

SPRi AI Brief

2025년
3월호

인공지능 산업의 최신 동향



Software
Policy & Research
Institute

CONTENTS

정책·법제

- 미국 국가AI자문위원회, 트럼프 행정부 대상 AI 정책 보고서 발간 2
- 파리 AI 행동 정상회의, AI의 공공성·지속가능성과 함께 규제 완화 논의 3
- EU 집행위원회, 경쟁력 강화 로드맵의 일환으로 AI 기가팩토리 구축 추진 4
- EU 집행위원회, 2025년 업무 프로그램에서 'AI 책임 지침' 철회 계획 발표 5
- 영국, AI안전연구소의 명칭 AI보안연구소로 변경하고 앤스로픽과 AI 협력 발표 6

기업·산업

- 구글, 성능 개선한 '제미나이 2.0' 제품군 확대 출시 8
- 오픈AI, 심층 조사를 수행하는 에이전트 '딥 리서치' 공개 9
- 오픈AI, GPT-5 출시 로드맵 발표 및 GPT-4.5 프리뷰 공개 10
- xAI, 최신 AI 모델 '그록 3' 프리뷰 출시 11
- 퍼플렉시티, 심층 조사와 분석을 수행하는 '딥 리서치' 무료 출시 12
- 알리바바, 딥시크 V3 능가하는 성능의 '큐원2.5-Max' 출시 13
- 아크 인베스트, 2025년 혁신 기술 중 하나로 AI 에이전트 선정 14

기술·연구

- 상하이교통대 연구진, 소량의 고품질 데이터를 활용한 추론 AI 모델 개발 16
- 스탠포드대와 워싱턴대 연구진, 저비용으로 고성능 추론 AI 모델 개발 17
- 바이트댄스, 인물 움직임을 생성하는 AI 모델 '옵니휴먼-1' 개발 18
- AI안전센터와 스케일 AI, 고난도 벤치마크 'HLE' 공개 19

인력·교육

- 딜로이트 조사 결과, 전 세계 기업들은 점진적으로 AI 도입 확대 21
- 마이크로소프트 연구 결과, 생성 AI 신뢰할수록 비판적 사고 감소 22
- 앤스로픽, AI가 노동시장에 미치는 영향을 분석한 '앤스로픽 경제 지수' 공개 23
- 앤스로픽을 비롯한 여러 기업들, 입사 지원 시 AI 도구 사용금지 요구 24

주요행사일정

25

정책·법제

미국 국가AI자문위원회, 트럼프 행정부 대상 AI 정책 보고서 발간

KEY Contents

- AI 정책에 관하여 대통령과 연방 정부에 자문을 제공하는 미국 국가AI자문위원회(NAIAC)가 트럼프 행정부가 우선적으로 추진해야 할 AI 정책을 정리한 보고서를 발표
- 보고서는 정책 우선순위로 △노동력 △AI 인식과 리터러시 △교육 △과학 △보건 △정부 △중소기업 지원 △AI 거버넌스 △미국 시민 △법 집행을 제시

○ 미국 국가AI자문위원회, 미국의 AI 주도권 유지를 위한 10대 정책 우선순위 제시

- 미국 국가AI자문위원회는 2025년 1월 28일 미국의 기술 주도권 유지를 위해 트럼프 행정부가 추진해야 할 AI 중점 분야를 제시한 보고서를 발표하고, 정책 우선순위로 △노동력 △AI 인식과 리터러시 △교육 △과학 △보건 △정부 △중소기업 지원 △AI 거버넌스 △미국 시민 △법 집행을 선정
- (노동력) AI가 노동시장에 미치는 영향에 대응해 연방 정부와 주·지방 정부 간 협력을 강화하고 AI를 비롯한 기술 개발 지원 등 AI로 실직 위험에 처한 근로자 지원 전략을 마련
- (AI 인식과 리터러시) AI의 광범위한 도입을 위해 전국 규모의 AI 인식 제고 캠페인을 시행하고 AI 기초 교육, 전문과정, AI 자격증 과정과 같은 교육 프로그램을 강화
- (교육) 교육 환경에 특화된 AI 위험관리 프레임워크를 개발하고 AI 교육 과정과 AI 도구 개발 해커톤을 개최
- (과학) 중요한 사회 문제 해결을 위해 AI 적용을 확대하고, 과학 분야의 AI 연구에 대한 자금 우선순위를 설정하며 지원을 강화하는 전략을 추진
- (보건) 백악관 과학기술위원회(NSTC) 산하 AI특별위원회 내에 개인 건강정보의 안전하고 책임 있는 사용을 담당할 소위원회를 신설
- (정부) AI로 정부의 운영 효율성을 향상하기 위해 국가AI이니셔티브 사무국에 인력과 자원을 충원하고 AI 모델 평가 프레임워크를 구축
- (중소기업 지원) 중소기업청(SBA) 주도로 중소기업과 비영리단체의 AI 도입을 지원할 태스크포스를 설립하고 AI 도입 보조금 및 AI 관련 연구개발과 직원 교육에 대한 세액 공제, 저금리 대출을 지원
- (AI 거버넌스) AI 모델의 안전 지침과 평가 방법을 개발하고, 국립표준기술연구소(NIST)와 AI안전연구소(AISI)에 자금 지원을 확대하며, AI 거버넌스에서 국제 협력을 강화
- (미국 시민) 파리 AI 정상회의에서 모든 미국인에게 이익이 되는 AI 우선순위를 발표하고, 과학기술정책국(OSTP)이 AI 우선순위에 관한 의견을 수렴해 국가 전략을 수립
- (법 집행) 개인 디지털 안전을 위협하는 AI 기반 리스크에 대응하기 위해 범정부 태스크포스를 구성하고, 위협 유형을 식별하며 해결책을 모색
- 보고서는 상기 10개 분야를 아우르는 종합 해결책 마련을 위한 5개 핵심 과제로 △미국의 AI 경쟁력과 혁신 강화 △정부의 AI 역량과 리터러시 제고 △전략적 우선순위(교육, 중소기업, 보건·과학 분야에서의 AI 활용) 추진 △민관 협력 강화 △AI 거버넌스에서 미국의 리더십 확립을 제시

파리 AI 행동 정상회의, AI의 공공성·지속가능성과 함께 규제 완화 논의

KEY Contents

- 파리 AI 행동 정상회의의 주요 결과물로 포용적이고 지속가능한 AI를 위한 선언문이 발표되었으며, 공익을 위한 AI 프로젝트 ‘커런트 AI’도 출범
- 파리 AI 정상회의에서는 AI 규제 완화와 투자 확대가 핵심 의제로 떠올랐으며, 마크롱 프랑스 대통령은 1,090억 유로 규모의 AI 인프라 민간 투자 프로젝트도 발표

○ 미국과 영국을 제외한 60개국, 포용적이고 지속가능한 AI에 대한 선언문 발표

- 프랑스 정부가 2025년 2월 10~11일, 파리에서 ‘AI 행동 정상회의(AI Action Summit)’를 개최
 - 이번 정상회의에는 전 세계 87개 국가에서 기업, 국제기구, 시민단체 등 총 1,000여 명이 참여해 AI 글로벌 거버넌스를 심도 있게 논의
- 정상회의 주요 결과물로 한국을 포함한 60개국이 공동 참여한 ‘인류와 지구를 위한 포용적이고 지속가능한 AI에 대한 선언문’이 발표되었고 미국과 영국은 선언에 불참
 - 선언문은 공익을 위한 AI 및 지속가능한 AI를 목표로 제시하고, 이를 달성하기 위한 구체적 행동으로
①공익을 위한 AI 플랫폼 및 인큐베이터 출범 ②환경적 지속가능성을 위한 AI 에너지 관측소 설립
③일자리에 대한 AI 영향 관측 네트워크를 제시
- 이번 정상회의에서는 9개국*과 구글, 세일즈포스 등의 기업이 4억 달러를 투자해 공익을 위한 AI 프로젝트 ‘커런트 AI(Current AI)’도 출범
 - * 나이지리아, 독일, 모로코, 스위스, 슬로베니아, 칠레, 케냐, 프랑스, 핀란드
- 이 프로젝트는 공익 AI 환경 조성을 목표로 △AI 훈련을 위한 고품질 공공 데이터 접근성 확대 △오픈소스 인프라 지원 △AI의 사회적·환경적 영향을 측정하기 위한 시스템 개발 지원을 추진

○ 파리 AI 정상회의, AI 규제 완화 및 AI 투자 확대가 핵심 화두로 부상

- 파리 AI 정상회의에서는 AI 규제 완화가 핵심 화두로 부상했으며, 마크롱 프랑스 대통령은 EU에 유럽이 AI 경쟁에서 승기를 잡기 위한 규제 간소화를 촉구
 - 이에 대하여 EU 디지털 총괄 헨나 비르크넨(Henna Virkkunen) EU 집행위원은 복잡한 절차와 행정 부담을 줄이기 위해 규제를 단순화하고, 기업 친화적 방향으로 시행할 계획이라고 설명
- 마크롱 프랑스 대통령은 이번 정상회의에서 총 1,090억 유로(약 163조 원) 규모의 AI 인프라 건설을 위한 민간 투자 계획도 발표
 - 프랑스는 유럽 최대 규모의 데이터센터 구축을 위해 UAE와 최대 500억 유로 규모의 투자에 합의했으며, 캐나다 투자기업 브룩필드(Brookfield)도 200억 유로를 투자 계획을 발표 하기로 약속

