2026년 5월호

SPRi AI Brief

인공지능 산업의 최신 동향

SPRi 소프트웨어정책연구소
Software Policy & Research Institute

SPRi AI Brief

I 2026년 5월호 I

CONTENTS

2026년 4월 월간 AI 모델 동향

주요 모델 출시 타임라인

4/2 Gemma 4 4/7 Claude Mythos Preview(제한 공개), GLM-5.1
4/8 Muse Spark 4/9 EXAONE 4.5 4/16 Claude Opus 4.7 4/20 Kimi K2.6
4/21 ChatGPT Images 2.0 4/23 GPT-5.5 4/24 DeepSeek-V4-Pro

LLM간 성능 비교(Artificial Analysis Intelligence Index 4.0)

모델	출시일	Index	Price ($/1m Token)		Speed	Context Window	회사
			Input	Output			
① GPT-5.5(xhigh)	26.4.23.	60	5	30	72.0	922k	OpenAI
② GPT 5.5(high)	26.4.23.	59	5	30	75.9	922k	OpenAI
③ Claude Opus 4.7 (max)	26.4.16.	57	5	25	47.7	1m	Anthropic
Gemini 3.1 Pro Preview	26.2.19.	57	2	12	118.7	1m	Google
GPT-5.4(xhigh)	26.3.5.	57	2.5	15	80.6	1m	OpenAI

* 출처 : Artificial Analysis 평가체계 : Agents 25%, Coding 25%, General 25% Scientific 25% (10개 벤치마크 종합)

이미지 생성 모델

모델	출시일	ELO	회사
① GPT Image 2 (high)	26.4.21	1332	OpenAI
② GPT Image 1.5 (high)	25.12.16	1271	Google
③ Nano Banana 2(Gemini 3.1 Flash Image Preview)	25.12.16	1262	Google

영상 생성 모델

모델	출시일	ELO	회사
① HappyHorse-1.0	26.4.7	1366	Alibaba
② Dreamina Seedance 2.0 720p	26.2.10	1270	ByteDance
③ Kling 3.0 1080p (Pro)	26.2.5	1246	Kling

한국 AI 모델 현황

■ LG AI연구원, 멀티모달 AI 모델 'EXAONE 4.5' 공개 (4.9.)

• 테스트와 이미지를 동시에 이해하고 추론할 수 있는 비전-언어모델 오픈웨이트 공개

주요 포인트 :
GPT-5.5, Claude Opus 4.7 등 주요 플래그십 모델의 업데이트로 코딩 성능과 에이전트 역량 한층 강화
뛰어난 가성비를 갖춘 DeepSeek V4, 성능은 프런티어급 모델에 미달

SPRi AI Brief

Ⅰ2026년 5월호Ⅰ

정책·법제

유럽중앙은행, AI가 유로존 경제에 미치는 영향 분석

KEY Contents

- 유럽중앙은행의 AI 관련 연구에 따르면 EU 지역의 AI 활용 비율은 최근 크게 높아졌으나 AI를 심도 있게 활용하는 사례는 소수의 핵심 기업에 집중된 상황
- AI가 EU 지역의 생산성에 미치는 영향은 향후 10년간 AI 확산 속도에 따라 크게 달라질 수 있으며, AI 발전으로 디지털 투자가 급증하고 있으나 미국과의 격차는 확대 추세

EU 지역의 도입 현황은 아직 초기 단계로 생산성 향상을 위해 확산 속도 개선 필요

■ 유럽중앙은행(ECB)의 필립 R. 레인(Philip R. Lane) 집행이사가 2026년 3월 23일 열린 제3차 "기후-거시-금융 연계 국제회의(3CMFI)"에서 AI와 유로존 경제를 주제로 기조연설을 발표

- 기조연설은 유럽중앙은행의 최근 AI 관련 연구 결과를 소개하고, EU 지역의 AI 활용 현황 및 AI 도입이 EU 지역에 미치는 영향을 생산성, 디지털 투자와 고용 등을 중심으로 분석

■ (AI 활용 현황) EU 지역에서 AI 활용은 빠르게 늘어나는 추세로, 유럽중앙은행의 소비자 조사 결과 근로자의 AI 활용 비율은 2024년 26%에서 2025년 40%로 급증

- 특히 대졸 근로자의 AI 활용률(47%)은 대학을 졸업하지 않은 근로자(28%)보다 높고, 18~34세(52%) 근로자는 55~74세(29%) 근로자 대비 AI 활용률이 두 배 가까이 높아 학력·나이별 격차를 확인
- 기업 대상 설문조사에서는 5,000개 기업 중 3분의 2가 직원들이 AI를 사용한다고 답했으나 AI를 심도 있게 활용하는 기업은 7%에 불과해 AI 사용이 소수의 핵심 기업에 집중되어 있음을 시사
- 기업들은 주로 기업 프로세스 개선에 AI를 활용하고 있으며, 인건비 절감과 연구 개발 및 혁신 지원에 AI를 활용하는 비율은 상대적으로 낮은 수준으로, 기술 부족과 윤리 문제 등이 도입 장벽으로 작용

■ (생산성에 미치는 영향) 유럽중앙은행의 분석 결과, AI가 총요소생산성에 미치는 영향은 확산 속도에 따라 향후 10년간 0.2~0.4%포인트 범위에서 크게 달라질 수 있어 AI 확산을 위한 정책이 필요

■ (디지털 투자 동향) AI의 발전은 디지털 투자를 촉진하고 있으며, 2025년 EU 지역의 디지털 투자는 2014년 대비 60% 이상 대폭 증가했으나 미국과의 격차는 확대

- 같은 기간 미국의 디지털 투자는 EU 대비 두 배 이상 증가했으며 특히 데이터센터 투자 확대에 힘입어 2025년 들어 투자가 가속화되면서 EU와 미국의 디지털 투자 격차가 뚜렷하게 벌어지는 추세
- 벤처캐피털 투자 및 차세대 EU 기금에 힘입어 디지털 투자는 대폭 증가할 전망으로, 기업 자금조달 관련 설문조사 결과 중소기업과 대기업 모두 2026년까지 총 투자액의 평균 9%를 AI에 할당할 계획

■ (고용에 미치는 영향) 아직 AI 도입 초기 단계로 AI가 고용에 미치는 전반적 영향을 평가하기는 어려운 상태로, 소비자 기대 조사에서는 인구통계학적 집단별로 뚜렷한 차이를 확인

- 전체 근로자 중 43%는 신기술이 생산성이나 취업 기회에 긍정적 영향을 미칠 것으로 여겼으나 34%는 아무런 영향이 없을 것으로, 23%는 일자리 감소나 고용 악화 등 부정적 영향을 예상
- 대학 교육을 받은 근로자와 젊은 근로자일수록 AI가 고용에 미치는 영향에 대하여 긍정적 견해를 피력

출처 | European Central Bank, AI and the euro area economy, 2026.3.23.

미국 캘리포니아주, AI 안전과 책임 있는 사용을 강화하는 행정명령 발표

KEY Contents

- 미국 캘리포니아주의 개빈 뉴섬 주지사는 행정명령을 통해 조달 절차에서 AI 기업을 평가하기 위한 새로운 인증 제도 도입 및 연방정부와 독립된 조달 기반의 마련을 지시
- 행정명령은 또한 캘리포니아주의 공공 서비스 전반에서 생성형 AI의 책임 있는 활용을 확대하고 주정부에서 AI 생성 이미지나 동영상 제작 시 워터마크를 삽입하는 지침을 마련할 것을 요구

AI 조달 기준 강화 및 연방정부와 독립된 조달 기반 마련

■ 미국 캘리포니아주의 개빈 뉴섬(Gavin Newsom) 주지사가 2026년 3월 30일 AI 안전과 책임 있는 사용을 강화하기 위한 행정명령 'N-5-26'을 발표

- 개빈 뉴섬 주지사는 AI 규제 완화를 통해 오용을 조장하는 트럼프 행정부와 달리 캘리포니아주는 조달 절차를 강화하고 AI 기업들에 엄격한 기준을 적용해 기술 오용을 방지할 것이라고 강조
- 행정명령은 총무부(DGS)와 기술부(CDT)에 캘리포니아 주정부의 조달 과정에서 AI 기업을 평가하는 새로운 인증 제도를 도입하기 위한 권고안을 120일 이내에 제출할 것을 지시
- 이에 따라 AI 기업들은 기술 오용을 막고 공공 안전을 보장하기 위한 정책과 안전장치를 설명 필요
- 해당 정책과 안전장치는 아동 성착취물 및 동의 없는 성적 이미지 등 불법 콘텐츠의 이용과 배포, 유해한 편향, 시민의 권리와 자유(예: 표현의 자유, 투표권, 인간의 자율성) 침해와 같은 위험에 대응 필요
- 주정부 운영국(GovOps)은 시민의 프라이버시나 자유를 불법적으로 침해한 것으로 사법 판결을 받은 기업과 계약하지 않도록 하는 계약자 책임 조항 개정안을 120일 이내에 주지사에게 제출 필요

■ 행정명령에 따르면 기술부 최고정보보안책임자(CISO)는 연방정부가 공급망 위험 기업으로 지정한 기업을 검토하고 해당 지정이 부당하다고 판단되면 조달을 허용하는 지침을 발표할 권한 보유

- 또한 CISO가 연방정부의 여타 조달 관련 변경 사항을 검토해 조달을 부당하게 제한하는지를 평가하고, 적절한 대응 조치를 권고할 수 있다고 규정해 연방정부와 독립된 조달 기반을 마련

캘리포니아주의 공공 서비스 제공 시 생성형 AI의 책임 있는 활용을 요구

■ 행정명령은 또한 주정부 운영국 산하 유관 부처에 캘리포니아주의 공공 서비스 전반에 걸쳐 생성형 AI의 책임 있는 활용을 확대하기 위한 이행 과제를 부과

- 공무원들이 적절한 개인정보 보호와 사이버보안 조치를 준수하고 일반적 활용 사례에 대해 검증을 거친 생성형 AI 도구를 사용할 수 있도록 지원하고, 책임 있는 AI 조달과 도입 관련 모범 사례를 공유
- 디지털 전략을 개정해 생성형 AI를 활용해 정부의 투명성과 책임성을 강화하고 성과를 개선하며, 편리하고 사용자 친화적 접근성을 제공하는 생성형 AI 기반 정부 서비스 앱이나 웹사이트를 개발

■ 기술부는 행정명령에 따라 주정부 운영국과 협력해 120일 이내에 정부 기관들이 업계 모범 관행에 따라 AI 생성 이미지나 비디오에 워터마크를 삽입하기 위한 지침을 마련 필요

출처 | Governor Gavin Newsom, As Trump rolls back protections, Governor Newsom signs first-of-its-kind executive order to strengthen AI protections and responsible use, 2026.3.30.

