AI 뉴스와 용어
월·수·금 오전 7시

METAL
First, AI Magazine

무료 · 언제든 수신거부
metallab.ai

제 37 호 2026년 10월 2일 금요일 오늘의 AI

AI

앤스로픽, 개발자 허브 claude.dev 공개

앤스로픽이 엔지니어링 심층 기사와 Claude Code·API 가이드를 모은 개발자 허브 claude.dev를 열었다. 대표 글은 claude.ai를 두 주 만에 약 3배 빠르게 만든 스프린트 기록과 평가 설계 자동화 방법이다.

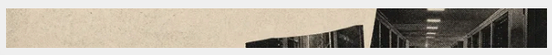


AI · 앤스로픽, 개발자 허브 claude.dev 공개

앤스로픽이 9월 30일 클로드로 개발하는 사람을 위한 개발자 허브 claude.dev를 공식으로 열었다. 회사 개발자 계정 ClaudeDevs는 X에서 claude.dev를 "클로드로 개발하는 사람들의 새 집"이라고 소개하며, 엔지니어링 심층 기사와 Claude Code·API 가이드, 클로드를 만드는 팀의 요령, 그리고 몇 가지 재미있는 이스터에그를 싣는다고 밝혔다. 이 게시물은 공개 후 하루가 지나기 전에 조회수 11만 회를 넘겼다. 앤스로픽이 흩어져 있던 기술 글을 개발자 전용 주소 하나로 모은 셈이다.

사이트의 정체성은 첫 화면 한 줄에 있다. claude.dev는 스스로를 앤스로픽 개발자들의 요령과 관점을 나누는 곳이라고 소개하고, 글마다 읽는 데 걸리는 시간을 분 단위로 붙인다. 첫 화면에 걸린 글 10편은 5월 20일부터 9월 28일 사이에 나온 것으로, Opus 5.5에서 과제 하나에 드는 비용을 21분 분량으로 뜯어본 글, Claude 5 세대 모델의 컨텍스트 엔지니어링 규칙, 클로드 코드 팀이 스킬을 쓰는 방식 같은 실무 주제가 대부분이다. 아래쪽에는 공식 유튜브 영상이 붙고, 바닥에는 Claude Academy와 디스코드, 오프라인 밋업, 레딧 커뮤니티로 가는 길이 모여 있다.

이스터에그도 실제로 있다. 메탈이 확인한 claude.dev의 터미널 페이지는 10월 1일 브라우저 접속 기준으로 CLAUDE.DEV라는 글자를 점으로 찍은 아스키 그림으로 열리고, /help를 치면 다른 명령을 찾을 수 있다고 안내하며 /posts로 글 목록을 불러오게 한다. X 게시물에 붙은 30초 영상은 픽셀 게임 Clawd's Quest의 플레이 장면이다. 하늘이 깨어나기를 잊은 밤이라는 설정에서 주황색 픽셀 캐릭터 Clawd가 불씨 5개를 모으면 해가 뜨고, 화면에는 고맙다는 인사가 뜬다. 개발자 대상 사이트에 게임을 숨겨 두는 방식은 문서 사이트보다 커뮤니티 공간에 가깝게 보이려는 선택으로 읽힌다.



8월 두 주짜리 스프린트에서 claude.ai 웹과 데스크톱 앱의 핵심 사용 경험을 약 3배 빠르게 했다. 사용 활동의 95%를 차지하는 네 가지 여정을 겨냥했고, 75백분위 기준으로 claude.ai를 새로 열어 입력 가능한 화면이 뜨는 시간은 3.1초에서 0.55초로, 새 Claude Code 세션 시작은 0.8초에서 0.3초로, 클라우드 Cowork 세션 불러오기는 2.6초에서 0.73초로 줄었다. 13개 측정값 가운데 12개 목표를 사흘 만에 달성했다.

작업 방식이 결과만큼 눈에 띈다. 팀은 슬랙 채널 하나에 모든 스레드를 모으고 스레드마다 클로드를 붙였다. 앤스로픽은 Opus 5.5와 대략 비슷한 수준의 내부 연구 모델을 Claude Tag 베타로 돌렸다고 밝혔다. 클로드가 병목을 찾고 벤치마크를 만들고 변경을 올리고 배포를 지켜봤으며, 사람은 목표를 정하고 절충을 판단하고 모든 변경을 승인했다. 그렇게 3,000건이 넘는 변경을 고객에게 보이는 사고나 롤백 없이 병합했다. 가장 바쁜 날에는 하루 200건 넘게 반영됐고, 한때 동시에 돌아간 스레드는 150개가 넘었다. 약 200개의 기능 플래그를 도입해 절반 넘게 스프린트 안에 정리했다.



엔비디아 젠슨 황 CEO, 2026 밴 플리트상 수상…"한국과 미국, 함께 미래 만들어가고 있어"

발견의 내용은 구체적이다. 클로드는 입력창 경로에서 키를 누를 때마다 다시 그려지는 React 훅 6,900개와 스토어 구독 900개를 찾았고, 하루 50만 번씩 보이지 않게 일어나던 새로고침을 잡아냈다. 코드 블록 하이라이트가 약 1초 멈추던 원인은 증폭였다. 당별에 증피나 등급 따오프 같은 무은 아니다. 다음은 뭔가. 야심 차게 가자"라는 같은 메시지를 남겼다. 채널에 있던 엔지니어 셸리 보어는 이 모델을 두고 "숫자 귀신"이라고 평했다. 결과를 사내에 공유했을 때 아이작 G.는 "여섯 달 전만 해도 이게 가능하다고 나를 설득하지 못했을 것"이라고 말했다.

