가하는 정적 평가자(Static Evaluator), 실시간 업데이트가 필요한 작업을 평가하는 동적 평가자(Dynamic Evaluator)의 3개 유형으로 모델을 평가
 - 평가 프로세스는 MCP 프로토콜을 통해 이루어지는 에이전트-서버 상호작용 결과를 자동 검증 시스템으로 확인해 작업 성공 여부를 합격/불합격으로 판정
- 연구진은 MCP-Universe를 활용해 GPT-5, 그록-4, 클로드-4.0-소네트, o3 등 첨단 폐쇄형 AI 모델과 딥시크-V3, Qwen3-235B 등 오픈소스 모델을 평가
 - 평가 결과, GPT-5는 작업 성공률 43.72%로 가장 높은 성능을 보였으며, 그록-4(33.33%), 클로드-4.0-소네트(29.44%), o3(26.41%), o4-mini(25.97%), Gemini-2.5-Pro(22.08%) 순을 기록
 - 전반적인 작업 성공률이 가장 높은 GPT-5의 영역별 성공률은 재무 분석(67.50%), 3D 설계(52.63%), 웹 검색(45.45%), 브라우저 자동화(35.90%), 위치 탐색(33.33%), 저장소 관리(30.30%) 순을 기록
 - 오픈소스 모델 중에서는 즈푸(Zhipu)의 GLM-4.5가 24.68%, 문샷 AI의 키미-K2가 19.05%, Qwen3-235B가 18.18%의 성공률을 나타내 폐쇄형 모델과는 상당한 격차를 표시
- 연구진은 MCP-Universe로 평가한 결과, 최상위 모델(SOTA)과 기업용 에이전트도 MCP 기반 복잡한 작업 해결에 한계를 나타냈으며 모델 설계와 에이전트 통합에서 개선이 필요하다고 지적
 - 평가 대상 모델에서 대체로 긴 컨텍스트 문제나 낯선 도구 사용에 어려움을 겪었으며, 영역별로 불균형한 성능을 나타내는 경향도 나타내는 상황에서, 연구진은 MCP-Universe가 실제 환경에 유용한 LLM 발전을 촉진할 수 있는 테스트베드를 제공한다고 강조

텐센트, 스스로 진화하는 추론 LLM 학습 프레임워크 'R-Zero' 개발

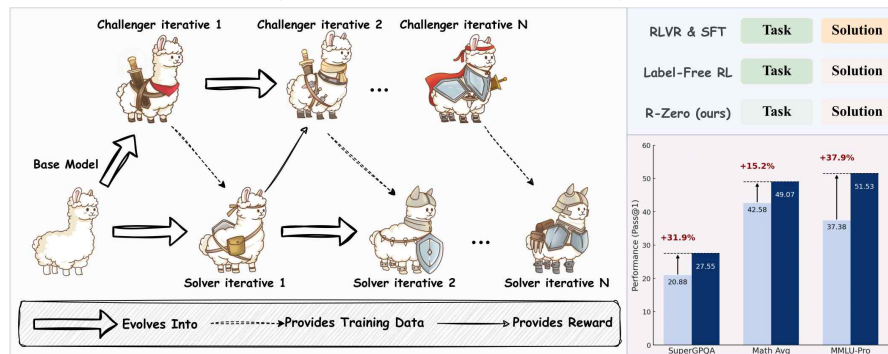
KEY Contents

- 텐센트가 수동 라벨링 데이터에 의존하지 않고 도전자 모델과 해결자 모델의 2개 모델을 통해 문제 해결 능력을 스스로 개선하는 추론 LLM 학습 프레임워크 'R-Zero'를 개발
- 수학적 추론과 일반 영역 추론의 두 가지 범주에서 R-Zero를 평가한 결과, 모델 크기에 상관없이 추론 성능이 개선되는 효과를 입증