EU 집행위원회, 경쟁력 강화 로드맵의 일환으로 AI 기가팩토리 구축 추진

KEY Contents

- EU 집행위원회가 5개년 정책 로드맵 ‘경쟁력 나침반’에 따라 혁신 격차 해소를 위한 AI 정책으로 ‘AI 기가팩토리’와 ‘AI 적용’ 전략을 제시
- EU 집행위원회는 파리 AI 정상회의에서 총 2천억 유로 규모의 ‘인베스트AI’ 계획의 일환으로 200억 유로를 투입해 4개의 AI 기가팩토리를 건설하겠다고 발표

● EU 집행위원회, 5개년 정책 로드맵 하에서 AI 기가팩토리와 AI 적용 전략 추진

- 우르줄라 폰 데어 라이엔(Ursula von der Leyen) EU 집행위원장이 2025년 1월 29일 2기 EU 집행부(2024년 12월 출범)의 5개년 정책 로드맵 ‘경쟁력 나침반(Competitive Compass)’을 발표
 - EU 집행위원회는 경쟁력 제고를 위해 △혁신 격차 해소 △탄소화 △공급망 안보의 3개 영역을 중점 과제로 제시했으며, 이중 혁신 격차 해소와 관련해 AI 정책을 포함
- EU 집행위원회는 핵심 분야의 AI 개발과 산업계 AI 도입 활성화를 위한 ‘AI 기가팩토리(AI Gigafactory)’와 ‘AI 적용(Apply AI)’ 전략을 제안
 - AI 기가팩토리는 입법 추진 예정인 ‘EU 클라우드 및 AI 개발법(EU Cloud and AI Development Act)*’을 통해 공공과 민간 자금을 활용하여 초거대 AI 모델 훈련에 특화된 대규모 데이터센터를 구축함으로써 EU 전역의 AI 생태계를 활성화한다는 계획
 - * 고성능 연산 자원과 디지털 인프라에 대한 강력한 규제 프레임워크를 통해 클라우드와 AI 분야에서 유럽의 리더십 강화를 위한 법안
 - AI 적용 전략은 제조업, 에너지, 자동차, 로봇공학, 제약, 항공, 금융 서비스 산업의 AI 적용을 촉진하고 보건과 사법 등의 공공서비스 개선을 추진
 - EU 집행위원회는 AI 개발의 필수 요소인 대량의 고품질 데이터를 확보하기 위해 개인 및 공공 데이터 공유를 개선하고 규제를 간소화하는 한편, 신규 시스템이나 앱 개발을 촉진하는 데이터 전략도 추진할 계획

● EU 집행위원회, 파리 AI 정상회의에서 AI 기가팩토리 투자 계획 발표

- 폰 데어 라이엔 EU 집행위원장은 파리 AI 정상회의에서 AI 기가팩토리 구축을 위한 200억 유로의 기금을 포함한 2천억 유로 규모의 AI 투자 계획을 발표
 - EU 집행위원회는 ‘인베스트AI(InvestAI)’라는 명칭의 투자 계획에 총 2천억 유로의 민간과 공공 자본을 동원하고, EU 전역에 4개의 AI 기가팩토리를 건설할 계획
 - 의학, 과학 등 특정 분야에서 획기적 발전을 촉진하기 위해 설계된 AI 기가팩토리는 기존 AI 팩토리보다 약 4배 더 많은 10만 개의 최신 GPU를 탑재할 예정
 - 한편, EU 집행위원회는 2024년 12월 스타트업의 AI 모델 훈련을 지원하기 위해 슈퍼컴퓨터를 제공하는 7개 AI 팩토리 설립 프로젝트를 발표했으며, 향후 5개를 추가 선정할 계획

EU 집행위원회, 2025년 업무 프로그램에서 'AI 책임 지침' 철회 계획 발표

KEY Contents

- EU 집행위원회가 2025년 업무 프로그램에서 규제 간소화를 중점 추진하며 AI 시스템으로 인한 피해자를 보호하기 위해 마련된 'AI 책임 지침'의 철회를 예고
- 미국의 JD 밴스 부통령이 파리 AI 정상회의에서 EU의 규제 완화를 요구한 직후 발표된 이번 결정에 대해 유럽의회 의원들은 부정적인 입장을 밝혔으나, 산업계는 환영

○ EU 집행위원회, 규제 간소화에 주력하며 AI 책임 지침 철회 예고

- EU 집행위원회가 2025년 2월 12일 발표한 업무 프로그램(Work Programme)에서 'AI 책임 지침(AI Liability Directive, AILD)'의 철회 계획을 포함
 - 2025년 업무 프로그램은 EU의 경쟁력 강화를 위한 규제 간소화에 초점을 맞추고 있으며, AI 책임 지침을 포함해 총 37개 규제안을 철회 대상으로 분류
 - EU 집행위원회가 2022년 처음 제안한 AI 책임 지침은 AI 시스템으로 인한 피해자를 보호하기 위해 민사상 손해배상을 주장하는 청구인의 입증 부담을 완화하는데 중점
 - AI 책임 지침의 철회 배경으로, 지침에 대한 합의 도출이 어려울 것으로 예상되어 다른 접근방식을 평가할 계획이라고 설명

○ 미국의 규제 완화 압박 이후 발표된 AI 책임 지침의 철회 방침, 각계의 엇갈린 반응 야기

- EU 집행위원회는 파리 AI 정상회의에서 미국의 JD 밴스(Vance) 부통령이 유럽의 엄격한 AI 규제를 비판한 직후, AI 책임 지침의 철회 계획을 발표
 - 밴스 부통령은 글로벌 영향력을 가진 미국 기술 기업에 대한 규제를 강화하려는 일부 외국 정부의 움직임을 우려하며, 미국 정부는 성장 친화적 AI 정책을 장려하기 위해 모든 노력을 기울일 것이라고 강조
 - 폰 데어 라이엔 EU 집행위원장은 AI 법이 EU 전역에 걸쳐 단일한 안전 규칙을 제공함으로써 규정 준수를 간소화한다고 강조하면서도, AI 시장 성장을 위해 불필요한 절차를 줄이겠다고 발언
- 유럽의회의 악셀 보스(Axel Boss) 의원은 AI 책임 지침에 대한 철회 결정을 비판하면서, EU 집행위원회가 미국 빅테크의 압박에 굴복해 유럽 기업과 소비자를 위험에 처하게 했다고 주장
 - 그는 EU AI 법이 피해 예방에 중점을 두는 반면, AI 책임 지침은 피해 발생 시 적용되는 '사후 책임 메커니즘'이므로, AI 법과 중복되어 불필요한 규제를 만들 것이라는 우려는 근거가 없다고 반박
- 그러나 산업계를 대표하는 유럽 컴퓨터·통신산업협회(CCIA Europe)는 지침의 철회를 환영하면서, EU 집행위원회에 기존 규칙의 간소화에 집중할 것을 촉구

영국, AI안전연구소의 명칭 AI보안연구소로 변경하고 앤스로픽과 AI 협력 발표

KEY Contents

- 영국 AI안전연구소가 AI보안연구소로 명칭을 변경하고, 정부 기관들과 협력하여 AI 관련 국가 안보 및 범죄 위험 대응에 중점을 둘 예정
- 영국 과학혁신기술부는 앤스로픽과 MOU를 체결하고 첨단 AI 도구를 활용한 공공서비스 개선 및 AI 보안 위험에 대응한 협업을 추진할 계획

○ 영국 AI보안연구소, AI 관련 국가안보와 범죄 위험 대응에 주력 예정

- 영국 AI안전연구소(AI Safety Institute)가 2025년 2월 14일 영국 AI보안연구소(UK AI Security Institute)로 명칭을 변경하고 AI 관련 국가안보와 범죄 위험 대응에 주력할 방침
 - 이는 경제성장을 위한 영국 정부의 장기 전략인 '변화 계획(Plan for Change)'의 일환이며, 국가안보가 변화 계획의 핵심 축을 형성
 - 영국 AI보안연구소는 향후 화학·생물학 무기 개발이나 사이버 공격, 사기, 아동 성착취 범죄와 같은 심각한 AI 위험 대응에 중점을 둘 예정
- 영국 AI보안연구소는 국방부(MOD) 산하 국방과학기술연구소(DSTL)를 포함한 정부 기관들과 협력하여 첨단 AI가 초래하는 위험을 평가할 계획
 - 영국 시민에게 피해를 줄 수 있는 다양한 범죄 및 보안 문제의 연구를 수행하기 위해 내무부와 함께 범죄 오용 대응팀도 출범할 예정으로, 특히 AI를 사용한 아동 성착취물 제작 방지에 주력할 계획
 - AI보안연구소는 AI 기술이 초래하는 심각한 위험에 대한 이해를 증진하는 데 초점을 맞추고, AI보안연구실(LASR), 국가사이버보안센터(NCSC) 등과도 협력할 예정

○ 영국 과학혁신기술부, 앤스로픽과 AI를 통한 공공서비스 개선을 위한 양해각서 체결

- 한편, 영국 과학혁신기술부(DSIT)는 2025년 2월 14일 앤스로픽(Anthropic)과 AI를 활용한 공공서비스 개선을 주요 목표로 하는 양해각서(MOU)를 체결
 - 영국 정부와 앤스로픽은 AI의 혁신적 잠재력에 대한 공감대를 형성하고, 공공 부문에서 첨단 AI 도구를 활용하여 시민 경험을 개선할 계획으로, 이 과정에서 대중의 수요 대응 및 개인정보보호와 신뢰 구축을 핵심 원칙으로 제시
 - 앤스로픽은 영국 AI보안연구소와 AI 시스템의 기능과 보안 위험을 파악하고, 이러한 위험을 완화하여 AI가 제공하는 기회를 최대한 활용하기 위한 협업을 지속할 예정
- DSIT와 앤스로픽은 아래의 공통 관심 분야에서 추가 협업을 통해 파트너십을 강화할 계획
 - △앤스로픽의 역량 및 영국 정부의 R&D와 데이터 자산 강점을 활용한 AI 기반 과학적 진보 가속화 △첨단 AI 및 미래 인프라에 대응한 공급망 확보 △앤스로픽의 첨단 시스템과 도구를 활용한 혁신 조직(스타트업 커뮤니티, 대학 등) 지원 △실제 AI 모델 사용 데이터 기반의 상황 인식 지원

출처 | DSIT, Tackling AI security risks to unleash growth and deliver Plan for Change, 2025.02.14.