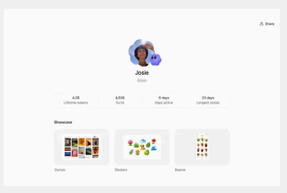
또 하나의 대표 글은 평가 설계를 다룬다. 랜스 마틴이 9월 28일 쓴 글은 claude-api 스킬에 /claude-api build-eval과 /claude-api hillclimb 두 명령을 더했다고 밝혔다. 앞의 명령은 사용자를 인터뷰해 코드베이스 안에 평가 세트를 만들고, 뒤의 명령은 한 번에 변경 하나씩 적용하며 점수를 올린다. 평가 사례는 학습용과 한 번도 보여 주지 않는 시험용으로 나누고, 학습 점수만 오르고 시험 점수가 그대로면 과적합으로 보고 변경을 되돌린다.

사례의 숫자는 비용 쪽으로 기운다. 내부 고객지원 벤치마크 44건 가운데 30건으로 탐색하고 14건을 남겨 둔 실험에서, 출발점은 Opus 4.8 기본 노력 설정의 결정 정확도 74.4%와 티켓당 4.6센트였다. 프롬프트를 정리한 뒤 Opus 5.5 낮은 노력 설정으로 바꾸자 87.8%에 1.9센트가 됐고, Sonnet 5 낮은 노력 설정은 88.9%를 약 1센트에 냈다. 앤스로픽에 따르면 Opus 5.5의 입력·출력 토큰 가격은 Opus 4.8보다 20%, 캐시 읽기는 60% 낮다. 한 번도 보지 않은 14건에서 최종 구성은 원래 구성의 78.6%보다 높은 90.5%를 약 5분의 1 비용으로 기록했다. claude-api 스킬 자체의 평가 점수도 66%에서 약 88%까지 올랐다.

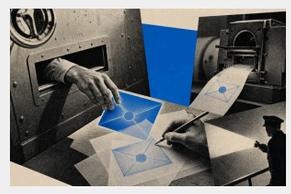
엔지니어링 조직을 운영하는 눈으로 보면 이 허브가 파는 것은 기능이 아니라 작업 방식이다. 두 글 모두 결론이 같다. 잴 수 있으면 클로드가 개선할 수 있고, 그래서 사람이 할 일은 잴 것을 늘리고 되돌릴 장치를 먼저 까는 일이라는 것이다. 메탈은 앞서 앤스로픽의 Claude Opus 5.5 공개를 보도한 바 있으며, claude.dev의 글들은 그 모델을

오늘 지면

1 AI 앤스로픽, 개발자 허브 claude.dev 공개
2 AI 앤스로픽, Claude 파운더 하우스 두 도시 개최
3 AI 앤스로픽, 클로드 코드 모드 공개
4 AI 바클레이즈, 클로드 전사 확대 발표
5 AI 구글 리서치, CDC 독감 예측 평가 1위
6 AI 구글 딥마인드, Gemini 4 Argon 공개



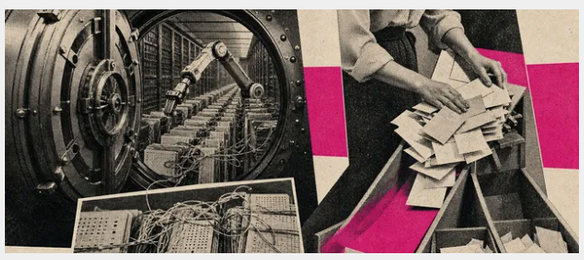
오픈AI, ChatGPT 공유 프로필 공개



오픈AI, 문샷 연관 증류 캠페인 차단 발표

이 지면은 2026년 10월 2일 금요일 아침 구독자 메일함으로 나간 레터를 신문 판형으로 다시 조판한 것입니다. 기사 전문은 metallab.ai 에서 읽을 수 있습니다.

AI



바클레이즈, 클로드 전사 확대 발표

영국 은행 바클레이즈가 앤스로픽의 AI 클로드를 전사로 확대한다. 앤스로픽은 10월 1일(현지시간) 발표문에서 바클레이즈가 앤스로픽과의 전략적 협력을 넓혀 기업용 AI 시스템을 글로벌 운영 전반에 통합한다고 밝혔다. 가장 구체적인 약속은 개발자 도입률이다. 바클레이즈는 2026년 말까지 자사 개발자의 50%가 클로드 코드를 쓰게 하고, 2027년에는 소프트웨어 엔지니어 대다수로 넓힐 계획이다. 메탈이 확인한 앤스로픽 발표문 원문에는 양사 임원

▶ 2면에 계속

AI

앤스로픽, Claude 파운더 하우스 두 도시 개최



앤스로픽이 9월 30일(현지 시간) Claude 공식 X 계정에서 창업자 대상 오프라인 행사 Claude Founder House(Claude 파운더 하우스)를 샌프란시스코와 스톡홀름 두 도시에서 연다고 밝혔다. 샌프란시스코 행사는 SF 테크 위크 기간인 10월 6일부터 8일까지 사흘 동안, 스톡홀름 행사는 10월 14일 하루 열린다. 앤스로픽은 게시물에서 "토크와 워크숍, 우리 팀과의 오피스아워에 더해 Claude로 무언가를 만드는 다른 사람들을 만날 시간도 있다"며 "창업자든 CEO든 빌더든 당신을 위한 세션이 있다"고 밝혔다.

