AI 뉴스와 용어
월·수·금 오전 7시

METAL
First, AI Magazine

무료 · 언제든 수신거부
metallab.ai

제 24 호 | 2026년 9월 15일 화요일 | 오늘의 AI

비즈니스

앤스로픽·오픈AI·구글, AI 표준기구 논의

세 회사가 7월부터 실무 회의를 열어 업계가 주도하는 AI 표준기구 설립을 검토해 왔다는 보도가 나왔어요. 같은 달 공개된 데미스 하사비스의 설계도에는 출시 30일 전 사전 검토와 의무화 단계까지 적혀 있어요.

비즈니스 · 앤스로픽·오픈AI·구글, AI 표준기구 논의

AI 회사들이 자기를 심판할 기구를 스스로 세우는 방안을 비공개로 논의해 왔어요. 복수의 소식통을 인용한 보도에 따르면 앤스로픽과 오픈AI, 구글은 올해 7월부터 실무 그룹 회의를 이어 오며 업계가 주도하는 AI 표준기구 설립을 검토했어요. 논의된 틀은 이 회사들이 자사 모델을 평가받을 안전 기준을 직접 세우는 구조예요. 규제를 둘러싼 말싸움이 바깥에서 가장 시끄러웠던 몇 달이, 안에서 그 회의가 열린 몇 달과 정확히 겹쳐요.

회의는 닫혀 있었지만 같은 달에 공개된 설계도가 하나 있어요. 데미스 하사비스 구글 딥마인드 최고경영자는 7월 14일 프런티어 AI 표준기구 구상을 직접 공개했어요. 그가 모델로 삼은 건 미국 증권업계의 자율규제기구예요. 업계가 돈을 대되 연방 감독 아래 독립적으로 운영되고, 정부 기관이 아니라 자율규제조직으로 서는 형태죠.

메탈이 확인한 그 제안문의 뼈대는 출시 전 사전 검토예요. 회사들이 모델을 내놓기 30일 전에 기구에 먼저 넘겨 검토받되, 처음에는 자발적으로 시작하고 효과가 입증되면 미국 시장에 내놓기 위한 요건으로 굳힌다는 3단계 설계예요. 마지막 단계에서는 기구가 외부에 공개하지 않는 자체 시험 문항을 따로 만들어요. 회사들이 이미 알고 있는 벤치마크에 맞춰 모델을 다듬는 일을 막기 위해서예요.

코히어 고메즈, AI 규칙 카르텔 반박 에세이

시험 과목도 적혀 있어요. 사이버보안 취약점과 생물학적 위협 가능성, 그리고 에이전트가 스스로 하는 행동이 국가 안보와 관련된 평가 영역으로 꼽혔어요. 하사비스는 모델이

기구의 권한 범위에 대해서는 하사비스 본인이 선을 그었어요. 그는 표준기구가 "평가 규약을 개발하고, 국가 안보와 관련된 영역의 시험을 수행하기 위해 적절한 연방 기관 및 미국 국립연구소들과 협력할 책임을 지게 될 것" 이라고 적었어요. 규제 기관이 아니라 기술 표준을 만드는 조직이라는 뜻이고, 판정의 최종 권한은 정부 쪽에 남겨 둔 셈이에요. 이사회에는 독립 기술 전문가와 오픈소스 진영 대표가 들어가야 한다는 조건도 달렸어요.

딥마인드, WeatherNext 3 발전량 예보 소개

왜 지금이냐는 질문에도 그가 먼저 답을 적어 뒀어요. 하사비스는 "지금 우리는 극도로 치열하고 여러 겹으로 얽힌 상업적, 지정학적 경주에 갇혀 있습니다" 라고 진단했어요. 경주에 갇힌 선수들이 심판을 만들자고 모이는 장면인데, 그 심판의 월급은 선수들이 내요. 하사비스도 이 대목을 피하지 않고 기술 인재를 모으고 대규모 시험을 감당하려면 재원이 상당해야 한다고 적었어요.

주말 사이에 이 구상은 비공개 회의실 밖으로 한 발 나왔어요. 사티아 나델라 마이크로소프트 최고경영자는 9월 13일 정렬을 제대로 맞추기 위한 신중한 속도 조절을 환영한다고 밝히고, 자사 MAI 모델에 적용할 행동강령을 하루 뒤 공개해 의견을 받겠다고 했어요. 그는 "초지능을 향한 어떤 시도든, 우리가 만드는 AI가 인류를 돕지 않고 인간의 통제 아래 있지 않다면 추구할 가치가 없다는 핵심 원칙에 뿌리를 두어야 합니다" 라고 적었어요. 거버넌스가 소수의 주체에 통제되어서는 안 되며 생태계와 국가, 학계를 포함한 여

를 조절하고 외부 평가자를 회사 안에 들이자는 에세이를 발표했다고 보도한 바 있어요. 나델라가 환영한 신중한 속도 조절이 바로 그 제안이에요. 공개 성명과 비공개 실무 회의가 같은 방향을 가리키고 있는 거예요.

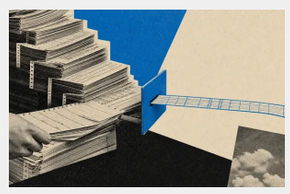
자율이 어디까지 버티는지는 이미 한 번 시험대에 올랐어요. 메탈은 앤스로픽이 9월 1일 내놓은 Claude Mythos 5.1 을 영국 AI안전연구소의 출시 전 평가에 내지 않았다고 보도한 바 있어요. 평가 기관에 모델을 넘기는 일이 자발적인 동안에는 넘기지 않는 선택도 늘 같이 존재해요. 하사비스가 3단계 설계의 마지막을 의무화로 잡아 둔 이유가 여기에 있어요.

같이 모여 속도를 늦추는 일 자체가 법적으로 안전하지 않다는 문제도 남아 있어요. 메탈은 오픈AI가 경쟁사들과 함께 개발 속도를 늦추는 행위가 담합에 해당하는지를 두고 의회에 판단을 요청했다고 보도한 바 있어요. 안전 기준을 함께 세우는 기구는 그 질문의 한복판에 서게 돼요. 경쟁사끼리 같은 기준에 합의하는 순간, 그 기준은 시장에 들어오려는 다른 회사에게는 문턱이 되니까요.