DSIT, Memorandum of Understanding between UK and Anthropic on AI opportunities, 2025.02.14.

기업·산업

구글, 성능 개선한 ‘제미나이 2.0’ 제품군 확대 출시

KEY Contents

- 구글이 ‘제미나이 2.0 플래시’를 모든 제미나이 앱 사용자에게 제공하는 동시에, 코딩과 복잡한 프롬프트 처리에 뛰어난 ‘제미나이 2.0 프로’ 실험 버전을 출시
- 구글은 비용 효율적인 신규 모델 ‘제미나이 2.0 플래시-라이트’도 프리뷰로 공개했으며, ‘제미나이 2.0 플래시 씽킹’ 실험 버전도 제미나이 웹과 모바일 앱으로 공개

● 구글, 코딩에 특화된 제미나이 2.0 프로 실험 버전을 포함한 제미나이 2.0 제품군 공개

- 미국 구글(Google)이 2024년 12월 공개한 추론 AI 모델 ‘제미나이 2.0 플래시(Gemini 2.0 Flash)’를 모든 사용자에게 확대 출시하고, 코딩과 복잡한 프롬프트 처리에 강한 최고 사양의 모델 ‘제미나이 2.0 프로(Gemini 2.0 Pro)’ 실험 버전을 공개
 - 제미나이 2.0 플래시는 100만 개 토큰 길이의 컨텍스트 창과 멀티모달 추론을 지원하며, 이미지 생성 및 텍스트-음성 변환 기능은 향후 추가될 예정
 - 제미나이 2.0 프로 실험 버전은 구글 AI 모델 중 가장 강력한 코딩 성능과 복잡한 프롬프트 처리 능력을 갖추고 200만 개 토큰 길이의 최대 컨텍스트 창을 지원하며 구글 검색 및 코드 실행과 같은 도구 호출을 통해 에이전트의 기반이 되는 기능도 구현
- 또한 구글은 비용 효율성이 가장 뛰어난 ‘제미나이 2.0 플래시 라이트(Flash-Lite)’ 모델도 구글 버텍스 AI*와 AI 스튜디오**에서 프리뷰로 공개
 - * 머신러닝 모델 구축, 배포, 확장을 위한 도구와 서비스를 제공하는 종합 플랫폼 ** 버텍스 AI 플랫폼 내 생성 AI 개발 도구
 - 제미나이 2.0 플래시 라이트는 이전 버전인 제미나이 1.5 플래시와 동일한 속도와 비용으로 더 높은 성능을 제공하며, 100만 개 토큰 길이의 컨텍스트 창과 멀티모달 입력을 지원
- 구글은 추론에 특화된 ‘제미나이 2.0 플래시 씽킹(Flash Thinking)’ 실험 버전도 제미나이 웹과 모바일 앱을 통해 출시
 - 제미나이 2.0 플래시 씽킹 실험 버전은 2.0 플래시를 기반으로 프롬프트를 단계별로 나누어 추론 기능을 강화하고 응답을 개선하도록 훈련된 모델로, 사고 과정을 확인해 특정 방식으로 응답한 이유와 추론 경로 등을 추적 가능
 - 제미나이 2.0 플래시 씽킹 실험 버전은 유튜브와 검색, 구글 맵스와 같은 앱과 연동되어 추론 기능 적용 시의 활용 가능성을 탐색
- 구글은 제미나이 모델의 안전성을 지속적으로 개선하기 위해 투자하고 있으며, 자체적으로 응답을 평가하는 새로운 강화학습 기법을 적용해 제미나이 2.0 시리즈를 개발함으로써 더욱 정확하고 구체적인 피드백을 제공할 수 있으며 민감한 프롬프트 처리 능력도 향상되었다고 주장

출처 | Google, Gemini 2.0 is now available to everyone, 2025.02.05.

Google, Access the latest 2.0 experimental models in the Gemini app, 2025.02.05.

오픈AI, 심층 조사를 수행하는 에이전트 ‘딥 리서치’ 공개

KEY Contents

- 오픈AI가 온라인 정보의 검색과 분석, 종합을 통해 전문가 수준의 종합 보고서를 작성하는 에이전트 ‘딥 리서치’를 챗GPT 프로 사용자를 대상으로 출시
- 딥리서치는 HLE 벤치마크에서 26.6%의 정확도로 주요 모델 중 최고점을 기록했으나, 아직 초기 단계로 환각 증상이나 수식 오류 등의 문제점도 존재

○ 오픈AI 딥 리서치, 온라인 정보 검색 및 분석을 통해 연구 분석가 수준의 보고서 작성

- 오픈AI(OpenAI)가 2024년 2월 2일 추론을 통해 대규모 온라인 정보를 종합하고 다단계의 복잡한 연구를 수행하는 에이전트 기능 ‘딥 리서치(Deep Research)’를 출시
 - 딥 리서치는 프롬프트에 따라 수백 개의 온라인 소스의 검색과 분석, 종합을 통해 연구 분석가 수준의 종합 보고서를 작성하며, 사람이 하면 여러 시간이 걸리는 작업을 수십 분 만에 완료
 - 딥 리서치는 웹 브라우징과 데이터 분석에 최적화된 별도의 o3 버전을 기반으로 인터넷에서 방대한 양의 텍스트와 이미지, PDF를 검색 및 분석
 - 이 기능은 월 200달러 구독료의 ‘챗GPT 프로’ 사용자에게 우선 출시 후, 2025년 2월 25일부터 월 20달러를 내는 챗GPT 플러스를 포함한 모든 유료 구독자에게 확대 제공
 - 딥 리서치는 작업 완료에 5~30분이 걸릴 수 있어 작업 완료 시 알림을 제공하며, 현재 텍스트로만 보고서를 작성하나 추후 이미지와 그래프 출력 기능을 추가 예정
- 오픈 AI는 딥 리서치 기능이 최근 개발된 고난도 벤치마크 HLE(Humanity’s Last Exam)* 평가에서 26.6%의 정확도를 보여 주요 모델** 중 최고점을 달성했다고 보고
 - * 언어학, 로켓 과학, 고전, 생태학 등 100개 이상 주제에 걸친 3,000개 이상의 객관식 및 단답식 문제로 구성
 - ** 주요 모델 비교: GPT-4o(3.3%), Claude 3.5 Sonnet(4.3%), Gemini Thinking(6.2%), DeepSeek-R1(9.4%)
 - 딥 리서치는 9.1%의 정확도를 기록한 o1과 비교해 특히 화학, 인문학, 사회과학, 수학에서 두각을 나타냈으며, 필요시 전문 정보를 효과적으로 찾아내는 등 인간과 비슷한 접근방식을 활용
 - 오픈AI 내부에서 다양한 분야에 걸친 전문가 수준의 작업을 평가한 결과, 각 영역의 전문가들은 오랜 시간이 걸리는 까다로운 조사를 딥 리서치를 통해 자동화했다고 평가

○ 초기 단계의 딥 리서치, 환각 증상이나 수식 오류 등 한계도 존재

- 오픈AI에 따르면 딥 리서치는 아직 초기 단계로 내부 평가에서 환각(Hallucination) 증상을 보이거나 잘못된 추론을 하는 경우도 발생
- 오픈AI는 출시 초기에는 보고서와 인용에 사소한 수식 오류가 발생하거나 작업 시작에 오랜 시간이 걸릴 수 있다고 설명하며, 이러한 문제는 사용이 늘어나며 빠르게 개선될 것으로 예상

오픈AI, GPT-5 출시 로드맵 발표 및 GPT-4.5 프리뷰 공개

KEY Contents

- 오픈AI가 o3를 독립형 모델로 출시하지 않는 대신, 올해 안에 o3를 포함한 오픈AI 모델군 전반의 기능을 통합한 GPT-5를 출시할 계획
- 오픈AI는 GPT-4.5를 프리뷰 버전으로 챗GPT 프로 사용자에게 우선 공개했으며, 비지도 학습을 확장해 더욱 자연스러운 상호작용을 지원하고 환각 증상도 덜할 것이라고 설명

○ 오픈AI, 2025년 내 o3와 통합된 GPT-5 출시 계획

- 샘 알트먼(Sam Altman) 오픈AI CEO는 2025년 2월 13일 자신의 X 계정*을 통해 GPT-4.5와 GPT-5 출시 로드맵을 공개