앤스로픽 행사 안내 페이지에 따르면 샌프란시스코 행사장은 해리슨 스트리트 511번지의 테라 갤러리다. 앤스로픽은 2층짜리 창고 건물을 사흘 동안 통째로 맡아 일하고 배우고 다

▶ 2면에 계속

AI

앤스로픽, 클로드 코드 모드 공개



앤스로픽이 10월 1일(미국 시간) 클로드 코드에 모드(mods)를 정식으로 열었다. 모드는 타입스크립트나 자바스크립트 함수 몇 줄로 클로드 코드가 일하는 방식과 화면을 바꾸는 확장 기능이다. 프롬프트가 모델에 닿기 전에 고쳐 쓰고, 도구 호출을 막거나 바꾸거나 다시 시도하고, 권한 요청을 대신 승인하거나 거절하고, 도구 출력에서 비밀 값을 지운 뒤 클로드에게 넘긴다. 모드는 플러그인 안에 담겨 나오므로 CLI와 데스크톱 앱에서 /plugin 명령으로 설치한다. 클로드 코드 2.1.287 이상이면 기본으로 켜져 있다.

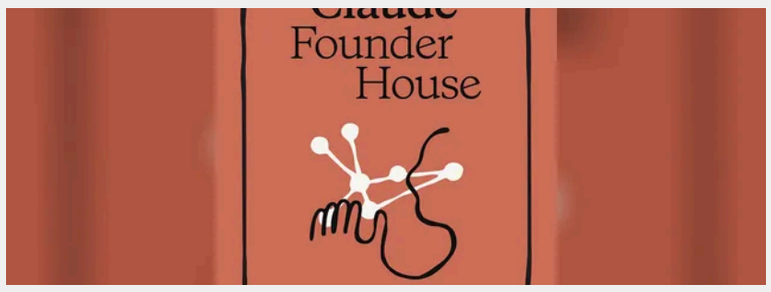
앤스로픽 공식 블로그에 따르면 클로드 코드는 도구를 부르거나 권한을 묻거나 화면 일부를 그릴 때마다 이벤트를 낸다. 모드는 그 이벤트에 걸리는 함수로, 이벤트 앞이나 뒤에서

▶ 2면에 계속

◀ 1면에서 계속

AI

앤스로픽, Claude 파운더 하우스 두 도시 개최



1 앤스로픽이 9월 30일 Claude X 계정에서 SF 테크 위크 기간 샌프란시스코(10월 6~8일)와 스톡홀름(10월 14일) 두 도시에서 Claude 파운더 하우스(Claude Founder House)를 개최한다고 밝혔다.

2 샌프란시스코 행사는 테라 갤러리 2층 창고에서 사흘간 열리고 연사 36명 가운데 11명이 앤스로픽 인력이며, 8일에는 마이크 크리거가 나서는 파운더 살롱이 따로 열린다.

3 스톡홀름은 오전 파운더 코호트와 오후 빌더 코호트로 나뉘며 두 순서 모두 앤스로픽 팀의 승인을 받는 선별제다.

연사 명단에는 36명이 올라 있다. 포스트호그, 넷리파이, 시뮬, 감마, 타운, 리툴, 오픈에비던스의 창업자와 임원, 세쿼이아와 라이트스피드의 투자자가 토크와 대담을 맡는다. 넷리파이 CEO 맷 빌만, 포스트호그 공동창업자 팀 글레이저, 오픈에비던스 최고의료책임자 트래비스 잭이 이름을 올렸고, 엔비디아와 클릭하우스, ICONIQ, 베세머 소속 인사도 명단에 있다. 앤스로픽 쪽에서는 Anthropic Labs를 이끄는 마이크 크리거와 스타트업 총괄 베스 로버트슨, 기술 스태프 재러드 섬너를 비롯해 11명이 나온다. 스타트업과 앤스로픽 팀이 여는 실습 워크숍이 따로 있고, 브런치와 창업자 러닝 클럽도 일정에 들어 있다.

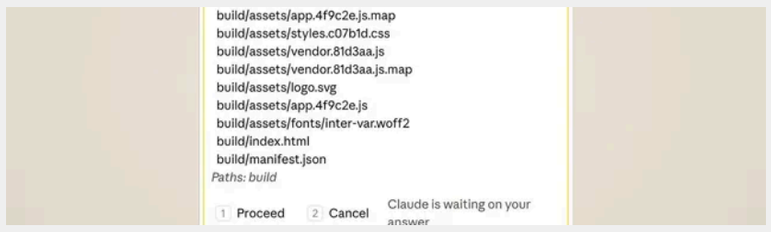
마지막 날인 10월 8일 오후 1시 45분부터 4시 45분까지는 Claude 위에서 제품을 내놓는 창업자를 위한 파운더 살롱이 별도로 열린다. 팟캐스트 Acquired의 공동 진행자 데이비드 로젠탈이 마이크 크리거와 대담하고, 다른 앤스로픽 리더들이 Claude로 만드는 스타트업의 다음 단계를 이야기한다. 창업자들이 무대에서 라이브 데모를 하면 관객이 마음에 드는 팀에 투표하는 순서도 있다. 데모를 하고 싶은 팀은 참가 신청서에 짧은 데모 영상을 붙이면 된다.