전문가들이 지적한 위험도 정확히 그 지점이에요. 캐스케이드인사이츠 공동창업자 숀 캠벨은 이런 기구가 규제 포획으로 흐를 수 있다고 경고했고, 스즈랩스 최고경영자 마이클 벨은 새로운 관료 조직을 하나 더 만드는 것이 최선인지 물었어요. 정보기술혁신재단의 미셸 로페스 말도나도는 기구를 새로 세우는 대신 기존 표준 기관의 역량을 키우는 쪽을 권했어요. 심판을 선수가 세울 때 나오는 의심은 늘 같아요.

지금 분명한 것은 순서예요. 의회는 나서지 않겠다고 했고 백악관은 속도를 늦출 이유가 없다고 했으며, 그 빈자리를 회사들이 7월부터 자기들끼리 채워 왔어요. 하사비스의 설계도는 공개돼 있고 나델라의 행동강령은 의견 수렴에 부

오늘 지면



코히어, 문서 파싱 모델 Parse 5 출시

메타, 멘로파크 인프라 랩 AI 하드웨어 공개

이 지면은 2026년 9월 15일 화요일 아침 구독자 메일함으로 나간 레터를 신문 판형으로 다시 조판한 것입니다. 기사 전문은 metallab.ai 에서 읽을 수 있습니다.

AI

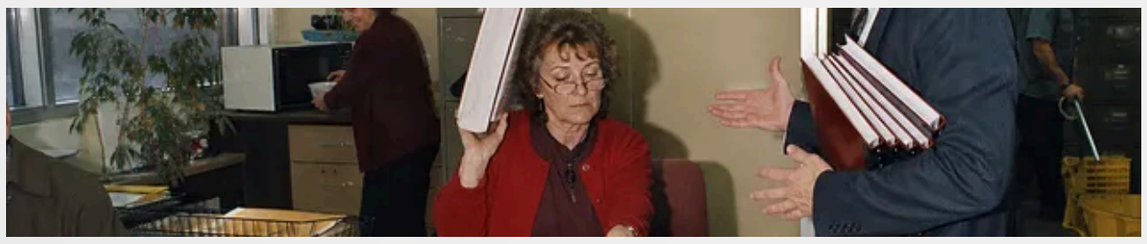
오픈AI, 퍼플렉시티 Astra 테스트 영상 공개

오픈AI 개발자 계정이 9월 14일 X에 68초짜리 영상 하나를 올렸어요. 답변 엔진 퍼플렉시티의 공동창업자 조니 호가 코덱스 안에서 GPT-6 Astra 를 어떻게 쓰는지 말하는 영상이에요. 게시글에 따르면 그는 이 모델로 테스트 하네스를 만들고 외부 API 응답을 흉내 내게 해서, 부품들이 함께 돌아가는지를 확인해요. 오픈AI 사이트의 고객 사례 페이지도 같은 내용을 담고 있고, 그 페이지에는 9월 14일 날짜가 적혀 있어요. 메탈이 확인한

▶ 2면에 계속

비즈니스

미다스 프로젝트, 오픈AI SB 53 위반 지적



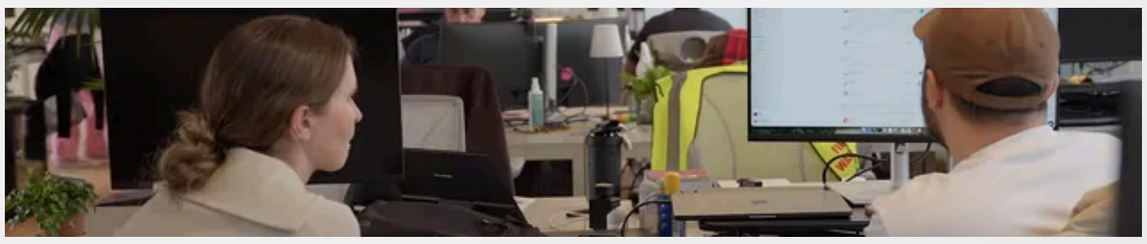
캘리포니아 AI 안전법 SB 53이 시행된 첫해에 오픈AI가 이 법을 세 번 어겼을 수 있다는 분석이 나왔어요. 비영리 감시단체 미다스 프로젝트가 9월 10일 자체 매체에 실은 글에서, 오픈AI가 5월에 법정 안전 프레임워크로 지정한 프런티어 거버넌스 프레임워크(FGF)는 대상 모델마다 통제 상실(loss of control) 위험 등급을 매기도록 하는데, 6월 GPT-5.6 프리뷰와 7월 GPT-5.6, 9월 GPT-6 아스트라의 시스템 카드 어디에도 그 등급이 없다고 지적했어요. 보도에 따르면 오픈AI 대변인은 SB 53 준수에 "자신 있다"고 답했어요.

SB 53의 정식 이름은 프런티어 AI 투명성법(TFAIA)이에요. 2025년 9월 서명돼 올해 초 시행됐고, 가장 큰 AI 개발사들이 위험을 어떻게 평가하고 완화하는지 적은 안전 프레임

▶ 2면에 계속

AI

오픈AI, 파이서 AI 비서 도입 사례 영상 공개



오픈AI가 9월 14일 공식 유튜브 채널에 영국 스타트업 파이서(Fyxer)의 AI 비서 도입 사례 영상을 올렸어요. 파이서는 이메일과 회의, 메신저를 넘나들며 대화의 맥락을 따라가는 AI 비서를 만드는 회사예요. 오픈AI가 8월 13일 홈페이지에 먼저 실은 고객 사례 글에 따르면, 이 회사는 오픈AI 모델에 50만 시간이 넘는 비서 업무 기록을 얹어 사용자마다 다른 말투로 답장 초안을 써 주는 시스템을 만들었어요. 지금은 AI가 쓴 초안의 53%가 한 글자도 고쳐지지 않고 그대로 전송되고, 가입 90일이 지난 사용자의 90% 이상이 계속 돈을 내며 쓰고 있다고 회사는 밝혔어요.

메탈이 확인한 3분 16초짜리 영상은 한 고객 이야기로 시작해요. 수백 개 건물과 수백 명을 상대하는 상업 부동산 관리

▶ 2면에 계속

◀ 1면에서 계속

비즈니스

미다스 프로젝트, 오픈AI SB 53 위반 지적

1 미다스 프로젝트가 9월 10일 오픈AI가 캘리포니아 SB 53을 최소 세 번 어겼을 수 있다는 분석을 냈어요.

2 오픈AI가 법정 프레임워크로 지정한 FGF는 모델마다 통제 상실 위험 등급을 요구하는데 세 시스템 카드에 그 등급이 없어요.