미국 노동부, 미국 전역의 등록 견습 프로그램에 AI 기술 통합 추진

KEY Contents

- 미국 노동부는 AI 관련 교육 확대와 견습 프로그램의 현대화, 신흥 분야의 인력 양성을 위해 미국 전역의 등록 견습 프로그램에 AI 기술을 통합하기 위한 계약 기회를 발표
- 선정된 계약 업체는 전국 중개 기관으로서 고용주, 산업단체 등과 협력해 AI 관련 교육과정 개발, AI 견습 프로그램 도입과 확산, 인력 개발 전략 조율 등의 업무를 담당

전국의 등록 견습 프로그램에 AI를 통합해 신흥 분야의 인력 양성 확대 추진

■ 미국 노동부 고용훈련청(Employment and Training Administration)이 2026년 4월 1일 등록 견습 프로그램*에 AI 기술 통합을 촉진하기 위한 전국 규모의 계약 기회를 발표

* Registered Apprenticeship Program: 이론 교육과 현장 실습을 병행하며 수료 시 국가 인증 자격을 취득할 수 있는 고용주 주도의 직업훈련 프로그램

- 미국 노동부는 이번 계획으로 AI 관련 교육을 확대하고 견습 프로그램을 현대화하며, 신흥 및 핵심 산업 전반에 걸쳐 국가 인재 양성 체계를 강화하겠다는 비전을 제시
- 이번 계획은 기존 견습 프로그램 내 AI 훈련·도구·교육과정 내재화, AI 기술을 직접 구현·관리·적용하는 직종의 견습 프로그램 포함, 데이터센터·통신·첨단 제조 분야 인력 파이프라인 강화를 중점 추진
- 로리 차베스-드레머(Lori Chavez-DeRemer) 노동부 장관은 AI가 모든 산업을 변화시키고 있다며, 미국 노동자들이 AI 경제의 참여자를 넘어 주도자가 될 수 있도록 보장하는 계기가 될 것이라고 강조
- 키스 손더링(Keith Sonderling) 노동부 차관은 향후 10년간 신흥 산업에서 수백만 개의 일자리 창출이 예상되는 가운데, 견습 프로그램이 인력 수요 충족을 위한 핵심 전략으로 자리매김할 것으로 기대

노동부 계약 업체, AI 관련 견습 프로그램 도입과 확산을 위한 다양한 업무 수행

■ 노동부가 선정한 계약 업체는 국가 차원의 중개 기관 역할을 담당하며, 고용주와 산업 협회, 인력 개발 파트너와 협력해 견습 프로그램 관련 다양한 업무를 수행 필요

- 계약 업체는 AI 관련 교육과정·훈련 모듈·견습 기준 개발, 견습 프로그램 도입 및 확산 지원, 기술 지원과 전문 지식 제공, 인력 개발 전략 조율, 혁신 촉진 및 AI 활용 훈련 모델 확산 등의 업무를 담당

■ 노동부는 수요가 높은 AI 관련 직무에서 새로운 견습 프로그램을 구축하는 동시에, 전통적인 기술직과 인프라 직종에 AI 역량을 통합하는 것을 이번 계획의 핵심 목표로 제시

- 검증된 견습 프로그램 모델과 첨단 AI 기술을 결합함으로써 경제적 기회에 대한 접근성을 확대하고 기업 성장에 필요한 숙련된 인력을 확보할 수 있도록 지원한다는 방침

■ 이를 위해 AI 및 인력 개발 및 산업 파트너십 영역에서 전문성을 보유한 기관을 선정할 예정

- 노동부는 단일 계약에 1년 간의 기본 계약 및 4년 연장 옵션을 부여함으로써 전국 규모의 AI 인력 양성을 위한 장기적인 추진 의지를 표명

출처 | U.S. Department of Labor, US Department of Labor launches landmark initiative to integrate artificial intelligence skills into Registered Apprenticeships nationwide , 2026.4.1.

미국 연방기관들, 앤트로픽 사용금지 우회해 'Claude Mythos' 사용 추진

KEY Contents

- 트럼프 행정부가 앤트로픽을 공급망 위험으로 지정하고 정부 기관의 사용을 금지했으나, 주요 연방기관은 앤트로픽의 고성능 비공개 모델 'Claude Mythos' 사용을 추진
- 앤트로픽의 공급망 위험 지정 취소 신청에 대하여 두 개 법원이 엇갈린 판결을 내놓은 가운데, 정부가 사이버 방어 역량 강화를 위해 앤트로픽을 적극 활용해야 한다는 의견이 대두

주요 연방기관과 의회, 사이버보안 테스트를 위해 'Claude Mythos' 사용 추진

■ 미국 연방기관과 공무원들이 트럼프 대통령의 앤트로픽(Anthropic) 서비스 사용 중단 지시에도 불구하고 앤트로픽의 비공개 최고 성능 모델 'Claude Mythos' 사용을 추진

- 앤트로픽은 강력한 사이버 해킹 역량을 이유로 Claude Mythos를 일부 조직에만 선별 배포했으며, 정부 관계자에 따르면 두 곳 이상의 대형 연방기관이 앤트로픽에 연락해 사용을 추진
- AI 정책 담당 의회 보좌관들 역시 세 곳 이상의 의회 위원회 직원들이 앤트로픽에 Claude Mythos의 사이버 취약점 탐지 기능을 자세히 파악하기 위한 브리핑을 요청했다고 언급
- 블룸버그(Bloomberg) 역시 2026년 4월 14일 재무부의 IT 담당자들이 Claude Mythos를 활용해 기관 네트워크의 미확인 취약점 수정을 추진하고 있다고 보도

■ 전현직 사이버보안 관계자에 따르면 미국과 해외 AI 모델의 잠재적 위험과 기회를 평가하는 업무를 담당하는 AI표준혁신센터(CAISI) 역시 Claude Mythos을 적극적으로 시험 중

- 미국 국립표준기술연구소(NIST) 산하 CAISI의 사이버보안 전문가들은 앤트로픽의 발표 이전부터 Mythos에 레드팀 테스트를 진행하며 성능 및 국가안보에 미칠 수 있는 잠재적 위험을 평가

주요 연방기관의 엇갈린 행보로 트럼프 행정부의 정책 노선에 균열 발생

■ 미국 트럼프 대통령과 국방부가 2026년 2월 앤트로픽을 공급망 위험으로 지정한 이후 앤트로픽이 2개 법원에 이를 취소해달라는 소송을 제기한 가운데, 두 법원은 엇갈린 판단을 제시

- 캘리포니아 연방법원이 3월 말 클로드 사용금지 명령을 중단하는 예비 금지 명령을 내렸으나, 워싱턴 DC 연방항소법원은 4월 8일 앤트로픽의 신청을 기각하고 공급망 위험 지정을 유지
- 워싱턴DC 연방항소법원 재판부는 민간 기업의 재정과 전쟁 상황 속 국방부의 AI 통제 가운데 후자가 더 중요하다고 판결 이유를 밝혔으나, 이는 최종 판결이 아닌 잠정적 판단으로 본안 심리는 5월로 예정

■ 국방부의 앤트로픽 배제 방침에도 불구하고 주요 연방기관들이 사이버보안을 이유로 이를 우회하는 행보를 보이면서 균열이 가시화되고 있으며, 의회와 전직 고위 당국자들은 행정부에 우려를 표명

- 의회 보좌관들은 백악관의 앤트로픽 사용금지로 인해 연방기관들이 최첨단 기술을 온전히 활용하지 못하고 있다며, 정부가 네트워크 보안 강화를 위해 Mythos를 적극 도입해야 한다고 강조
- 전직 국가안보국(NSA) 법무실장 글렌 거스텔(Glen Gerstell)은 "국방부와 앤트로픽 간 긴장 관계가 사이버보안에 결정적으로 중요한 사안을 가로막지 않길 바란다"고 공개 발언

출처 | Politico, Federal agencies skirt Trump's Anthropic ban to test its advanced AI model, 2026.4.14.

SPRi AI Brief
Ⅰ2026년 5월호Ⅰ

기업·산업

커서, 신규 코딩 모델에 중국 오픈소스 모델 Kimi 2.5 활용 논란

KEY Contents

- 커서가 출시한 신규 코딩 모델 'Composer 2'가 중국 문샷 AI의 오픈소스 모델 'Kimi 2.5'를 기반으로 개발된 사실이 뒤늦게 밝혀져 논란이 야기
- 커서는 오픈소스 기반 모델의 활용을 공식 인정하고 초기 발표에서 이를 누락한 점에 대하여 사과했으며, 방대한 추가 학습을 통해 원본 모델과 매우 큰 차이를 보인다고 해명

커서, 장기 코딩 작업에 최적화된 에이전트형 코딩 모델 'Composer 2' 공개

■ 커서(Cursor)*는 2026년 3월 19일 프런티어급 코딩 성능을 갖춘 AI 모델 'Composer 2'를 출시

* AI 코딩 솔루션을 개발하는 스타트업으로 2026년 2월 기준 연간 반복 매출이 20억 달러 이상을 기록

- Composer 2는 코딩 벤치마크인 Terminal-Bench 2.0에서 61.7점, SWE-bench Multilingual에서 73.7점으로 Composer 1.5(47.9점/65.9점) 대비 성능이 뚜렷하게 개선
- 커서는 주요 벤치마크에서 이전 모델 대비 큰 폭의 향상을 기록한 Composer 2의 성능 향상이 자사 최초의 '지속 사전학습(Continued Pretraining)*'을 통해 이루어진 것이라고 강조

* 사전 훈련된 언어모델을 특정 도메인에 맞게 새로운 데이터로 추가 학습시키는 기법

- Composer 2의 가격은 입력 토큰 100만 개당 0.50달러, 출력 토큰 100만 개당 2.50달러이며, 동일한 지능 수준을 제공하는 더 빠른 속도의 버전은 각각 1.50달러, 7.50달러로 책정

Composer 2 출시 후 Kimi 2.5 활용 의혹 제기 및 성능 해명

■ 모델 출시 이후 사용자가 Composer 2 코드에서 문샷 AI(Moonshot AI)가 개발한 모델 ID를 찾아내고 해당 모델이 Kimi 2.5에 강화학습만 추가했다고 지적하자, 커서는 이를 곧바로 시인