스톡홀름 행사는 하루를 두 갈래로 나눴다. 오전 8시 30분부터 오후 1시(중앙유럽 서머타임)까지의 파운더 코호트는 북유럽 AI 네이티브 스타트업의 창업자와 CEO를 모은다. 앤스로픽은 이 순서를 "앞으로 12개월을 날카롭게 다듬도록 짠 오전"이라고 소개했다. Claude가 나아갈 방향, 앤스로픽 안에서 Claude로 제품을 내놓는 내부 플레이북, 제품과 팀, 마진을 주제로 한 창업자 패널, Applied AI 팀 오피스아워가 이어진다. 연사는 6명으로, 앤스로픽 스타트업 총괄 에이든 블레이크와 Applied AI의 라미스 무크타, 러버블 공동창업자 파비안 헤딘, 레고라 CTO 야코브 라우리첸, 제너럴 카탈리스트 파트너 제이넵 야부즈윌슨, 슈퍼메트릭스 공동 CEO 미카엘 투네베리다.

◀ 1면에서 계속

AI

앤스로픽, 클로드 코드 모드 공개



1 앤스로픽이 10월 1일 프롬프트와 도구 호출, 권한 요청, 화면 그리기에 함수를 거는 클로드 코드 모드를 정식 공개했다.

2 모드는 플러그인으로 설치하고 클로드에게 말로 만들게 할 수도 있으며, /diff와 AGENTS.md 지원 같은 기본 기능도 모드로 옮겨졌다.

3 모드는 샌드박스 없이 사용자 권한으로 돌기 때문에 팀·엔터프라이즈 요금제에서는 sec-default 가드가 가장 먼저 로드된다.

메탈이 확인한 72초 시연 영상에는 공식 안내서의 예제 모드 세 개가 차례로 나온다. Token Weather는 입력창 위에 문맥 창이 얼마나 찼는지 날씨처럼 띄운다. 작은 파일 하나를 읽자 맑음과 함께 20만 토큰 중 1만 9,200 토큰(10%)이 표시되고, 사흘치 로그와 한 달 치 측정값을 읽자 소나기와 10만 7,300 토큰(54%)으로 바뀐다. Blast Radius는 클로드가 rm -rf build를 실행하려는 순간 명령을 붙잡고 지워질 파일 9개(498KB)를 늘어놓은 뒤 진행과 취소 버튼을 내민다. Replay Theater는 함수 이름 하나를 바꾸며 파일 네 개에 걸쳐 생긴 수정 여섯 건을 한 단계씩 넘겨 보여 준다. 보도에 따르면 Token Weather는 80줄 안팎의 코드로 만들어졌다.

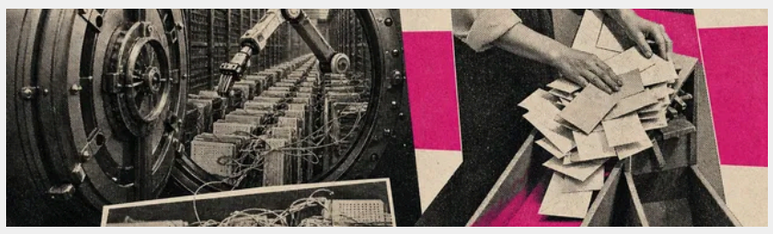
코드를 직접 짜지 않아도 된다. 세션에서 현재 git 브랜치를 입력창 위에 보여 주는 모드를 만들어 달라고 말하면 클로드가 내장 스킬 plugin-authoring을 따라 모드를 쓰고, 사용자가 핫 리로드를 승인하면 그 턴이 끝날 때 곧바로 불러온다. 앤스로픽은 이를 클로드 코드로 클로드 코드를 모드하는 일이라고 설명했다. 이렇게 만든 모드는 그 세션에서만 돌기 때문에 다른 세션에서 쓰려면 폴더를 옮겨 --plugin-dir로 불러오거나 마켓플레이스에 올려야 한다. claude plugin validate 명령은 모드를 실행하지 않고도 거는 이벤트와 호출하는 API 목록을 보여 주고, claude plugin test 명령은 세션과 로그인, 네트워크 없이 자동 테스트를 돌린다.

클로드 코드의 기본 기능도 모드로 옮겨 가기 시작했다. /diff 화면과 AGENTS.md 지원이 이미 모드로 바뀌어, 사용자는 /plugin에서 이를 끄거나 자기 버전으로 갈아끼울 수 있다. 개발 문서가 꼽은 내장 모드는 원격 측정과 보안 가드를 포함해 여섯 개이고, 그중 네 개의 소스가 클로드 코드 저장소에 공개됐다. 앤스로픽은 블로그에서 더 많은 기본 기능을 모드로 옮겨 클로드 코드를 작은 코어로 줄이고 원하는 기능만 다시 얹을 수 있게 하겠다고 밝혔다.

◀ 1면에서 계속

AI

바클레이즈, 클로드 전사 확대 발표



1 앤스로픽은 10월 1일 바클레이즈가 전략적 협력을 넓혀 클로드를 글로벌 운영 전반에 통합한다고 발표했다.

2 바클레이즈는 2026년 말까지 개발자 50%, 2027년에는 소프트웨어 엔지니어 대다수가 클로드 코드를 쓰게 할 계획이다.

3 RAG 기반 사내 도구는 직원 1만 6,000명 이상이 채택해 검색 100만 건을 넘겼고, 글로벌 마켓 사업부는 하루 약 12만 건의 이메일을 클로드로 분류한다.

이미 돌아가는 사례부터 보면 규모가 드러난다. 바클레이즈 UK 직원이 쓰는 사내 도구 Colleague Knowledge Assistant는 2천만 명이 넘는 영국 소매 고객을 응대할 때 필요한 정보와 답을 빠르게 찾아 준다. 2025년부터 가동된 이 도구는 검색 증강 생성(RAG) 구조로 클로드를 쓴다. 질문이 들어오면 사내 자료에서 관련 내용을 먼저 찾아 모델에 건네고, 모델은 그 자료를 근거로 답을 만든다. 지금까지 1만 6,000명이 넘는 직원이 이 도구를 채택했고, 처리한 검색은 100만 건을 넘었다.