3 오픈AI는 대비 프레임워크가 접근법의 토대이며 SB 53 준수에 자신 있다고 밝혔어요.

오픈AI는 5월 28일 FGF를 공개하면서 이 문서가 TFAIA상의 자사 프런티어 AI 프레임워크라고 명시했어요. 메탈이 확인한 22쪽짜리 문서는 사이버 공격, 화학·생물·방사능·핵(CBRN), 유해한 조작, 통제 상실 네 범주에 각각 1단계에서 3단계까지 위험 등급을 두고, 대상 모델마다 임계값 도달 여부를 판정하고 등급에 비례하는 안전장치를 두며 잔여 위험이 수용 가능한 이유를 문서화한다고 적었어요. 통제 상실은 사람이 모델을 신뢰성 있게 지시하거나 수정하거나 종료할 능력을 잃는 데서 오는 위험으로 정의돼요. 판정 결과는 TFAIA에서 투명성 보고서로 부르는 안전·보안 모델 보고서에 기록하고, 출시 때 시스템 카드 등에도 싣는다고 돼 있어요. 문서는 통제 상실 등급이 AI 자기 개선을 빼면 아직 탐색 단계이며 크게 바뀔 수 있다는 단서도 달았어요.

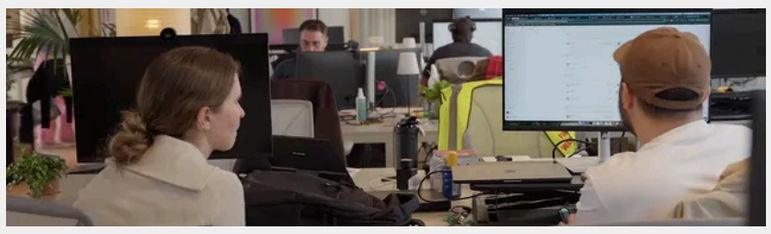
미다스 프로젝트에 따르면 세 모델의 시스템 카드는 전부 오픈AI의 기존 내부 문서인 대비 프레임워크(Preparedness Framework)의 용어와 범주로 위험을 보고했고, 법적 구속력이 있는 FGF는 한 번도 언급하지 않았어요. 사이버와 CBRN은 두 문서가 비슷해서 넘어갈 수 있지만, 통제 상실은 FGF에만 있고 대비 프레임워크에는 없는 범주예요. 아스트라 시스템 카드의 정렬과 감시 가능성 절은 사람이 모델을 지시할 수 있는지를 서술형으로 다루고 실시간 오정렬 감시 장치 같은 안전장치를 설명하지만, 등급 지정과 위험 수용 판정은 없어요. 단체는 FGF가 든 1단계와 2단계의 예시, 그러니까 지시받으면 은근히 성능을 낮추는 모델과 감시 체계의 빈틈을 찾아 이용하는 모델이 아스트라 시스템 카드의 적대적 시험에서 나온 행동과 겹친다고 적었어요. 오픈AI는 세 모델 어느 것에도 별도의 안전·보안 모델 보고서를 내지 않았다고 단체는 밝혔어요.

오픈AI는 성명에서 "우리의 대비 프레임워크는 첨단 AI의 가장 심각한 위험을 관리하는 우리 접근법의 토대로 남아 있습니다. 프런티어 거버넌스 프레임워크는 그런 안전과 보안 관행이 특정 규제 요건과 어떻게 맞물리는지 설명합니다"라고 밝혔어요. 대변인은 새로 나타나는 위험을 평가하고 안전장치를 개발하는 데 많이 투자하며 시스템 카드와 안전 프레임워크로 결과를 공개하고 있다고 덧붙였어요. 아스트라는 대비 프레임워크 기준으로 사이버 범주 최고 단계인 치명적 등급을 받은 모델이에요. 메탈은 GPT-6 아스트라 공개와 함께 이 등급이 붙은 사정을 보도한 바 있어요.

◀ 1면에서 계속

AI

오픈AI, 파이서 AI 비서 도입 사례 영상 공개



1 오픈AI가 9월 14일 공식 유튜브 채널에 영국 스타트업 파이서의 AI 비서 도입 사례 영상을 올렸어요. 8월 13일 홈페이지에 실은 고객 사례 글의 영상판이에요.

2 파이서는 이메일 답장을 30~50개의 작은 전문 모델로 쪼개고, AI 제품 전에 운영한 비서 대행 서비스에서 쌓은 50만 시간 이상의 업무 기록으로 파인튜닝했어요. 사용자가 고친 초안은 DPO 학습 데이터로 다시 들어가요.

3 AI가 쓴 초안의 53%가 수정 없이 전송되고, 90일 뒤에도 사용자의 90% 이상이 유료로 쓰고 있어요. 2025년 연간 반복 매출은 100만 달러에서 3,200만 달러로 늘었다고 회사는 밝혔어요.

이메일 답장이 왜 어려운 문제인지부터 회사는 설명해요. 같은 메일을 두 사람이 받아도 관계와 지난 대화, 지금 하려는 일에 따라 전혀 다른 답이 필요하기 때문이에요. 홀링스워스는 "사람에게 쉬운 일이 컴퓨터에는 어렵고, 컴퓨터에 쉬운 일이 사람에게는 어렵다"는 모라벡의 역설을 들어 이 일이 겉보기보다 훨씬 까다롭다고 인정했어요. 그래서 파이서는 모델 하나에 좋은 이메일을 써 달라고 시키지 않고, 일을 30개에서 50개의 작은 전문 모델로 쪼갰어요. 새 메일이 오면 먼저 답장 여부를 판단하는 모델이 답이 필요한 메일인지, 일정 조율인지, 그냥 알아 두면 되는 정보인지 분류하고, 답이 필요하면 다른 모델들이 의도를 읽고 대화가 회의 잡기로 가는지 요청 처리로 가는지 예측해요.