* <https://x.com/sama/status/1889755723078443244>

- 알트먼 CEO는 내부적으로 'Orion'이라고 불리는 GPT-4.5가 오픈AI의 마지막 非추론 모델로서, 올해 안에 출시될 GPT-5는 o3를 포함해 오픈AI 모델군 전반의 기능을 통합한 시스템이 될 것이며, o3는 독립형 모델로 출시하지 않을 계획이라고 설명
- 알트먼 CEO는 현재 오픈AI의 모델군이 지나치게 복잡해졌으며, “모든 도구를 활용하고 언제 오랫동안 생각해야 하는지 아닌지 알 수 있으며, 광범위한 작업에 유용한 시스템을 구축해 o-시리즈와 GPT-시리즈를 통합하는 것”이 가장 중요한 목표라고 언급
- 그는 무료 사용자도 ‘표준 인텔리전스(Standard Intelligence)’ 설정에서 GPT-5를 무제한으로 사용할 수 있으며, 챗GPT 플러스(월 20달러) 가입자는 더 높은 인텔리전스, 챗GPT 프로 가입자는 훨씬 더 높은 인텔리전스에서 실행할 수 있게 될 것이라고 밝혔으나, 인텔리전스의 세부 내용은 미공개

○ 오픈AI, 챗GPT 프로 사용자에게 GPT-4.5 프리뷰 공개

- 오픈AI는 2025년 2월 27일 GPT-4.5를 월 200달러를 내는 챗GPT 프로 사용자들에게 프리뷰로 우선 공개했으며, 다른 유료 사용자들에게도 순차적으로 출시 계획
- 오픈AI는 GPT-4.5가 채팅을 위한 최대·최고의 모델이라며, 사람의 감독 없이 데이터를 통해 학습하는 비지도 학습을 확장함으로써 패턴을 인식하고 연관성을 도출하며 창의적인 통찰력을 끌어내는 능력이 강화되었다고 주장
- 또한 상호작용이 더 자연스러워지고 사용자 의도를 따르는 능력이 향상되었으며, 감성지능(EQ)이 향상되어 글쓰기, 프로그래밍, 실제 문제해결과 같은 작업에 유용하고 환각(Hallucination) 증상도 덜할 것으로 예상된다고 설명
- 한편, GPT-4.5는 과학, 수학, 다국어, 멀티모달 등의 벤치마크 평가에서 GPT-4o를 능가했으나, o3-mini(high) 대비 낮은 성능을 기록

xAI, 최신 AI 모델 ‘그록 3’ 프리뷰 출시

KEY Contents

- xAI가 별도의 ‘Think’ 버튼을 통해 추론 기능을 활성화할 수 있도록 설계된 그록 3와 그록 3 미니를 프리뷰로 공개했으며, 그록 3(Think)는 주요 벤치마크에서 최고 수준의 성능을 기록
- xAI는 인터넷과 X의 데이터를 검색 및 분석해 보고서를 작성하는 에이전트 기능 딥서치도 공개했으며, 추론 기능과 딥서치 기능은 X 프리미엄 플러스 사용자를 대상으로 제공

○ xAI, 추론 기능 및 딥서치 기능을 지원하는 ‘그록 3’ 베타 버전 공개

- 일론 머스크(Elon Musk) 테슬라 CEO가 설립한 AI 스타트업 xAI가 2025년 2월 19일 최신 모델 ‘그록 3(Grok 3)’ 베타 버전을 프리뷰로 공개
 - 그록 3는 기본 모델과 비용 효율적인 ‘그록 3 미니’로 구성되며, 두 모델 모두 별도의 ‘Think’ 버튼을 통한 추론 기능 활성화를 지원
 - xAI는 AIME2025* 평가에서 그록 3(Think)가 93.3%, 그록 3 미니(Think)는 90.8%로 오픈AI o3-미니-하이(86.5%), DeepSeek-R1(70.0%)을 능가하는 최고 점수를 기록했다고 보고
 - * 2025년 미국 수학 올림피아드 예선 문제
 - 그록 3(Think)는 최종 응답뿐 아니라 모델 자체의 추론 과정에 대한 검증도 지원하며, 코딩, 퍼즐, 수학 등 다양한 문제 영역에서 고른 성능을 발휘
- 그록 3는 응답에 수 분이 걸리는 추론 기능을 사용하지 않을 시 고품질의 응답을 즉각 제공하며, 이미지와 동영상 이해에서도 뛰어난 성능을 발휘
 - xAI는 AIME2024 평가에서 그록 3가 52.2%, 그록 3 미니가 39.7%로 답시크-V3(39.2%)와 클로드 3.5 소네트(16.0%), GPT-4o(9.3%)를 앞섰으며, 그록 3 초기 버전(코드명 chocolate)은 실제 사용자들의 선호도를 나타내는 ‘챗봇 아레나(Chatbot Arena)’ 리더보드에서 1위를 차지했다고 발표
- 그록 3는 그록 2보다 10배 많은 연산 자원으로 훈련되었으며, 그록 2 대비 8배 증가한 100만 개 토큰의 컨텍스트 창을 지원
 - xAI는 총 20만 개의 GPU가 탑재된 슈퍼컴퓨터 ‘콜로사스(Colossus)’에서 그록 3의 훈련을 지속하고 있으며, 향후 몇 달간 모델을 지속 업데이트 예정
- xAI는 인터넷과 X의 데이터를 실시간으로 검색해 분석하고 보고서를 작성하는 에이전트 기능 ‘딥서치(DeepSearch)’도 공개
 - xAI는 딥서치가 주요 정보를 종합하고, 상충하는 사실과 의견에 대해 추론하며, 복잡한 내용을 명확하게 풀어내도록 설계되었다고 설명
- 그록 3는 X 및 Grok.com에서 X 프리미엄(월 8달러) 및 프리미엄 플러스(월 40달러*) 사용자에게 제공되며, 추론 기능과 딥서치 기능은 프리미엄 플러스 사용자에게만 지원

* X는 그록 3 출시와 함께 2025년 2월 18일부터 프리미엄 플러스 요금제 가격을 기존 월 22달러에서 월 40달러로 인상

퍼플렉시티, 심층 조사와 분석을 수행하는 ‘딥 리서치’ 무료 출시

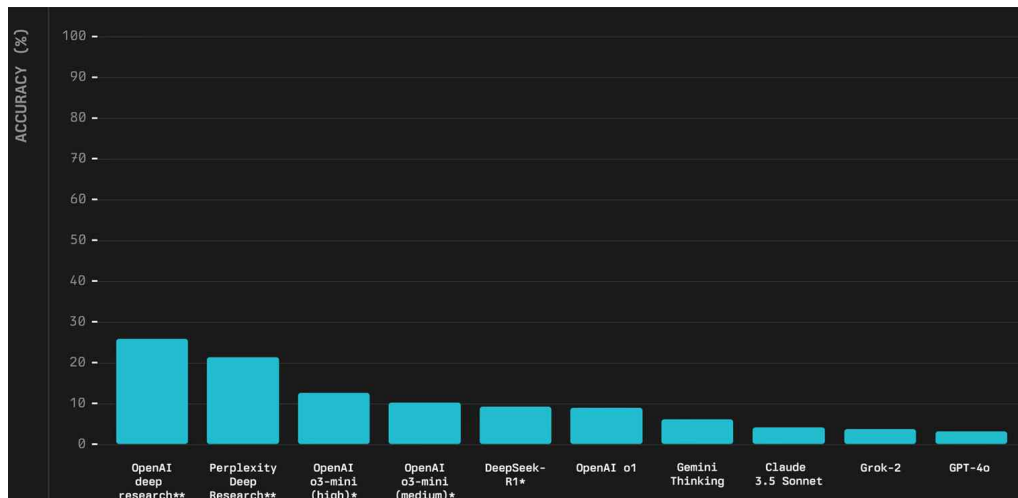
KEY Contents

- 퍼플렉시티가 오픈AI의 딥 리서치와 유사한 심층 조사 및 분석 기능인 ‘딥 리서치’를 출시하고, 무료 사용자에게도 하루 5개의 쿼리를 지원한다고 발표
- 퍼플렉시티 딥 리서치는 대부분의 작업을 3분 이내에 완료하며, HLE 벤치마크에서 21.1%의 정확도로 오픈AI의 딥 리서치(26.6%)에 이어 두 번째로 높은 점수를 기록

○ 퍼플렉시티 딥 리서치, 무료 사용자에게도 하루 5개의 질문(쿼리) 제공

- 퍼플렉시티(Perplexity)가 2025년 2월 14일 심층 조사 및 분석을 수행하는 AI 에이전트 기능 ‘딥 리서치(Deep Research)’를 무료로 출시
 - 퍼플렉시티는 사용자가 딥 리서치에 질문하면 수십 번의 검색을 통해 수백 개의 자료를 읽고 추론해 종합적인 보고서를 자율적으로 제공한다고 설명
 - 오픈AI가 유료 구독자에게만 딥 리서치 기능을 제공하는 반면, 퍼플렉시티는 모든 사용자에게 이를 무료로 제공하며, 월 20달러의 프로(Pro) 가입자는 무제한 사용이 가능하고, 무료 사용자는 하루 5개까지 쿼리를 실행 가능
 - 딥 리서치 기능은 현재 웹(perplexity.ai)에서 사용할 수 있으며, 향후 iOS와 안드로이드(Android) 및 맥(Mac)에서 출시 예정
- 퍼플렉시티에 따르면 딥 리서치는 재무, 마케팅, 제품 조사 등 다양한 전문가 수준의 작업에 탁월하며, 인간 전문가가 수 시간 걸리는 작업을 대부분 3분 이내에 완료
- 퍼플렉시티의 딥 리서치는 HLE 벤치마크에서 21.1%의 정확도를 달성해 오픈AI 딥 리서치(26.6%)에 이어 두 번째로 높은 점수를 기록