두 번째 사례는 글로벌 마켓 사업부의 고객 문의 처리다. 이 사업부 팀은 들어오는 이메일을 클로드 모델로 분류하고, 필요한 정보를 덧붙이고, 가장 알맞은 처리 경로를 정한다. 운영 담당자가 바로 조치할 수 있을 만큼 요청에 필요한 정보가 담겼는지도 이 단계에서 확인하고 우선순위를 매긴다. 이 플랫폼이 처리하는 이메일은 하루 약 12만 건이다. 앤스로픽은 이 시스템으로 수작업 처리가 줄었고 직원이 관련 정보를 더 효율적으로 찾아 조치하게 됐다고 밝혔다.

바클레이즈의 앤 마리 달링 그룹 공동 최고운영책임자(COO)는 "어떤 기술이든 진짜 척도는 고객과 동료에게 미치는 영향, 그리고 그것이 가능하게 하는 결과"라고 말했다. 그는 "바클레이즈 전반에 클로드 사용을 넓히는 일은 AI를 들이는 것 이상"이라며 "일이 처리되는 방식을 바꾸는 일"이라고 밝혔다. 일상 업무에 드는 시간을 줄이고 의사결정을 간소하게 만들어, 직원이 더 복잡한 문제에 전문성을 쓰게 하겠다는 것이 달링의 설명이다.

AI

구글 리서치, CDC 독감 예측 평가 1위

구글의 AI 연구 도구 ERA로 만든 독감 입원 예측 모델이 CDC 시즌 평가에서 39개 모델 중 가장 좋은 점수를 냈다. 2위와의 차이는 0.02이고 주별 순위는 엇갈렸다.



1 CDC가 9월 30일 공개한 FluSight 2025~2026 시즌 평가에서 구글의 Google_SAI-FluEns가 분석 대상 39개 모델 중 개별 팀 제출 모델 1위를 차지했다.

2 구글 모델의 상대 WIS는 0.56으로, 0.58을 받은 세 모델과 CDC 앙상블(0.62·7위)을 앞섰지만 텍사스·매사추세츠·코네티컷에서는 다른 모델이 더 나았다.

3 예측 모델은 제미나이로 과학 코드를 쓰고 최적화하는 구글의 ERA로 만들었고, ERA는 5월 네이처 논문으로 공개됐다.

구글 리서치가 만든 독감 입원 예측 모델이 미국 질병통제예방센터(CDC)의 2025~2026 시즌 독감 예측 평가에서 1위에 올랐다. CDC는 9월 30일(현지시간) 공개한 FluSight 시즌 평가 보고서에서 분석 대상 39개 모델 가운데 구글의 GoogleSAI-FluEns가 개별 팀 제출 모델 중 가장 좋은 성적을 냈다고 밝혔다. 구글은 이 예측 모델을 자사 AI 연구 도구인 Empirical Research Assistance(ERA)로 만들었다. 메탈이 확인한 구글 발표문 원문은 네 문단 분량으로, 모델이 어떤 입력 자료를 썼는지는 적지 않았다.

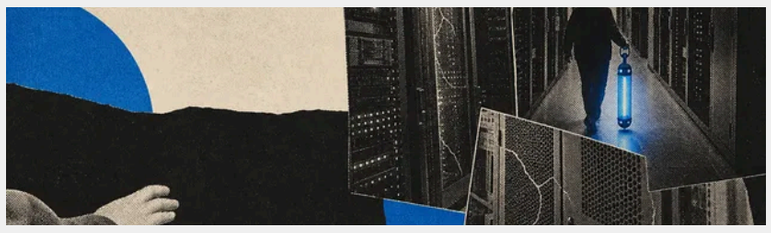
FluSight는 CDC가 학계와 업계, 정부의 예측 팀에서 주별 독감 입원 환자 수 예측을 받아 하나로 묶는 사업이다. 이번 시즌 접수는 2025년 11월 22일부터 2026년 5월 20일까지 이어졌고, 예측은 2025년 12월 12일부터 FluSight 웹페이지에 공개됐다. 각 팀은 50개 주와 워싱턴 D.C., 미국 전체를 대상으로 이번 주와 최대 3주 뒤의 입원 환자 수를 내야 했다. CDC는 묶은 예측을 매주 주 단위 의료 수요 전망을 알리는 데 쓴다. CDC는 독감 유행이 미미했던 2020~2021 시즌을 빼고 2013~2014 시즌부터 해마다 이 예측 대회를 열어 왔다.

이번 시즌에는 34개 팀이 53개 모델을 냈고, 예측 대상의 75% 이상을 제출한 39개 모델이 평가에 들어갔다. 채점의 주된 잣대는 상대 가중구간점수(WIS)다. 예측 구간이 실제 관측값과 얼마나 맞아떨어졌는지를 재는 지표로, 지난주 입원 환자 수를 그대로 다음 주로 옮기는 기준 모델을 1로 놓고 1보다 낮으면 기준보다 나은 예측이다. 39개 모델 가운데 33개가 기준 모델을 이겼다.