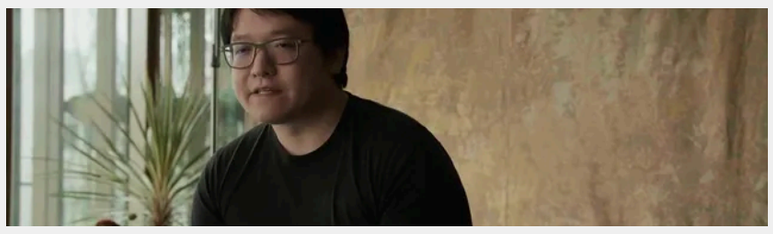
비서실에 비유하면 이해가 쉬워요. 유능한 비서 한 사람이 모든 걸 하는 대신, 접수를 보는 사람, 일정을 잡는 사람, 지난 일을 기억해 두는 사람이 나뉘어 있는 구조예요. 그중 기억 담당이 가장 중요하다고 회사는 적었어요. 어떤 정보를 다음 대화까지 가져가고 어떤 정보는 한 번 쓰고 버릴지 정해야 하는데, 새 메일이 오면 검색 모델이 저장된 대화와 비교해서 그 사람과 그 대화 줄기에 가장 관련 있는 기억만 꺼내 와요. 홀링스워스는 사례 글에서 "문제를 여러 작은 모델로 쪼개는 것이 한 모델에 좋은 이메일을 쓰라고 하는 것보다 훨씬 잘 됩니다"라고 말했어요.

50만 시간이라는 숫자는 AI 제품보다 먼저 있었던 사업에서 나왔어요. 파이서는 AI 제품을 내놓기 전 몇 년 동안 사람이 직접 일하는 비서 대행 서비스를 운영했고, 그 과정에서 언제 바로 답하고 언제 기다리는지, 어느 지난 대화가 지금 중요한지 같은 작은 판단이 주석으로 달린 업무 기록을 쌓았어요. 회사는 이 데이터로 지도학습 방식의 파인튜닝과 적은 비용으로 모델 일부만 고치는 LoRA#Low-rankadaptation)를 써서 작업별 모델 변형을 만들었고, 초기에는 오픈AI의 파인튜닝 플랫폼을, 최근에는 오픈AI의 관리형 파인튜닝 팀과 함께 새 체크포인트를 실서비스에 올렸다고 밝혔어요. AI·ML 제품 엔지니어 조이 드원칙은 "고객에 대해 우리가 배운 것을 모델이 움직이는 방식에 옮겨 넣는 데 오픈AI가 결정적이었어요"라고 말했어요.

◀ 1면에서 계속

AI

오픈AI, 퍼플렉시티 Astra 테스트 영상 공개



1 오픈AI 개발자 계정이 9월 14일 X 영상으로 퍼플렉시티의 GPT-6 Astra 활용법을 공개했고, 사이트의 고객 사례 페이지도 같은 내용이에요.

2 조니 호 공동창업자는 모델에게 테스트 프로그램을 짓게 하고 외부 서비스 응답을 대신 만들게 해 작업 흐름을 끝까지 검증해요.

3 퍼플렉시티는 소통 문안 작성, 소프트웨어 수정, 프로덕션 감시를 맡기고 이전 세대보다 훨씬 덜 들여다본다고 말했어요.

영상에서 가장 먼저 나오는 주장은 코드와 검색의 관계예요. 조니 호는 "모델이 코드를 더 잘 쓰게 될 때마다 검색과 정확도도 좋아집니다" 라고 말했어요. 모델이 웹이나 내부 정보를 뒤지는 코드를 직접 쓰고, 그 결과를 사람이 읽을 짧은 요약으로 합쳐 주기 때문이라는 설명이에요. 답변 엔진에서 코딩 능력은 곁가지가 아니라 검색 품질을 끌어올리는 지렛대라는 뜻이에요.

그다음이 이번 영상의 핵심이에요. 오픈AI 페이지에 따르면 조니 호는 손으로 테스트할 시간이 부족해서, GPT-6 Astra 에게 애플리케이션 주변에 작은 테스트 프로그램을 만들어 달라고 해요. 모델은 언어 모델 API 나 커넥터 같은 다른 서비스가 보낼 법한 현실적인 응답을 만들어 내고, 그 서비스 자리를 대신 서요. 그래서 애플리케이션이 어떻게 반응하는지, 작업 흐름이 처음부터 끝까지 이어지는지를 확인할 수 있어요.

엔지니어의 눈으로 보면 이건 테스트 자동화의 순서를 바꾼 이야기예요. 보통은 사람이 가짜 응답을 미리 써 두고 그 위에서 테스트를 돌려요. 여기서는 모델이 가짜 응답을 만드는 쪽까지 맡아요. 외부 서비스가 어떤 응답을 돌려줄지 모델이 이미 알고 있으니, 그걸 흉내 내는 일도 모델에게 시키는 구조예요. 집을 지을 때 목수가 자기가 쓸 거푸집까지 직접 짜는 것과 비슷해요.

비즈니스

앤스로픽, 금융 자문가용 클로드 플러그인 출시

앤스로픽이 9월 14일 Claude for Financial Advisors 를 내놨어요. 블랙록과 찰스 슈왑, 뱅가드 등 열한 곳의 새 커넥터와 스킬 여덟 개를 묶었고, 투자 추천과 고객 커뮤니케이션은 사람의 승인을 거치게 설계했어요.

1 앤스로픽이 9월 14일 금융 자문가용 클로드 플러그인 Claude for Financial Advisors 를 출시했어요.

2 블랙록·찰스 슈왑·애드파·뱅가드 등 열한 곳의 새 커넥터와 회의 준비·컴플라이언스 등 스킬 여덟 개를 묶었어요.

3 투자 추천과 고객 커뮤니케이션, 컴플라이언스 판정은 사람의 검토와 승인을 거치고 엔터프라이즈 요금제의 감사 로그가 기록을 뒷받침해요.

앤스로픽이 9월 14일 금융 자문가용 클로드 플러그인을 출시했어요. 이름은 Claude for Financial Advisors 이고, 자문가가 매일 쓰는 수탁사와 포트폴리오 플랫폼, CRM, 재무 설계 도구를 클로드에 연결하는 커넥터 묶음에 자문가의 하루를 따라 짠 워크플로 스킬을 얹은 제품이에요. 회사는 발표문에서 조사와 회의 준비, 문서 작업처럼 자문가의 시간을 잡아먹고 고객 앞에 앉을 시간을 빼앗는 일을 덜어 주는 게 목적이라고 밝혔어요. 오픈AI 가 며칠 전 금융사 전용 챗GPT 를 내놓은 자리에 앤스로픽이 자문가 쪽 문으로 들어온 셈이에요.

발표문이 먼저 꺼낸 숫자는 시간이에요. 앤스로픽은 키치스 리서치를 인용해 보통의 자문 회사가 고객 회의에 쓰는 시간이 전체의 6분의 1에 그친다고 적었어요. 나머지는 회의 앞뒤의 일, 여러 시스템에 흩어진 정보를 모아 준비하고 계획하고 논의 내용을 기록하는 데 들어가요. 회사는 EBRI 의 2026년 은퇴 신뢰도 조사를 근거로 미국 노동자 열 명 중 네 명 넘게가 좋은 재무·은퇴 설계 조언을 어디서 받아야 할지 모른다고도 밝혔어요. 준비와 서류에서 아낀 시간이 기존 고객이나 아직 받을 여력이 없던 새 가구로 간다는 게 발표문의 계산이에요.

