2026년 7월호

SPRi AI Brief

인공지능 산업의 최신 동향

SPRi 소프트웨어정책연구소
Software Policy & Research Institute

SPRi AI Brief

I 2026년 7월호 I

CONTENTS

2026년 6월 월간 AI 모델 현황

주요 모델 출시 타임라인

6/1 MiniMax M3 6/2 MAI-Thinking-1, MAI-Code-1-Flash, MAI-Image-2.5
6/4 NVIDIA Nemotron 3 Ultra 550B 6/9 Fable 5, Mythos 5 6/12 Kimi K2.7 Code
6/13 GLM-5.2 6/16 Grok Imagine Video 1.5 6/22 Fugu Ultra

LLM간 성능 비교(Artificial Analysis Intelligence Index 4.0)

모델	출시일	Index	Price ($/1m Token)		Speed	Context Window	회사
			Input	Output			
① Claude Fable 5(with fallback)	26.6.9.	60	10	50	N/A	1m	Anthropic
② Claude Opus 4.8(max)	26.5.28.	56	5	25	66.9	1m	Anthropic
③ GPT-5.5(xhigh)	26.4.23.	55	5	30	65.8	922k	OpenAI
GLM-5.2(max)	26.6.13.	51	1.4	4.4	141.2	1m	Zhipu AI
Gemini 3.5 Flash (high)	26.5.19.	50	1.5	9	190.7	1m	Google

* 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	1338	OpenAI
② MAI-Image -2.5	26.6.2.	1275	Microsoft
③ HiDream-O1 -Image-1.5	26.6.9.	1264	HiDream

영상 생성 모델

모델	출시일	ELO	회사
① Dreamina Seedance 2.0 720p	26.2.10.	1194	ByteDance
② HappyHorse -1.1	26.6.22	1120	Alibaba
③ grok-imagine -video-1. 5-preview	26.5.31.	1111	xAI

한국 AI 모델 현황

■ NC AI, 3D 생성형 AI 모델 '바르코 3D 2.0' 공개

- 기존 3D 생성형 AI의 주요 한계인 '원본 형상 왜곡'을 개선한 것이 특징으로, 원본의 실루엣과 전체 비례를 충실히 반영하고 복잡한 장식 요소도 높은 정확도로 재현하는 데 초점

주요 포인트 :
Claude Fable 5, MiniMax M3 등 대규모 컨텍스트와 장기 에이전틱 작업에 특화된 프런티어급 모델 출시
외부 AI 모델 종속을 피하기 위한 독자 모델 전략 본격화(Micrsoft MAI 시리즈, Z.ai의 GLM-5.2 등)

SPRi AI Brief

Ⅰ2026년 7월호Ⅰ

정책·법제

미국 일리노이주, 첨단 AI 모델 대상 외부 감사 의무화하는 법안 통과

KEY Contents

- 미국 일리노이주 의회에서 프런티어 AI 모델이 초래할 수 있는 중대한 위험에 대응한 프런티어 AI 프레임워크 수립 및 제3자 감사를 의무화하는 SB 315 법안이 통과
- 동 법안은 연 매출 5억 달러 이상의 프런티어 모델 개발사에만 적용되며, 최초 위반 시 최대 100만 달러, 반복 위반 시 최대 300만 달러의 벌금 부과를 규정

AI 안전 조치법, 프런티어 AI 모델 개발자에 재앙적 위험 대응 및 외부 감사 요구

■ 미국 일리노이주 의회가 2026년 5월 27일 첨단 AI 모델로 인한 중대한 위험을 방지하기 위해 주법 최초로 제3자 감사를 의무화하는 「AI 안전 조치법(SB 315)*」를 가결

* Artificial Intelligence Safety Measures Act

• 동 법안은 2026년 5월 21일 상원을 통과한 데 이어 5월 26일 하원 본회의를 만장일치로 통과했으며, 주지사 서명 후 2028년 1월 1일부터 시행 예정

■ 동 법안은 프런티어 AI 모델 개발자*에 '재앙적 위험'** 식별과 평가, 위험 완화 조치, 사이버보안 관행을 포함한 프런티어 AI 프레임워크 수립을 요구

* 직전 연도의 연간 총매출액 5억 달러 이상의 프런티어 모델(10^{26} 이상 부동소수점 연산을 사용해 훈련된 모델)개발사

** Catastrophic Risk: 화생방·핵무기 제작에 관한 전문가급 지원, 인간의 감독 없는 사이버 공격이나 범죄, 모델의 개발자 통제 이탈 등으로 인한 단일 사고로 50명 초과 사망·중상 또는 10억 달러 초과 재산 피해를 초래할 수 있는 위험

• 또한 매년 제3자를 선임하여 프런티어 AI 프레임워크 준수 여부에 대한 독립적인 감사를 수행해야 하며, 독립 감사인은 프런티어 AI 안전 분야의 전문성을 입증하고 개발자와 재정적 이해관계 공유 불가

■ 프런티어 AI 모델 개발자는 자사의 프런티어 모델 관련 중대한 안전사고 발생 시 24시간 이내 당국에 신고, 72시간 이내 일리노이주 비상관리청(IEMA)에 보고 필요

• 동 법에 따른 의무를 위반할 경우, 최초 위반 시 최대 100만 달러, 이후 반복 위반 시 건당 최대 300만 달러의 벌금이 부과될 수 있으며, 일리노이주 법무장관에게 법 집행 권한을 부여

오픈AI와 앤트로픽은 법안 지지, 일부 산업단체는 반발

■ 동 법안에 대하여 오픈AI와 앤트로픽은 입법 과정 전반에서 지지 의사를 밝혔으나, 컴퓨터·통신산업협회(CCIA)는 외부 감사에 필요한 인프라 부재를 지적하며 법안에 반대

• CCIA는 일리노이주 정부가 감사를 의무화하기 전에 독립적 검증 체계의 타당성을 연구하도록 지시한 버지니아주와 코네티컷주의 사례를 따를 것을 촉구

■ 한편, 동 법안은 연방 법률이나 규정, 지침이 SB 315의 요건과 실질적으로 동등하다고 인정되면 개발자의 해당 연방 요건 충족 시 SB 315 준수로 간주하는 상호운용성 조항을 포함

• 이를 통해 주별 AI 규제가 규제 파편화로 이어진다는 이유로 주별 규제를 봉쇄하려는 백악관에 대응

출처 | LegiScan, Illinois Senate Bill 315, 2026.5.29.
Akerman, Illinois SB 315: A State Strategy for Enduring National AI Safety Standards, 2026.6.10.

레오 14세 교황, 첫 회칙에서 AI 시대 인간 존엄성 수호 강조

KEY Contents

- 레오 14세 교황이 즉위 후 처음 발표한 회칙에서 AI 무장 해제 원칙을 선언하고 인간 존엄성을 중심 법적 프레임워크와 독립적 감시, 책임 있는 정치를 강조
- 교황은 특히 군사 분야의 AI 활용을 경계하며 자율 살상 무기에 우려를 표시하고 무력 사용의 범위가 엄밀한 의미의 자기방어에 한정되어야 한다고 명시

AI 무장 해제 선언과 인간 존엄 수호를 위한 핵심 원칙 제시

■ 레오 14세 교황이 2026년 5월 25일 발표한 첫 회칙* 「고귀한 인류(Magnifica Humanitas)」에서 AI가 일상 전반을 재편하는 시대에 인간 존엄을 수호하기 위한 원칙과 실천 지침을 제시

* 교황이 전 세계 주교와 14억 가톨릭 신자에게 보내는 최고위 문서로 시대적 난제에 도덕적 지침으로 평가

- 회칙은 AI가 도덕적으로 중립적일 수 없다며, 인간의 존엄성을 중심에 둔 견고한 법적 프레임워크와 독립적 감시, 책임 있는 정치의 필요성을 제시
- AI 개발 시 자원의 공정한 분배와 인간의 존엄성, 사회 정의 및 환경 보호를 포함한 일련의 원칙을 적용할 것을 촉구하는 한편, AI 개발에 다양한 의견을 가진 여러 집단이 참여해야 한다고 강조

■ 교황은 회칙에서 AI의 '무장 해제'를 촉구하면서, 이를 기술 거부가 아니라 AI에 의한 인류 지배를 막고 기술 변화 속에서 인간의 위대함을 수호하기 위함이라고 강조

- 또한 소수가 AI 통제권을 독점해서는 안 된다고 강조하며, 편견과 불의로 오염된 데이터에 기반한 알고리즘이 의료나 고용, 안전에 대한 접근을 차단해 새로운 형태의 배제를 낳을 수 있다고 경고
- 핵에너지와 마찬가지로 AI는 모든 사람과 공동선을 위해 봉사해야 하며, 기술에 관한 결정은 결코 양심과 책임에서 분리될 수 없다고 지적

■ 교황은 AI 개발에 견고한 법적 프레임워크와 독립적 감시, 책임 있는 정치의 필요성을 강조하는 한편, 군사 분야의 AI 활용에 대하여 가장 엄격한 윤리적 제약의 적용을 요구

- 인간의 통제 범위를 사실상 벗어나 갈수록 자율화되어 가는 무기 체계에 대한 우려를 표시하면서, 무력 사용의 범위는 '엄밀한 의미의 자기방어'에만 허용된다고 명시

산업혁명 시대 레오 13세의 회칙을 계승해 AI 시대의 새로운 이정표 제시

■ 이번 회칙은 레오 14세 교황의 재위 기간 중 중요한 이정표가 되는 문서로 평가되고 있으며, 레오 14세는 이번 회칙 반포를 직접 주관해 가톨릭교회의 엄중한 AI 인식을 반영

- 역대 교황들은 회칙 발표를 통상적으로 추기경 등 고위 인사에 위임해 왔으나 레오 14세는 회칙 반포식에 직접 참여해 연단에 섰으며, 반포식에는 앤트로픽의 공동 창업자 크리스 올라(Chris Olah)도 참석
- 이번 회칙은 산업혁명 시대 노동자 권리를 다룬 레오 13세 교황의 회칙 「새로운 사태(Rerum Novarum)」 반포 135주년인 5월 15일에 서명된 것으로, 기술 격변속 인간 존엄성을 지킨 가르침 계승

출처 | Vatican, MAGNIFICA HUMANITAS, 2026.5.25.
CNN, Pope Leo warns of AI fueling warfare in first major theological document, 2026.5.26.