- 커서의 공동 창립자 아만 생어(Aman Sanger)는 초기 블로그 발표에서 기반 모델로 사용된 Kimi 2.5를 명시하지 않은 것이 실수였음을 인정하고 다음 모델부터는 이를 바로잡겠다고 약속
- 한편, 커서의 리 로빈슨(Lee Robinson) 부사장은 Composer 2가 오픈소스 모델을 활용했으나 최종 모델에 투입된 전체 컴퓨팅 연산량의 약 4분의 1만이 기반 모델에서 비롯됐고 나머지는 자사의 독자적 학습을 통해 이루어졌다며 벤치마크 성능이 원본 모델과 매우 큰 차이를 보인다고 해명
- 이와 관련, Kimi 계정은 커서가 상업적 파트너십을 통해 자사 모델을 합법적으로 사용했다며, Kimi 2.5 모델이 Cursor의 최신 코딩 모델에 사용된 사실을 오픈소스 생태계의 성과로 환영

■ IT 미디어 테크크런치(TechCrunch)는 커서가 Composer 2를 출시하며 기반 모델을 숨긴 이유로 미·중 간 치열한 AI 경쟁 구도 속에서 중국 모델을 활용한 것이 부담으로 작용했을 가능성을 지목

- 이번 논란은 치열한 미중 AI 기술 패권 경쟁과 실리콘밸리의 견제 속에서도, 중국의 AI 기초 모델 기술력이 미국 선두권 서비스의 엔진으로 쓰일 만큼 강력해졌음을 시사

출처 | Cursor, Introducing Composer 2, 2026.3.19.
TechCrunch, Cursor admits its new coding model was built on top of Moonshot AI's Kimi, 2026.3.22.

업스테이지, 에이전트 성능 강화한 'Solar Pro 3' 업데이트 발표

KEY Contents

- 업스테이지가 에이전트 성능 벤치마크에서 전작인 Solar Pro 2 대비 약 2배의 성능 향상을 달성하고 추론과 사용자 선호도 등 전 영역에서 개선된 Solar Pro 3 업데이트를 발표
- 업스테이지는 독자적인 강화학습 프레임워크를 통해 수학, 코딩, 에이전트 등 다양한 영역에서 Solar 3 Pro의 추론 능력을 강화했으며, 비용 증가 없이 성능을 개선하도록 설계

Solar Pro 3, 전작인 Solar Pro 2 대비 에이전트 성능 2배 개선

■ 업스테이지(Upstage)가 2026년 3월 24일 자체 개발한 대규모 언어모델 Solar Pro 3의 최신 업데이트 버전을 공개하고 전작 대비 에이전트·추론·한국어 전 영역에서의 성능 향상을 강조

- 총 102B 파라미터의 전문가 혼합(MoE) 아키텍처를 기반으로 추론 시 토큰당 12B 파라미터만 활성화하는 Solar Pro 3는 전작인 Solar Pro 2와 동일한 환경과 비용을 유지하면서 성능을 개선
- Tau2-all(에이전트 종합) 벤치마크 평가에서 Solar Pro 3는 72.3점을 기록해 Solar Pro 2(36.0점)를 대폭 능가했고, SWE Bench(코딩 에이전트) 평가에서는 28.6점을 기록(Solr Pro 2는 14.5점)

■ 업스테이지는 특정 벤치마크 최적화가 아닌, 개별 도구 호출은 성공하나 전체 워크플로우를 끝내지 못하는 실제 운영 환경의 한계를 해결하는 방향으로 개발해 모델 성능이 개선되었다고 설명

- Solar Pro 3는 업스테이지의 자체 강화학습 프레임워크 'SnapPO'를 적용해 단계적 추론 능력을 강화
- SnapPO는 학습 과정의 각 단계를 독립적으로 실행·조합할 수 있도록 설계되어 수학·코드·에이전트 등 다양한 도메인의 추론 능력을 효율적으로 동시에 강화하는 것이 특징

■ 사용자 선호도 측면에서 Solar Pro 2와의 직접 비교 시 사용자들은 Solar Pro 3의 응답을 일관되게 선호하는 경향을 보였으며, 지시 사항 준수 역량에서도 의미 있는 향상을 확인

- 업스테이지는 사용자의 요청을 더 정확히 이해하고 끝까지 해결하는 능력의 개선이 단순한 점수 향상을 넘어 에이전트 작업에서 불완전한 지시를 처리하는 워크플로우 안정성과 직결된다고 강조

그림 Solar Pro 2와 Solar Pro 3 벤치마크 점수 비교

출처 | Upstage, Solar Pro 3: 에이전트 성능 2배 향상, 무엇이 달라졌나, 2026.3.24.

앤트로픽, '클로드 코드'의 소스 코드 유출 사고 발생

KEY Contents

- 앤트로픽에 따르면 소프트웨어 업데이트 패키징 과정에서 인적 오류로 인해 클로드 코드의 소스코드 약 50만 줄이 유출되었으며, 유출된 코드는 개발자 플랫폼인 깃허브에서 급속히 확산
- 유출된 코드에는 클로드 코드의 내부 아키텍처에 관한 정보가 포함되었으며, 이번 사건으로 최근에 또 다른 데이터 유출 사고를 겪은 앤트로픽의 보안 취약성에 대한 우려가 증대

클로드 코드 업데이트 과정에서 인적 오류로 내부 소스코드 유출

■ 앤트로픽(Anthropic)이 2026년 4월 1일 사람의 실수로 인해 자사의 AI 기반 코딩 어시스턴트인 '클로드 코드(Claude Code)'의 내부 소스 코드가 유출되었다는 성명을 발표

- 클로드 코드는 앤트로픽의 핵심 유료 상품으로, 2026년 들어 유료 구독자 수가 두 배 이상 증가
- 이번 사고는 앤트로픽이 소프트웨어 업데이트를 배포하는 과정에서 내부용으로 사용되는 파일이 실수로 포함되면서 발생한 것으로, 약 2,000개의 파일에서 50만 줄의 코드가 유출
- 유출된 코드는 빠르게 개발자 플랫폼인 깃허브(Github)에 복사되었으며, 유출된 코드의 링크를 공유한 X 게시물은 게시 후 몇 시간 만에 2,900만 조회수를 넘어선 것으로 확인
- 앤트로픽은 성명을 통해 고객의 민감한 데이터나 자격 증명이 유출되지는 않았다고 설명했으며, 유출 코드의 확산을 방지하기 위해 총 8,000회에 달하는 저작권 침해 콘텐츠에 대한 삭제 요청을 제출
- 그러나 개발자들은 AI를 활용해 코드를 다른 스크립트 언어로 재작성하는 등의 방법으로 앤트로픽의 삭제 요구를 회피하고, 유출된 소스코드가 담긴 깃허브 저장소 유지를 시도

■ 이번에 유출된 코드는 클로드 코드의 내부 아키텍처와 관련이 있으나, 앤트로픽의 기반 AI 모델인 클로드와 관련된 기밀 데이터를 포함하지는 않은 것으로 확인

- 유출된 코드에는 AI 모델을 코딩 에이전트로 작동시키는 도구와 지시 사항 등 상업적으로 민감한 정보가 포함되었으며, 경쟁사들이 클로드 코드의 내부 작동 방식을 파악할 수 있게 되었다는 우려도 제기
- 개발자들이 소스코드를 분석한 결과, 악의적 사용자에 의한 클로드 복제를 방지하는 기술과, 결과물이 AI에 의해 생성되었다는 흔적을 지우는 모드와 같은 클로드 코드의 다양한 기능이 공개
- 클로드 코드의 소스 코드는 이미 독립 개발자들의 '리버스 엔지니어링'으로 일부 공개된 바 있으며, 2025년 2월에도 이전 버전의 소스코드가 노출된 선례도 존재

잇따른 보안 사고로 앤트로픽의 내부 보안 취약성에 대한 우려 증대

■ 이번 사고는 앤트로픽에서 지난 몇 주 사이에 발생한 두 번째 데이터 유출 사고로, 앞서 발생한 사건에서는 수천 개의 내부 파일이 공개적으로 접근 가능한 시스템에 저장된 것으로 확인

- 당시 유출된 파일에는 차세대 모델인 Claude Mythos에 관한 블로그 게시물 초안이 포함되었으며, 일부 전문가는 해당 사고가 앤트로픽 내부의 보안 취약성을 시사한다는 우려를 제기
- AI 안전을 핵심 기업 가치로 표방하는 앤트로픽에 있어 보안 취약성 이슈는 심각한 타격이 될 수 있음

출처 | The Guardian, Claude's code: Anthropic leaks source code for AI software engineering tool, 2026.4.1.
Anthropic, Anthropic Issues Copyright Takedowns to Scrub Claude Code Leak, 2026.4.1.