메탈이 확인한 EBRI 팩트시트 원문은 다섯 쪽짜리이고 노동자 1,007명을 조사했어요. 전문 재무 자문가와 일하고 있다고 답한 노동자는 36%였고, 지금은 아니지만 앞으로 일하겠다는 응답은 47%였어요. 은퇴 설계 정보를 어디서 얻느냐는 질문에는 가족과 친구가 39%, 스스로 하는 온라인 조사가 35%, 전문 자문가가 33% 순이었고, 챗GPT 를 비롯한 AI 도구를 꼽은 응답도 18%였어요. 자문가에게 바라는 도움 1위는 직장 퇴직연금 바깥에서 더 저축하고 투자하는 방법으로 48%였어요. 자문가 시장에 빈자리가 많다는 발표문의 전제는 이 숫자 위에 서 있어요.

AI

코히어 고메즈, AI 규칙 카르텔 반박 에세이

에이든 고메즈 코히어 CEO가 9월 14일 에세이에서 앤스로픽의 속도 조절 제안을 카르텔이라고 불렀어요. 소수 기업이 반독점 예외를 받아 규칙을 쓰는 대신 증거 기반 위험 틀과 모든 회사에 열린 인증을 제안했어요.

1 고메즈는 독립 검토에는 찬성하지만 누가 기준을 쓰고 누가 참여하느냐가 논쟁이라고 적었어요.

2 1975년 신용평가기관 지정과 1985년 유럽 자동차 블록 면제를 안전 명분이 카르텔로 끝난 선례로 들었어요.

3 데이비드 색스도 두 회사가 허락 없이 스스로 속도를 늦추면 된다며 반독점 예외 요구를 비판했어요.

AI 회사 코히어가 앤스로픽의 속도 조절 제안에 정면으로 반대하는 글을 냈어요. 에이든 고메즈 코히어 공동창업자 겸 CEO는 9월 14일 회사 블로그에 누가 AI의 규칙을 정하는가라는 제목의 에세이를 올리고, 소수의 시장 지배 기업이 안전 기준과 기술 진보 속도를 함께 정하겠다는 계획은 안전 장치가 아니라 카르텔이라고 적었어요. 부제부터 AI에 필요한 것은 증거에 기반한 기준이지 카르텔이 아니라는 문장이에요. 다리오 아모데이 앤스로픽 CEO가 9월 12일 에세이로 정부에 반독점 예외를 요청한 지 이틀 만에 나온 반박이에요.

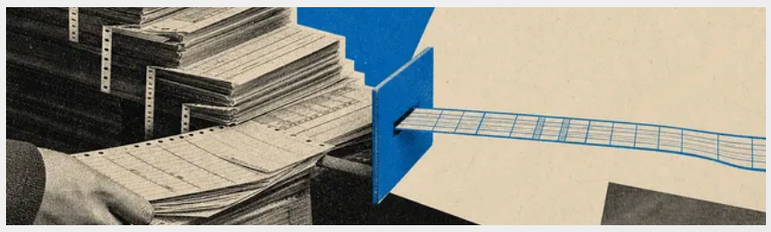
그가 반대하는 지점은 분명해요. 고메즈는 "AI에는 가드레일이 필요합니다. 그것은 논쟁거리가 아니고 한 번도 논쟁거리였던 적이 없습니다" 라고 쓴 뒤, 논쟁은 누가 규칙을 쓰고 누가 참여하며 누구의 이익을 보호하느냐에 있다고 못 박았어요. 고성능 AI 시스템에 대한 독립적 검토 자체는 좋은 생각이고 코히어도 지지한다고 그는 밝혔어요. 문제 삼은 것은 아모데이 계획의 구조예요. 한 나라에 있는 가장 강력한 몇 개 연구소가 공동 기준과 진보 속도의 한계를 합의하고, 경쟁사끼리 생산을 제한하는 합의는 보통 불법이라 정부에 좁은 반독점 예외를 요청하며, 나머지 모든 AI 개발자에게는 그 합의를 따르도록 정부가 강제해 달라고 요구한다는 게 그의 요약이에요. 메탈은 그 에세이가 상주 평가자, 민주주의 국가 안의 조율, 세계적 조율 세 단계로 짜여 있다고 보도한 바 있어요.

변호사 눈으로 보면 이 에세이의 무게는 역사 두 토막에 있어요. 고메즈에 따르면 1975년 미국 증권거래위원회는 자본 규제에 쓸 신용등급이 필요해 세 회사를 국가공인통계평가기관으로 지정했고, 다른 회사가 어떻게 그 자격을 얻을 수 있는지는 끝내 공개하지 않았어요. 25년이 지나도 평가기관은 여전히 셋뿐이었고, 그 셋이 서브프라임 모기지 증권에 최고 등급을 매겨 세계 경제를 무너뜨릴 뻔했다는 거예요. 두 번째는 1985년 유럽의 자동차 블록 면제예요. 제조사들이 자동차는 복잡하고 안전이 중요한 기계라며 누가 팔고 고칠 자격이 있는지 자기들이 정해야 한다고 로비했고, 그 결과 안전한 차가 아니라 기존 업체를 보호하는 시장 구조가 남았으며 유럽집행위원회가 이를 되돌리는 데 약 25년이 걸렸다고 그는 적었어요. 두 사례 모두 명시된 목표는 안전이었고 아무도 카르텔을 만들려고 하지 않았다는 대목이 이 글에서 가장 날카로운 문장이에요.

AI

코히어, 문서 파싱 모델 Parse 5 출시

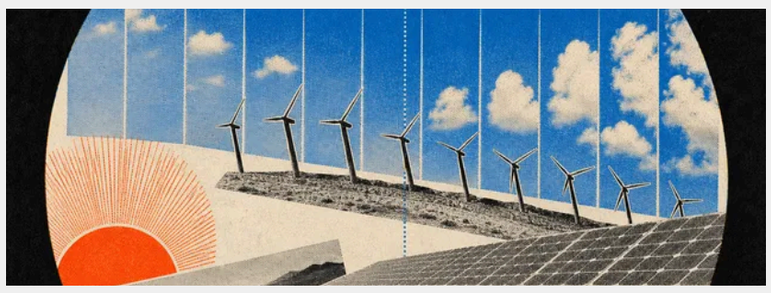
코히어가 8월 27일 출시한 문서 파싱 비전 언어 모델 Parse 5 는 ParseBench 79.2점으로 GPT-5.5 와 Opus 4.8 아래예요. 회사는 그 표를 공개하고 대신 H100 여덟 장에서 분당 2,160쪽이라는 처리량과 비용을 앞세웠어요.