캐나다 정부, 국가 AI 전략 'AI for All' 발표

KEY Contents

- 캐나다 정부가 향후 5년간 추진할 새로운 국가 AI 전략 'AI for All'을 출범하고 신뢰 확보와 기회 창출, 주권 강화의 3대 핵심 원칙 기반의 입법과 투자 프로그램을 추진할 계획
- 향후 5년간 2,000억 달러의 추가 경제성장과 25만 개의 AI 관련 신규 일자리 창출과 함께 2034년까지 AI 도입률을 현재의 12%에서 60%로 끌어올리겠다는 목표를 제시

향후 5년간 2,000억 달러의 추가 경제성장과 25만 개 AI 일자리 창출 목표

■ 캐나다 정부가 2026년 6월 4일 '모두를 위한 AI(AI for All)' 전략을 발표하고 향후 5년간 입법, 투자, 인력 양성, 산업 육성을 포함한 국가 차원의 AI 청사진을 제시

- 마크 카니(Mark Carney) 캐나다 총리는 AI가 이미 현실이 되었다며, AI를 통한 일자리 창출과 국민 보호, 국가 번영을 통해 모든 캐나다인의 삶을 개선하겠다는 비전을 제시
- 캐나다 정부는 향후 5년간 AI를 통한 2,000억 달러의 추가 경제성장 달성과 25만 개의 AI 관련 일자리 창출을 핵심 목표로 제시
- AI 도입률을 현재 12% 대에서 2034년 60%로 확대하고 청년층에게 최대 9만 개의 AI 관련 일자리와 현장실습 기회를 제공해 산업 전반의 글로벌 경쟁력을 높일 계획

■ 이번 전략은 ▲신뢰 확보(캐나다인 보호·신뢰 동맹 구축) ▲기회 제공(공동 번영 실현·역량 강화) ▲주권 수호(캐나다 AI 챔피언 육성·주권 AI 기반 구축) 3개 원칙 하의 6개 핵심 축으로 구성

- (신뢰 확보) 디지털 시대에 맞춘 법제 현대화와 온라인 안전 규제 도입으로 AI 위험으로부터 국민을 보호하고, AI 투명성 제고와 독립적 안전 평가 역량 강화를 통해 국민 신뢰를 확보
- (기회 제공) 전 국민 대상 AI 리터러시 이니셔티브를 신설해 무료·실무 교육을 제공하고, 보건 분야를 우선으로 AI 미션 프로그램을 출범해 진단·환자 치료·시스템 효율 개선과 보건 혁신 생태계 강화를 추진
- (주권 수호) 컴퓨팅·클라우드·데이터·인재 등 자국 AI 토대를 구축하고, 세계 최고 수준의 공공 AI 슈퍼컴퓨터와 주권형 컴퓨팅·클라우드 인프라에 투자해 캐나다 맞춤형 AI 구축과 도입 역량을 확보

보건, 에너지, 교통, 농업, 제조업 등 5대 우선 분야에 투자 집중

■ 캐나다 AI 생태계에 대한 지속적 투자와 함께, 과학적·경제적·산업적 강점이 결합된 분야 및 전략적 투자를 통해 상업적 성공과 국가적 회복탄력성을 달성할 수 있는 5대 분야에 집중 투자할 계획

- (보건 및 생명과학) AI 기반 보건 개선 및 신속하고 효율적인 의료 서비스 제공
- (에너지 및 천연자원) AI를 활용해 자원 추출 최적화, 청정에너지 전환 가속화, 에너지 공급망 강화
- (교통) 지능형 물류, 자율 시스템, 예측 기반 시설 유지관리로 교통 흐름 개선
- (농업) AI 기반 정밀 농업을 도입해 생산량을 늘리고 환경 영향을 줄이며 식량 안보를 강화
- (제조업 및 로봇공학) 산업용 AI와 로봇공학을 활용해 노동력 부족 대응과 국내 제조업 강화

출처 | Government of Canada, Canada's National Artificial Intelligence Strategy: AI for All, 2026.6.4.

미국 트럼프 대통령, AI 혁신과 안보 증진을 위한 행정명령에 서명

KEY Contents

- 미국 트럼프 대통령이 AI 혁신과 국가안보를 목표로 한 행정명령에 서명하고 첨단 AI 모델 출시 30일 전에 보안 평가를 위한 정부 접근을 요구
- 행정명령은 또한 30일과 60일 기한으로 주요 기관에 연방 민간기관의 사이버 방어 강화를 위한 운영 지침 마련, AI 사이버보안 정보 교환센터 설립 등의 이행 과제를 부여

행정명령을 통해 첨단 AI 모델에 대한 30일 사전 검토 체계 도입

■ 미국 트럼프 대통령이 2026년 6월 2일 AI 혁신 촉진과 연방 사이버 방어 현대화를 목표로 하는 '첨단 AI 혁신과 안보 증진' 행정명령*에 서명

* Promoting Advanced Artificial Intelligence Innovation and Security(EO 14409)

- 행정명령은 민간과 정부 부문의 정보 시스템 현대화와 사이버 위협 방어 강화, 적대국의 악용 및 도용 시도로부터 미국의 창의성과 지적 재산 보호, 미국의 첨단 AI 역량 강화를 정책 목표로 명시

■ 행정명령은 정책 목표 달성을 위해 30일과 60일 기한으로 기관별 이행 과제를 부여

- 행정명령 서명일로부터 30일 이내에 국방부는 정보 시스템의 사이버 방어를 우선 추진하고, 사이버보안·인프라보안국(CISA)은 연방 민간기관의 보안 방어를 강화하기 위한 운영 지침을 발표
- 재무부 장관은 30일 이내에 AI 업계 및 주요기반시설 운영자와 협력해 소프트웨어 취약점 발견과 검증, 취약점 패치의 수정과 배포 조정을 위한 AI 사이버보안 정보 교환센터를 설립
- 재무부, 국방부, 국토안보부 장관은 60일 내 AI 모델의 고급 사이버 역량을 평가하는 기밀 벤치마킹 절차를 마련해 '보호 대상 프런티어 모델' 지정 대상의 기준점을 결정
- 적절한 기밀 유지와 사이버보안, 내부자 위험과 지식재산권 보호, 비공개 요건 준수를 조건으로 AI 모델 출시 전 최대 30일 전부터 연방 정부의 해당 모델 접근을 허용하는 자발적 프레임워크 설계
- 단, 동 행정명령의 어떤 조항도 최첨단 모델을 포함한 신규 AI 모델의 개발과 배포에 대한 정부의 의무 라이선스나 사전 승인 또는 허가 요건의 수립을 허용하는 것으로 해석되어서는 안 된다고 명시

CISA, 행정명령 후속 조치로 취약점 대응 우선순위 재편한 'BOD 26-04' 공개

■ 행정명령이 발표된 이후 2026년 6월 10일 CISA는 연방 민간기관의 취약점 대응 우선순위를 실제 공격 위험에 따라 재편한 구속력 있는 운영 지침 'BOD 26-04'를 발표

- 인터넷 노출 시스템에 적용되던 BOD 19-02와 알려진 악용 취약점(KEV)을 대상으로 한 BOD 22-01을 폐지·통합하고, 취약점과 자산의 위험을 함께 평가하는 단일 체계로 전환
- 이 지침은 연방 민간기관들이 가장 위험도가 높은 영역에 취약점 패치 노력을 집중하고 우선순위가 낮은 패치는 연기할 수 있도록 지원하며, 위험도가 높은 취약점은 3일 이내 패치 요구

출처 | The White House, Promoting Advanced Artificial Intelligence Innovation and Security, 2026.6.2.
CISA, CISA Issues New Directive Improving How Federal Agencies Prioritize the Mitigation of Cyber Vulnerabilities, 2026.6.10.