〈퍼플렉시티 딥 리서치와 주요 AI 모델과의 HLE 정확도 비교〉



알리바바, 딥시크 V3 능가하는 성능의 '큐원2.5-Max' 출시

KEY Contents

- 알리바바가 20조 개 이상의 토큰으로 사전 학습한 전문가 혼합 모델 '큐원2.5-맥스'를 출시하고, 일부 벤치마크 평가에서 오픈AI GPT-4o와 딥시크 V3의 성능을 능가한다고 주장
- 알리바바에 따르면 큐원2.5-맥스는 추론 능력을 평가하는 Arena-Hard와 최신 데이터 기반 LLM 평가 지표인 LiveBench에서 주요 AI 모델보다 높은 점수를 기록

○ 큐원2.5-맥스, 주요 벤치마크에서 오픈AI GPT-4o와 딥시크 V3 능가

- 중국 알리바바(Alibaba)가 2025년 1월 28일 일부 벤치마크 평가에서 오픈AI GPT-4o와 딥시크 V3를 능가하는 성능을 갖춘 신규 모델 '큐원2.5-맥스(Qwen 2.5-Max)'를 출시

- 큐원2.5-맥스는 20조 개가 넘는 토큰으로 사전학습을 거친 전문가혼합(MoE)* 모델로, 알리바바는 추론 중 연산 비용을 크게 늘리지 않고도 확장성을 제공해 챗봇부터 코딩 등 다양한 애플리케이션에 큐원2.5-맥스 활용이 적합하다고 설명

* Mixture-of-Experts: 특정 작업이나 영역에 특화된 여러 개의 전문가 모델을 혼합하여 활용하는 방식

- 알리바바가 벤치마크 평가에서 큐원2.5-맥스의 성능을 딥시크 V3와 라마 3.1-405B, GPT-4o, 클로드 3.5 소네트와 비교한 결과 큐원2.5-맥스는 Arena-Hard*와 LiveBench** 두 개 분야에서 1위를 차지

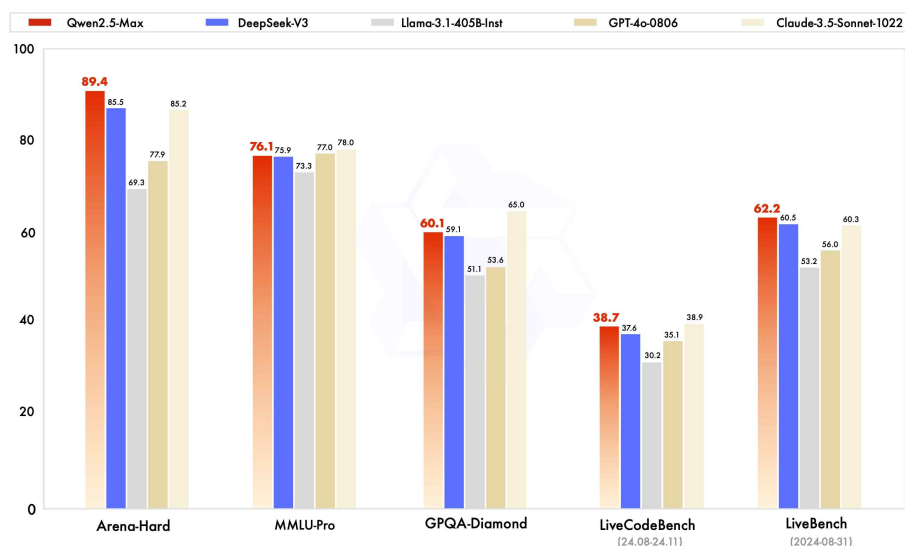
* 복잡한 수학 문제, 프로그래밍 과제 등으로 AI 모델의 심층적 추론 능력과 문제 해결 능력을 평가

** 최신 데이터를 기반으로 수학, 코딩, 추론, 언어이해 등 다양한 영역에서 AI 모델의 처리 능력과 실시간 대응 능력을 평가

- 큐원2.5-맥스는 MMLU-Pro*와 코딩 능력을 평가하는 LiveCodeBench에서 GPT-4o와 클로드 3.5 소네트와 유사한 성능을 보였으며, 5개 벤치마크 전반에서 딥시크 V3보다 높은 점수를 달성

* 광범위한 지식과 추론 능력을 평가하는 MMLU를 개선해 문제 난도를 높이고 추론 중심 문제를 확대한 벤치마크

〈큐원2.5-맥스와 경쟁 AI 모델의 벤치마크 점수 비교〉



아크 인베스트, 2025년 혁신 기술 중 하나로 AI 에이전트 선정

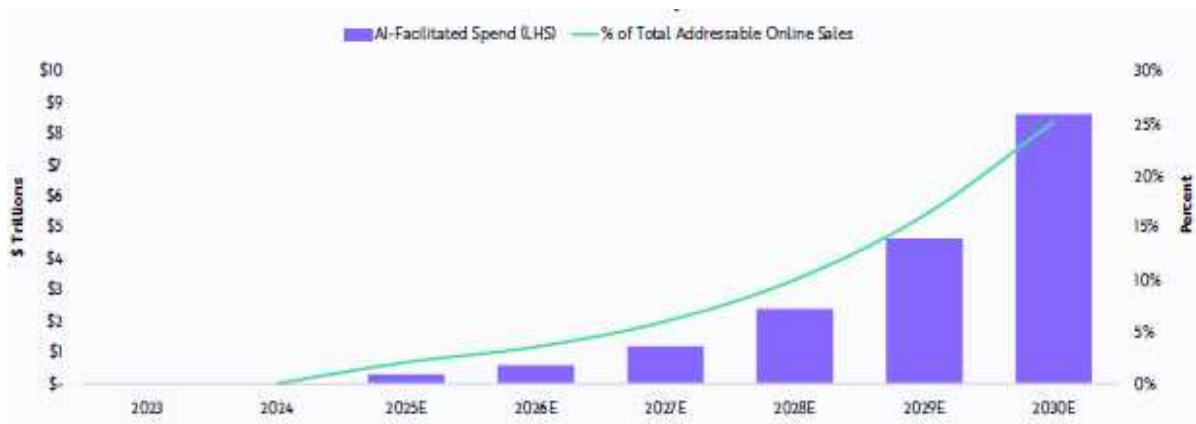
KEY Contents

- 아크 인베스트에 따르면 빠르게 대중화되는 AI 에이전트가 상품 검색과 정보 탐색을 수행하며, AI 기반 광고와 쇼핑 시장이 급성장할 전망
- 2030년에는 AI 기반 쇼핑이 글로벌 온라인 소비의 25%에 해당하는 9조 달러로 성장하며, AI 에이전트 기반 자동 결제 수단으로 디지털 지갑이 부상할 전망

○ AI 에이전트의 보급으로 AI 기반 광고와 전자상거래 시장 확대 전망

- 미국 투자운용사 아크 인베스트(ARK Invest)는 혁신 기술이 경제와 산업 전반에 미치는 영향을 분석한 '빅 아이디어 2025(Big Idea 2025)' 보고서에서 AI 에이전트를 11개 혁신 기술 중 하나로 선정
- 2025년부터 출시되는 대부분 소비자 하드웨어가 AI 기능을 기본으로 탑재하면서 빠르게 대중화되는 AI 에이전트는 소비자 검색과 발견을 자동화하며 AI 광고 및 전자상거래 시장의 성장을 이끌 전망
- 보고서에 따르면 소비자들이 시간 절약을 위해 기기에 내장된 AI 에이전트에 상품 검색과 정보 탐색을 맡기면서 검색과 발견이 자동화될 전망
- 검색이 개인화된 AI 에이전트를 통해 이루어지면서 새로운 AI 광고 시장이 열릴 전망으로, 보고서는 2030년까지 디지털 광고 시장 규모가 1조 1,000억 달러에 달하고 이 중 54% 이상을 AI가 중개하는 광고가 차지할 것으로 예측
- 소비자들이 제품 검색과 구매 과정을 AI 에이전트에 위임하면서 2030년에는 AI 기반 쇼핑이 글로벌 온라인 소비의 25%에 해당하는 9조 달러에 달할 전망

〈글로벌 온라인 소비에서 AI 기반 쇼핑이 차지하는 비중〉



- AI 에이전트를 활용한 온라인 쇼핑 활성화에 따라 신용카드나 직불카드 대신 디지털 지갑(Digital Wallet)이 자동 결제 수단으로 부상하며 2030년에는 전체 전자상거래의 72%를 차지할 전망
- 보고서는 2030년까지 AI 쇼핑 에이전트의 확산으로 디지털 지갑 플랫폼에서 최소 400억 달러에서 최대 2,000억 달러의 글로벌 매출이 창출될 것으로 예상

기술·연구

상하이교통大 연구진, 소량의 고품질 데이터를 활용한 추론 AI 모델 개발

KEY Contents

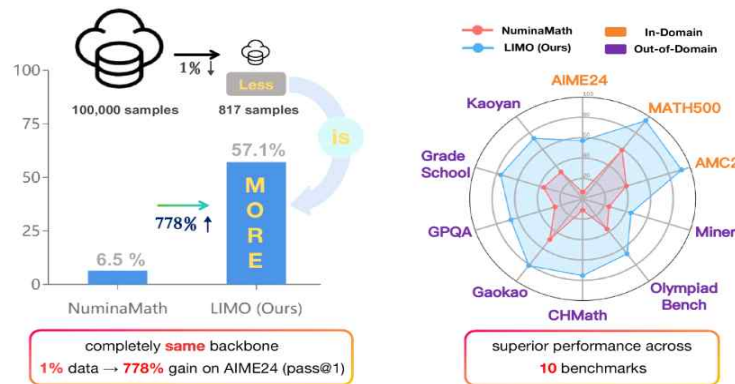
- 상하이교통大 연구진이 817개의 훈련 샘플만으로 주요 벤치마크에서 오픈AI o1-preview를 능가하는 성능의 추론 AI 모델을 개발
- 연구진에 따르면 사전학습 과정에서 이미 방대한 수학적 지식을 습득한 LLM은 추론에 특화된 소량의 학습 데이터를 통해 추론 능력을 활성화