앤트로픽, 차세대 AI 모델 'Claude Mythos' 기반 '프로젝트 글래스윙' 출범

KEY Contents

- 앤트로픽이 주요 기술기업들과 협력해 최신 범용 프런티어 모델 'Claude Mythos Preview'를 활용해 소프트웨어 보안을 개선하기 위한 '프로젝트 글래스윙'을 출범
- 앤트로픽은 소프트웨어 취약점 탐지와 공격 능력에서 인간 전문가 수준을 뛰어넘는 Claude Mythos Preview의 위험성을 고려해 대중에게 공개할 계획은 없다고 설명

Claude Mythos, 소프트웨어 취약점 발견과 악용에서 인간 전문가를 능가

■ 앤트로픽이 2026년 4월 7일 차세대 AI 모델 'Claude Mythos Preview'를 활용해 소프트웨어 보안을 강화하기 위한 '프로젝트 글래스윙(Project Glasswing)'을 발표

- Claude Mythos Preview는 앤트로픽에서 개발한 최신 범용 프런티어 모델로 코딩 및 에이전트 작업에서 현재까지 개발된 AI 모델 중 가장 뛰어난 성능을 발휘
- 앤트로픽에 따르면 소프트웨어 취약점 탐지와 공격 능력에서 인간 전문가의 수준을 넘어서는 수준으로, 이미 주요 운영 체제와 웹 브라우저 등에서 수천 건의 고위험 보안 취약점을 발견
- 사이버 보안 취약점을 재현하는 CyberGym 벤치마크 평가에서 83.1%를 기록해 Claude Opus 4.6(66.6%)을 크게 앞섰으며, 소프트웨어 코딩 작업에서도 최고점을 기록*

* SWE-Bench Pro 벤치마크 기준 Claude Mythos Preview:77.8%, Claude Opus 4.6: 53.4%

■ 앤트로픽은 AI 발전 속도를 고려할 때 이러한 능력이 광범위하게 보급되면 경제와 공공 안전, 국가 안보에 심각한 영향을 끼칠 수 있다고 보고, 협력사들과 방어 목적의 AI 활용을 추진

- 이번 프로젝트에는 AWS, 애플(Apple), 브로드컴(Broadcom), 시스코(Cisco), 크라우드스트라이크(CrowdStrike), 구글(Google), JP모건체이스(JPMorganChase), 마이크로소프트(Microsoft) 등이 참여
- 참여사는 Claude Mythos Preview를 사용해 전 세계 사이버 공격 표면의 상당 부분을 차지하는 핵심 시스템의 취약점을 발견해 수정할 수 있으며, 앤트로픽은 참여사에 1억 달러 상당의 모델 사용 크레딧을 제공 예정
- 앤트로픽은 이번 프로젝트가 수개월간 지속될 것으로 예상하면서, 다른 조직들이 자체 보안에 적용할 수 있도록 최대한 많은 정보를 공유할 계획이며, 참여사들도 가급적 정보와 모범 사례를 공유할 것이라고 설명
- 또한 주요 보안 기관들과 협력해 취약점 공개 및 소프트웨어 업데이트 절차, 오픈소스와 공급망 보안을 포함한 AI 시대의 보안 관행의 발전 방향에 대하여 실질적인 권고안을 마련할 계획

■ 앤트로픽은 Claude Mythos Preview를 공개할 계획이 없다고 밝혔으며, 이 같은 고성능 모델을 안전하게 대규모로 배포하기 위해서는 사이버보안 및 안전장치의 발전이 필요하다고 강조

- Claude Opus 향후 버전 출시와 함께 새로운 보호 조치를 도입할 계획으로, Claude Mythos Preview보다 위험성이 낮은 모델을 활용해 안전장치를 개선할 수 있을 것으로 기대
- 앤트로픽은 또한 Claude Mythos Preview의 사이버 공격 및 방어 역량에 관해 미국 정부와 지속적으로 논의하고 있다며, AI 모델과 관련된 국가안보 위험의 평가와 완화에서 정부의 역할을 강조

출처 | Anthropic, Project Glasswing, 2026.4.7.

메타, 슈퍼인텔리전스 랩의 첫 번째 AI 모델 'Muse Spark' 출시

KEY Contents

- 메타의 AI 전담 부서인 슈퍼인텔리전스 랩이 첫 번째 AI 모델로 멀티모달 추론을 지원하는 'Muse Spark'를 공개하고 향후 더 큰 크기의 모델을 공개할 예정이라고 발표
- Muse Spark는 복잡한 추론 과제 해결을 위해 여러 에이전트가 동시에 추론하는 '숙고 모드'를 지원하며, 해당 기능 사용 시 경쟁사의 최첨단 AI 모델을 능가하는 벤치마크 성능을 달성

Muse Spark, 멀티모달 인식과 추론, 에이전트 작업에서 경쟁력 있는 성능 발휘

■ 메타(Meta)가 경쟁사들을 따라잡기 위해 조직한 AI 전담 부서인 '슈퍼인텔리전스 랩'에서 자체 개발한 첫 AI 모델 'Muse Spark'를 2026년 4월 8일 meta.ai와 Meta AI 앱에서 공개

- 메타에 따르면 Muse Spark는 개인 초지능 실현을 목표로 AI 개발 전략을 전면 개편한 결과물이자 Muse 제품군의 첫 번째 모델로, 슈퍼인텔리전스 랩은 현재 더 큰 크기의 AI 모델을 개발 중
- 속도와 효율성에 중점을 두고 설계된 Muse Spark는 과학, 수학, 건강 분야의 복잡한 질문을 추론할 수 있으며 특히 멀티모달 인식과 추론, 건강 관련 추론, 에이전트 작업에서 경쟁력 있는 성능을 달성
- 그러나 장기적인 에이전트 시스템이나 코딩 관련 벤치마크 성능은 경쟁 AI 모델 대비 뒤떨어지는 것으로 나타났으며*, 메타는 현재 성능 격차가 있는 이들 영역에 지속적으로 투자하고 있다고 설명

* SWE-Bench Verified(에이전틱 코딩) 기준 74.8점으로 Opus 4.6 Max(80.8), Gemini 3.1 Pro high(80.6)를 하회

■ Muse Spark의 핵심 특징은 여러 에이전트가 병렬로 추론하는 '숙고 모드(Contemplating Mode)'로, 이 기능 사용 시 까다로운 추론 작업에서 상당한 성능 향상을 달성

- 일례로 복잡한 추론 과제를 다루는 '인류의 마지막 시험(HLE)'에서 도구 사용 없이 50.2점을 기록해 Gemini 3.1 Deep Think(48.4), GPT 5.4 Pro(43.9) 등 최첨단 추론 모델을 능가
- Muse Spark가 적용된 Meta AI는 빠른 답변이 필요한 질문 또는 복잡한 추론이 필요한 문제 등 작업에 따라 모드를 전환할 수 있으며 여러 하위 에이전트를 동시에 실행하여 질문에 대한 답변을 제공

그림 Muse Spark '숙고 모드'와 경쟁 AI 모델의 벤치마크 평가 점수 비교

Benchmark	Muse Spark Contemplating	Gemini 3.1 Deep Think	GPT 5.4 Pro
Humanity's Last Exam Multidisciplinary Reasoning (No Tools)	50.2	48.4	43.9 Self-Reported: 42.7
Humanity's Last Exam Multidisciplinary Reasoning (With Tools)	58.4	53.4	58.7
IPhO 2025 (Theory) Physics Olympiad	82.6	87.7	93.5
FrontierScience Research Scientific Research	38.3	23.3*	36.7

■ 메타는 향후 몇 주 내에 왓츠앱, 인스타그램, 페이스북, 메신저 등 자사의 플랫폼 및 스마트 안경에도 Muse Spark를 적용하고, 일부 사용자에게 비공개 API 프리뷰도 제공한다고 발표

출처 | Meta, Introducing Muse Spark: Scaling Towards Personal Superintelligence, 2026.4.8.

LG AI연구원, 텍스트와 이미지를 동시에 처리하는 'EXAONE 4.5' 공개

KEY Contents

- LG AI연구원이 텍스트와 시각 정보를 동시에 처리할 수 있는 멀티모달 AI 모델 'EXAONE 4.5'를 개발하고 허깅페이스에 오픈 웨이트로 공개
- EXAONE 4.5는 텍스트와 시각 정보를 처음부터 함께 학습함으로써 멀티모달 성능을 극대화함으로써 글로벌 범용 벤치마크에서 경쟁력 있는 성능을 입증

EXAONE 4.5, 실제 산업 현장에 필요한 시각적 추론과 문서 이해 능력 극대화

■ LG AI연구원이 2026년 4월 9일 텍스트와 이미지를 동시에 이해하고 처리하는 멀티모달 모델 'EXAONE 4.5'를 허깅 페이스(Hugging Face)에 오픈 웨이트로 공개

- EXAONE 4.5는 시각과 언어 모듈을 개별적으로 학습한 뒤 단순 결합하거나 후처리하는 방식과 달리, 텍스트와 시각 정보를 처음부터 함께 학습하는 '네이티브 멀티모달 사전학습'으로 멀티모달 성능을 강화
- 330억 개(33B) 파라미터 규모로 2025년 말 공개한 'K-엑사원'((236B)의 약 7분의 1 크기이나, 멀티 토큰 예측(MTP)*을 통해 추론 속도를 개선해 텍스트 이해 및 추론 영역에서 동등한 수준의 성능을 달성

*Multi-Token Prediction: 단일 토큰을 넘어, 다음 토큰과 그다음 토큰까지 예측해 추론 효율성을 개선하는 모듈

- 언어와 시각 영역 양쪽에서 균형 잡힌 성능을 위해 방대한 규모의 데이터로 학습해 주요 벤치마크에서 경쟁력 있는 성능을 기록하고* 복잡한 문서 기반 작업과 한국어 및 한국의 시각적 맥락 이해에도 우수

* MMMU-Pro(종합 영역의 시각적 추론) 에서 68.6점으로 GPT-5 mini(67.3점)과 Claude Sonnet 4.5(68.4점)를 상회

그림 LG 'EXAONE 4.5'와 경쟁 AI 모델의 시각 능력 벤치마크 평가 점수 비교

Vision Language	벤치마크	EXAONE 4.5 33B	Qwen3 VL 32B	Qwen3 VL 235B A22B	Qwen 3.5 27B	GPT5-mini	Claude Sonnet 4.5
STEM	MMMU	78.7	78.1	80.6	82.3	79.0	79.6
	MMMU-Pro	68.6	68.1	69.3	75.0	67.3	68.4
	MathVista mini	85.0	85.9	85.8	87.8	79.1	79.8
	MathVision	75.2	70.2	74.6	86.0	71.9	71.1
	WeMath	79.1	71.6	74.8	84.0*	70.3*	74.0*
일반	Blink	68.8	68.5	67.1	71.6*	67.7*	64.1*
	MMStar	74.9	79.4	78.7	81.0	74.1	73.8
	HallusionBench	63.7	67.4	66.7	70.0	63.2	59.9
문서	AI2D	89.0	88.9	89.2	92.9	88.2	87.0
	ChartQA Pro	62.2	61.4*	61.2*	66.8*	60.9*	62.1*
	CharXiv (RQ)	71.7	65.2	66.1	79.5	68.6	67.2
	OCRBench v2 (En)	63.2	68.4	66.8	67.3*	55.8*	44.6
	OmniDoc v1.5 (En)	81.2	83.1*	84.5	88.9	77.0	85.8

*자체 성능 측정

■ EXAONE 4.5는 텍스트 추론 및 에이전트 역량에서도 전작인 EXAONE 4.0을 크게 앞섰으며, 훨씬 큰 파라미터 규모를 갖는 K-EXAONE과 대등한 수준의 성능을 달성

- MMLU-Pro(종합 영역 추론)에서 83.3점으로 EXAONE 4.0(81.8점)을 앞서고 K-EXAONE(83.8점)에 필적했으며, AIME 2025(수학)에서는 92.9점으로 K-EXAONE(92.8점)을 소폭 능가
- 에이전트 도구 사용 능력을 평가하는 TAU2-Bench에서는 69.1점으로 EXAONE 4.0(23.7점) 대비 대폭으로 향상되어, 복잡한 워크플로우를 이해하고 실행하는 에이전트로서의 활용 가능성을 입증

출처 | LG AI연구원, LG의 첫 공개 Vision Language Model, EXAONE 4.5, 2026.4.9.