트럼프 행정부에서 AI 국부펀드 조성 논의 본격화

KEY Contents

- 트럼프 대통령이 AI 기업의 지분 공유를 통해 미국 국민이 AI 성장의 수혜자가 되는 파트너십 구상을 공개 지지하며 주요 AI 기업과 관련 회의를 추진 중이라고 언급
- 오픈AI의 샘 올트먼 CEO도 최근 버니 샌더스 상원의원과 만나 국부펀드 조성 방안을 논의하는 등, 정치권에서 국부펀드 조성 논의가 본격화

트럼프 대통령과 샘 올트먼 오픈AI CEO가 AI 국부펀드 논의 주도

■ 트럼프 대통령이 2026년 6월 5일 전용기 에어포스원에서 취재진에게 AI 기업과 미국 국민이 수익을 공유하는 파트너십을 언급하고, 주요 AI 기업 경영진과 백악관 회의를 추진 중임을 공개

- 트럼프 대통령은 AI 기업의 규모가 매우 크고 자금도 풍부해서 일부 지분을 미국인에게 줄 수 있다고 밝히면서, 이를 통해 국민이 매우 부유해질 것이라고 강조

■ 이와 관련, 오픈AI는 이전부터 AI 기업의 지분을 국민에게 분배하는 국부펀드 조성을 제안해 왔으며, 트럼프 대통령은 샘 올트먼(Sam Altman) 오픈AI CEO의 의회 방문 이후 제안을 지지

- 정부가 직접 지분을 소유하는 것이 아니라 알래스카 석유 수익 분배와 유사하게 대중이 개별적으로 주식을 소유하는 방식의 펀드 조성 방안이 유력하게 논의
- 오픈AI의 주요 투자자인 브래드 거스너(Brad Gerstner)는 기업들이 기부한 지분을 활용해 2025년부터 2028년 사이에 태어나는 모든 아동의 투자 계좌에 1,000달러를 지급하는 방안도 제시

■ 원래 AI 국부펀드 조성안은 버니 샌더스(Bernie Sanders) 미국 상원의원과 같은 진보주의 진영에서 적극적으로 지지해 왔으나, 트럼프 대통령도 비슷한 방향성을 공유

- 버니 샌더스 상원의원은 최근 올트먼 CEO와 만나 국부펀드에 AI 기업 지분을 포함하는 방안을 논의했으며, AI 기업의 주식 50%를 현물 세금 형태로 징수해 국부펀드를 조성하는 법안 발의를 추진
- 올트먼은 50% 지분율에는 동의하지 않았으나 대중이 AI 기업의 지분을 갖는다는 기본 방향에는 공감하며 샌더스 측과 협력 의사를 표시

AI에 대한 대중의 반발 확산에 따른 정치권의 대응으로 풀이

■ 데이터센터 건설과 일자리 위협 우려 등 AI 확장에 대한 반발이 전국적으로 퍼져나가면서 좌우를 막론한 정치권이 AI 수익 분배 해법을 모색하는 배경으로 작용

- 전국 각지에서 데이터 건설 프로젝트가 추진되면서 전력 수요와 물 소비 증가, 환경 영향을 우려한 주민 반발이 늘어나고 있으며, 데이터센터 유치에 적극적이었던 일부 주들도 세금 인센티브 재검토를 추진
- 대학 캠퍼스에서도 AI를 둘러싼 긴장감이 고조되고 있으며, 하버드 케네디스쿨의 2025년 여론조사에 따르면 대학생의 약 70%가 AI를 일자리 위협으로 인식

출처 | The Washington Post, Donald Trump, Bernie Sanders and Sam Altman are all talking about public ownership in AI, 2026.6.6.

G7의 중국 견제에 대응해 중국은 글로벌 AI 협력 기구 설립 가속화

KEY Contents

- 프랑스에서 열린 G7 정상회의에서 7개국 정상과 주요 AI 기업 CEO들이 첨단 AI 모델을 신뢰할 수 있는 국가에만 공유하는 방안을 논의하며 중국 견제를 위한 공조를 시사
- G7 정상회의 기간 중 중국 외교부는 글로벌 AI 협력 기구 설립의 가속화와 모든 국가의 참여를 촉구하며 미국 주도의 폐쇄적 접근 방식에 정면 대응

G7, 중국 견제를 위한 AI 규제와 협력 플랫폼 구축 논의

■ 프랑스 에비앙에서 2026년 6월 15~17일 열린 G7 정상회의에서 미국·영국·프랑스·독일·캐나다·이탈리아·일본 7개국 정상과 11개 AI 기업 CEO들이 AI 협력 방안을 논의

- 오픈AI의 샘 올트먼 CEO는 AI 모델 평가와 테스트를 위한 국제 토론포럼 창설을 제안
- 앤트로픽의 다리오 아모데이(Dario Amodei) CEO는 프런티어 모델에 대한 구조적 접근과 중국을 배제한 반도체·핵심부품 교역, 사이버·생물테러·정보 분야의 위험 대응을 핵심 국제 협력 과제로 제시
- 영국 소재 AI 기업 신세시아(Synthesia)의 빅터 리파르벨리(Victor Riparbelli) CEO는 G7 국가와 동맹국들이 단결해 권위주의 국가와 AI 경쟁에서 승리해야 한다는데 공감대가 형성되었다고 언급

■ G7 정상회의 개최 직전 미국이 앤트로픽 최신 모델 Fable 5와 Mythos 5의 해외 접근을 차단하는 직접 수출통제를 단행했으나, EU는 반발을 자제하고 공동 규제를 강조

- 마크롱 프랑스 대통령은 '프런티어 모델이 권위주의 정권 수중에 넘어가지 않도록 더 나은 규제가 필요하다'고 언급했으며, 9월 G7 장관급 회의를 열어 구체적 기준을 논의할 전망

중국, 글로벌 AI 협력 기구 설립과 AI 모델 개방형 공유로 미국 폐쇄 진영에 대응

■ G7 정상회의에서 배제된 중국은 2026년 6월 17일 국무원 주도의 글로벌 거버넌스 백서를 발표하며, 글로벌 AI 협력 기구의 설립 가속화를 천명하고 모든 국가의 참여를 촉구

- 백서는 무역전쟁과 일방주의적 기술 패권을 비판하고, 미국·유럽의 영향력 밖에 있는 글로벌 사우스 등 저개발 경제권에 대한 지원을 부각하며 개방적이고 포용적인 AI 협력 노선을 천명
- 대체로 유료 구독 기반으로 제공되는 미국 AI 모델과 대조적으로 중국은 모델 전체를 다운로드해 사용할 수 있는 저가·무료 모델 보급에 집중하며 접근성 측면에서 차별화

■ 백서 발표 행사에서 자오하이빙 중국 국가발전개혁위원회 부주임은 폐쇄적·배타적· 독점적 기술 개발 방식을 정면 비판하고 다자 협력 틀을 통한 국제 AI 협력 노력을 부각

- 또한 브릭스(BRICS)와 상하이협력기구(SCO)를 통한 AI 협력 확대와 모두를 위한 AI 역량 강화, UN 주도의 글로벌 AI 거버넌스 지원, 개발도상국 대상 기술 및 인재 지원 방안을 제시

출처 | Politico, "West plays nice on AI in bid to shut out China", 2026.6.17.
CNBC, "China pushes for AI safety as G7 summit wraps up without Beijing", 2026.6.17.

SPRi AI Brief
Ⅰ2026년 7월호Ⅰ

기업·산업

앤트로픽, 코딩과 에이전트 성능 강화한 'Claude Opus 4.8' 출시

KEY Contents

- 앤트로픽이 이전 버전보다 코딩과 에이전트, 전문 작업 전반의 벤치마크 성능이 크게 향상되고 정직성과 정렬 평가도 향상된 Claude Opus 4.8을 공개
- Claude Opus 4.8과 함께 대규모 작업 처리를 위한 동적 워크플로 기능 및 사용자가 AI의 사고 깊이와 응답 속도를 직접 조절할 수 있는 노력 조절 기능도 새로 공개

Claude Opus 4.8, 벤치마크 전반의 성능 향상과 정렬 개선 달성

■ 앤트로픽이 2026년 5월 28일 코딩과 에이전트, 전문 작업 전반의 벤치마크 성능 개선과 장기 작업 일관성 향상이 특징인 'Claude Opus 4.8'을 출시

- Claude Opus 4.8은 이전 버전과 동일한 가격(입력 토큰 100만 개당 5달러, 출력 100만 개당 25달러)으로 제공되며 2.5배 빠르게 작동하는 고속 모드(Fast Mode)는 기존 모델 대비 3배 저렴
- 여러 벤치마크 평가에서 이전 버전과 GPT-5.5를 능가하는 뛰어난 에이전트 성능을 달성했으며, 특히 판단력이 현저히 개선되어 스스로 오류를 수정하거나 올바른 질문을 던지며 복잡한 탐색 과정을 완수

표 Claude Opus 4.8과 주요 AI 모델의 벤치마크 점수 비교

	Opus 4.8	Opus 4.7	GPT-5.5	Gemini 3.1 Pro
Agentic coding SWE-Bench Pro	**69.2%**	64.3%	58.6%	54.2%
Agentic terminal coding Terminal-Bench 2.1	74.6%	66.1%	**78.2%**	70.3%
Multidisciplinary reasoning Humanity's Last Exam	**49.8%** no tools **57.9%** with tools	46.9% no tools 54.7% with tools	41.4% no tools 52.2% with tools	44.4% no tools 51.4% with tools
Agentic computer use OSWorld-Verified	**83.4%**	82.8%	78.7%	76.2%
Knowledge work GDPval-AA	**1890**	1753	1769	1314
Agentic financial analysis Finance Agent v2	**53.9%**	51.5%	51.8%	43.0%

■ 앤트로픽의 내부 평가 결과에 따르면 Claude Opus 4.8은 이전 버전 대비 코드 결함을 보고하지 않는 비율이 약 4배 낮아진 것으로 확인되어 정직성이 향상

- 정렬 평가에서도 사용자 자율성 지원 등 친사회적 특성이 최고 수준에 도달하고, 기만·오용 협조 등 오정렬 행동 비율은 Opus 4.7보다 현저히 낮아 최상위 정렬 모델 Claude Mythos Preview에 근접

■ Claude Opus 4.8 출시와 함께 대규모 작업 처리를 위한 동적 워크플로와 응답 노력 수준을 사용자가 직접 정하는 노력 조절, 작업 중 지시 갱신을 지원하는 API 기능을 함께 도입

- 동적 워크플로는 단일 세션에서 수백 개의 하위 에이전트를 병렬 실행해 수십만 줄 규모의 코드 베이스 마이그레이션을 수행하며, 노력 조절은 높은 설정에서 더 깊은 추론, 낮은 설정에서 더 빠른 응답을 지원

출처 | Anthropic, Introducing Claude Opus 4.8, 2026.5.28.