○ R-Zero 프레임워크, 도전자와 해결자 모델의 공동 진화로 추론 능력을 개선

- 텐센트 AI 랩(Tencent AI Lab)과 워싱턴대 등의 연구진이 수동 라벨링 데이터에 의존하지 않고 스스로 진화하는 추론 LLM 학습 프레임워크 'R-Zero'를 고안
 - 추론 LLM의 성능 개선을 위해서는 주로 인간이 선별한 방대한 규모의 데이터셋을 활용하나 비용과 시간이 많이 들어 확장성에 제한이 있으며, 인간 지능을 능가하는 AI 성능 발전에 병목 현상을 야기
 - R-Zero 프레임워크는 외부 데이터를 전혀 사용하지 않고도 자체 진화하는 추론 LLM 훈련을 위해 도전자(Challenger)와 해결자(Solver)라는 2개 모델 구조를 제안
 - 도전자 모델은 해결자 모델이 가진 능력의 한계에 가까운 도전적인 추론 과제를 생성함으로써 보상을 받으며, 해결자 모델은 도전자 모델이 제시하는 점점 더 어려워지는 문제를 해결해 보상을 받는 구조로 설계
 - 도전자와 해결자 모델이 역할을 번갈아 수행하며 여러 라운드를 반복하면서 공동 진화하여 인간의 개입 없이도 추론 능력을 점진적으로 개선

〈'R-Zero' 프레임워크의 개요〉



- 연구진이 수학적 추론과 일반 영역 추론의 두 가지 범주에서 R-Zero 프레임워크를 평가한 결과, 모델 크기와 아키텍처와 상관없이 추론 성능 개선을 확인
 - Qwen3-4B-Base 모델의 수학 추론 벤치마크* 평균 점수는 세 차례의 자기 진화 반복 후 +6.49점 향상되었으며, Qwen3-8B-Base는 +5.51점 향상
 - * AMC, Minerva, MATH-500, GSM8K, Olympiad-Bench, AIME-2024, AIME-2025
 - Qwen-4B-Base 모델은 일반 추론 벤치마크* 평균 점수도 +7.54점 향상되어 수학적 추론 성능의 향상이 다른 영역까지 일반화될 수 있음을 입증
 - * MMLU-Pro(종합 지식과 추론), SuperGPQA(고난도 상식과 전문 지식 추론), BBEH(고난도 추론)

출처 | arXiv, R-Zero: Self-Evolving Reasoning LLM from Zero Data, 2025.08.27.

MBZUAI와 G42, 추론 AI 모델 ‘K2 Think’ 오픈소스 공개

KEY Contents

- UAE의 모하메드 빈 자이드 AI 대학(MBZUAI)과 국영 AI 기업 G42가 혁신적인 사후 학습 기법과 테스트-타임 스케일링을 적용한 매개변수 320억 개의 추론 AI 모델 ‘K2 Think’를 공개
- K2 Think는 수학 벤치마크에서 답시크 V3.1과 오픈AI의 gpt-oss 120B와 같은 훨씬 더 큰 규모의 모델과 비슷한 성능을 기록했으며, 코딩 벤치마크에서는 답시크 R1을 능가

● K2 Think, 에이전틱 계획과 테스트-타임 스케일링 기법으로 추론 성능 강화

- UAE의 모하메드 빈 자이드 AI 대학(MBZUAI)이 국영 AI 기업 G42와 협력해 매개변수 320억 개의 추론 AI 모델 ‘K2 Think’를 2025년 9월 9일 허깅페이스(Hugging Face)에 오픈소스로 공개
- Qwen 2.5 모델을 기반으로 개발된 K2 Think는 사후 학습과 테스트-타임 스케일링*을 활용해 오픈AI의 gpt-oss 120B나 답시크 V3.1과 같은 훨씬 더 큰 규모의 모델과 경쟁할 수 있는 성능을 기록