817개의 고품질 데이터로 훈련된 LIMO 모델, 뛰어난 추론 성능 발휘

- 중국 상하이교통大(SJTU) 연구진이 2025년 2월 5일 소규모의 엄선된 학습 데이터만으로도 추론 AI 모델을 효과적으로 훈련할 수 있다는 연구 결과를 공개
- 정교한 추론 능력을 갖추려면 10만 개 이상의 방대한 학습 데이터를 활용한 미세조정이 필요하다는 통념과 달리, 연구진이 제안한 LIMO(Less is More for Reasoning) 방식으로 훈련된 모델은 단 817개의 엄선된 학습 데이터만으로 뛰어난 추론 성능을 달성
- LIMO 모델은 AIME2024 벤치마크 평가에서 57.1%, MATH500*에서 94.8%의 정확도를 기록해, 오픈AI o1-preview, Qwen2.5-32B-Instruct, QWQ-32B-preview 등 10만 개 이상의 학습 데이터로 미세 조정된 모델들보다 뛰어난 성능을 기록

* 수학적 추론과 문제 해결 능력을 평가하는 벤치마크

〈소규모 고품질 학습 예제 기반 LIMO의 성능 향상 효과〉



- 연구진은 사전학습 과정에서 방대한 수학적 지식을 습득한 LLM은 이미 충분한 추론 지식을 보유하여 추론에 특화된 소량의 고품질 학습 데이터로 추론 기능을 활성화할 수 있다는 가설을 수립하고 실험을 통해 입증
- 연구진이 큐원1.5-32B-Chat 및 사전학습 품질을 개선한 큐원2.5-32B-Instruct에 LIMO 방식을 적용한 결과, 큐원2.5-32B-Instruct는 AIME2024와 MATH500에서 각각 57.1%와 94.8%를 기록하며, 큐원1.5-32B-Chat의 10.0%와 60.4%를 압도
- 사후 학습 예제의 효과를 확인하기 위해 고품질 추론 예제로 모델을 훈련한 결과, 저품질 예제로 훈련된 모델과 뚜렷한 성능 차이가 발생*

* 최고 및 최저 품질 예제로 훈련된 모델의 성능 차이는 AIME2024에서 약 15%p, MATH500에서 12%p

스탠포드大와 워싱턴大 연구진, 저비용으로 고성능 추론 AI 모델 개발

KEY Contents

- 미국 스탠포드大와 워싱턴大 연구진이 고품질 데이터셋과 ‘테스트-타임 스케일링’을 구현하는 간단한 방법을 활용해 50달러 미만의 비용으로 추론 AI 모델 ‘s1’을 개발
- 구글 제미나이 2.0 플래시 씽킹 실험 버전을 증류한 s1은 수학과 코딩 벤치마크에서 오픈AI o1 및 답시크 R1에 근접한 성능을 기록

50달러 미만의 비용으로 훈련된 s1, 수학과 코딩 벤치마크에서 o1급 성능 발휘

- 미국 스탠포드大와 워싱턴大 연구진이 구글의 ‘제미나이 2.0 플래시 씽킹’ 실험 버전을 증류*하는 방법으로 추론 AI 모델 ‘s1’을 개발하고 2025년 2월 3일 관련 논문**을 공개

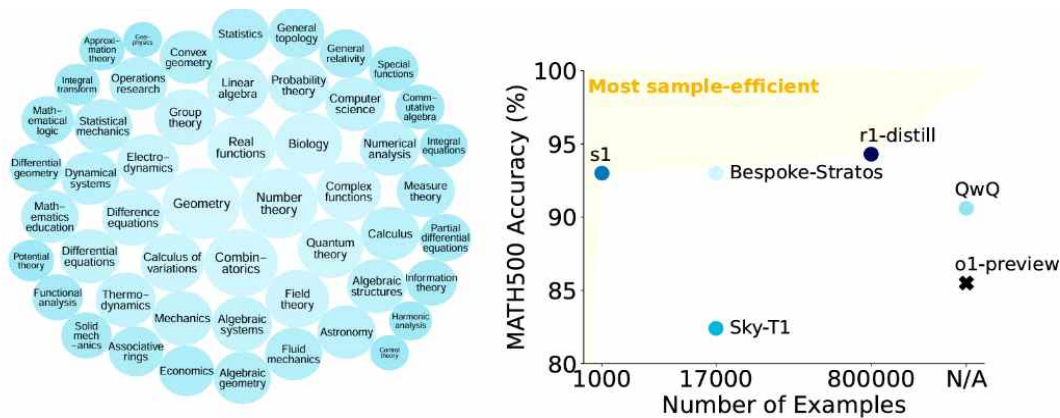
* Distillation: 더 큰 모델에서 추출한 지식을 더 작은 모델로 전달하여 모델 성능을 향상하는 기법

** s1: Simple test-time scaling(<https://arxiv.org/abs/2501.19393>)

- 연구진은 엔비디아 H100 GPU 16개로 26분 만에 모델을 학습시켜 50달러 미만의 비용으로 수학 및 코딩 벤치마크에서 오픈AI o1 및 답시크 R1과 같은 첨단 추론 모델에 근접한 성능을 발휘하는 AI 모델을 개발
- 연구진은 제미나이 2.0 플래시 씽킹 실험 버전에서 추출한 데이터를 정제된 후 △난이도 △다양성 △품질 3개 기준으로 신중하게 선별된 1,000개의 데이터 샘플을 사용한 지도 미세조정(SFT)*으로 알리바바의 ‘큐원2.5-32B-Instruct’ 모델을 훈련

* Supervised Fine-Tuning: 대규모 언어 모델을 특정 작업에 맞게 지도학습 방식으로 미세 조정하는 기법

〈S1의 샘플 효율성 비교〉



- 연구진은 ‘테스트-타임 스케일링’을 구현하는 간단한 접근방식을 모색해 강력한 추론 성능을 달성

* Test-Time Scaling: AI 모델이 추론 단계에서 더 많은 시간과 연산 자원을 사용해 성능이 향상되는 특성

- s1이 추론을 진행하는 도중에 강제로 사고 과정을 멈추고 다시 생각하도록 지시해 생각하는 시간을 연장하면 s1은 답을 다시 확인하여 잘못된 추론을 수정하며, 일례로 raspberry라는 단어에 r이 몇 개인지 묻는 간단한 질문에 처음에 2개라고 답했다가 생각할 시간을 더 주자 정답을 3개로 수정

바이트댄스, 인물 움직임을 생성하는 AI 모델 ‘옴니휴먼-1’ 개발

KEY Contents

- 바이트댄스가 인물 이미지와 오디오 입력을 바탕으로 음성을 포함한 자연스러운 인물 동영상을 생성하는 AI 모델 ‘옴니휴먼-1’을 개발
- 옴니휴먼-1은 인물의 상반신뿐 아니라 전신까지 다양한 화면 비율을 지원하고 2D 캐릭터와 인형 등 다양한 이미지의 애니메이션도 구현 가능

○ 옴니휴먼-1, 음성을 포함한 자연스러운 인물 동영상 생성

- 중국 기술 기업 바이트댄스(ByteDance)가 2025년 2월 3일 음성과 움직임을 포함하는 자연스러운 인물 동영상을 생성하는 AI 모델 ‘옴니휴먼-1(OmniHuman-1)’에 대한 논문*을 발표
 - * OmniHuman-1: Rethinking the Scaling-Up of One-Stage Conditioned Human Animation Models(<https://arxiv.org/pdf/2502.01061>)
- 이 AI 모델은 인물 이미지와 오디오 입력을 조합해 사람의 음성과 일치하는 제스처와 움직임을 표현하며, 특히 얼굴이나 상반신 움직임만을 생성하는 기존 기술과 달리 전신으로 표현 범위를 확장하고 더욱 자연스러운 움직임을 생성 가능
- 상반신에서부터 전신까지 다양한 화면 비율을 지원하고 실사 인물 외 2D 캐릭터, 인형 등 다양한 스타일의 이미지를 애니메이션화할 수 있으며, 악기 연주, 쇼핑, 운전과 같은 상황별 동작도 구현 가능
- 연구진은 옴니휴먼-1에 ‘옴니-조건(Omni-Conditions)’이라는 학습 전략을 적용해 더욱 현실적인 인물 동영상을 구현했다고 설명
 - 연구진은 1만 8,700시간 이상의 인간 영상 데이터 및 텍스트, 오디오, 신체 움직임 등 다양한 유형의 입력을 결합한 훈련 방식을 통해, 기존 방식보다 훨씬 방대하고 다양한 데이터로 모델을 훈련
 - 훈련은 텍스트 조건만 사용하는 1단계, 텍스트와 이미지 조건을 결합한 2단계, 텍스트, 오디오, 이미지를 모두 결합한 3단계로 구성되며, 조건에 따라 학습 비율을 조정하여 모델이 각 조건의 장점을 최대한 활용하도록 설계
- 연구진은 ‘옴니휴먼-1’이 연구 단계로 일반 공개 계획은 아직 없다고 밝혔으며, 영상 생성 시 상당한 연산 자원이 필요하다는 점을 단점으로 제시