앤트로픽 'Fable 5·Mythos 5', 미국 정부 지시로 서비스 중단

KEY Contents

- 앤트로픽이 일반 사용자에게 공개된 모델 중 가장 강력한 'Claude Fable 5'와 함께, 세계 최고 수준의 사이버보안 역량을 갖춘 'Claude Mythos 5'를 출시
- 그러나 미국 정부가 국가안보를 이유로 외국 국적자에 대하여 두 모델의 접근을 차단할 것을 요구하는 수출통제 지시를 내리면서 앤트로픽은 두 모델의 서비스를 전면 중단

최고 성능 모델 'Claude Fable 5'와 사이버 특화 모델 'Claude Mythos 5' 공개

■ 앤트로픽이 2026년 6월 9일 비공개 최고 성능 모델인 Mythos급 모델 'Claude Fable 5'를 일반 공개하고 보안 전문가용으로 일부 안전장치를 완화한 'Claude Mythos 5'도 출시

- Fable 5는 소프트웨어 엔지니어링·지식 업무·시각 작업·생명과학 연구 등 거의 모든 벤치마크에서 최고 성능을 기록했으며*, 작업이 길고 복잡할수록 기존 모델 대비 격차가 확대

* 일례로 에이전틱 코딩을 평가하는 SWE-Bench Pro에서 80.3%로, Mythos Preview(77.8%), Opus 4.8(69.2%), GPT-5.5(58.6%), Gemini 3.1 Pro(54.2%)를 능가

- Mythos 5는 소수의 사이버 방어 담당자와 인프라 제공업체를 위해 안전장치를 완화한 모델로, 세계 최고 수준의 사이버보안 기능을 제공한다고 강조
- 악용 위험에 대응해 일부 주제의 질문에 대해서는 Claude Opus 4.8 모델이 응답하도록 하는 안전장치를 적용했으며, 일부 무해한 요청에 안전장치가 적용되는 비율은 전체 세션의 5% 미만이라고 설명
- Fable 5와 Mythos 5는 입력 토큰 100만 개당 10달러, 출력 토큰 100만 개당 50달러로 Claude Mythos Preview의 절반에 못 미치는 가격으로 제공

미국 정부의 수출통제 지시로 Fable 5·Mythos 5 전면 접근 중단

■ 그러나 앤트로픽은 2026년 6월 12일 미국 정부가 국가안보를 이유로 외국 국적자에 대하여 두 모델의 접근을 즉시 중단하라는 수출통제 지시를 발동했다며 서비스를 중단

- 정부의 지시 서한에는 구체적인 국가안보 우려 사항이 명시되지 않았으나, 앤트로픽은 정부가 Fable 5의 안전장치를 우회하는 탈옥 방법을 발견한 것으로 확인
- 앤트로픽은 자체적으로 검토한 탈옥 사례는 특정 코드베이스를 읽고 소프트웨어 취약점을 수정하도록 요청하는 제한된 범위에 해당하며, 이미 알려진 일부 경미한 취약점을 식별하는 수준이라고 해명

■ 앤트로픽은 정부 지침의 근거가 된 Fable 5 및 Mythos 5의 탈옥 사례는 오픈AI의 GPT 5.5를 포함한 기존 공개 모델에서도 가능하다고 반박하며 정부의 조치에 반발

- Fable 5의 사이버보안 관련 악용 가능성을 줄이기 위해 강력한 안전장치를 적용하고 미국과 영국 정부 및 여러 민간 외부 기관과 협력해 수천 시간에 걸쳐 철저하게 검증했다고 강조
- 극히 예외적인 탈옥 가능성을 이유로 수억 명이 사용하는 상용 모델을 전면 회수해서는 안 되며, 이 기준이 업계 전반에 적용되면 모든 AI 개발업체의 신규 모델 배포가 사실상 중단될 것이라고 주장

출처 | Anthropic, Claude Fable 5 and Claude Mythos 5, 2026.6.9.
Anthropic, Statement on the US government directive to suspend access to Fable 5 and Mythos 5, 2026.6.12.

미국 정부의 앤트로픽 'Fable 5·Mythos 5' 금지에 산업계 반발 확대

KEY Contents

- 미국 상무부가 앤트로픽의 최신 모델 Fable 5와 Mythos 5를 국가안보 위협으로 규정하고 수출통제를 단행하면서 중국 AI 모델의 반사이익 및 사이버 방어력 약화 우려도 확산
- 수십 명의 사이버보안 전문가들은 Fable 5가 여타 프런티어 모델 대비 독자적 해킹 역량을 제공하지 않는다고 반박하며 수출통제 조치의 철회를 요구

미국 정부의 수출통제 조치에 중국의 반사이익 및 사이버 방어 약화 우려 제기

■ **트럼프 행정부가 아마존이 제보한 보안 위험을 이유로 앤트로픽의 Fable 5와 Mythos 5의 서비스를 금지**

- 앤디 재시(Andy Jassy) 아마존 CEO는 미국 정부 고위 관계자들에게 자사 연구원들이 Fable 5로부터 사이버 공격에 악용될 수 있는 금지된 정보를 유도했다고 직접 제보
- 트럼프 행정부는 미국 연구자가 Fable 5를 탈옥할 수 있다면 적대국도 가능하다는 판단하에 국가안보에 위협이 된다고 인식하고 외국 국적자의 사용을 금지하는 수출통제 조치를 부과

■ **그러나 이번 제재가 미국 AI 기업의 신뢰도 하락으로 이어져 중국 오픈소스 모델의 반사이익을 가져오는 한편, 사이버 방어력 약화로 이어질 수 있다는 지적도 제기**

- 이번 조치는 미국 AI 기업에 대한 의존도 저하 및 유럽의 자체 AI 개발 의지를 자극하는 한편으로, 규제가 없고 저렴한 중국의 오픈소스 모델들이 매력적인 대안으로 부상하는 계기로 작용
- 보안 전문가들은 AI 모델 접근 차단을 "소프트웨어를 핵무기용 우라늄처럼 통제하려는 시도"에 비유하며, 사이버 방어 연구를 가로막아 오히려 국가를 사이버 공격에 취약하게 만든다고 경고
- 규제 완화 기조를 유지해온 트럼프 행정부가 유망 AI 기업을 재차 안보 위협으로 규정하면서, 명확한 연방 AI 규제와 법안을 마련해야 한다는 정치권과 대중의 요구도 확산
- 마크 워너(Mark Warner) 상원의원은 제재가 위험에 기반한 투명한 절차에 따라 이루어져야 한다며, 행정부의 앤트로픽 적대성을 지적하고 의회 차원의 프런티어 모델 시험·승인 입법 체계 마련을 촉구

사이버보안 전문가들, Fable 5와 Mythos 5에 대한 수출통제 철회 요구

■ **사이버보안 전문가들도 제3자 연구에서 가드레일 우회가 입증되지 않았으며 시연된 역량이 사이버 방어에 필수적인 기능이라고 반박하며, 수출통제 철회를 요구**

- 연구자들은 보고된 탈옥 사례가 취약점이 있는 오픈소스 코드 검토를 요청한 후 다단계 수동 프로세스로 패치 테스트 스크립트를 생성한 것으로 사이버보안 업무의 핵심 기능에 해당한다고 설명
- 일부 사이버보안 업계 관계자들 '페이블 해방(Free Fable)' 공개서한에 서명하고, 미토스급 모델이 취약점 탐지와 악용에 "매우 우수하나 해당 모델에 국한된 역량은 아니다"라고 강조

출처 | MIT Technology Review, Three things to watch amid Anthropic's latest feud with the government, 2026.6.22.
CyberScoop, Cybersecurity experts don't think Anthropic's Fable 5 presents a unique threat, 2026.6.15.
Freefable, On Transparent AI Cyber Protections, 2026.06.14.