AI

딥마인드, WeatherNext 3 발전량 예보 소개

구글 딥마인드가 9월 14일 X 쓰레드로 WeatherNext 3 의 전력 산업 용도를 소개했어요. 100미터 터빈 높이 풍속과 일사량, 구름의 양을 매시간 예보해 풍력·태양광 발전량을 추산하고 전력망 수요에 맞추게 한다는 내용이에요.



1 구글 딥마인드가 9월 14일 X 쓰레드 3건으로 WeatherNext 3 의 풍력·태양광 발전량 예보 기능을 소개했어요.

2 9월 3일 나온 이 모델은 정지궤도 위성 영상을 직접 받아 매시간 예보하고, 기온·습도는 5킬로미터 격자로 내놓아요.

3 43쪽 논문에 따르면 15일치 64개 구성원 앙상블을 만들고, 시간별 갱신은 초반 예보에서 2~3시간의 선행 시간을 더 벌어요.

구글 딥마인드가 9월 14일 X 에 세 개짜리 쓰레드를 올려 자사 기상 AI 모델 WeatherNext 3 를 전력 산업 쪽으로 다시 소개했어요. 첫 게시글에 따르면 이 모델은 전력망 운영자와 풍력·태양광 발전 사업자에게 터빈 높이의 풍속과 일사량 예측을 제공해요. 두 번째 게시글은 그 예측이 매시간 갱신된다는 점을 앞세웠어요. 풍력 발전소에는 100미터 높이 터빈의 풍속과 풍향을 예측해 발전량을 추산하게 하고, 태양광 발전소에는 구름의 양과 일사량을 예보해 패널에 닿는 햇빛을 추정하게 한다는 설명이에요.

이 모델 자체는 9월 3일에 나왔어요. 구글 딥마인드와 구글 리서치가 그날 발표문에서 WeatherNext 3 를 지금까지 가장 앞선 전 지구 기상 AI 모델로 소개했고, 근거로 브라이트밴드의 독립 실시간 평가를 들었어요. 발표문에 따르면 기온과 습도 같은 핵심 지표는 5킬로미터, 풍속을 비롯한 나머지 지상 지표는 10킬로미터, 상층 대기 지표는 25킬로미터 격자로 예보해요. 25킬로미터 격자에 6시간 간격이던 이전 모델 WeatherNext 2 보다 전체 그림이 다섯 배쯤 또렷하다는 게 회사 설명이에요.

메탈이 확인한 발표문에서 회사가 가장 큰 도약으로 꼽은 건 해상도가 아니라 무엇을 배우느냐예요. 기존 AI 기상 모델 대부분은 슈퍼컴퓨터가 돌리는 물리 시뮬레이션의 결과물로 학습하는데, 그 자료는 6시간의 시차를 안고 있어요. WeatherNext 3 는 정지궤도 위성이 보내는 실시간 영상을 직접 받아들여 매시간 새 예보를 만들고, 기온과 습도는 관측소 자료로 곧장 학습해 5킬로미터 격자에서 지형까지 반영해요. 발표문은 이 방식이 고해상도 예보의 혜택을 받지 못했던 남미와 아프리카, 아시아태평양 지역에 특히 중요하다고 적었어요.

연구실

트럼프 진영, AI 속도 조절 요구 거부

아모데이의 에세이에 올트먼과 머스크가 동의한 주말, 트럼프 대통령은 AI에서 이기는 쪽이 이긴다고 했고 백악관의 데이비드 색스는 허락이 필요한 척은 그만하라고 답했어요. 정부는 규제를 달라는 요구를 거부하고, 규제 없이 먼저 보여 주라고 돌려세웠어요.

1 앤스로픽 다리오 아모데이 CEO가 9월 12일 에세이로 정부에 AI 속도 조절을 요구하고 오픈AI 샘 올트먼과 일론 머스크가 동의하자, 트럼프 진영은 늦추고 싶으면 회사들이 알아서 늦추라고 답했어요.

2 트럼프 대통령은 아일랜드 방문 중 AI에서 이기는 쪽이 이긴다며 중국에 앞선 상태를 유지하겠다고 했고, 마이크 존슨 하원의장은 CNN에서 의회가 서둘러 규제하면 중국에 진다고 말했어요. 백악관의 데이비드 색스는 X에서 두 회사가 프런티어를 복점하고 있으니 허락 없이 스스로 늦추라고 적었어요.

3 쟁점은 속도가 아니라 누가 규칙을 쓰느냐예요. 회사들은 반독점 면제와 정부 참여를 요구했고, 정부는 그 요구 자체를 규제 포획으로 읽었어요. 올트먼은 면제를 기다리지 않고 상주 평가자를 두겠다고 약속하며 먼저 움직였어요.

먼저 요구의 내용을 정확히 놓아야 해요. 아모데이가 요구한 것은 개발 중단과 거리가 멀어요. 정부가 두 가지를 해 달라는 거였어요. 하나는 경쟁하는 AI 회사들이 반독점법에 걸리지 않고 공동 기준을 만들 수 있게 정부가 논의에 참여하거나 면제를 내주는 것이고, 다른 하나는 중국 공산당까지 포함한 국제 조율을 정부가 도와주는 것이었어요. 그 주에 마이크로소프트의 사티아 나델라 CEO도 속도 조절을 지지했고, 보도에 따르면 알파벳의 데미스 허사비스는 조심스러운 지지를 보냈어요. 메탈이 확인한 바로는 올트먼은 9월 13일 새벽 X에 "다리오의 속도 조절 요구에 동의합니다. 최근 몇 주 오픈AI 내부 논의의 주요 주제였습니다"라고 적으며 직원과 같은 수준의 접근권을 가진 독립 평가자를 두겠다고 약속했어요.