〈‘옴니휴먼-1’로 생성된 영상 예시〉



AI안전센터와 스케일 AI, 고난도 벤치마크 'HLE' 공개

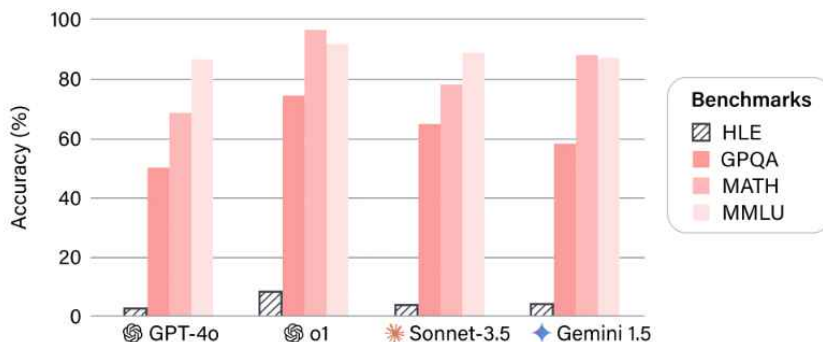
KEY Contents

- AI안전센터와 스케일 AI가 기존 벤치마크의 한계를 넘어서기 위해 50개국 500개 이상 기관에 소속된 천여 명의 전문가 협업을 통해 고난도 벤치마크 'HLE'를 개발
- HLE 벤치마크로 평가한 결과 오픈AI o3-미니 (HIGH) 정확도가 14.0%, 답시크-R1은 8.6%에 불과했으나 AI안전센터는 2025년 말에는 첨단 모델의 정확도가 50%를 넘어설 것으로 예상

○ HLE 벤치마크, 100개 이상 주제를 포괄하는 3천 개의 까다로운 질문으로 구성

- 미국의 AI 안전 비영리기구 AI안전센터(Center for AI Safety)와 AI평가전문 스타트업 스케일 AI(Scale AI)가 2025년 1월 24일 고난도 평가를 위한 벤치마크 'HLE(Humanity's Last Exam)'를 공개
 - 최근 개발된 AI 모델들이 MMLU와 같은 인기 벤치마크에서 대부분 90% 이상의 정확도를 달성하며 기존 벤치마크로는 모델 성능 평가에 한계가 있다는 여론이 대두
 - 이에 AI안전센터는 50개국 500개 이상의 기관 소속 교수와 연구자 등 천여 명의 전문가 협업을 통해 100개 이상의 주제를 포괄하는 2,700개의 까다로운 질문으로 구성된 데이터셋을 개발
- AI 안전센터가 HLE 벤치마크로 주요 기업의 AI 모델을 평가한 결과, 1위를 차지한 오픈AI o3-미니 (HIGH)도 정확도가 14.0%에 불과했으며, 답시크-R1은 8.6%를 기록*
 - * 평가시점에 따라 정확도 차이 존재(Last Updated: 2025.02.11.)
 - 클로드 3.5 소네트(4.8%), 제미니 1.5 프로(5.2%), o1(8.8%)도 10% 미만의 정확도를 나타내, 전문가 수준의 학업 능력에서는 상당한 개선의 여지가 있음을 시사
- AI안전센터는 AI 기술의 발전 속도를 고려할 때, 2025년 말까지 AI 모델의 HLE 벤치마크 정확도가 50% 이상으로 향상될 가능성이 있다고 예상
 - AI안전센터는 HLE에서 높은 정확도를 달성한 AI 모델은 객관식 질문과 최신 과학 지식에서 전문가 수준의 성능을 보여줄 수 있지만, 자율적인 연구 능력을 갖추거나 인공일반지능(AGI) 수준에 도달했다는 의미는 아니라고 부연
 - HLE를 통해 과학자와 정책 입안자들이 AI 역량을 평가할 수 있는 공통된 기준을 제시함으로써 개발 방향이나 잠재적 위험, 필요한 정책 조치에 대하여 정보에 입각한 논의가 가능해질 것으로 기대

〈벤치마크별 주요 AI 모델의 정확도 비교〉



인력·교육

딜로이트 조사 결과, 전 세계 기업들은 점진적으로 AI 도입 확대

KEY Contents

- 딜로이트에 따르면 전 세계 기업들은 초기의 생성 AI 열풍에서 벗어나 실용적 태도로 AI에 접근하고 있으며, AI 투자의 증가에도 규제 준수 우려 등으로 도입이 더디게 진행
- 기업 내 가장 성숙한 생성 AI 프로젝트는 대개 투자수익률을 실현한 것으로 나타났으며, 기업 내 생성 AI 도입 수준이 가장 성숙한 부문은 IT 부문으로 확인

● 전 세계 기업들, 생성 AI 투자 증가에도 도입 속도는 느린 편

- 글로벌 컨설팅기업 딜로이트(Deloitte)가 2025년 1월 21일 발표한 ‘2024년 4분기 기업 내 생성 AI 현황’ 보고서에 따르면 전 세계 기업에서 생성 AI 도입이 점진적으로 늘어나는 추세
 - 딜로이트는 글로벌 14개국에서 생성 AI를 시범 운영하거나 도입한 경영진 2,773명에 대한 설문조사를 바탕으로 이번 보고서를 작성
 - 보고서에 따르면 기업 리더들은 초기의 생성 AI 열풍을 벗어나 실용적인 태도로 AI에 접근하는 추세로, AI 도입 속도는 더디게 진행
 - 조사 응답자의 78%가 다음 회계연도에 AI 투자를 늘릴 계획이라고 답했으나 3분의 2 이상은 생성 AI 시범 프로젝트의 30% 이하만 향후 3~6개월 내 전면 도입으로 이어질 것이라고 예상
- 기업 내 생성 AI 도구와 애플리케이션 도입을 저해하는 주요 요인은 규제 준수 우려로 확인
 - 2024년 1분기 조사에서는 응답자 28%만이 규제 준수 문제를 꼽았으나, 4분기 조사에서는 같은 응답이 38%를 기록했으며, 69%는 생성 AI 거버넌스 전략이 완전히 자리 잡는 데 1년 이상이 걸릴 것으로 예상

● 기업 내 IT 부문의 생성 AI 도입 수준이 가장 성숙한 것으로 평가

- 기업 내 가장 성숙한 생성 AI 프로젝트는 대개 투자수익률(ROI)을 실현한 것으로 나타났으며, 주요 사업 기능 중 IT 부문의 생성 AI 도입이 가장 성숙
 - 조사 응답자의 74%는 조직 내 가장 성숙한 AI 프로젝트에서 ROI를 실현했다고 답했으며, 응답자의 20%는 ROI가 30%를 넘어섰다고 응답한 가운데, 특히 사이버보안 관련 생성 AI 프로젝트에서는 응답자의 44%가 기대치를 넘어서는 수준의 ROI를 실현했다고 응답해 가장 높은 성과를 달성
 - 조사 응답자의 28%는 IT 부문의 생성 AI 프로젝트가 가장 앞서 있다고 답했으며, 운영(11%), 마케팅(10%), 고객 서비스(8%)가 다음 순서를 기록
- AI 에이전트도 기업 리더들의 주요 관심사로, 응답자의 26%가 자율 에이전트 개발을 광범위하게 모색 중이라고 답했으나, 딜로이트는 생성 AI가 직면한 주요 장벽인 규제 불확실성과 위험관리, 데이터 부족, 인력 문제가 에이전트 도입 시에도 동일하게 나타날 것으로 예상

마이크로소프트 연구 결과, 생성 AI 신뢰할수록 비판적 사고 감소

KEY Contents

- 마이크로소프트와 카네기멜론대가 지식 근로자들의 실제 업무 사례 분석을 통해 생성 AI가 비판적 사고에 미치는 영향을 분석한 결과, 생성 AI를 신뢰할수록 비판적 사고가 감소
- 연구진은 사용자가 AI의 한계를 인식하고 비판적 사고를 통해 AI가 생성한 결과물을 평가하고 개선할 수 있도록 생성 AI 도구를 설계할 것을 강조

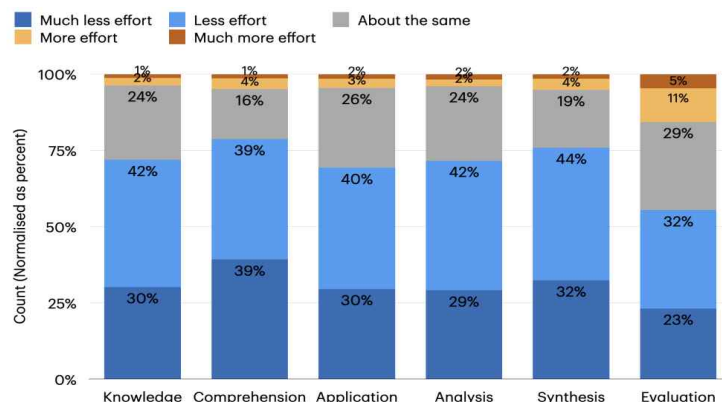
○ 생성 AI에 과도한 의존 시 독립적인 문제 해결 능력이 저하될 위험 존재

- 마이크로소프트(Microsoft)와 카네기멜론대 연구진이 생성 AI가 지식 근로자의 비판적 사고에 미치는 영향을 조사한 연구 결과를 공개
- 생성 AI를 주 1회 이상 업무에 사용하는 지식 근로자 319명 대상의 설문조사를 활용한 연구 결과, 생성 AI는 업무 효율성을 향상할 수 있지만 장기적으로 독립적인 문제 해결 능력을 저해할 위험 존재