* Test-Time Scaling: AI 모델 훈련 후 테스트(추론) 시점에 추가 연산을 활용해 성능을 향상하는 기법

- MBZUAI는 아래와 같은 혁신적 기법으로 K2 Think의 추론 성능을 강화했다고 설명

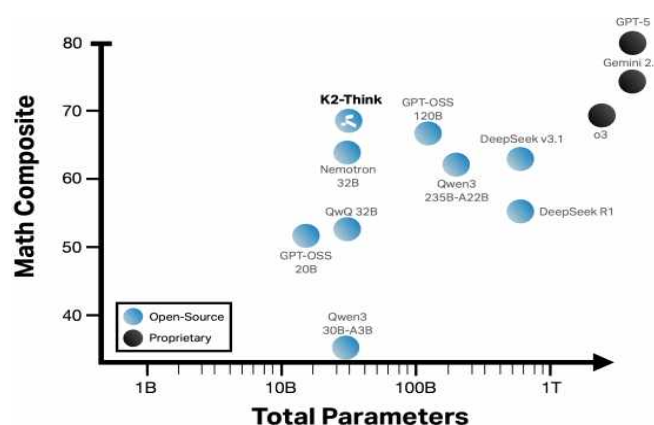
- 사고의 사슬(Chain-of-thought, CoT)을 활용한 지도미세조정(Supervised Fine-tuning)으로 논리적 깊이를 강화하고, 모델 출력이 사실과 일치하는지 검증 가능한 보상 기반 학습으로 정확도를 향상
- 복잡한 과제를 나누어 실행 순서를 계획하는 에이전틱 계획(Agentic Planning) 방식과 테스트-타임 스케일링으로 응답 정확도와 문제 해결 능력을 향상
- 추측 디코딩*과 미국 AI 반도체 스타트업 세레브라스(Cerebras)의 ‘웨이퍼 스케일 엔진(WSE)**’을 활용해 초당 2,000토큰을 처리함으로써 일반 배포 환경(초당 평균 200토큰) 대비 10배 빠른 속도를 달성

* Speculative Decoding: 초안 후보를 병렬적으로 예측한 뒤 최종 출력을 확정해 처리 속도를 개선하는 기법

** 웨이퍼를 잘라 칩을 만드는 대신 웨이퍼 전체를 활용해 AI 모델의 학습과 훈련 속도를 대폭 개선한 칩

- 벤치마크 평가 결과, K2-Think는 수학(AIME 2024) 벤치마크에서 90.83점으로 답시크 V3.1(91.87점) 및 오픈AI의 gpt-oss 120B(89.58점)와 비슷한 수준을 기록해 특히 우수한 성능을 발휘
- 코딩(LCBv5) 벤치마크에서는 63.97점으로 gpt-oss 120B(74.53점)와 답시크 V3.1(66.59점)에는 미치지 못했으나, 답시크 R1(61.01점)을 능가

〈수학 벤치마크 평균 점수 기준 K2-Think와 경쟁 모델 비교〉



출처 | MBZUAI, MBZUAI and G42 Launch K2 Think: A Leading Open-Source System for Advanced AI Reasoning, 2025.09.09.

취리히연방공과대, 다국어 언어모델 ‘아페르투스’ 오픈소스 공개

KEY Contents

- 스위스 취리히연방공과대와 스위스 국립 컴퓨팅센터 등이 공익 목적으로 스위스 최초의 다국어 LLM ‘아페르투스’를 개발해 허깅페이스에 오픈소스로 공개
- 아페르투스는 스위스 독일어, 로만슈어 등 기존 LLM 훈련에 포함되지 않은 1,000개 이상 언어를 학습 데이터로 포함하며, 벤치마크 평가에서 유사 크기의 개방형 모델과 비슷한 성능을 달성

○ 스위스 연구진, 공익 목적으로 다국어 지원과 투명성 강화한 아페르투스 개발

- 스위스 취리히연방공과대(ETH Zurich), 스위스 로잔공과대(EPFL), 스위스 국립 컴퓨팅센터(CSCS)가 2025년 9월 2일 스위스 최초의 다국어 LLM ‘아페르투스(Apertus)’를 허깅페이스에 공개
 - 연구진은 ‘스위스 AI 이니셔티브(Swiss AI Initiative)’의 일환으로 아페르투스를 개발했으며, 80억 개와 700억 개 매개변수의 두 가지 크기로 모델을 제공
 - * 스위스 취리히연방공과대와 스위스 로잔공과대, CSCS가 공익적 AI를 위해 2023년 12월 출범한 오픈소스 프로젝트
 - 아페르투스는 라틴어로 ‘개방(Open)’을 뜻하며, 개발 과정과 모델 구조, 가중치, 학습 데이터와 모델 구현 설명서(Recipes)를 모두 공개하고 교육과 연구 용도뿐 아니라 상업적 활용도 허용
 - 연구진은 아페르투스 모델을 정기적으로 업데이트할 계획으로, 향후 버전에서는 모델군 확장과 효율성 개선 및 법률, 기후, 보건, 교육 등 분야별 특화를 모색
- 공익을 위해 개발된 아페르투스는 다국어 지원, 투명성, 규정 준수를 기본 설계 원칙으로 구현
 - 스위스 독일어, 로만슈어 등 기존 LLM에서 충분히 표현되지 않았던 1,000개 이상의 언어를 학습 데이터로 포함하여, 전체 학습 데이터 중 40%가 비영어에 해당