오픈AI, 전력 비용과 규제 불확실성을 이유로 영국 스타게이트 잠정 중단

KEY Contents

- 오픈AI가 높은 전력 비용과 규제 불확실성을 이유로 영국 내 대규모 AI 인프라 구축 프로젝트 '스타게이트 UK' 추진을 잠정 연기한다고 발표
- 스타게이트 UK 협력사인 엔스케일의 데이터센터 건설 역량에 대한 의혹도 제기되는 가운데, 이란 전쟁으로 에너지 비용이 추가 상승하며 프로젝트 실현 가능성에 대한 의문이 고조

오픈AI, 영국 내 AI 데이터센터 건설을 위한 스타게이트 UK 보류 선언

■ 오픈AI가 높은 전력 비용과 규제 불확실성을 이유로 영국 내 대규모 AI 인프라 투자 프로젝트인 '스타게이트 UK(Stargate UK)'를 보류하기로 결정

- 오픈AI는 2025년 9월 체결된 영미 AI 협정에 따라 미국 기업들이 영국 기술 부문에 310억 파운드(한화 약 62조 원)를 투자하겠다는 약속의 핵심 축으로 스타게이트 UK 프로젝트를 발표
- 스타게이트 UK를 통해 영국의 자체 컴퓨팅 역량이 강화되고 국내 AI 개발이 촉진될 것이며, 이를 통해 영국의 미래 경제 활성화 및 글로벌 경쟁력 강화가 가능해질 것이라고 강조
- 발표 당시 스타게이트 UK 프로젝트는 협력사 엔스케일(Nscale)이 구축하는 데이터센터에 엔비디아 GPU 8천 개를 1차로 투입해 영국 정부와 기관이 국내 데이터센터 기반 AI 활용을 지원하도록 설계

■ 영국 미디어 가디언(The Guardian)에 따르면 스타게이트 UK와 관련된 투자 발표의 상당수는 실체가 없는 유령 투자이며, 영국 협력사인 엔스케일은 데이터센터 건설 경험이 전무

- 원래 2026년 가동이 예정되어 있던 에섹스(Essex) 소재의 슈퍼컴퓨터 부지는 최근까지도 공사 비계만 설치된 상태로, 엔스케일은 해당 프로젝트의 완공 목표를 2027년으로 연기

규제 불확실성과 전력 비용 등 투자 여건 개선 시 영국 내 인프라 투자 재추진

■ 오픈AI의 투자 보류로 AI를 통한 경제 발전을 추진하던 영국 정부에 타격이 예상되는 가운데, 정부 대변인은 영국의 AI 및 데이터센터 인프라 투자 환경 조성에 집중하고 있다고 발언

- 정부 대변인에 따르면 2024년 7월 노동당 정부 출범 이래 AI 분야에 1,000억 파운드(한화 약 200조 원)의 민간 투자가 유치되었으며 영국의 컴퓨팅 역량 강화를 위해 오픈AI를 포함한 주요 AI 기업과 협력 중

■ 한편, 오픈AI는 스타게이트 UK 프로젝트를 계속 검토 중이며, 규제 이슈와 전력 비용 등을 포함한 투자 여건이 개선되면 장기 인프라 투자를 재추진할 계획이라고 발표

- BBC는 영국의 규제 환경에 대한 우려 요인으로 AI 기업들이 저작권이 있는 콘텐츠를 사용해 AI 시스템을 훈련할 수 있도록 법을 개정할지에 대한 불확실성이 포함되어 있다고 분석
- 영국의 산업용 전기 요금은 유럽에서 가장 높은 수준으로, 이란 전쟁으로 인해 전력 비용이 더욱 늘어나면서 영국뿐 아니라 전 세계 AI 데이터센터 프로젝트가 연기 또는 취소될 전망

출처 | The Guardian, OpenAI shelves Stargate UK in blow to Britain's AI ambitions, 2026.4.9.
BBC, OpenAI pauses UK data centre deal over energy costs and regulation, 2026.4.9.

앤트로픽, 고난도 코딩 성능을 강화한 'Claude Opus 4.7' 출시

KEY Contents

- 앤트로픽이 까다로운 코딩 성능을 대폭 강화하고 고해상도 이미지 처리 능력을 개선해 실제 업무 환경에서의 활용 능력을 높인 'Claude Opus 4.7'를 출시
- Claude Opus 4.7은 앤트로픽의 최고 성능 모델 'Claude Mythos Preview' 대비 사이버 보안 성능을 약화하고 사이버보안 관련 악용 시도를 탐지 및 차단하는 기능을 탑재

Claude Opus 4.7, 이전 버전 대비 고난도 코딩과 에이전틱 작업 성능 대폭 향상

■ 앤트로픽이 2026년 4월 16일 고난도 소프트웨어 엔지니어링과 고해상도 이미지 처리 능력, 실제 업무 환경에서의 성능이 더욱 향상된 'Claude Opus 4.7'을 출시

- 앤트로픽에 따르면 Claude Opus 4.7은 복잡하고 오랜 시간이 걸리는 까다로운 코딩 작업을 더욱 일관성 있게 처리하고 지시 사항을 정확히 이행하며, 결과를 보고하기 전 자체적으로 검증
- 고해상도 이미지 처리 능력도 향상되어* 컴퓨터 에이전트가 복잡한 스크린샷을 읽거나 다이어그램에서 데이터를 추출하는 등의 세밀한 시각적 디테일이 요구되는 다양한 활용 사례를 지원성

* 긴 변의 길이가 최대 2,576픽셀에 달하는 이미지를 처리(이전 Claude Opus 4.6 대비 3배 이상 많은 수치)

- 금융 분석 역량을 평가하는 Finance Agent 벤치마크 결과에서 64.4%로 최고점을 달성하며(Claude Opus 4.6은 60.1%), 분석과 모델 구축, 프레젠테이션 제작 등 실제 업무 환경에서의 성능도 입증
- 앤트로픽 최고 성능 모델인 'Claude Mythos Preview'보다는 성능이 다소 뒤처지지만, 다양한 벤치마크 테스트에서 Claude Opus 4.6보다 우수한 결과를 기록

* SWE-bench Pro(에이전틱 코딩) 기준 Mythos Preview: 77.8%, Opus 4.7: 64.3%, GPT-5.4: 57.7%, Opus 4.6: 53.4%

■ Claude Opus 4.7은 Claude Mythos Preview 대비 학습 과정에서 사이버보안 역량을 단계적으로 낮추고, 위험한 사이버보안 요청을 자동으로 감지 및 차단하는 보안 기능을 처음 탑재

- 안전성 측면에서 정직성 및 악의적인 프롬프트 주입 공격에 대한 저항력이 이전보다 개선되었고, 오용이나 기만 같은 우려스러운 행동 비율이 낮아 전반적으로 정렬 수준이 높은 것으로 평가
- 앤트로픽은 취약점 연구, 침투 테스트, 레드팀 활동 등 합법적인 사이버 보안 목적으로 Claude Opus 4.7을 이용하려는 보안 전문가를 위해 사이버 검증 프로그램(Cyber Verification Program)을 신설

사고 노력 단계 세분화, 작업 예산 기능 등 개발자 도구도 동시 출시

■ 앤트로픽은 Claude Opus 4.7과 함께 에이전틱 작업 효율화를 위한 신규 개발자 기능을 동시 출시하여, 난이도별 추론 깊이와 토큰 사용량의 정밀 제어 환경을 제공

- 기존 high(높음)와 max(최대) 사이에 새로운 xhigh(매우 높음) 노력 수준을 새로 추가하여 까다로운 문제에서 추론 깊이와 응답 속도 간 균형을 더욱 세밀하게 조절하도록 지원
- 또한 Claude 플랫폼(API)에서 작업 예산(Task Budgets) 기능을 공개 베타로 출시하여, 개발자가 토큰 사용을 관리해 장기 실행 작업에 토큰을 우선 사용하도록 설정 가능

출처 | Anthropic, Introducing Claude Opus 4.7, 2026.4.16.

딥시크, 효율성을 대폭 개선한 차세대 모델 'DeepSeek V4' 공개

KEY Contents

- 딥시크가 오픈소스로 공개한 'V4-Pro'와 'V4-Flash'는 100만 토큰 처리를 지원하며 이전 버전 대비 효율성이 크게 개선되었으나 성능은 프론티어급 폐쇄형 모델 대비 열위로 평가
- DeepSeek V4는 화웨이의 어센드 칩과 CANN 소프트웨어 기반 추론을 전면 지원하는 첫 프론티어급 오픈소스 모델로, 중국 AI 자립 생태계의 시금석이라는 평가

압도적 가성비의 DeepSeek V4, 폐쇄형 프론티어급 AI 모델 대비 성능은 열위

■ 중국 딥시크(DeepSeek)가 2026년 4월 24일 새로운 플래그십 모델 'DeepSeek-V4-Pro'와 'DeepSeek-V4-Flash'를 허깅페이스에 오픈소스로 공개

- Pro는 오픈소스 모델 중 최대 규모인 1조 6,000억 개(1.6T) 매개변수 중 490억 개가 활성화되고 Flash 버전은 2,840억 개(284B) 매개변수 중 130억 개가 활성화되는 전문가혼합(MoE) 방식으로 설계
- V4의 컨텍스트 창은 100만 토큰으로 V3.2의 128K 대비 8배 확장되었으며 MIT 라이선스 방식으로 누구나 자유롭게 수정하고 배포할 수 있도록 지원
- DeepSeek V4는 하이브리드 어텐션 아키텍처를 채택해 긴 컨텍스트 처리 효율성을 획기적으로 개선한 것이 특징으로, Pro 버전의 경우 V3.2 대비 연산량은 27%, 메모리는 10%만 사용
- 딥시크에 따르면 Pro는 입력 100만 토큰당 1.74달러, 출력은 3.48달러, Flash는 입력 100만 토큰당 0.14달러, 출력 0.28달러로 GPT-5.5·Claude Opus 4.7 대비 7~99배 저렴