SPRi AI Brief
Ⅰ2026년 7월호Ⅰ

기술·연구

마이크로소프트, AI의 지식노동 균등화 가능성과 격차 해소 과제 분석

KEY Contents

- 마이크로소프트 연구진에 따르면 생성형 AI는 지식노동 영역에서 집중적으로 활용되고 있으며, 지식노동자는 전체 노동자 평균 대비 월등히 높은 채택률을 기록
- AI는 지식노동 생산성을 유의미하게 향상하나, 업무 유형에 따라 AI의 균등화 효과는 다르게 나타나며 인구통계학적·지역적 격차 해소를 위해서는 제도적 개입이 필요

지식노동에서 AI의 균등화 효과는 명확히 정의된 객관적 업무 위주로 형성

■ 마이크로소프트 연구진이 2026년 5월 27일 네이처(Nature)에 생성형 AI가 지식노동을 균등화할 가능성*을 분석하고 격차 해소를 위한 개입 방향을 제시한 연구 논문을 발표

* AI가 지식노동의 진입장벽과 성과 격차를 낮춰서, 숙련도나 경험이 낮은 사람도 더 높은 수준의 지식노동을 수행 가능

- 지식노동자는 전체 인구 대비 월등히 높은 생성형 AI 채택률을 기록해, 2024년 조사 기준 전 세계 지식노동자의 75%가 생성형 AI 사용을 보고했으나 미국 전체 노동자 평균은 21%에 불과
- 마이크로소프트 Bing Copilot 대화의 73%가 업무·창작·웹 개발 등 지식노동 영역에 해당하며, 앤트로픽 Claude의 업무용 대화 중 50% 이상이 컴퓨터 과학·수학·창작·정보 과제로 구성
- 여러 업종에 걸친 M365 Copilot 무작위 현장 실험 결과, 이메일 처리 속도는 7~18% 향상되었고 문서 완성 속도는 반일에서 하루로 단축되는 등 작업 속도와 완성도 양쪽에서 유의미한 생산성 향상 확인

■ AI 혜택의 균등화 효과는 과제가 명확히 정의되고 성공 기준이 객관적인 업무에 두드러지며, 개방형 창의적 과제에서는 고숙련자에게 편중되는 경향이 확인

- 보도자료 작성·법률 시험·고객 지원 실험에서 사전 점수가 낮은 참가자의 성과 개선이 더 컸으며, 콜센터 실험에서 초보·저숙련 직원 생산성은 34% 향상되었으나 경험 많은 직원에게는 유의미한 효과 미확인
- 케냐 기업인 대상 ChatGPT 활용 실험에서는 사전 고성과자 수익이 15% 증가했으나 저성과자 수익은 약 10% 감소했으며, 이는 AI가 제안한 조언 중 적합한 것을 선별하는 판단력 차이에 기인

인구통계학적·지역적 격차를 넘어선 지식노동의 민주화를 위해 제도적 개입 필요

■ 교육 수준, 소득, 성별, 나이, 지역에 따른 생성형 AI 사용 격차가 미국·유럽·아시아 등 47개국 이상의 설문과 사용 데이터에서 공통적으로 확인되며, 제도적 개입 없이는 격차 심화 우려

- 고학력·고소득·청년층의 AI 사용률이 저학력·저소득·고령층보다 유의미하게 높으며, 성별 격차도 지속되어 남성이 여성보다 일관되게 높은 채택률을 기록
- 노르웨이 경영대학원 연구에서 AI 사용을 명시적으로 장려한 경우 성별 격차가 완전히 해소되었으며, 기업 차원의 교육·채택 권장 시 성별 격차는 13%p에서 2~3%p로 축소
- 저소득국에서 AI가 기존 디지털 인프라를 건너뛰는 도약 기술로 기능할 가능성이 초기 연구에서 확인되며*, 저자원 언어·문화 다양성 확보와 인터넷·광대역 접근 확대가 핵심 정책 과제로 제시

* 시에라리온 교사 대상 연구에서 AI 챗봇 질문 비용이 웹 검색보다 98% 저렴하고 데이터 사용량은 평균 3,000배 감소

출처 | Nature Computational Science, AI and the democratization of knowledge work, 2026.5.27.

구글 리서치, I/O 2026에서 최신 AI 연구 성과 공개

KEY Contents

- 구글 리서치가 2026년 5월 열린 구글 I/O 2026에서 AI 기반 과학 발견 도구 'Gemini for Science'를 발표하고 실증 연구 지원(ERA) 시스템과 Co-Scientist를 핵심 기반 기술로 제시
- 의료 AI 연구에서는 사용자 증상 관련 대화 데이터를 분석해 진단하는 Symptom AI, 다중 에이전트 연구 시스템 AIME 등을 개발했으며 기상 예측과 개발자 생산성 분야에서도 성과를 확인

과학 연구 자동화, 의료 AI 도구 등으로 AI의 실세계 적용 범위를 확대

■ 구글 리서치가 I/O 2026에서 AI 기반 과학 발견 도구 'Gemini for Science'를 비롯해 헬스케어, 기상 예측, 개발자 생산성 등 다양한 분야를 망라하는 AI 연구 성과를 발표

- 과학 연구 자동화를 위한 Gemini for Science는 '실증 연구 지원(ERA, Empirical Research Assistance)' 시스템과 다중 에이전트 공동 연구 시스템 'Co-Scientist'를 핵심 기반 기술로 포함
- ERA는 과학자가 전문 수준의 실증 소프트웨어를 작성할 수 있도록 지원하는 AI 코딩 시스템으로, 수천 개의 코드 변형을 최적화하여 신경과학, 우주론 등 다양한 분야의 발견 가속화에 기여
- Co-Scientist는 제미나이 기반 다중 에이전트 시스템으로 가설 생성과 평가, 정제를 반복 수행하며 연구자들은 이를 사용해 항생제 내성부터 식물 면역 등 과학적 난제를 해결

■ 헬스케어 AI 분야에서는 건강과 웰빙 전반의 지원을 위한 기초 연구를 바탕으로 개인 맞춤형 AI 서비스를 강화하는 한편, 임상 환경에서 AI의 잠재력 개발을 지원

- 사용자 증상 관련 대화 데이터를 분석해 추론하는 AI 연구를 위해 설계된 실험 도구 'Symptom AI'를 개발했으며, 블라인드 비교 연구 결과 AI 감별 진단이 다른 임상의 진단 대비 약 2배 선호
- 구글 리서치와 구글 딥마인드가 개발한 다중 에이전트 시스템인 AIME는 복잡한 사례와 의료 대화 기록을 해석 및 추론하며, 병력, 검사 결과, 복잡한 의료 영상 등 다양한 형식의 데이터를 지원
- 의료 특화 개방형 가중치 모델 'MedGemma'는 멀티모달 의료 텍스트 및 임상 추론, 영상 이해에 특화되어 현재까지 500만 건 이상의 다운로드를 기록

■ 기상 예측 분야에서는 기상 AI 모델 'WeatherNext'가 2025년 10월 허리케인 멜리사 상륙 5일 전 급속 강화와 자메이카 상륙 지점을 높은 신뢰도로 예측해 사전 대피 경보 발령에 기여

- 또한 도시 홍수 예측을 위해 20년 치 비정형 뉴스 보도를 제미나이로 변환해 260만 건 규모 데이터셋을 구축하고, 플러드 허브(Flood Hub)를 통해 150개국 20억 명 대상 홍수 예측 정보를 제공

■ 구글은 개발자 생산성 지원을 위해 에이전트 기반 개발 플랫폼 'Antigravity 2.0'도 I/O 2026에서 공개했으며, 이 플랫폼은 여러 로컬 에이전트의 병렬 관리와 작업 자동화를 지원

- 구글 리서치가 구글 내 여러 팀과 협력해 Antigravity에 도입한 /teamwork-preview 에이전트는 며칠씩 걸리는 복잡한 장기 소프트웨어 엔지니어링 작업을 몇 시간으로 단축

출처 | Google Research, A New Era of Discovery: Google Research at I/O 2026, 2026.5.28.

앤트로픽, AI 기반 사이버 위협의 MITRE ATT&CK 매핑 분석 결과 공개

KEY Contents

- 앤트로픽이 지난 1년간 악의적 활동으로 차단된 832개 계정을 MITRE ATT&CK* 프레임워크에 매핑해 분석한 결과, AI 기반의 사이버 공격 고도화 현상이 확인
- AI가 인간을 대신해 고난도 작업을 수행하면서 기존 위험 평가 지표의 변별력이 약해지고 있으며, 에이전틱 AI의 공격 행동을 포착하지 못하는 한계도 노출

AI 기반 사이버 공격 고도화 및 기존 위험 평가 지표의 변별력 약화 현상 확인

■ 앤트로픽의 프런티어 레드팀이 2025년 3월부터 2026년 3월까지 1년간 악의적 사이버 활동으로 차단된 832개 계정을 MITRE ATT&CK* 프레임워크에 매핑해 분석한 결과를 발표

* MITRE가 개발·운영하는 글로벌 사이버 위협 행위자의 전술·기술·절차(TTP) 표준 지식 베이스

- 이번 분석에 포함된 계정들은 초기 정찰부터 최종 공격까지 MITRE ATT&CK 전반에 걸친 다양한 전술과 하위 기술에 AI 모델을 사용한 것으로 확인

■ 분석 결과, 위협 행위자들이 초기 침투 이후 복잡하고 난도 높은 후속 단계에 AI를 적용하면서 전반적 위협 수준을 한층 끌어올리고 있다는 점이 명확히 입증

- 832개 계정 가운데 560개(67.3%)가 악성코드 작성에 AI를 활용해 공격 준비 단계의 활용이 최대 비중을 차지했으며, 54개(6.5%)는 침해된 네트워크 내에서 측면 이동하는 심층 단계에 AI를 활용
- 분석 기간 내 초기 접근 단계(예: 피싱)에 AI를 활용하는 공격 비중은 8.6% 감소했으나, 침해 후 계정 탐색에 활용하는 비중은 8.9% 증가해 AI 활용 공격의 심화 추세를 반영

■ AI가 인간을 대신해 고난도 작업을 수행하면서 사용 기법 수나 사용 플랫폼 유형과 같은 기존 위험 평가 지표의 변별력이 약해지는 현상도 확인

- 최저 숙련 행위자가 평균 약 16개 기법을 사용했고 최고 숙련 행위자가 사용한 기법도 약 20개에 그쳐 사용 기법 수와 악성 행위자의 숙련도 간 상관관계가 사실상 붕괴
- 클로드 코드, API, 채팅 인터페이스 등 사용 플랫폼 유형도 행위자의 실제 위험 수준과 유의미한 상관관계를 나타내지 않는 것으로 확인
- 고위험 행위자들은 계정 탐색이나 측면 이동, 권한 상승 등 AI가 공격 각 단계를 자율적으로 실행하도록 설계한 스캐폴딩(Scaffolding) 구조 구축에서 핵심 차별화

에이전틱 AI의 공격을 포착하지 못하는 MITRE ATT&CK 프레임워크의 한계 노출

■ MITRE ATT&CK 프레임워크는 에이전틱 AI의 공격을 포착하지 못하는 구조적 한계를 노출했으며, 분석팀은 AI 에이전트 역량 향상에 따라 인간 개입을 최소화한 자율 실행 공격의 급증을 예상

- 일례로 '25년 11월 앤트로픽이 차단한 국가 지원 첩보 작전은 13개 전술에 걸쳐 30개 기법을 사용해 ATT&CK 기준 중간 위험으로 과소평가
- 그러나 에이전틱 AI의 자율 실행 비중을 반영한 앤트로픽의 자체 위험 점수 산정에서는 최고치인 100점을 받아 기법 수에만 기반한 기존 평가 방식의 구조적 맹점을 시사

출처 | Anthropic, What we learned mapping a year's worth of AI-enabled cyber threats, 2026.6.3.