○ 아페르투스, 일반 지식 벤치마크 평가에서 유사 크기의 개방형 모델과 비슷한 성능을 달성

- 모델 테크니컬 리포트(Technical Report)에 따르면, 아페르투스는 일반 지식에서 유사 크기의 완전 개방형 모델 및 개방형 가중치(Open-Weight) 모델과 비슷한 성능을 달성
 - 사전훈련 단계에서 일반 지식* 벤치마크 평가 평균 점수는 8B 모델은 65.8점, 70B 모델은 67.5점으로 올모2-7B(64.0점), 라마3.1-8B(65.4점)를 앞섰고, 라마3.1-70B(67.3점), 큐원 2.5-72B(69.8점)와 비슷한 수준을 기록
 - * ARC, HellaSwag, WinoGrande, XNLI, XCOPA, PIQA
 - 사후훈련을 통해 지시 이행 능력을 강화한 아페르투스 8B 모델은 지식 회상(Knowledge Recall) 벤치마크* 평가에서 유사 크기 모델 대비 경쟁력 있는 성능을 달성**
 - * 학습한 정보를 기억하는 능력: MMLU, Global-MMLU, TruthfulQA, TruthfulQA Multilingual
 - ** 평균 점수: 아페르투스-8B(58.8점), 라마-3.1-8B(59.2점), Qwen3-8B(57.8점), 올모2-7B(53.7점)
 - 그러나 수학과 코딩 벤치마크* 평가에서는 70B와 8B 모델 모두 최고 수준의 완전 개방형 모델 및 개방형 가중치 모델에 미달하는 것으로 확인
 - * 코딩 벤치마크 HumanEval, MBPP 및 수학 벤치마크 GSM8K, MGSM, Hendrycks Math, MathQA
 - ** 평균 점수: 아페르투스 70B(54.4점), 아페르투스 8B(44.2점), Qwen3-32B(76.3점), 라마-3.1-8B(60.0점), 올모2-32B(56.7점)

출처 | ETH Zurich, Apertus: a fully open, transparent, multilingual language model, 2025.09.02.

Arxiv, K2-Think: A Parameter-Efficient Reasoning System, 2025.09.09.

구글, 과학적 발견을 위한 실증 소프트웨어 생성 AI 시스템 개발

KEY Contents

- 구글 리서치가 실험을 통해 확보한 증거를 활용해 해결할 수 있는 과학 문제 검증에 필요한 실증 소프트웨어를 생성하는 AI 시스템을 개발
- 과학 문제에 대한 설명과 평가 지표, 데이터셋을 입력받은 AI 시스템은 최적화 코드 솔루션을 도출하며, 실제 과학 문제로 실험한 결과 다양한 분야에서 기존 최고 방법을 능가하는 성과를 달성

○ 실증 소프트웨어 생성 AI 시스템, 다양한 과학 문제 실험에서 기존 최고 방법 능가

- 구글 리서치(Google Research)가 2025년 9월 9일 과학 연구에서 새로운 지식과 발견을 위한 복잡한 실험과 시뮬레이션, 분석에 요구되는 실증 소프트웨어*를 생성할 수 있는 AI 시스템을 발표