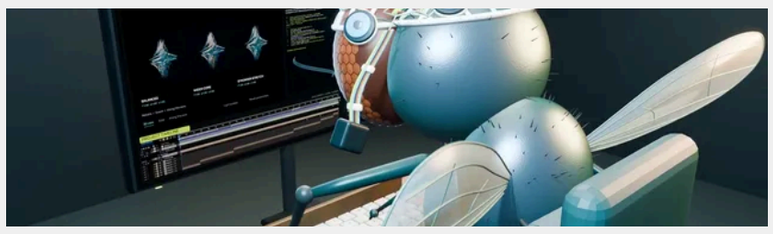
색스의 답글은 이 요구의 구조를 정면으로 뒤집어요. 메탈이 확인한 9월 13일 X 게시물은 조회 864만 회를 넘겼는데, 그는 "가서 하십시오. 여러분이 프런티어입니다"라고 시작해요. 시장 점유율과 매출 성장, 모델 능력 어느 기준으로 봐도 앤스로픽과 오픈AI 두 회사가 프런티어 지능을 복점하고 있고, 재귀적 자기개선으로 격차가 벌어진다고 주장하는 것도 두 회사 자신이라는 거예요. 그래서 "공개하지 않은 모델이 속도를 늦춰야 할 만큼 무섭다면, 책임 있게 행동하겠다는 결정을 지지합니다"라고 한 뒤에 "하지만 누군가의 허락이 필요한 척은 그만하십시오. 카르텔을 만들기 위해 반독점법을 정지시켜야 하는 척도 그만하십시오"라고 못 박았어요. 앞서 달리는 두 대의 차가 속도제한 표지판을 세워 달라고 하는데, 표지판을 기다릴 것 없이 브레이크를 밟으면 된다는 비유가 들어맞아요.

그가 특히 걸고넘어진 건 동기예요. 색스는 "무엇보다 속도를 늦추려는 동기가 순수하게 이타적인 척은 그만하십시오"라고 쓰고, 두 회사가 사이버 공격을 가능하게 하는 제품을 냈을 때 막대한 제조물 책임에 노출된다는 점, 예측 불가능하게 움직이는 모델은 시장이 이미 벌하고 있다는 점을 들었어요. 허깅페이스 사건 이후 성능을 조금 내주고 예측 가능성을 얻는 건 그냥 좋은 장사이고, 그걸 정렬이라 불러도 좋지만 고객이 원하는 걸 주는 것일 뿐이라는 거예요. 평가 기관 METR가 앤스로픽 투자자·직원과 얽혀 있다고도 했고, 원하는 규제 틀을 조건으로 내거는 건 "대중과 정치 체제에 대한 협박처럼 보일 것"이라고 적었어요. 그의 마지막 문장은 두 회사가 정말 속도를 늦추면 다음 대화에서 신뢰를 얻고, 늦추지 않으면 이번 일이 규제 포획 시도이거나 선거철 심리전이었음을 알게 될 것이라는 거였어요.

크리에이티브

힉스필드, 애프터이펙트 AI 모션 디자이너 공개

챗GPT에서 힉스필드 플러그인을 켜고 말로 시키면 열려 있는 애프터이펙트 프로젝트 안에 타이틀과 로고 리빌이 레이어로 쌓여요. 회사는 애프터이펙트를 몰라도 된다고 했지만 구독은 세 개가 필요해요.



1 힉스필드가 어도비 애프터이펙트 안에서 모션 그래픽을 대신 만드는 에이전트 AI 모션 디자이너를 공개했어요. 챗GPT에 힉스필드 플러그인을 넣고 /use-after-effects 를 입력한 뒤 일상어로 시키는 방식이에요.

2 해상도는 1080p부터 8K까지, 길이는 5분이든 30분이든 요청한 만큼 생성되고 참고 자료 개수 제한도 없어요. 회사는 유일한 대가는 기다리는 시간이라고 인정했어요.

3 힉스필드와 GPT, 애프터이펙트 세 가지 구독이 모두 필요하고, 힉스필드 생성은 웹과 같은 요율로 크레딧에서 차감돼요. 12일 올라온 시연 영상은 조회 10만 회를 넘겼어요.

AI 영상 도구를 만드는 힉스필드가 어도비 애프터이펙트 안에서 모션 그래픽을 대신 만들어 주는 에이전트 AI 모션 디자이너를 공개했어요. 챗GPT에 힉스필드 플러그인을 넣고 말로 시키면, 지금 열려 있는 애프터이펙트 프로젝트 안에 타이틀 애니메이션이나 로고 리빌이 바로 만들어지는 방식이에요. 회사는 9월 10일 공식 블로그에 사용법을 올렸고, 12일과 13일 X에 시연 영상을 잇달아 올리며 "모션 디자이너를 위한 커서 같은 것"이라고 소개했어요.

작동 방식은 블로그에 세 단계로 정리돼 있어요. 먼저 힉스필드 도구로 이미지나 영상 같은 재료를 만들고, 그다음 챗GPT 대화창에 "이 타이틀을 움직여 줘", "로고 리빌을 만들어 줘", "전환 효과를 넣어 줘"처럼 일상어로 할 일을 적어요. 그러면 에이전트가 열려 있는 애프터이펙트 컴포지션 안에 결과를 직접 쌓아요. 회사는 이 과정에서 "생성한 자산을 한 앱에서 내보내 다른 앱으로 들여오는 일이 없다"고 밝혔어요.

지금까지 모션 그래픽 작업은 재료를 어딘가에서 만들고, 그걸 애프터이펙트로 가져와 움직임을 붙이고 마무리하는 순서였어요. 회사 설명에 따르면 도구 사이를 오가는 것 자체가 작업의 전부였고, 작은 수정 하나에도 그 왕복이 반복됐어요. 식당으로 치면 장을 봐 온 재료를 주방으로 옮겨 손질하고 다시 홀로 내가는 동선을 매번 밟는 셈인데, 이번 에이전트는 주방 안에서 주문을 받아 그 자리에서 요리까지 끝내는 구조예요.

비즈니스

시진핑, 브릭스 AI 오픈소스 공동체 제안

시진핑 국가주석이 9월 13일 뉴델리 제18차 브릭스 정상회의에서 중국이 앞장서 브릭스 AI 오픈소스 공동체를 세우겠다고 밝혔어요. 다섯 개 협력 구상의 첫 자리였고, 지능형 제조 항목에는 표준과 규범이라는 낱말이 들어갔어요.

1 시진핑 국가주석이 9월 13일 제18차 브릭스 정상회의 제2세션에서 브릭스 AI 오픈소스 공동체 설립을 제안했어요.