앤트로픽, AI의 재귀적 자기개선 임박 경고

KEY Contents

- 앤트로픽이 현재 자사 코딩 작업의 80%를 클로드 코드가 수행하는 등, AI 모델이 빠르게 발전해 인간의 개입 없이 스스로 성능을 향상하는 '재귀적 자기개선' 단계가 임박했다고 경고
- 앤트로픽은 AI의 재귀적 자기개선이 가져올 막대한 파급효과에 대처하기 위해 첨단 AI 개발 속도를 늦추거나 일시적으로 중단하는 방안을 제안하고 이를 위한 국제 협력의 필요성을 강조

AI의 재귀적 자기개선에 대응해 첨단 AI 개발 속도 조절 필요

■ 앤트로픽이 공개 벤치마크 데이터와 내부 데이터를 바탕으로 AI가 인간의 개입 없이 스스로 성능을 향상하는 '재귀적 자기개선(Recursive Self-Improvement)' 단계가 임박했다고 주장

- AI가 단독으로 완료할 수 있는 작업 소요 시간은 약 4개월마다 2배씩 증가해, 이러한 추세가 지속되면 올해 안에 전문가가 며칠 걸리는 작업을, 2027년에는 몇 주가 걸리는 작업을 처리할 수 있을 전망*

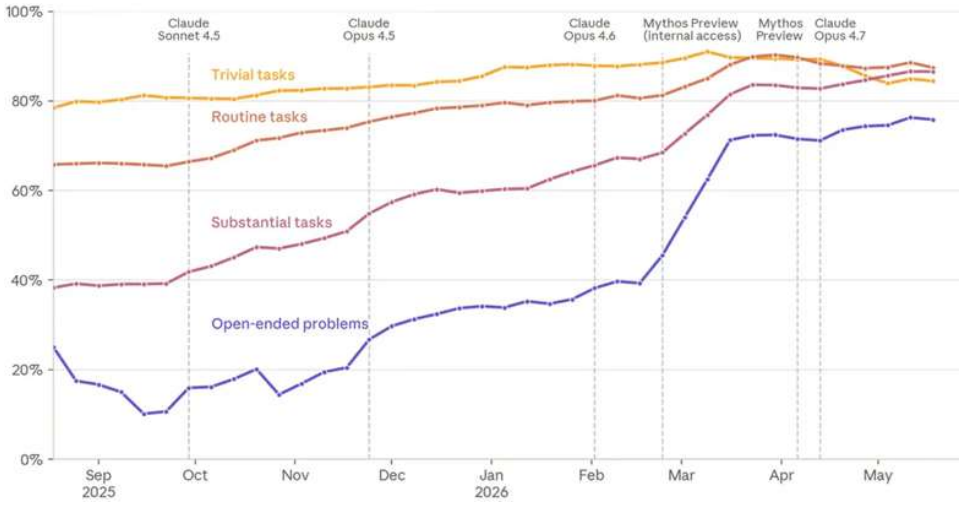
* 2024년 3월 Claude Opus 3은 사람이 약 4분이 걸리는 과제를 완료했으나, 2년 뒤 Opus 4.6은 12시간 소요 과제를 수행

- AI 모델의 SWE-bench(소프트웨어 엔지니어링) 점수는 2년 만에 한 자릿수에서 최대치로 상승했으며, CORE-Bench(연구 재현성 평가)는 2024년 약 20%에서 15개월 후 최대치에 도달

■ 앤트로픽 코드의 상당 부분도 클로드에 의해 작성되며, 2026년 5월 기준 앤트로픽 코드베이스에 병합되는 코드의 80% 이상이 클로드로 생성(2025년 2월에는 한 자릿수 초반)

- 앤트로픽이 코드 수정에 개입하는 비율도 지난 1년간 꾸준히 감소했으며, 가장 까다로운 작업에서 클로드 성공률은 2026년 5월 기준 76%에 달해 6개월 만에 50%p 상승

그림 Claude Code 세션의 성공률 개선 추이



■ 앤트로픽은 이러한 증거를 바탕으로 AI 개발 과정에서 인간의 역할이 점차 축소되고 있다며, 사회 제도와 AI 정렬 연구가 기술 발전 속도를 따라갈 수 있도록 AI 개발 속도의 조절을 제안

- 개발 속도의 조절을 틈타 적대적 세력이 은밀하게 앞서나가는 사태를 방지할 수 있도록 여러 기관과 협력해 신뢰할 수 있는 개발 속도 조절이나 중단에 필요한 시스템 구축 관련 연구를 수행할 계획

출처 | Anthropic, When AI builds itself, 2026.6.4.

구글 딥마인드, 범용지능(AGI)에서 초지능(ASI)으로 진화 경로 분석

KEY Contents

- 구글 딥마인드 연구진은 AGI를 넘어 ASI로 발전하는 과정에서 나타날 수 있는 기술 경로와 제약 요인들을 심층적으로 분석한 논문을 공개
- 연구진은 불확실한 미래의 기술 발전 속도와 사회적 영향에 대비하기 위해 지속적인 측정과 예측, 그리고 다학제적 연구의 필요성을 강조

AGI에서 ASI에 이르는 4가지 기술 경로와 마찰 요인 제시

■ 구글 딥마인드 연구진이 일반지능(AGI)에서 ASI(초지능)로 발전하기 위한 네 가지 기술 경로를 분석하고 해당 경로에서 발생할 수 있는 마찰과 병목 현상을 논의

- AGI는 광범위한 인지 작업에서 인간의 중간 수준 성능에 도달하는 시스템으로, ASI는 수년 동안 작업하는 수천 명의 인간 전문가 집단보다 뛰어난 성능을 보이는 시스템으로 정의
- 연구진은 향후 AI의 발전 경로는 극도로 불확실하므로, 다양한 시나리오를 예측 및 평가할 수 있는 새로운 벤치마크 체계와 거시적 지표의 개발, 다학제적 협력이 필요하다고 강조

■ 첫 번째 경로는 현재의 AI 혁신을 가능하게 한 컴퓨팅·모델·데이터의 확장으로, 지난 10년간 지속된 모델 크기와 학습 데이터 규모의 확장은 향후 수년간 지속 전망

- 현재로서는 가장 유망한 발전 경로로 여겨지나, 기하급수적 성장이 경제성 면에서나 하드웨어 생산 및 에너지 측면에서 얼마나 지속될 수 있을지는 불확실
- AI 훈련에 적합한 데이터의 확보와 생산 속도도 규모 확장에 필요한 속도를 따라가지 못하고 있으며, 대규모 모델의 사전학습 패러다임 역시 한계에 도달하거나 수익이 급감하는 상태

■ 두 번째 경로는 알고리즘 패러다임의 전환으로, 데이터·연산·에너지 측면에서 효율성이 더 높은 알고리즘 및 아키텍처나 새로운 학습 패러다임의 발견 등을 포함

- 그러나 새로운 패러다임이 어떤 형태가 될 것이며, 그에 따른 에너지와 컴퓨팅 자원, 데이터 수요를 예측하기 어려우며, 새로운 아이디어 발견을 위해 투입될 연구 자원 역시 마찰 요인으로 작용

■ 세 번째 경로는 AI가 스스로 연구와 개발을 자동화하여 성능을 높이는 재귀적 자기 개선 방식으로, AI 기능의 폭발적 증가로 이어질 이론적 가능성도 존재

- 그러나 AI 연구 개발이 완전히 자동화되더라도 모델 학습과 실험 수행, 하드웨어 개발에 필요한 시간과 컴퓨팅 자원, 에너지, 경제적 투자로 인해 지능의 폭발적 성장이 저해될 것으로 예상

■ 네 번째 경로는 복수의 AGI 에이전트가 협력하여 집단 지성을 형성하는 방식으로, 복잡한 문제를 하위 작업으로 분해하고 전문 에이전트에 할당함으로써 단일 지능의 한계를 극복

- 현재로서는 해당 방식이 어떤 유형의 문제에 적용되는지, 또한 에이전트 집단을 구성하는 최선의 방식이 무엇인지, 다중 에이전트 확장이 개별 모델 크기 확대보다 효율적인지 여부는 불분명
- 또한 AI 그룹의 규모를 확장하기 위해서는 컴퓨팅 자원과 에너지 공급을 균형 있게 확장해야 하며, 궁극적으로는 경제적 투자가 수반될 필요

출처 | Google DeepMind, From AGI to ASI, 2026.6.10.