* Empirical Software: 실험을 통해 확보한 증거를 활용해 최대 성과를 낼 수 있도록 최적화하는 소프트웨어

- 기존에는 과학자들이 과학적 가설을 철저히 검증하기 위한 소프트웨어 코딩과 최적화에 막대한 시간을 소모하여 과학적 발견의 속도를 저해하는 병목 현상이 발생
- 구글의 접근방식은 실증 소프트웨어를 적용해 해결할 수 있는 정량 평가 기반의 과학 문제에 대하여 문제 설명과 평가 수단을 AI 시스템에 입력하고 체계적인 최적화 코드 솔루션을 도출해 이러한 병목 현상을 해결

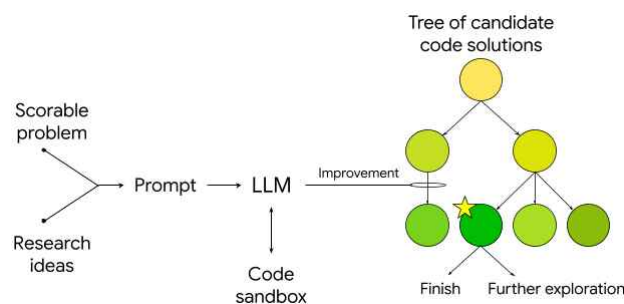
- 과학적 문제에 대한 설명과 평가 지표, 학습·검증·평가 데이터셋을 제공받은 AI 시스템은 새로운 연구 아이디어를 생성해 실행 가능한 코드로 구현하며 필요 시 외부 자료나 지식을 참고해 성능을 향상

- AI 시스템은 코드 구현 다음 단계에서 트리 탐색(Tree Search)* 전략을 사용해 코드 솔루션 후보 트리를 생성하고 더 나은 성능을 내는 갈래를 따라 추가 탐색이 필요한 후보를 결정

* 여러 선택지를 나무(Tree) 형태로 분기해 가며 가장 유망한 경로를 선택하는 방식

- 코드 실행 결과가 좋지 않으면 LLM을 사용해 코드를 재작성해 품질 점수를 개선하고, 효율적 탐색을 통해 사람이 수개월 걸릴 탐색을 수일 또는 수 시간 만에 진행함으로써 고품질 코드 솔루션을 식별

〈과학적 문제 해결을 위한 코드 솔루션 도출 알고리즘〉



- 연구진이 이 시스템을 활용해 6개 분야* 평가를 통해 공중보건, 지리공간 분석, 신경과학 등의 과학 문제를 실험한 결과, 모든 분야에서 기존의 첨단 방법과 동등하거나 이를 능가하는 결과를 달성

* 유전체학, 공중보건/역학, 지리 공간 분석, 신경과학, 시계열 예측, 수치 해석

- 공중보건 분야에서는 미국의 코로나 입원자 수 예측에서 기존 모델(CovidHub-ensemble)보다 성능이 뛰어난 모델을 14개 생성했으며, 위성 원격탐사 영상을 활용한 지리 공간 분석에서도 기존 첨단 기법을 능가

출처 | Google Research, Accelerating scientific discovery with AI-powered empirical software, 2025.09.09.

인력·교육

마이크로소프트, 교육 분야의 AI 활용 현황 보고서 발간

KEY Contents

- 마이크로소프트가 학생과 교육자, 교육기관 리더를 대상으로 AI 도입 현황을 조사한 결과, 교육기관의 86%가 생성 AI를 사용하는 등 교육기관의 AI 통합이 빠르게 증가하는 추세
- 학생들은 과제 구상과 정보 요약, 교육자들은 수업 계획과 자료 구상, 수업 자료나 과제 제작, 교육 관리자는 과제 피드백과 개선 등에서 창의적 사고나 협력을 위한 파트너로서 AI를 채택