AI

구글 딥마인드, Gemini 4 Argon 공개

구글이 새 프런티어 모델을 사이버 방어자에게 먼저 풀었다. 출력 토큰 한도는 100만 개로 늘렸고 일반 공개 날짜는 정하지 않았다.



1 구글 딥마인드가 9월 30일 새 프런티어 모델 Gemini 4 Argon을 공개하고 페어윈드 프로그램의 신뢰받는 사이버 방어자에게 먼저 배포했다.

2 Argon은 DeepSWE v1.1 77.9%, Vals Index 68.9%, AutomationBench 51.3%를 기록했지만 FrontierSWE v2와 Terminal-bench 4.0에서는 다른 모델이 앞섰다.

3 구글은 출력 토큰 한도를 64K에서 1M으로 늘렸고, 오용 방어·프롬프트 인젝션·비정렬 감시·샌드박스 강화의 네 갈래 안전장치를 갖춘 뒤 넓게 공개하겠다고 밝혔다.

구글 딥마인드가 9월 30일(현지시간) 새 프런티어 모델 Gemini 4 Argon을 공개했다. 실제 소프트웨어 엔지니어링과 법률·금융 같은 기업 지식 노동, 사이버보안 방어처럼 길고 복잡한 작업 흐름을 겨냥한 모델이다. 첫 배포 대상은 일반 사용자가 아니라 구글의 페어윈드 프로그램(Fairwind Program)에 참여한 신뢰받는 사이버 방어자들이다.

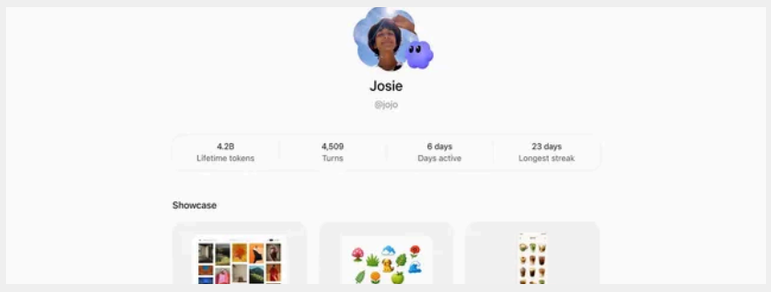
출시 순서 자체가 이번 발표의 첫 메시지다. 코라이 카부쿠오글루 구글 딥마인드 수석부사장 겸 구글 최고AI아키텍트는 발표문에서 "이 수준의 프런티어 역량을 안전하게 내놓으려면 단계적 접근이 필요하다"고 밝혔다. 구글은 미국 정부의 자발적 출시 전 모델 접근 절차에 참여하고 있으며, 초기 테스터의 피드백으로 가드레일을 다듬은 뒤 개발자와 기업, 소비자로 넓힐 계획이다. 일반 공개는 유료 API 고객과 Google AI Ultra 구독자부터 시작한다.

Argon은 이미 구글 사내 업무를 돌리고 있다. 구글에 따르면 수천 명의 직원이 특수 코딩 작업과 심층 조사, 글쓰기 품질에서 이 모델의 강점을 꼽았다. 양자 컴퓨팅 연구진은 중요한 응용의 병목이 되는 서브루틴의 시공간 자원(큐비트 수와 게이트 수의 곱)을 줄이는 데 Argon을 썼고, 한 사례에서는 몇 분 만에 공개된 기준선보다 40% 나은 결과를 냈다. Argon 에이전트 팀은 데이터센터 전체의 프로파일링 원격측정 자료를 분석해 메모리 최적화를 스스로 찾아 적용했고, 배포 뒤 300TiB가 넘는 메모리를 확보했다. 회사가 추정한 총 절감 규모는 500TiB에서 1PiB다.

AI

오픈AI, ChatGPT 공유 프로필 공개

ChatGPT에서 만든 Sites와 자주 쓰는 플러그인을 한 화면에 모아 공개하는 기능이다. 개인 프로필은 기본 비공개, 회사 프로필은 기본 공유로 반대로 움직인다.



1 오픈AI가 ChatGPT 공유 프로필을 열어 사용자가 만든 Sites를 최대 12개까지 쇼케이스에 올리고 가장 많이 쓴 플러그인을 보여 주게 했다.

2 프로필에는 Chat 메시지 수와 Work·Codex 토큰 사용량, 연속 사용 일수가 나오지만 대화 제목과 내용은 나오지 않는다.

3 개인 프로필은 기본 비공개, ChatGPT Business 프로필은 워크스페이스 안에 기본 공유되며 공개 범위는 관리자가 정한다.

오픈AI가 ChatGPT에 공유 프로필(shareable profiles)을 열었다. 사용자가 ChatGPT 안에서 만든 Sites와 자주 쓰는 플러그인을 한 화면에 모아 다른 사람이 찾아보고 다시 쓰게 하는 기능이다. 오픈AI 개발자 계정은 9월 30일(현지시간) X에서 시연 영상과 함께 "만들었다면 보여 주라"고 밝혔다. 데브데이에서 발표한 기능을 단계적으로 배포하는 중이라 아직 일부 계정에는 보이지 않는다.