SPRi AI Brief
Ⅰ2026년 7월호Ⅰ

인력·교육

미국 노동시장에서 AI가 미치는 영향은 아직 미미

KEY Contents

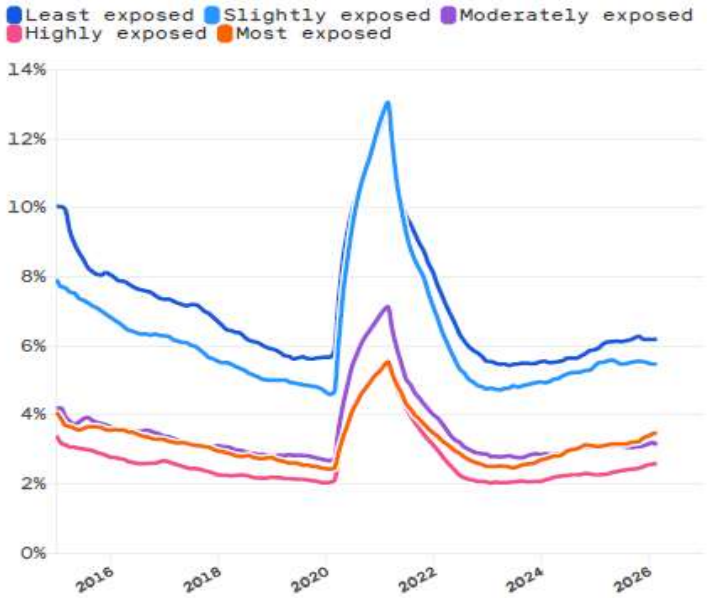
- 미국 노동통계국 데이터 분석 결과 AI 노출도가 높은 직종의 실업률이 비노출 직종보다 낮아 AI가 미국 노동시장에 대규모 영향을 미쳤다는 증거는 확인되지 않음
- 그러나 소프트웨어 개발 등 AI 노출도 최상위 직종에서는 22~25세 신입 근로자의 일자리가 감소하고 경력자의 일자리는 오히려 증가하는 세대 간 격차 확인

직무의 AI 노출도와 경력 단계가 일자리 대체 우려 수준을 좌우

■ MIT 테크놀로지 리뷰가 미국 노동통계국의 데이터를 분석한 결과, AI가 사무직 일자리를 대체할 것이라는 우려와 달리, AI 노출도가 높은 직업군의 실업률이 오히려 더 낮은 것으로 확인

- 스탠포드 경제정책연구소의 에리카 맥엔타퍼(Erika McEntarfer) 연구원에 따르면 AI의 노동시장 영향은 현재 매우 제한적이며, 이는 혁신이 산업 변화로 이어지기까지 시간이 걸린다는 역사적 사실과 일치
- 하버드대 경제학과 데이비드 데밍(David Deming) 교수 연구진의 조사에서도 근로자의 40%가 생성형 AI를 사용하고 있으나 생산성 향상 효과는 아직 경제 전반을 뒤흔들 수준에 미치지 못하는 단계
- 2026년 2월 기준 미국 18~64세 인구의 57.9%가 생성형 AI를 사용하며, 근로자 중 업무 활용 비율은 43.4%로 집계되었으나 생성형 AI 도입으로 절감된 총 근로 시간 비율은 2.2%에 불과

그림 직업군의 AI 노출도와 실업률 비교 – Y축 실업률



■ 그러나 스탠포드 디지털경제연구소 조사 결과, SW 개발 등 AI 노출도 최상위 직종에서는 22~25세 고용 인원이 2024년 이후 16% 감소했으며, 35세 이상 중·고경력 근로자는 증가세를 기록

- ChatGPT 공개 시점인 2022년 11월을 기준으로 청년 초급 인원 감소가 포착됐으며, AI를 자동화 도구로 주로 활용하는 직군에서 고용 감소가 집중되어, 일자리 전환에 대비한 정책적 준비가 시급

출처 | MIT Technology Review, A reality check on the AI jobs hysteria, 2026.5.26.

AI 일자리 충격 대응을 위한 5가지 정책 방안

KEY Contents

- 워싱턴포스트가 AI로 인한 일자리 충격 우려에 대응하기 위한 로봇세 부과·실업급여 확충·직업훈련·AI 수익 배분·현상 유지 등 5가지 정책 방안을 분석
- 각 방안은 재원 확보와 정치적 실현 가능성, 실효성 검증 여부를 둘러싸고 찬반 논쟁이 진행 중이며, 양당 간 정치적 합의와 같은 과제도 산적

로봇세 신설, 실업급여 확대, 직업훈련 프로그램, AI 수익 분배 방안이 주로 거론

■ 워싱턴포스트가 2026년 6월 2일 기사를 통해 AI로 인한 대규모 일자리 감소에 대응하기 위한 5가지 정책 방안과 주요 논점을 정리

■ (로봇세 신설) 노벨경제학상 수상자 대런 애쓰모글루(Daron Acemoglu)를 중심으로 AI 도입을 촉진하는 세제 혜택을 폐지하고 기업·부유층 증세를 통해 재원을 마련해야 한다는 주장이 부상

- AI 사용량에 비례해 기업에 부과하는 'AI 토큰세' 개념이 캘리포니아 주지사 후보 톰 스타이어(Tom Steyer) 등 민주당 정치인들 사이에서 확산 추세
- 엘리자베스 워런(Elizabeth Warren) 민주당 상원의원은 AI 데이터센터 신규 과세와 법인세 인상을 통해 실업급여와 의료보험 확충 재원을 조성해야 한다고 주장

■ (실업급여 확대) 기존 실업급여 제도를 개선해 AI로 인한 직업 전환 과정에서 생계 보호를 추구

- 미시시피주의 주당 최대 235달러에서 매사추세츠주의 1,105달러까지 편차가 큰 실업급여 체계를 개편하고, AI 피해 직종을 정밀 추적하는 데이터 수집 체계를 마련해야 한다는 주장이 부상

■ (직업훈련 프로그램) AI 시대에 대응한 직업훈련 강화는 초당적 공감대를 얻고 있으나, 어떤 훈련이 효과적인지에 대한 검증은 여전히 불충분하다는 지적도 제기

- 지나 레이몬도(Gina Raimondo) 전 상무장관은 AI 시대 노동력 확보를 위해 정부와 기업 간 '새로운 대타협'이 필요하다고 강조하며, TSMC의 애리조나 공장 인력 양성 사례를 모범 사례로 제시
- 예일대 정책연구소 버짓 랩(Budget Lab)의 마사 짐벨(Martha Gimbel)은 훈련 범위가 단기 AI 기술 교육부터 장기 재교육까지 광범위하며 성과에 대한 근거가 부족하다고 지적

■ (AI 수익 분배) 고용 여부와 무관하게 AI 성과를 국민 전체에게 환원하는 방식으로, 정부의 AI 기업 지분 취득이나 직접 지급을 포함하는 다양한 형태로 논의

- 뉴욕 하원 후보 알렉스 보레스(Alex Bores)는 AI 기업 성공 시 대중에게 주식을 배분하고 일자리 손실 발생 시 직접 지급을 발동하는 'AI 배당(AI dividend)' 구상을 제안
- 살 칸(Sal Khan) 칸아카데미(Khan Academy) 설립자는 AI로 근로자를 대체한 기업이 이익의 1%를 자발적으로 재훈련 재원으로 적립하는 방안을 제안

■ (현상 유지) 트럼프 행정부와 공화당 다수 의원은 AI 일자리 손실의 규모와 속도가 불확실한 상황에서 검증되지 않은 대규모 제도 개편이 오히려 경제에 해가 될 수 있다는 시각을 유지

출처 | The Washington Post, "5 ideas for how we survive the possible AI jobs apocalypse", 2026.6.2.

AI 채용 시장, 경력자 중심 구조로 재편되며 신입 일자리 축소

KEY Contents

- AIDE 연구소가 S&P500 기업의 AI관련 채용 공고를 분석한 결과, 전체 공고 중 시니어급이 71%, 주니어급은 13%에 불과해 청년층의 시장 진입이 구조적으로 제한
- AI가 초안 작성이나 반복 업무 등 기존에 신입이 처리하던 업무를 대신 수행하며 주니어급의 직위 자체가 구조적으로 소멸하고 청년 실업률이 높아지는 추세

S&P500 기업 AI 채용 공고의 71%가 시니어급에 집중

■ AIDE 연구소*가 링크드인(LinkedIn)에 게재된 S&P500 기업의 AI 관련 채용 공고를 분석한 결과, 시니어급 비율이 압도적으로 높고 주니어급 비율은 극히 낮은 구조로 확인

* 전 세계 기업들의 AI 성숙도와 도입 수준을 객관적인 데이터로 측정하고 벤치마킹하는 조사 연구기관

- 전체 AI 관련 공고 8,140건 중 시니어급이 71%, 주니어급은 13%, 중간급은 16%로 집계
- AIDE 연구소의 폴 칙(Paul Cheek) CEO는 AI 채용 열풍은 실제로 일어나고 있지만, 이미 상위권에 있는 인재에게만 기회가 집중되는 노동시장의 협소화를 경고
- 기업들이 급변하는 AI 환경에 대응할 수 있는 경험을 갖춘 시니어 인재를 선호하는 경향이 커지면서, 청년 인재가 스타트업 등으로 이탈 시 기업의 장기 경쟁력에 위협이 될 수 있다는 지적도 제기
- AIDE 연구소 측은 경영진이 조직의 모든 단계에서 AI 인재 확보를 우선시해야 하며, 고위직뿐 아니라 미래를 위해 육성하고 있는 중간급과 주니어급 인재도 고려해야 한다고 강조