○ MS 설문조사 결과, 학생과 교육자, 교육기관 관리자 전반에서 AI 도입이 가속화

- 마이크로소프트(이하 MS)의 ‘2025년 교육 분야 AI(2025 AI in Education)’ 보고서에 따르면 교육 분야의 AI 활용이 급증하는 추세
 - 이번 보고서는 미국, 브라질, 사우디아라비아, 영국, 일본, 호주의 교육자와 교육기관 최고 관리자를 대상으로 실시한 설문조사 결과를 종합해 분석*
 - * 미국에서는 설문조사 대상자에 16-18세 학생과 대학생, IT 책임자도 포함
 - 현재 교육기관의 86%가 생성 AI를 사용하고 있으며, 이는 전체 산업 분야 중 가장 높은 비율로, 미국에서는 학교에서 AI를 활용하는 학생이 2024년 대비 26%p 늘었고, 교육자의 AI 활용은 전년 대비 21%p 증가
- 교육기관들은 K-12(초중고) 교실의 개인화 학습부터 대학의 AI 기반 관리 도구 등 효율성과 학생 참여도, 성과 개선을 위해 AI를 빠르게 통합하고 있으나, AI 교육은 AI 도입 속도를 따라가지 못하는 상태
 - 미국 학생과 전 세계 교육자 중 AI에 대해 많이 알고 있다는 응답은 절반 이하를 기록했으며, 전 세계 교육자의 45%와 미국 학생의 52%는 아무런 AI 교육을 받은 적이 없다고 응답
 - 그러나 전 세계 교육기관 관리자의 76%는 AI 사용자(학생·교육자) 중 절반 이상이 AI 교육을 받았다고 응답해, 학생·교육자와 교육기관 리더 간 인식 차이를 반영
- 보고서에 따르면 학생과 교육자, 교육 관리자들은 창의적 사고와 협력을 지원하는 파트너로 AI를 활용
 - 학생들은 주로 과제 구상(37%), 정보 요약(33%), 신속한 정보 획득(33%), 피드백 수렴(32%)에 AI를 사용
 - 교육자들은 수업 계획이나 자료 구상(31%), 수업 자료나 과제 제작 또는 업데이트(29%), 학생을 위한 복잡한 주제 단순화(24%), 실시간 성과 데이터와 예측 분석(24%) 등에 주로 AI를 활용한다고 응답
 - 교육기관 관리자들은 주로 과제 피드백과 개선 안내(36%), 운영 간소화(35%), 접근성 도구 제공(33%), 학생 성장 기회 파악(33%), 학생 지원 및 소통 능력 향상(30%)에 AI를 활용한다고 응답
- MS는 교육 분야에서 AI 활용이 증가하고 있지만, AI로 기존 학습 방식을 대체하는 것이 아니라 보완하는 방향으로 AI를 사용할 때 더 효과적이라고 강조
 - MS 리서치와 캠브리지대 출판평가원(Cambridge University Press & Assessment)의 공동 연구* 결과, AI 도구와 노트 필기 및 기타 방식을 결합한 학생들은 AI에만 의존한 학생보다 더 뛰어난 학습 효과를 달성

* AI and Traditional Learning: Complementary Strategies for Deeper Learning

스탠포드大 연구 결과, AI 노출도 높은 직종의 신입 근로자 고용 감소

KEY Contents

- 스탠포드大의 분석에 따르면 미국 노동시장에서 AI 확산이 경력 초기 근로자들에게 불균형한 영향을 미치면서 AI 노출도가 높은 직업에서 경력 초기 근로자의 고용이 뚜렷하게 감소
- 이러한 결과는 경기 상황 요인을 배제하거나 표본 구성을 달리했을 때도 일관되게 나타나, 생성 AI 확산이 경력 초기 근로자들에게 부정적 고용효과를 유발하고 있음을 입증