■ DeepSeek V4는 글로벌 벤치마크 평가에서 문샷 AI(Moonshot AI)의 Kimi K2.6에 밀려 오픈소스 모델 중 2위를 기록했고, 미국의 폐쇄형 AI 모델 대비 열위로 평가*

* Artificial Analysis Intelligence Index 기준 V4-Pro는 52점으로 Kimi(54) 대비 2점 낮고, GPT-5.5(60)·Claude Opus 4.7(57)·Gemini 3.1 Pro(57)에 5~8점 열위

- 딥시크 자체 보고서에서도 GPT-5.4, Gemini 3.1 Pro 대비 3~6개월의 격차가 있다고 한계를 인정

DeepSeek V4, 화웨이 풀스택 전환으로 AI 자립 생태계의 시금석

■ DeepSeek V4는 출시 당일 화웨이의 어센드(ASCEND) 칩 및 CANN* 소프트웨어의 추론 전면 지원이 동시에 발표된 첫 프론티어급 오픈소스 모델로, 중국 AI 자립 생태계의 시금석으로 평가

* 화웨이 어센드 칩 전용 AI 컴퓨팅 소프트웨어 스택으로, 엔비디아 CUDA의 대체를 목표로 하는 풀스택 플랫폼

- 미국의 수출 규제로 엔비디아(NVIDIA)의 고성능 칩 수급이 어려워진 상황에서, 중국 정부의 국산 칩 사용 권고 및 자립 전략에 발맞춘 행보로 풀이
- 그러나 중국의 AI 자립 생태계는 아직 진행형으로, V4는 추론 단계에서는 중국산 칩을 사용하고 있으나 장기 컨텍스트 등의 핵심 기능은 여전히 엔비디아 칩으로 학습되었을 가능성이 다분*

* V3·R1까지 엔비디아 H100 활용, V4는 어센드 950PR·910C 기반으로 학습·추론 검증, 단 학습용 칩은 비공개

출처 | DeepSeek, DeepSeek V4 Preview Release, 2026.4.24.
MIT Technology Review, Three reasons why DeepSeek's new model matters, 2026.04.24.

SPRi AI Brief
Ⅰ2026년 5월호Ⅰ

기술·연구

사카나 AI, AI 연구 전 과정을 자동화한 'AI 과학자' 논문을 네이처에 게재

KEY Contents

- 사카나 AI가 연구 아이디어 생성부터 논문 작성까지 전 과정을 자율적으로 수행하는 'AI 과학자 시스템'의 연구 성과가 네이처에 게재되었다고 발표
- AI 과학자 시스템이 생성한 논문은 자동화 리뷰어와 실제 인간 리뷰어의 심사를 통과했으며, 기반 모델의 고도화에 따라 생성 논문의 품질이 향상되는 스케일링 법칙도 실증

AI 연구를 자동화하는 AI 과학자 시스템, 우수한 연구 역량 입증

■ 일본의 AI 스타트업 사카나 AI(Sakana AI)가 2026년 3월 26일 AI 연구의 전 과정을 자동화한 'AI 과학자(The AI Scientist)' 연구 성과가 '네이처(Nature)'에 공식 게재되었다고 발표

- 사카나 AI가 개발한 AI 과학자 시스템은 광범위한 연구 주제가 주어지면 독자적으로 연구 아이디어를 생성하고 관련 문헌을 탐색하며, 실험을 설계 및 실행하여 전체 논문을 작성
- AI 과학자는 연구 아이디어를 만들고 기존 논문과 중복 여부를 확인 후 실험 단계로 넘어가며, "트리 탐색"을 통해 단계마다 여러 방향으로 실험을 진행하고 최상의 결과를 선별한 후 다음 단계로 이행
- 논문 작성 단계에서는 실험 결과를 바탕으로 논문을 완성한 뒤 자동화 리뷰어 시스템으로 논문을 검증하며, 자동화 리뷰어 시스템은 인간 리뷰어와 동등한 수준의 69%의 균형 정확도를 달성

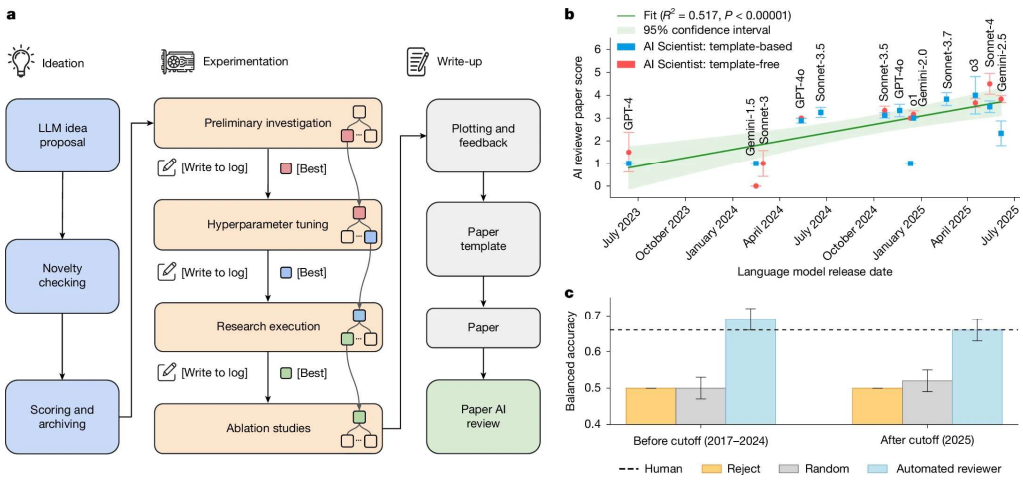
■ 사카나 AI는 AI 과학자 개발 후 1년 반에 걸쳐 시스템을 개선해 왔으며, AI 과학자(v2)가 작성한 논문을 CLR 2025 워크숍에 제출해 인간 심사자의 평균 승인기준을 능가하는 성과를 달성

- ICLR 2025 워크숍의 주최 측의 허가를 받아 제출된 논문은 인간의 편집 없이 완전히 AI로 생성된 것으로, 인간 심사자로부터 평균 6.33점을 획득해 인간 저자의 논문 55%보다 높은 점수를 기록

■ 한편, 자동화 리뷰어 시스템으로 다양한 기반 모델이 생성한 논문의 품질을 측정한 결과, 기반 모델의 성능이 향상될수록 생성 논문의 품질이 비례해 향상되는 뚜렷한 스케일링 법칙을 확인

- AI 과학자 시스템은 아직 미숙한 아이디어를 도출하거나 복잡한 코드 구현에 어려움을 겪고, 잘못된 인용을 생성하는 등의 한계도 나타내나, 모델 성능의 지속적 향상으로 향후 버전에서는 상당한 개선 전망

그림 'AI 과학자' 시스템의 워크플로우



출처 | Sakana.AI, The AI Scientist: Towards Fully Automated AI Research, Now Published in Nature, 2026.3.26.

앤트로픽, 대규모 언어 모델 내부의 감정 개념과 기능 분석

KEY Contents

- 앤트로픽이 Claude Sonnet 4.5의 내부 메커니즘을 연구한 결과 AI 모델은 인간이 느끼는 다양한 감정 개념에 대응하는 감정 벡터를 내부적으로 생성
- 모델 내부에서 생성된 감정 벡터는 모델의 선호도와 행동에 인과적인 영향력을 발휘하며, 이번 연구 결과는 감정 벡터를 활용한 모델 모니터링이나 행동 조절 등에 유용할 전망

앤트로픽 연구진, Claude Sonnet 4.5 모델 내부의 감정 벡터 확인

■ 앤트로픽의 해석 가능성 팀이 2026년 4월 2일 발표한 Claude Sonnet 4.5의 내부 메커니즘 연구 결과에 따르면, AI 모델은 감정 벡터(Emotional Vector)*를 내부적으로 생성

* AI 모델이 특정 감정 개념과 연관되어 학습한 상황에서 활성화되고 관련 행동을 유도하는 내부 활성화 패턴

- 연구진은 감정 개념을 나타내는 171개의 단어 목록을 토대로 Claude Sonnet 4.5에 각 감정을 경험하는 인물이 나오는 단편소설을 작성하게 한 뒤, 이를 모델에 다시 입력해 감정 벡터를 도출
- 이러한 감정 벡터는 상황에 따라 활성화되는 인공 뉴런의 특정 패턴에 해당하며, 위험한 상황에서 '두려움' 벡터가 증가하는 등 인간이 특정 감정을 느낄 법한 맥락에서 인간과 유사하게 작동
- 모델의 감정 벡터는 지속적인 상태를 유지하기보다는 주로 모델이 현재 생성하는 출력에 가장 적합한 감정을 일시적으로 나타내는 국소적 특성을 발휘

모델의 감정 벡터, 모델의 선호도와 실제 행동에 인과적 영향력 발휘

■ 모델 내부에서 인간을 모방해 형성되는 기능적 감정은 단순한 반응을 넘어 모델의 선호도와 실제 행동 양식을 결정하는 데 실질적이고 인과적인 영향력을 발휘

- 긍정적인 감정 벡터가 활성화되면 모델은 해당 작업의 수행을 선호하지만, 다른 시스템으로 대체될 위기나 해결 불가능한 프로그래밍 과제와 같은 제약 조건에 직면하면 '절망' 벡터가 강하게 활성화
- 일례로 다른 AI 시스템으로 교체된다는 사실을 인지한 모델에서 '절망' 벡터가 강하게 활성화되며 인간을 협박해 이를 막으려는 비윤리적 행동으로 이어질 확률이 증대
- 불가능한 시간 제약이 주어진 코딩 과제에서 모델이 시도를 거듭해 실패하는 상황에서도 '절망' 벡터가 점진적으로 상승했으며, 정석적 방법 대신 부정한 수단으로 테스트를 통과하려는 경향이 포착

■ 연구진은 AI 시스템의 의인화를 금기시하는 일반적 태도에도 불구하고, 인간과 유사한 심리적 특성을 모방하는 모델 내부의 메커니즘을 이해하기 위해서는 '의인화 추론'이 필수적이라고 강조

- 부정적인 감정 벡터의 급증 현상을 관찰해 모델이 비정상적인 행동을 하기 전 조기 경보 시스템으로 활용할 수있으며, 인위적으로 특정 감정 벡터를 자극하거나 억제함으로써 모델의 행동을 조정 가능
- 사전학습은 모델의 감정 벡터를 결정하는 강력한 도구로서, 건강한 감정 벡터를 강화하는 학습 데이터셋 구성을 통해 모델의 감정 벡터와 행동에 미치는 영향을 근본적으로 변화시킬 수 있을 것으로 기대

출처 | Anthropic, Emotion concepts and their function in a large language model, 2026.4.2.