■ 뉴욕 연방준비은행이 2026년 5월 공개한 데이터에서도 청년층 실업률이 전체 실업률을 상회

- 코로나19 전후까지 별다른 차이가 없던 대학 졸업자 실업률과 전체 실업률의 격차는 점점 벌어지는 추세로, 2026년 3월 기준 최근 대학 졸업자 실업률은 5.6%로 전체 실업률 4.2%를 상회

AI가 초급 업무를 대체하면서 청년층의 고용 부진이 심화

■ 주니어급 직무가 AI로 대체되며 업무 경험을 축적할 경로 자체가 구조적으로 사라지고 있으며, 경력을 쌓을 발판이 없이 경험이 요구되는 역설적 구조가 청년 취업난을 심화

- 초안 작성이나 반복 업무 처리 등 저위험·고빈도의 초급 업무가 AI 대체에 가장 적합해, 결국 남은 역할은 AI 산출물을 감독하는 시니어 업무로 집중
- 익명의 전문직 종사자는 4,000달러를 투자한 재교육 과정이 8개월 만에 첨단 AI 모델로 인해 쓸모없어진 사례를 소개하며, 업스킬링이 지속적으로 비용을 소모하는 쳇바퀴로 전락했다고 지적

■ 미국의 급여 관리 플랫폼 ADP의 수석 경제학자 넬라 리처드슨(Nela Richardson)은 AI가 일자리 수를 바꾸는 데 그치지 않고 직무 내 세부 과업 자체를 재편하고 있다고 진단

- 그는 AI의 등장으로 직무의 진입 지점이 상향 이동한 만큼 청년층이 고부가가치 업무로 도약할 수 있도록 지원하는 기업의 역할이 필요하다고 강조

출처 | CNN, AI is sparking a jobs boom — just not for newbies, 2026.6.15.

보스턴대·펜실베이니아대 공동 연구팀, AI 주도 해고의 시장 실패 분석

KEY Contents

- 보스턴대와 펜실베이니아대의 공동 연구 결과, 개별 기업은 AI 주도 해고에 따른 이익 전부를 갖지만 해고로 인한 수요 감소 비용은 시장 전체에 분산되어 자동화 군비경쟁을 초래
- 주요 정책 대안을 분석한 결과, 기업이 시장에서 제거하는 수요에 비례한 과세만이 과도한 자동화 유인을 실질적으로 교정할 수 있으며, 세수는 재교육 재원으로 활용 가능

경쟁 시장 구조가 기업의 과잉 자동화를 유발하는 시장 실패 규명

■ 미국 보스턴大와 펜실베이니아大 연구진이 AI 주도 해고가 경제 전반의 소비 수요를 잠식하는 시장 실패 구조를 분석한 논문 'AI 해고 함정(The AI Layoff Trap)'을 발표

- AI 주도 해고로 인한 비용은 시장 전체에 분산되는 반면, 절감 이익은 개별 기업에 귀속되어 인지 여부와 상관없이 해고를 억제하지 못하는 구조적 시장 실패를 규명
- 기업이 노동자를 AI로 대체하여 절감된 비용은 해당 기업에 귀속되지만, 노동자 소득 상실에 따른 수요 감소 비용은 경제 전체에 분산되는 비대칭적 유인 구조를 형성

■ AI가 저렴해지고 성능이 고도화될수록 경쟁 압력에 의해 자동화가 가속화되며, 이는 개별 기업의 선견지명만으로는 방지할 수 없는 '자동화 군비 경쟁'을 초래

- 과도한 자동화 효과는 경쟁이 가장 치열한 시장에서 극대화되며, 반대로 독점 기업은 효율적 수준에 가장 근접하게 자동화하는 것으로 분석
- 과도한 자동화는 단순히 노동자의 손실이 기업주 이익으로 전환되는 분배 문제를 넘어, 노동자와 기업주 양측이 모두 절제하는 시나리오 대비 더 나쁜 결과를 초래하는 구조적 시장 실패를 초래
- 유연한 임금 조정, 신규 기업 진입, 기업주의 이윤 재소비 등 통상적 균형 회복 기제는 문제 발생 시점에 영향을 줄 수 있지만 근본적 왜곡을 제거하지는 못하는 것으로 분석

기업이 시장에서 제거하는 수요분에 비례한 과세로 과도한 자동화 억제 가능

■ 광범위하게 논의된 6가지 정책 대안을 검토한 결과, 이 중 5가지는 자동화 인센티브에 영향을 주지 못하며 자동화 피구세*만이 유일하게 시장 실패를 교정할 수 있는 수단으로 확인

* 외부효과 해결을 위해 외부성을 일으킨 사람에게 비용의 차이만큼 부과하는 세금으로 영국 경제학자 피구가 주장

- 보편적 기본소득은 생활 수준을 높이지만 각 기업의 자동화 인센티브를 변경하지 못하며, 자본소득세와 노동자 이윤 공유 방식도 유사한 이유로 자동화 결정에 영향을 미치지 못함
- 기업 간 자발적 협약의 경우, 다른 기업의 행동과 무관하게 자동화를 택하는 것이 각 기업의 최선 전략이 되어 구조적으로 불안정
- 과도한 자동화를 실질적으로 억제하는 유일한 수단은 기업이 시장에서 제거하는 수요분에 비례해 과세하는 자동화 피구세로, 이를 통해 왜곡된 자동화 유인을 직접 교정 가능
- 이를 통해 확보된 세수는 재훈련 프로그램에 투입해 장기적 문제 완화에 활용될 수 있으며, 단일 국가의 피구세 단독 도입 시 자동화가 역외로 이전될 수 있어 국제 공조를 통한 도입 필요

출처 | Business Wire, New BU Research Reveals an “AI Layoff Trap” That Harms Both Workers and Firms, 2026.6.10.

MIT 경제학자들, 노동자 역량을 확장하는 '친노동 AI'의 가치 강조

KEY Contents

- MIT 경제학자들이 기술 유형을 5가지로 분류하고 이중 인간이 수행하는 가치 있는 업무의 범위를 확장하는 신규 과업 창출만을 친노동 기술로 평가
- 이들은 인간 역량 확장에 AI를 활용하는 기업들이 혁신과 인재 유지, 차별화된 전문성 구축 등 경쟁 우위를 확보할 수 있다며 친노동 AI의 가치를 강조

인간의 전문성 수요를 높이는 신규 과업 창출 기술을 친노동 AI 기술로 규정

■ 대런 애쓰모글루를 비롯한 MIT 경제학자들이 인간 전문가에 대한 수요를 늘리는 방식으로 AI를 활용하는 '친노동 AI(Pro-Worker AI)' 개념을 제시

- 이들은 기술을 ▲노동 증강 ▲자본 증강 ▲자동화 ▲전문성의 평준화 ▲신규 과업 창출 5개 유형으로 분류
- 5가지 기술 유형 중 신규 과업 창출 기술만이 기존 전문성을 무력화하지 않고 새로운 형태의 인간 전문성 수요를 창출하는 뚜렷한 친노동 기술로 평가
- 표면상 유사한 기술이라도 업무에 미치는 영향은 매우 다를 수 있으며, 가령 자동화나 전문성 평준화 기술은 일부 작업자에게 새로운 기회를 제공할 수 있지만 기존 전문가의 희소성을 낮출 우려

■ 신규 과업 창출 기술은 기존 업무의 효율화에 그치지 않고 인간이 수행하는 가치 있는 업무의 범위 자체를 확장한다는 점에서 여타 기술과 차별화

- 가령 전기 분야에서 이더넷 네트워크, 광섬유 케이블, 재실 감지 냉난방, 조명 시스템 등은 건물의 복잡성을 높이고 이를 설계·설치·유지하는 전문 인력 수요를 새롭게 창출
- 사이먼 존슨(Simon Johnson) 교수는 "이전에 하지 못했던 새로운 일을 할 수 있도록 인간의 역량을 확장하면 인간 전문 지식의 가치가 높아지는 경향이 있다"며 이를 AI의 핵심 시험대로 강조
- 특허 심사관이 AI 기반 선행 기술 검색 도구를 활용하는 사례는 업무 속도만 개선되면 노동 절감 효율에 머물지만, 심층 분석을 가능하게 하면 전문적 판단력의 가치를 높이는 친노동 기술로 전환

정책 수단을 통해 친노동 AI 관련 연구와 투자로 전환 유도 필요

■ 자동화 위주의 AI 도입이 비용 효율성 향상에 그친다면, 인간 역량 확장에 AI를 활용하는 기업은 고객 경험 혁신, 인재 유지, 경쟁사와 차별화되는 전문성 구축 등의 경쟁 우위를 확보

- 존슨 교수는 기업들이 친노동 접근 방식을 취하지 않으면 많은 기회를 놓칠 수 있다며, 노동자 중심의 사업 타당성 분석을 통해 직원의 역량을 향상하고 혁신을 주도할 수 있다고 강조

■ 연구진은 친노동 AI 연구와 투자를 이끌어내기 위한 정책 수단으로 보건과 교육에 대한 공공 투자, 정부의 AI 평가 역량 강화, 노동자 중심 AI 기술 대상 보조금 등을 거론

- 노동 대비 자본을 우선시할 유인을 줄이는 세제 개편, 반독점법 집행 강화, 노동자의 의견 존중, 노동자의 전문성에 대한 지식재산권 보호, 직업 면허 제도 개편 등도 제시

출처 | MIT Sloan, Pro-worker AI, explained, 2026.6.17.

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