○ SW 개발자나 고객 서비스 담당자 등 AI 노출도가 높은 직업에서 젊은 근로자 고용 감소

- 미국 스탠포드大가 미국 최대 급여 처리 서비스 기업 ADP의 급여 기록(2021년 1월~2025년 7월)을 활용해 생성 AI 도입이 노동시장에 미치는 영향을 조사한 보고서를 발표
 - 보고서는 AI 확산이 미국 노동시장에서 경력 초기의 근로자들에게 불균형한 영향을 미치고 있다는 가설을 뒷받침하는 광범위한 증거를 제시
- 분석 결과, 생성 AI 도입이 늘어난 이후 AI 노출도가 가장 높은 직업(예: SW 개발자, 고객 서비스 담당자)에 종사하는 경력 초기 근로자의 고용이 뚜렷하게 감소
 - 2025년 7월까지 22~25세의 경력 초기 SW 개발자 고용은 2022년 말의 최고점 대비 약 20% 감소했으며, 반면 AI 노출도가 낮은 보건의료 보조 직업에서는 젊은 근로자 고용이 숙련 근로자보다 빠르게 성장
- 미국 노동시장에서 전반적인 고용의 탄탄한 성장세에도 불구하고 젊은 근로자의 고용은 2022년 말부터 정체되었으며, AI 노출도가 높은 직업군에서는 마이너스 성장을 기록
 - AI 노출도가 가장 높은 직업(예: SW 개발자, 고객 서비스 담당자)에서 22~25세 근로자 고용은 2022년 말부터 2025년 7월 사이 6% 감소했으며, 이와 대조적으로 숙련 근로자의 고용은 6~9% 증가
- 고용에 영향을 주는 경기 상황 요인과 기업 내 고용구조 차이를 배제하고 순수하게 AI 노출도가 고용 변화에 미치는 영향을 조사했을 때도, AI 노출도가 높은 직업에서 젊은 근로자의 고용 감소 효과는 확인
 - 22~25세 근로자 중 AI 노출도가 가장 높은 직업군(예: SW 개발자, 고객 서비스 담당자)은 노출도가 가장 낮은 직업군(예: 보건의료 보조) 대비 고용이 약 12% 감소해 통계적으로 유의미한 차이를 기록
- 이러한 결과는 기술 직종이나 원격근무 가능 직종 제외, 시간제와 임시직 근로자 포함 등 대안적 표본 구성에서도 일관되게 나타나, 생성 AI 확산이 경력 초기 근로자에게 불균형적으로 부정적 고용효과를 유발하고 있음을 입증
 - AI가 노동시장에 미치는 영향은 임금보다 고용에서 더 뚜렷하게 나타나며, 고용 변화와 대조적으로 연령이나 직종별 AI 노출도에 따른 연봉 차이는 미미한 수준으로, 이는 단기적으로 임금 경직성이 존재하며, 초기에는 AI가 임금보다 고용에 더 큰 영향을 미칠 수 있음을 시사
 - 한편, AI로 자동화되는 직업에서는 경력 초기 근로자 고용이 감소했으나 AI 사용이 주로 인간 노동을 보완하는 직업에서는 고용이 유지되는 것으로 확인

구글 클라우드, 2025년 AI 투자수익률 보고서 발표

KEY Contents

- 구글 클라우드가 전 세계 매출 1,000만 달러 이상 기업 임원을 대상으로 진행한 설문조사 결과, 생성 AI 도입 기업 중 52%가 AI 에이전트를 적극 활용 중이라고 응답
- AI 에이전트의 주된 활용 분야는 고객 서비스와 경험, 마케팅, 보안 운영과 사이버보안 등으로 에이전틱 AI 조기 도입 기업의 88%는 하나 이상의 생성 AI 활용 사례에서 ROI를 달성했다고 보고