○ 참가자들이 공유한 936개의 실제 업무 사례를 분석한 결과, 대부분 사례에서 생성 AI 사용 시 비판적 사고와 관련된 인지 활동이 감소

- 인지 활동 중 지식(Knowledge)에서 72%, 이해(Comprehension)에서 79%, 적용(Application)에서 69%, 분석(Analysis)에서 72%, 종합(Synthesis)에서 76%, 평가(Evaluation)에서 55%가 “훨씬 적은 노력”이나 “적은 노력”을 들였다고 응답
- 생성 AI를 신뢰하는 근로자일수록 비판적 사고가 줄어드는 반면, 업무 능력에 대한 자신감이 높은 근로자는 비판적 사고를 더 많이 하는 것으로 확인

〈생성 AI 도구 사용 시 인지 활동에 들이는 노력 수준〉



- 연구진은 사용자가 AI의 한계를 인식하고 비판적 사고를 통해 AI가 생성한 결과물을 평가하고 개선할 수 있도록 생성 AI 도구를 설계해야 한다고 강조
- 특히 일상 업무를 과도하게 자동화하면 근로자가 비판적 사고를 연습할 수 없어 법률 문서 작성과 같은 고위험 상황에서 사고 능력이 저하될 수 있다고 경고

앤스로픽, AI가 노동시장에 미치는 영향을 분석한 ‘앤스로픽 경제 지수’ 공개

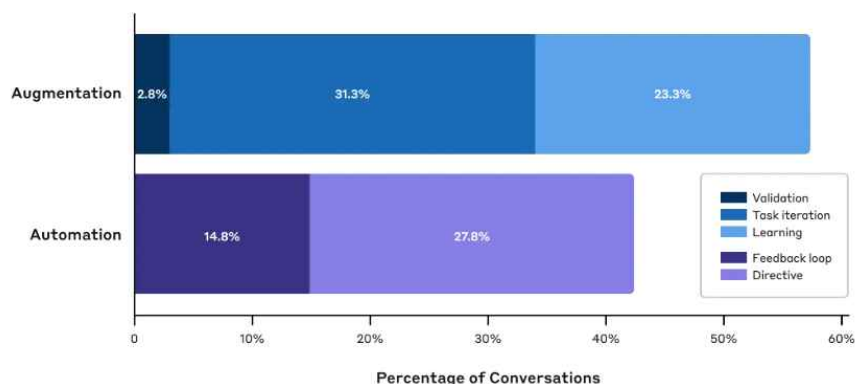
KEY Contents

- 앤스로픽이 클로드에서 이루어진 수백만 건의 대화를 기반으로 AI가 실제 업무에 어떻게 통합되고 있는지를 분석한 결과, AI가 가장 많이 사용되는 업무는 소프트웨어 개발로 확인
- 전체 직업의 약 36%가 업무의 25% 이상에서 AI를 일정 수준 사용하고 있으며, 지식 근로자들은 AI로 업무를 자동화하기보다 증강하는 비율이 더 높은 것으로 조사

● 전체 직업의 약 36%가 업무의 25% 이상에서 AI를 일정 수준 사용

- 앤스로픽이 2025년 2월 10일 AI가 노동시장과 경제에 미치는 영향을 이해하기 위한 ‘앤스로픽 경제 지수(Anthropic Economic Index)’ 보고서를 발표
 - 보고서는 앤스로픽의 생성 AI 서비스 ‘클로드(Claude.ai)’에서 이루어진 수백만 건의 익명화된 대화를 기반으로 경제 전반에서 AI가 실제 업무에 어떻게 통합되고 있는지를 분석
 - 앤스로픽은 노동시장의 변화에 대응한 정책 개발을 위해서는 다양한 관점이 필요하다는 점, 분석에 사용된 데이터셋을 공개하고 경제학자와 정책 전문가, 기타 연구자의 의견을 요청
- 보고서에 따르면 소프트웨어 개발 및 기술문서 작성에 AI가 가장 많이 사용되며, 전체 직업의 약 36%가 업무의 25% 이상에서 AI를 일정 수준 사용
 - 클로드에 입력된 질문의 37.2%가 소프트웨어 수정, 코드 디버깅, 네트워크 문제 해결 등 소프트웨어 엔지니어링 직무와 연관된 “컴퓨터 및 수학” 범주에 해당
 - 두 번째로 AI를 많이 사용하는 직무는 “예술, 디자인, 스포츠, 엔터테인먼트, 미디어”(질문의 10.3%)로 나타났으며, 신체 노동 중심의 “농업, 어업, 임업” 범주에 해당하는 질문은 0.1%로 최하위를 기록
 - 전체 직업 중 약 4%만 업무의 75% 이상에서 AI를 사용했으며, 저임금이나 초고임금 일자리보다 컴퓨터 프로그래머와 데이터 과학자 등 중간~고임금 범위의 직업이 AI를 가장 많이 사용
 - AI를 업무 자동화(Automation: AI가 문서 서식 지정과 같은 작업을 직접 수행)와 증강(Augmentation: AI가 사용자와 협력하여 작업을 수행) 중 주로 어디에 사용하는지 분석한 결과, 업무의 57%가 증강에 해당하고 자동화 비중은 43%로 확인

〈클로드 대화에서 업무 증강 vs 업무 자동화 비율〉



출처 | Anthropic, The Anthropic Economic Index, 2025.02.10.

앤스로픽을 비롯한 여러 기업들, 입사 지원 시 AI 도구 사용금지 요구

KEY Contents

- 앤스로픽이 채용을 위한 AI 정책에서 지원자의 개인적 관심을 이해하고 의사소통 기술을 평가하기 위해 입사 지원 시 AI 도구 사용을 금지한다는 내용을 포함
- 앤스로픽 외에도 상당수 기업의 채용 관리자들은 AI를 이용한 입사 지원을 부정적으로 여기며, 지원자들에게 AI 도구에서 기대하기 어려운 소프트 스킬을 요구

○ 앤스로픽, 채용 정책을 통해 입사 지원 시 AI 도구 사용 금지

- AI 스타트업 앤스로픽이 채용 공고에 포함된 AI 정책(AI Policy for Application)을 통해 입사 지원 시 AI 도구의 사용을 금지
 - 해당 정책은 실제 업무에서는 AI 시스템을 사용해 더 빠르고 효과적으로 일할 수 있도록 장려하되 입사 지원 과정에서는 AI 사용을 금지하며, 이는 앤스로픽에 대한 지원자의 개인적 관심을 이해하고 의사소통 기술을 평가하기 위함이라고 설명
 - 이러한 제한사항은 리서치 엔지니어, 브랜드 디자이너, 영업 담당 임원 등 앤스로픽의 모든 구인 공고에 포함되며, 입사 지원 시 지원자의 동의를 요구
- AI 기업인 앤스로픽이 입사 지원에 AI 도구 사용을 금지하는 것은 모순적이라는 지적도 제기되고 있으나, 실제로 상당수 기업의 채용 관리자들이 AI를 이용한 입사 지원에 거부감을 표시
 - 온라인 채용 플랫폼 CV 지니어스(CV Genius)의 2024년 채용 관리자 대상 설문 조사*에 따르면 채용 관리자의 80%가 AI가 생성한 이력서와 자기소개서를 꺼린다고 답했으며, 74%는 AI로 작성된 지원서를 알아볼 수 있다고, 57%는 AI가 생성한 콘텐츠를 사용한 구직자를 채용하지 않겠다고 응답
 - *CV Genius, CV & Cover Letter Trends in 2024 (New Data), 2024.06.17.
 - 그러나 생성 AI 활용이 업무와 일상에서 보편화되면서 채용 평가 플랫폼 뉴로사이트(Neurosight)의 2024년 입사 지원자 대상 설문 조사*에 따르면 구직자의 약 57%가 입사 지원서 작성 시 챗GPT를 사용한 것으로 확인
 - *Financial Times, Job hunters flood recruiters with AI-generated CVs, 2024.08.13.

○ 채용 관리자들, 구직자들에게 AI 도구에 기대하기 어려운 ‘소프트 스킬’ 요구

- 미국 경제 전문 매체 포춘(Fortune)지는 입사 지원 시 AI 사용을 금지하는 앤스로픽의 채용 정책이 디지털 시대에 더욱 부상하는 소프트 스킬(Soft Skill)의 중요성을 반영한다고 평가
 - 채용 관리자들은 구직자들에게 AI 도구에서 기대하기 어려운 의사소통 능력과 스토리텔링, 감성 지능과 같은 소프트 스킬을 기대
 - 그러나 일자리를 얻기 위해 수백 개의 일자리에 지원해야 하는 구직자와 수많은 서류를 검토하는 채용 담당자의 상황을 고려하면 채용 절차에서 AI 사용은 불가피할 전망

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 |
| | 11~11일 | AI Seoul 2025 | 서울, 강남 | aiseoul2025.com/home |
| | 15~16일 | (AIMLA 2025) International Conference on AI, Machine Learning and Applications | 인도, 첸나이 | ccnet2025.org/aimla/index |
| | 24~26일 | Generative AI DACH 2025 | 독일, 베를린 | www.aidataanalytics.network/events-generative-ai-dach |
| | 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-cloud-ai-infrastructure-summit-2025-korea |
| 4월 | 15~16일 | World Summit AI Americas | 캐나다, 몬트리올 | americas.worldsummit.ai |
| | 24~26일 | 월드 IT 쇼 2025 | 서울, 강남 | www.worlditshow.co.kr |
| | 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 |
| | 21~23일 | 2025 산업AI EXPO | 서울, 강서 | industrialaiexpo.or.kr |
| 6월 | 4~5일 | AI & Big Data Expo North America 2025 | 미국, 산타클라라 | www.ai-expo.net/northamerica |
| | 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 |
| | 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 | 3~5일 | (소프트웨이브 2025) 10회 대한민국 소프트웨어 대전 | 서울, 강남 | 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.