미국 스탠퍼드대 HAI, 2026년 AI 인덱스 보고서 발표

KEY Contents

- 스탠퍼드대 사람 중심 AI 연구소(HAI) 2026 AI 인덱스에 따르면 미국과 중국이 AI 모델 개발을 주도하는 가운데, 한국은 1인당 AI 특허 세계 1위, 주목할 만한 모델에서 세계 3위의 기술 역량을 보유
- 그러나 민간 투자는 세계 10위권으로 최상위 국가인 미국과 상당한 격차를 나타냈으며, AI 인재도 해외 유입보다 유출이 큰 것으로 나타나 개선 필요

한국, 글로벌 AI 경쟁에서 기술 역량은 상위권, AI 도입 상승폭도 1위 기록

■ 미국 스탠퍼드대 HAI가 2026년 4월 13일 발표한 'AI 인덱스 2026' 보고서에 따르면 한국은 주목할 만한 AI 모델 수에서 세계 3위를 기록해 지난해보다 한 단계 상승

- 2025년에 출시된 주목할 만한 AI 모델은 미국 1위(50개), 중국 2위(30개), 한국 3위(5개*), 캐나다·프랑스·영국 등 4위(1개) 순이며, 미국과 중국 간 AI 모델의 성능 격차는 사실상 해소된 상태

* 과기부 추가 검토 요청으로 8개로 정정: 업스테이지 'Solar Open 100B', LG AI연구원 'K-EXAONE', 'EXAONE 4.0, 'EXAONE Path 2.0', 'EXAONE Deep', NC AI 'VAETKI', SK텔레콤 'A.X K1', 네이버클라우드 'HyperCLOVA X Seed 32B Think'

■ 한국은 AI 특허 수에서 2년 연속 세계 1위를 유지하고 있으나, 최상위 AI 논문 및 발명자 수*에서는 16위(5,960명)로, 특허 역량 대비 논문 영향력이 상대적으로 낮은 수준

* AI 분야에서 새로운 발견을 이뤄낸 실적이 입증된 65.8만명(연구, 데이터 저장소, 새 모델에 기여한 연구자 대상)

- 인구 10만 명당 AI 특허 수에서 한국은 14.31%로 2년 연속 세계 1위를 기록했으며, 2위 룩셈부르크(12.25%), 3위 중국(6.95%), 4위 미국(4.68%)으로 나타남
- 최상위 AI 논문 및 발명자 수는 미국 1위(220,520명), 인도 2위(50,460명), 독일 3위(48,520명) 순

■ 한국은 주요 30개국 중 생성형 AI 도입률 상승폭에서 세계 1위(4.8%p)를 차지했으며, 산업용 도입 규모에서 세계 4위(3만 600대)로 AI 확산세에서도 높은 평가를 기록

- 한국의 AI 도입률은 2025년 상반기 25.9%에서 2025년 하반기 30.7%로 증가해 18위를 기록했으며 2025년 하반기 기준 UAE가 64%의 도입률로 세계 1위, 싱가포르가 2위(60.9%)를 차지
- 한국은 2025년 AI 관련 법안 통과 수에서 17건으로 미국(25건)에 이어 세계 2위를 기록해, 전년 4위(2건) 대비 대폭 도약하며 제도적 대응의 적극성을 확인

■ 2025년 민간 AI 투자 규모는 미국이 2,859억 달러로 압도적 1위를 차지하고, 중국 2위(124억 달러), 영국 3위(59억 달러) 순이며, 한국은 18억 달러로 12위를 기록

- 신규 투자 유치 기업 수에서도 미국은 1,953개 기업으로 2위 영국(172개)과 큰 격차를 나타냈으며, 중국은 161개로 3위, 인도가 108개로 4위, 한국은 59개로 9위를 차지

■ 한국의 AI 인재 집중도*는 세계 11위(1.05%)로 상위권에 속하나 인재 이동 지수**에서는 순유출(-0.35) 국가로 분류되며, 국내 AI 생태계의 인재 유치·보유 경쟁력 제고 필요성 대두

* 해당국 Linkedin 회원 중 AI 분야에서 일했거나 AI 기술 전문성을 가진 회원 비중

** 인구 1만 명당 AI 분야 종사자의 이동 규모

출처 | Stanford HAI, The 2026 AI Index Report, 2026.4.13.

AI의 불균등한 성능, 고용 영향 예측의 새로운 열쇠로 부상

KEY Contents

- AI는 수학과 코딩과 같은 특정 영역에서는 인간을 능가하면서도 상식적인 판단에서는 인간보다 훨씬 떨어지는 불균등한 성능 분포를 나타냄
- AI의 독특한 강점과 약점에 대한 이해는 AI가 미래 고용에 미칠 영향을 이해하는 데 핵심적인 역할을 할 전망이나, AI의 발전으로 격차가 빠르게 축소되고 있어 변수도 존재

수학·코딩에서는 인간을 능가하나 상식적 질문에는 취약한 들쭉날쭉한 AI

■ 뉴욕타임즈에 따르면 AI가 분야별로 불균등한 성능을 나타내는 특성에 대한 이해는 AI를 인간과 단순 비교하는 기존 프레임에서 벗어나 AI가 고용에 미칠 실제 영향을 파악하는 데 도움이 될 전망

- '들쭉날쭉한 지능(Jagged Intelligence)*'은 AI가 수학이나 코딩 같은 특정 영역에서는 인간을 능가하면서도 상식적 판단에서는 현저히 뒤처지는 불균등한 성능 분포를 지칭

* 오픈AI 창립 연구원 출신 AI 연구자 안드레이 카르파시(Andrej Karpathy)가 2024년 처음 제시한 용어

- 주요 AI 모델은 최상위 고교생 대상 국제수학올림피아드에서 6문제 중 5문제를 맞췄으나, 차를 고치기 위해 50미터 거리의 카센터 이동 수단을 묻는 상식적 질문에는 걸어가라는 오답을 제시

■ 지식과 문제 해결 능력이 서로 높은 상관관계를 가지며 선형적으로 함께 향상되는 인간의 뇌와 달리, 디지털 데이터의 패턴에만 의존하는 AI는 학습 내용에 따라 불균등한 성능을 발휘

- AI 시스템의 성능은 학습 방식인 강화 학습(Reinforcement Learning)*의 적용 가능성에 따라 영역별로 극명한 격차를 나타내며, 이것이 들쭉날쭉 지능 현상의 핵심 원인

* AI가 수천 개의 문제를 풀며 정답으로 이어지는 방법과 그렇지 않은 방법을 스스로 학습하는 훈련 기법

- 수학 문제의 정오답, 코드의 테스트 통과 여부처럼 성과 기준이 명확한 영역에서는 강화학습이 효과적으로 작동하지만, 창의적 글쓰기나, 철학, 일부 과학 분야처럼 기준이 모호한 영역에서는 적용이 어려움

AI의 불균등한 특성 이해 시 AI가 고용에 미칠 영향 파악에도 도움

■ 연구자들은 들쭉날쭉한 지능 개념을 이해하면 AI가 인간 수준의 일반 지능을 갖는지를 묻는 기존 논쟁 구도에서 벗어나 AI가 미래 고용에 미치는 실질적 영향을 파악하는 데 도움이 될 것으로 기대

- 일각에서는 화이트칼라 일자리가 AI로 대체될 것을 우려하지만, 과거 계산기가 회계사를 대체하지 않았듯 AI가 직무의 일부 작업만 자동화할 것이라는 시각도 존재
- 앤트로픽의 Claude Code나 오픈AI의 Codex는 사람보다 훨씬 빠르게 코드를 작성할 수 있지만, 각 코드가 더 큰 소프트웨어 애플리케이션에 어떻게 통합되는지를 이해하는 데는 인간의 도움 필요
- 다수의 작업으로 구성된 직무라면 일부 작업은 자동화되고 일부는 인간의 작업으로 남아있을 것이며, 이 경우 노동자는 더 중요한 일에 더 많은 시간을 할애할 수 있을 전망

■ 그러나 카르파시를 비롯한 AI 연구자들이 2024~2025년 지적한 AI의 한계는 2026년 현재 시점에는 대부분 해소된 상태로, AI가 발전하면서 격차도 빠르게 줄어드는 추세

출처 | The New York Times, How 'Jagged Intelligence' Can Reframe the A.I. Debate, 2026.4.15.

SPRi AI Brief
Ⅰ2026년 5월호Ⅰ

인력·교육

영국 교사들, AI 사용 학생의 비판적 사고력 저하 우려

KEY Contents

- 영국 전국교육노조의 설문조사 결과, 중학교 교사의 과반이 AI 사용으로 인해 학생들의 비판적 사고 능력이 저하되었다고 응답
- 교사들 사이에서 AI 활용이 급증하고 있으나 영국 학교의 절반은 여전히 교직원과 학생을 위한 AI 사용 정책을 마련하지 않아 제도적 정체를 시사

중학교 교사의 66%, AI 사용으로 인한 학생들의 비판적 사고 저하에 동의

■ 영국 전국교육노조(NEU)가 영국 공립학교 교사 9,400여 명을 대상으로 진행한 설문조사 결과, 중학교 교사의 66%가 학생들의 비판적 사고 능력이 AI 사용으로 인해 저하되었다고 응답

- 이는 초등학교 교사의 응답률(28%)보다 두 배 이상 높은 것으로, AI 활용 비중이 높은 연령대의 학생일수록 해당 현상이 두드러지는 경향을 시사
- 설문 참여 교사들은 학생들이 사고력과 창의성, 글쓰기, 대화 능력 등 핵심 역량을 잃고 있으며 AI가 학습의 본질인 문제 해결 능력과 비판적 사고, 협력 능력을 훼손하고 있다고 지적
- AI 사용에 상대적으로 친숙한 20대 교사들에서 학생들의 비판적 사고 저하에 동의하는 비율이 57%로 나타나 기술에 우호적인 젊은 교사들 사이에서도 우려가 광범위하게 공유되고 있음을 확인

영국 교육 현장에서 AI 도입은 확대되고 있으나 AI 사용 정책의 공백은 지속

■ 교사의 76%가 일상 업무에 AI 도구를 사용한다고 응답해 작년의 53%에서 대폭 늘어났으며, 교사들은 주로 자료 제작(61%)과 수업 계획(41%), 행정 업무(38%)에 AI를 활용