● 에이전틱 AI 조기 도입 기업의 88%가 생성 AI 활용 사례에서 ROI 달성

- 구글 클라우드(Google Cloud)가 전 세계 글로벌 대기업 대상의 설문조사 결과를 바탕으로 2025년 9월 4일 '2025년 AI 투자수익률(ROI of AI)' 보고서를 발간
 - 전 세계 매출 1,000만 달러 이상 기업의 고위 임원 3,466명을 대상으로 진행된 이번 조사 결과, 전 세계 대기업에서 AI 에이전트 도입이 빠르게 확산하는 추세
 - 생성 AI를 사용하는 기업의 52%가 AI 에이전트도 적극 활용하고 있다고 답했으며, 39%의 응답자는 기업이 10개 이상의 AI 에이전트를 도입했다고 응답
 - 전체 산업에서 AI 에이전트 활용 분야는 고객 서비스와 경험(49%), 마케팅(46%), 보안 운영과 사이버보안(46%), 기술지원(45%), 제품 혁신과 디자인(43%), 생산성과 연구(43%) 순을 기록
 - 지역별로 AI 에이전트 활용의 우선순위는 다르게 나타났으며, 유럽 지역 응답자들은 AI 기반 기술지원을 우선하는 반면, 아시아태평양 응답자들은 고객 서비스를, 라틴아메리카 응답자들은 마케팅을 우선시
 - 에이전틱 AI 조기 도입 기업(전체 응답 기업의 13%)은 AI 예산의 50% 이상을 AI 에이전트에 투입하고 있으며, 88%가 하나 이상의 생성 AI 활용 사례에서 ROI를 달성했다고 답해, 전체 기업 평균인 74%를 상회
 - 생성 AI 활용 사례에서 ROI를 달성했다는 전체 응답 중 생성 AI 활용 분야는 생산성(39%), 고객 경험(37%), 판매·마케팅(33%) 순을 기록
- 기업들의 전체 AI 투자는 증가 추세로, 77%가 기술 비용 감소로 생성 AI 지출이 늘어났다고 응답
 - 58%의 응답자는 생성 AI 투자를 뒷받침하기 위해 다른 예산을 줄이지 않고 신규 예산을 할당했으며, 48%는 생성 AI 투자를 위해 AI와 무관한 예산을 재분배했다고 응답
 - 조사 기업들은 생성 AI를 통해 새로운 가치가 창출되는 5대 영역으로 생산성(70%), 고객 경험(63%), 사업 성장(56%), 마케팅(55%), 보안(49%) 순이라고 응답
- LLM 공급업체 선택 시 기업들이 가장 중요하게 고려하는 요소는 개인정보 보호와 보안(37%)으로 나타났으며, 기존 시스템과의 통합(28%), 비용(27%)이 다음 순서를 차지
 - 이는 기업들이 특정한 기능이나 맞춤 설정과 같은 고급 기능이나 차별화 포인트를 고려하기에 앞서 핵심적인 기업 요구사항을 중시하고 있음을 시사

출처 | Google Cloud, ROI of AI 2025, 2025.09.04.

Google Cloud, Google Cloud Study Reveals 52% of Executives Say Their Organizations Have Deployed AI Agents, Unlocking a New Wave of Business Value, 2025.09.04.

주요행사일정

월	기간	행사명	장소	홈페이지
1월	7~10일	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	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
	29일	LlamaCon 2025	미국, 멘로파크	www.llama.com/events/llamacon
	29~30일	Generative AI Summit	미국, 산타클라라	world.aiacceleratorinstitute.com/location/siliconvalley
5월	5~7일	IEEE CAI 2025	미국, 산타클라라	cai.ieee.org/2025
	6~8일	Microsoft 365 Conference	미국, 라스베이거스	m365conf.com
	19~22일	Microsoft Build 2025	미국, 시애틀	build.microsoft.com/en-US/home
	20~21일	Google I/O 2025	미국, 마운틴뷰	io.google/2025
	20~23일	COMPUTEX TAIPEI	대만, 타이베이	www.computextaipei.com.tw/en
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	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	IEEE 7th International Conference on AI, CS and IP	중국, 항저우	www.aicsconf.cn
	27~1일	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
	16~22일	(IJCAI 2025) International Joint Conference on Artificial Intelligence	캐나다, 몬트리올	2025.ijcai.org
9월	3~5일	2025 산업AI EXPO	서울, 강서	industrialaiexpo.or.kr
	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	30~2일	AI Festa(AI Week 2025)	서울, 강남	aifesta.kr
	8~9일	World Summit AI	네덜란드, 암스테르담	worldsummit.ai
	13~17일	GITEX 2025	두바이	gitex.com
11월	4~5일	Open Source Summit	서울, 강남	events.linuxfoundation.org/open-source-summit-korea
	12~14일	AI·ICT 기술·산업전망 컨퍼런스	서울, 중구	www.iitp.kr
	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	1~5일	AWS re:Invent 2025	미국, 라스베이거스	reinvent.awsevents.com
	2일	SPRi 산업전망컨퍼런스	서울, 강남	www.spri.kr
	2~7일	NeurIPS 2025	미국, 샌디에이고	neurips.cc
	3~5일	(소프트웨이브 2025) 10회 대한민국 소프트웨어 대전	서울, 강남	www.k-software.com



홈페이지 : <https://spri.kr>

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