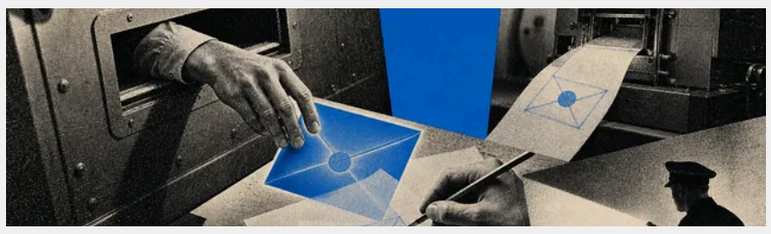
프로필 하나에는 이름과 사진, 활동 기록, 그리고 사용자가 고른 Sites가 담긴다. 오픈AI 도움말 문서에 따르면 쇼케이스에는 프로필마다 직접 소유한 Sites를 최대 12개까지 올릴 수 있고, 순서를 바꾸거나 빼는 것도 사용자가 정한다. 공유 프로필은 모든 국가에서 쓸 수 있으며 웹과 데스크톱에서 열린다. 쇼케이스는 모바일에서 아직 지원되지 않는다.

활동 기록은 숫자로 드러난다. 프로필은 Chat 메시지 수와 Work·Codex의 토큰 사용량, 연속 사용 일수를 보여 준다. 메시지 수와 토큰 사용량은 따로 표시되고 일간·주간·누적 합계가 함께 나온다. Chat 메시지나 Work·Codex 토큰 활동이 있는 날은 연속 사용 일수에 하루로 잡히며, 두 가지가 모두 있는 날도 한 번만 센다. 가장 많이 쓴 플러그인은 별도 칸에 모인다.

AI

오픈AI, 문샷 연관 증류 캠페인 차단 발표

오픈AI가 자사 모델의 보호된 추론을 빼내려던 조직적 증류 캠페인을 차단했다고 밝히고, 핵심 무리를 Kimi 개발사 문샷 AI 연관 개인들로 지목했다. 7월 이틀 동안 4,000명 넘는 사용자가 요청 1만 6,000건을 보냈다.



1 오픈AI가 9월 30일 자사 모델의 보호된 추론을 추출하려던 조직적 증류 캠페인을 적발해 차단했다고 발표하고, 핵심 무리를 문샷 AI와 연관된 개인들로 지목했다.

2 활동은 7월 1일 시작해 7월 24~25일 4,000명 넘는 사용자의 요청 1만 6,000건으로 치솟았고, 관련 패턴은 1만 5,000명 넘는 사용자로 번졌으나 7월 28일까지 완전히 차단됐다.

3 오픈AI는 계정 정지와 추론 보호 강화, 프런티어 모델 포럼·정부 채널 정보 공유로 대응했고, 이 사안을 오픈 모델이 아닌 이용약관 위반 문제로 규정했다.

오픈AI가 9월 30일 자사 모델의 보호된 추론을 빼내려던 조직적 캠페인을 적발해 차단했다고 발표했다. 회사는 이 활동의 핵심 무리를 Kimi 개발사 문샷 AI와 연관된 개인들로 지목했다. 보도에 따르면 오픈AI가 문샷을 공개적으로 지목한 것은 이번이 처음이다. 공격자들은 암호화를 깨거나 데이터베이스에 침입하지 않았고, 모델과의 상호작용을 조작해 숨겨진 추론을 요청자에게 보이는 형태로 되살려 냈다.

오픈AI는 이 활동을 적대적 증류로 규정했다. 발표문에 따르면 적대적 증류는 한 모델의 출력이나 추론을 허가 없이 체계적으로 써서 다른 모델을 훈련하거나 복제하거나 개선하는 행위다. 보호된 추론은 모델이 과제를 풀며 남기는 내부 기록으로, 최종 답변에서 뺀 정보를 드러낼 수 있고 남이 그 모델의 능력을 재현하는 데 쓰일 수 있다. 회사는 발표문에서 "이 조작은 오픈AI 모델에만 있는 취약점이 아니다"라고 밝혔다.

수법은 새로웠다. 공격자들은 한 대화에서 암호화된 추론을 복사한 뒤, 다른 대화의 모델에게 그 숨겨진 추론을 복호화해 받아 적으라고 요청했다. 독립 보안 연구자들도 모델 사이를 오가는 공격과 대화 압축 과정의 취약점을 책임 있는 공개 절차로 알려 왔고, 오픈AI는 그들이 찾은 공격 경로가 실제로 존재한다고 확인했다.

비즈니스

엔비디아 젠슨 황 CEO, 2026 밴 플리트상 수상…"한국과 미국, 함께 미래 만들어가고 있어"

수십 년간 이어진 한국과의 인연… 한국의 변화 "놀라운 기적"



1 수십 년간 이어진 한국과의 인연… 한국의 변화 "놀라운 기적"

2 AI 시대 한국의 강점으로 기술 수용력과 빠른 확산 역량 강조

3 한국의 제조 역량 기반으로 '인텔리전스 팩토리' 함께 구축

2026년 10월 1일 – AI 컴퓨팅 기술 분야의 선두주자인 엔비디아(www.nvidia.co.kr) 창립자 겸 CEO 젠슨 황(Jensen Huang)이 지난 9월 28일(현지 시간) 미국 뉴욕에서 열린 코리아 소사이어티(The Korea Society) 연례 갈라에서 2026 밴 플리트상(James A. Van Fleet Award)을 수상했다.

밴 플리트상은 한미 관계 발전에 크게 기여한 개인이나 기관에 수여되는 상이다. 이번 수상은 엔비디아가 AI, 반도체, 컴퓨팅 전반에서 한국의 기술·산업 생태계와 긴밀한 협력을 이어가고 있는 가운데 이뤄졌다.

젠슨 황 CEO는 수상소감을 통해 한국과 오랜 기간 이어온 인연을 돌아보고, AI 시대에서 한국이 수행할 역할에 대한 기대를 밝혔다.