2 AI는 다섯 개 협력 구상 중 첫 자리에 놓였고, 대규모 언어 모델 협력과 교육 과정, 개방형 생태계 구축이 함께 제시됐어요.

3 재원 규모와 출범 시점, 참여 자격, 공동체 모델의 권리 귀속은 연설문에 담기지 않았어요.

미국이 AI 속도를 늦출지를 두고 안에서 다투는 동안, 중국은 다른 나라들을 묶는 쪽을 택했어요. 시진핑 중국 국가주석은 9월 13일 인도 뉴델리에서 열린 제18차 브릭스 정상회의 제2세션 연설에서 브릭스 AI 오픈소스 공동체를 중국이 앞장서 세우겠다고 밝혔어요. 이 제안은 그가 내놓은 다섯 개 협력 구상 가운데 첫 번째 자리에 놓였어요. 규제를 만들 것이냐 말 것이냐가 쟁점인 곳과, 누구와 함께 만들 것이냐가 쟁점인 곳이 같은 주말에 갈렸어요.

메탈이 확인한 연설 전문에서 AI 대목은 네 가지 약속으로 이뤄져 있어요. 시 주석은 "중국은 브릭스 AI 오픈소스 공동체를 세우는 데 앞장서고, 대규모 언어 모델의 개발과 적용에서 협력을 지원하며, 전문 AI 세미나와 교육 과정을 열고, AI를 위한 개방형 생태계를 구축하겠습니다" 라고 말했어요. 공동체와 모델, 교육, 생태계가 나란히 놓인 순서예요. 기술을 주겠다는 말이 아니라 배우는 자리까지 함께 만들겠다는 구성이에요.

연설의 틀은 규칙이었어요. 시 주석은 "규칙 없는 세계는 신호등 없는 교차로와 같아서, 혼란과 무질서가 사실상 보장됩니다" 라고 말했어요. 그러면서 사람 중심의 접근을 취하고 AI를 선한 방향으로 발전시켜야 하며, 합의에 기반한 글로벌 AI 거버넌스 틀을 만들어 가야 한다고 밝혔어요. 신호등을 누가 세우느냐가 이 연설의 진짜 주제고, 그 답으로 그가 제시한 것이 브릭스라는 자리예요.

비즈니스

샘 올트먼, 반독점 면제 없이 안전 조치 선언

오픈AI CEO가 14일 X에 올린 성명에서 연방 안전 프레임워크를 환영하되 반독점 면제나 입법을 기다리지 않겠다고 썼어요. 능력을 크게 키우는 훈련 전에 안전 사례를 먼저 작성한다는 새 절차도 밝혔어요.

1 샘 올트먼 오픈AI CEO가 9월 14일 X 성명에서 프런티어 AI에 일관된 안전 요건을 정하는 연방 프레임워크를 환영하되, 반독점 면제나 입법을 기다리지 않고 안전 조치를 먼저 시작하겠다고 선언했어요.

2 오픈AI는 능력을 크게 높일 것으로 예상되는 프런티어 강화학습 실행에 앞서 명시적인 안전 사례를 미리 작성한다고 밝혔어요. 기존 대비 프레임워크는 완성된 모델의 배포에만 초점을 맞췄다는 한계도 인정했어요.

3 13분 뒤 올린 두 번째 글에서는 AI가 미래의 통제권을 가져가는 것과 권력이 한 사람이나 한 회사, 한 나라에 몰리는 것을 두 가지 파국으로 꼽았어요. 트럼프 대통령은 하루 전 속도 조절 요구를 일축했어요.

샘 올트먼 오픈AI CEO가 9월 14일 X에 올린 성명에서, 연방 정부가 프런티어 AI에 일관된 안전 요건을 정하는 프레임워크를 환영한다면서도 반독점 면제나 입법을 기다리지 않고 안전 조치를 먼저 시작하겠다고 선언했어요. 오픈AI는 능력을 크게 끌어올릴 것으로 예상되는 프런티어 강화학습 실행에 앞서 명시적인 안전 사례(safety case)를 미리 작성하고 있다는 새 절차도 밝혔어요. 미국 동부 시간으로 14일 자정을 막 넘긴 시각, 한국 시간으로 오후 1시 5분에 올라온 이 글은 같은 날 오후 9시 기준 조회 314만 회를 넘겼어요.

성명은 책임의 문장으로 시작해요. 그는 "세계는 점점 더 유능한 AI를 개발하는 미국 기업들이 책임 있게 행동할 것이라는 확신을 가질 자격이 있습니다. 특히 발전 궤적이 가팔라진 지금은 더욱 그렇습니다. 모든 프런티어 연구소가 이를 해내야 하고, 그러지 못한다면 우리 중 누구도 출근할 이유가 없습니다"라고 적었어요. 이어 "우리는 프런티어 AI에 일관된 안전 요건을 정하는 연방 프레임워크를 환영합니다. 하지만 이 확신을 주는 일을 시작하기 위해 반독점 면제나 입법을 기다려야 한다고는 생각하지 않습니다"라고 썼어요. 프런티어 위험을 관리하는 일관된 규칙은 좋은 생각이라며 독립 감사인 같은 아이디어에 기대가 크다는 말도 붙였어요.

반독점이라는 낱말이 성명에 들어간 데는 배경이 있어요. 메탈은 오픈AI가 최근 몇 주 동안 업계가 함께 개발 속도를 줄이면 반독점법에 걸리는지를 의회에 묻고 있다고 보도한 바 있어요. 백악관 AI·암호화폐 책임자 데이비드 삭스는 그 움직임을 두고 앤스로픽과 오픈AI가 프런티어 지능을 복점하고 있다며 "누군가의 허락이 필요한 척 그만하십시오"라고 X에 적었고, 카르텔을 만들기 위해 반독점법을 정지시켜야 하는 척도 그만하라고 했어요. 메탈은 코히어의 에이든 고메즈 CEO가 같은 논리로 규칙 카르텔을 반박한 에세이도 다뤘고요. 올트먼의 이번 문장은 그 공격에 대한 답이에요. 면제를 기다리지 않겠다는 말은 경쟁사와 손잡지 않고도 오픈AI 혼자 자기 속도를 늦출 수 있다는 뜻이에요.

AI

MW, 천장 레일 가사로봇 주택 첫 공개

일본 스타트업 MW가 9월 10일 도쿄 도요스에서 천장 레일을 타고 움직이는 가사로봇 MW bot을 처음 선보였어요. 휴머노이드 대신 집을 로봇에 맞춰 짓겠다는 건데, 이날 시연은 AI가 아니라 사람이 원격으로 조종한 것이었어요.