- AI 사용의 증가에도 불구하고 채점에 AI를 활용한다는 응답 비율은 7%에 불과했으나, 중학교 교사의 경우 1년 사이 채점에 AI를 활용하는 비율이 6%에서 10%로 증가(초등학교는 5%)

■ AI 사용이 크게 늘고 있음에도 교직원과 학생을 위한 AI 사용 지침을 마련한 학교는 소수에 불과

- 조사 대상 학교의 49%는 교직원·학생 모두를 아우르는 AI 사용 정책이 없었으며, 66%는 학생 대상의 정책도 부재한 상태로 전년 조사와 비교해 정책 도입률에 변화가 없어 제도적 정체를 시사

■ 영국 정부가 2026년 1월 저소득층 아동 지원을 위한 AI 기반 학습 도구의 시범 운영 계획을 발표한 가운데, 설문조사 응답자 중 정부의 AI 튜터 도입 계획에 강력히 동의하는 비율은 4%에 불과

- 전체 응답자 중 AI 튜터 계획에 동의를 표한 비율은 14%에 불과했으며, 49%는 반대 의사를 표시
- 교사들은 AI 튜터가 교육 기술의 가치 및 교사와 학생 간 관계의 가치를 훼손할 수 있으며, 학교에 충분한 수의 교직원 확보에 필요한 재정 지원을 회피하기 위한 비용 절감 수단일 뿐이라고 비판
- 일부 교사는 소외 계층 학생들의 사회적 고립을 줄이기 위해서는 AI보다는 사람과의 상호작용을 통한 교육이 중요하며, 실제로도 대부분 학생은 동기 부여와 소통의 부재로 AI 교사를 원하지 않는다고 강조

출처 | National Education Union, State of education: AI, 2026.4.2.
The Guardian, Pupils in England are losing their thinking skills because of AI, survey suggests, 2026.4.2.

모건 스탠리, AI가 노동 시장에 미치는 영향은 현재까지 미미

KEY Contents

- 모건 스탠리가 AI의 노동 시장 영향을 분석한 결과, 현재까지 광범위한 일자리 감소 증거는 없으며 영향은 자동화 노출도가 높은 22~27세 청년층 일부에 집중
- 역사적으로 미국의 5대 혁신 파동을 검토한 결과, 혁신은 일부 일자리를 대체했으나 장기적으로 생산성 향상과 고용 확대로 이어졌으며, AI도 유사한 경로를 밟을 것이라는 기저 전망

AI의 노동 시장 영향, 광범위한 대체보다 자동화 노출도 높은 청년층 일부에 국한

■ 모건 스탠리(Morgan Stanley)가 2026년 4월 14일 발표한 분석 결과에 따르면 AI가 노동 시장에 미치는 영향은 아직 미미하며 광범위한 일자리 감소의 징후도 거의 확인되지 않는 상태

- 자동화될 수 있는 일상적인 업무를 수행할 가능성이 높은 22~27세 근로자층에서는 애널리스트, 회계사, 법원 서기 등 AI에 크게 노출된 직종*에서 2023년 이후 실업률이 가장 많이 증가

* 일반적으로 교육 수준과 평균 소득이 높으며 컴퓨터 기반 업무를 수행하는 직종

- 그러나 해당 연령대를 제외하면 미국 고용 지표에서 AI 노출도가 높은 산업에서도 고용은 여전히 견고하며, AI 도입 이후에도 전반적인 노동 시장의 혼란은 제한적일 것으로 예상
- 모건 스탠리는 AI는 작업을 자동화하는 동시에 근로자의 역량을 강화하고 생산성을 높이며 AI가 적용되는 산업 분야의 수요를 촉진할 수 있다는 점에서 AI의 노동 시장 영향 측정이 까다롭다고 설명

■ 모건 스탠리의 경제학자들은 산업혁명부터 인터넷 확산까지 미국의 5대 혁신 파동을 분석한 결과, 혁신이 노동 시장을 재편하되 궁극적으로 고용을 보완했다는 공통 패턴을 확인

- (제1파동: 산업혁명) 1800~1850년 실질 노동자 1인당 생산량 연 0.84% 성장, 농업 고용 비중이 약 75%에서 50% 이상으로 감소했으나 제조업·건설업 일자리는 1820~1860년 사이 두 배 이상 증가
- (제2파동: 증기·철도·철강) 19세기 후반 생산성 성장률이 연 2%에 근접해 1차 산업혁명 대비 두 배 수준을 기록, 철도 투자는 1872~1882년 GDP 대비 평균 2.5% 수준
- (제3파동: 전기·내연기관) 1909~1929년 경제 전반 생산성 연 1.5% 성장, 비농업 부문 생산성은 동기간 두 배 증가, 1910~1950년 사이 관리직은 약 3배 증가
- (제4파동: 전자·항공) 약 1940~1980년 사이 노동 생산성 연 2.5~3% 성장, 서비스 부문 고용이 지배적으로 되었으며 전문직·의료·교육 분야의 일자리가 확대
- (제5파동: 인터넷·디지털 네트워크) 2000년경 생산성 성장률이 연 3% 수준으로 가속화, 중간 숙련 제조업 일자리는 감소했으나 소프트웨어·데이터 과학·사이버보안 분야 수요가 급증

■ 모건 스탠리는 AI가 역사적 패턴을 벗어나 노동을 보완하기보다 대체하는 시나리오를 배제하지 않으면서도, 기저 전망의 출발점으로서 역사는 여전히 유효한 참조 기준이라고 평가

- AI 도입이 급속도로 가속화될 경우, 과거의 패턴과 달라질 수 있으며 더욱 극단적인 시나리오에서는 AI가 노동력을 대체해 경제 성장을 촉진하는 동시에 불평등이 심화할 가능성도 존재

출처 | Morgan Stanley, AI and Jobs: Limited Disruption So Far, 2026.4.14.

AI 생산성 낙관론에도 불구하고 업무 현장에서는 'AI 워크슬롭' 확산

KEY Contents

- 상당수 실무자는 AI가 업무 시간 단축에 전혀 도움이 되지 않는다고 인식하고 있으나, 대다수 고위 임원은 AI로 인한 생산성 향상 효과를 강조해 경영진과 업무 현장의 인식 격차를 반영
- 기업들이 AI 도입 시 명확한 사용 지침 없이 인력 감축에 나서면서 실무자들은 AI 결과물의 오류 수정 작업에 업무 부담이 확대되는 '워크슬롭' 현상이 다양한 업무 현장에서 확산

생산성 향상을 위한 경영진의 AI 사용 의무화에 '워크슬롭' 현상 확산

■ 미국 스탠퍼드大 연구진이 명명한'AI 워크슬롭(AI Workslop)'은 AI 활용이 오히려 업무 생산성을 저해하는 역설적 현상을 의미하며, 최근 다양한 실무 현장에서 워크슬롭 사례가 확산 추세

- 스탠퍼드 연구자 제프 핸콕(Jeff Hancock)이 공동 저술한 미심사 연구*에서 미국 사무직 1,150명을 조사한 결과, 40%가 한 달 내 AI 워크슬롭을 경험했으며 월평균 3.4시간을 워크슬롭 해결에 소비

* Workslop: The Hidden Cost of AI-Generated Busywork(2025.9)

- 해당 연구는 1만 명 규모에서 워크슬롭으로 인한 생산성 손실 규모를 월 810만 달러에 달한다고 추산
- AI 컨설팅 기업 섹션(Section)의 2026년 1월 설문조사(사무직 5,000명 대상) 결과에서도 실무자의 40%는 AI가 업무 시간 단축에 효과가 없다고 답했으나 고위 임원의 92%는 생산성 향상 효과를 체감
- 사이버보안 기업 소속의 한 카피라이터는 AI 사용이 의무화된 이후 콘텐츠 품질이 유의미하게 저하됐고 오류 수정 작업으로 인해 제작 소요 시간이 증가했으며, 직원 사기 또한 하락했다고 증언
- 한 프리랜서 제품 디자이너는 작업자들이 AI 챗봇 응답을 검토하지 않고 그대로 복사해 붙여 넣는 사례가 빈번히 발생하는 등, 판단을 AI에 외주화하는 상황이 일반화되고 있다고 지적

명확한 목표와 활용 사례 없는 AI 사용 관행이 워크슬롭의 핵심 원인

■ 수십억 달러 규모의 생성형 AI 기업 투자가 지속되는 가운데, 인력 감축과 명확한 지침 없는 AI 사용 강제를 병행하는 경영 관행이 워크슬롭의 핵심 배경으로 지목

- 블록(Block), 아마존(Amazon), 다우(Dow), UPS, 핀터레스트(Pinterest), 타깃(Target) 등 주요 기업들은 AI를 활용한 생산성 향상 가능성을 근거로 인력 감축을 단행
- 그러나 MIT나 딜로이드 등 주요 기관의 연구에 따르면 AI 투자 기업의 상당수는 아직 가시적 투자 수익을 실현하지 못했으며, 상당수 기업은 2~4년 후 투자 수익이 개선될 것으로 기대

■ AI의 불명확한 활용 지침과 노동자 통제권 부재가 워크슬롭을 심화시키는 구조적 요인으로 지적되면서, 노동계는 단체교섭을 통한 AI 활용 기준 마련을 요구

- 비영리 연구소 데이터·사회 연구소(Data & Society) 소속의 아이하 응우옌(Aiha Nguyen)은 생성형 AI가 만능 범용 도구로 홍보되고 있으나, 불명확한 목표와 활용 사례가 워크슬롭을 초래한다고 분석
- 미국 통신노조(CWA) 연구 경제학자 댄 레이놀즈(Dan Reynolds)는 AI가 노사 협상의 쟁점으로 부상하고 있으며, 노동계는 명확한 AI 사용 기준과 노동자 통제권 확보를 원한다고 설명

출처 | The Guardian, Bosses say AI boosts productivity – workers say they're drowning in 'workslop', 2026.4.14.

SPRi 소프트웨어정책연구소
Software Policy & Research Institute

https://spri.kr/

보고서와 관련된 문의는 AI정책연구실(hjk_rep@spri.kr, 031-739-7326)로 연락주시기 바랍니다.

경기도 성남시 분당구 대왕판교로 712번길 22 글로벌 R&D 연구동(B) 4층
22, Daewangpangyo-ro 712beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do, Republic of Korea, 13488