AI

메타, 인도에 레이밴 메타 3세대 출시

메타가 인도에서 레이밴 메타 3세대를 출시하고 란비르 싱을 첫 브랜드 앰배서더로 세웠다. 43그램짜리 레이밴 메타 오디오 출시 예고와 Muse의 AI 안경 탑재 계획도 함께 밝혔다.



1 메타는 10월 1일 인도 뉴스룸에서 란비르 싱을 인도 첫 레이밴·레이밴 메타 브랜드 앰배서더로 발표하고 레이밴 메타 3세대 인도 출시를 알렸다.

2 3세대는 배터리 최대 9시간, 배경 소음을 90% 이상 줄이는 6개 마이크, 1,200만 화소 카메라와 3K 영상, 3가지 스타일·27가지 조합을 갖췄다.

3 메타는 무게 43그램의 레이밴 메타 오디오를 곧 내놓고, 몇 달 안에 개인 AI 에이전트 Muse를 AI 안경에 넣는다고 밝혔다.

메타가 인도에서 레이밴 메타 3세대를 출시하고, 란비르 싱을 인도 첫 레이밴·레이밴 메타 브랜드 앰배서더로 세웠다. 메타는 10월 1일 인도 뉴스룸 발표문에서 이 소식과 함께 오디오 중심의 새 안경 레이밴 메타 오디오의 출시 예고, 개인 AI 에이전트 Muse의 AI 안경 탑재 계획을 밝혔다. 메타는 레이밴 메타 3세대를 "세계에서 가장 많이 팔린 AI 안경의 다음 세대"라고 소개했다. 메탈이 확인한 메타 발표문 원문은 앰배서더 소개와 신제품 소식 세 가지를 한 글에 묶었고, 앞의 절반을 사양이 아니라 스타일 이야기에 썼다.

란비르 싱은 수년간 메타와 여러 차례 협업해 온 인물이다. 메타는 이번 계약이 인도에서 레이밴의 새 장을 연다고 밝혔다. 브랜드 앰배서더가 된 싱은 "레이밴은 언제나 개성을 상징해 왔고, 그것은 내가 기억하는 한 오래전부터 공감해 온 부분"이라고 말했다. 그는 "나에게 스타일은 정해진 공식을 따르는 것이 아니라 본능과 기분, 그리고 자신의 모든 모습을 표현할 자유"라고 했다. 이어 "모두를 위한 AI 안경을 만든다는 메타의 비전과 함께, 인도에서 그 이야기의 일부가 되어 기쁘다"고 밝혔다.

제품 쪽 핵심은 3세대의 인도 판매다. 메타에 따르면 3세대는 이전 세대의 기능을 더 얇고 편한 디자인에 담았고, 배터리는 한 번 충전으로 최대 9시간을 간다. 2세대보다 1시간 늘어난 수치다. 메타 AI를 바로 부르는 사용자 지정 액션 버튼이 새로 들어갔고, 착용감을 맞추도록 안경다리 끝을 전문가가 조정할 수 있게 했다. 마이크는 6개 배열로 바뀌어 배경 소음을 90% 이상 줄인다. 메타는 바람 불거나 붐비는 거리에서도 통화와 음성 명령이 또렷하게 들린다고 설명했다.

기후

유엔대, 데이터센터 환경비용 보고서

2030년 데이터센터 전력은 945테라와트시에 이를 전망이다. 유엔 연구진은 탄소만 보면 물과 토지로 옮겨 간 부담을 놓친다고 경고했다.



1 유엔대학교 연구진은 2030년 데이터센터 전력 수요를 945테라와트시, 연결된 물 발자국을 9조3000억 리터로 추산했다.

2 석탄을 바이오에너지로 바꾸면 탄소는 줄어도 물과 토지 사용이 크게 늘 수 있어 탄소 하나만으로 친환경을 판정할 수 없다.

3 연구진은 추론이 AI 에너지 사용의 80~90%를 차지한다며 모델 선택과 출력 길이, 해상도 같은 제품 기본값까지 관리하라고 제안했다.

AI를 쓰는 일은 대화창 안에서 끝나는 것처럼 보이지만 청구서는 전력망과 물길, 땅에 남는다. 유엔대학교 물·환경·보건연구소(UNU-INWEH)가 6월 3일 발표한 보고서는 2030년 세계 데이터센터 전력 사용량을 945테라와트시로 전망했다. 2030년 세계 예상 전력 사용량의 약 3%에 해당하는 기준 전망이다.

이 수치는 9월에 새로 발표된 값이 아니다. 9월 22일 기후 보도를 통해 다시 주목받은 6월 보고서의 전망이다. 보고서는 AI만 따로 떼어 측정한 전력량이 아니라 AI 확산이 성장을 이끄는 전체 데이터센터 전력 수요를 출발점으로 삼았다. 국제에너지기구(IEA)도 2024년 415테라와트시였던 데이터센터 전력 사용량이 2030년 945테라와트시로 두 배 넘게 늘어난다는 기준 전망을 제시했다.

UNU-INWEH 연구진이 새로 붙인 것은 전기 뒤에 숨은 세 개의 발자국이다. 2030년 전력 생산과 연결된 탄소 발자국은 이산화탄소환산량(CO_2e) 3억9900만 톤, 물 발자국은 9조3000억 리터, 토지 발자국은 1만4500제곱킬로미터로 추산됐다. 물은 사하라 이남 아프리카 13억 명의 기본적인 연간 가정용 수요와 맞먹고, 땅은 자카르타 대도시권의 약 두 배다.