AI 뉴스와 용어
월·수·금 오전 7시

LAB
METAL
어려운 AI, 쉽고 재미있게

무료 · 언제든 수신거부
metallab.ai

제 9 호 | 2026년 8월 23일 일요일 | 위클리

앤스로픽, 클로드 배우는 무료 교육 사이트 '아카데미' 열었다

AI 입문자부터 개발자까지, 제품별·개념별 강좌를 한 곳에 모았다

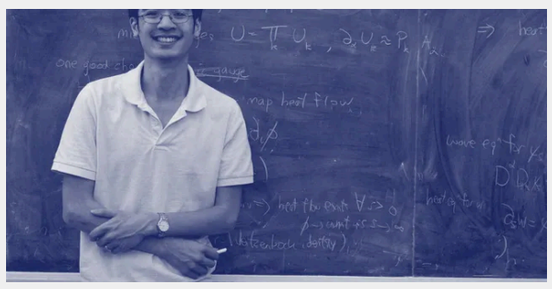


· 앤스로픽, 클로드 배우는 무료 교육 사이트 '아카데미' 열었다

AI가 뭔지도 잘 모르겠다"는 사람과 "클로드를 매일 쓴다"는 사람이 같은 사이트에서 각자 다른 강좌를 고를 수 있게 됐다. 앤스로픽은 8월 20일 X 공식 계정을 통해 클로드 아카데미(Claude Academy)가 문을 열었다고 알렸다. 회사는 게시물에서 "강좌와 튜토리얼은 무료이며 누구나 이용할 수 있다"고 밝혔다.

앤스로픽은 클로드를 만드는 AI 스타트업이다. 최근 몇 달 사이 이 회사는 대화형 챗봇 하나였던 클로드를 작업을 통째로 맡기는 클로드 코워크, 팀 채널에 태그해 업무를 넘기는 클로드 태그, API로 제품에 심는 클로드 플랫폼까지 늘려왔다. 실제로 지난 8월 19일과 20일에는 클로드 매니지드 에이전트에 메모리·도메인 제어·세션뷰어 기능을 추가하며 개발자용 플랫폼도 계속 확장했다. 제품이 갈래를 치는 만큼 어디서부터 배워야 할지 안내가 필요해졌고, 클로드 아카데미가 그 창구를 맡는다.

사이트는 크게 두 갈래다. 하나는 제품별 사용법이다. 대화창에서 문제를 풀고 문서를 쓰는 기본형 클로드닷에이아이, 작업 전체를 맡기고 결과물만 받는 클로드 코워크, 터미널·IDE·브라우저에서 코드를 짜는 클로드 코드, 팀 채널에 클로드를 태그해 업무를 넘기는 클로드 태그, API와 콘솔·MCP로 자사 제품에 클로드를 심는 클로드 플랫폼까지 다섯 카드로 나뉜다.



테렌스 타오 "AI, 괴델 이후 수학 최대 위기 부를 수도"

다른 한 갈래는 AI 자체를 이해하는 'AI 프루언시' 섹션이다. 자료로 먼저 돌려볼 수 있고, API로 제품을 만드는 개발자는 클로드 플랫폼 카드에서 MCP·콘솔 활용법을 확인할 수 있다.

페이지 하단에는 2026년 7월 8일 시행된 개인정보 처리방침과 이용정책 링크가 걸려 있어, 강좌 이용 중 입력한 정보가 어떻게 처리되는지도 함께 확인할 수 있다.

교육 콘텐츠를 별도 사이트로 떼어낸 것 자체가 신호다. 그동안 새 AI 도구는 릴리스 노트 한 줄이나 블로그 글 하나로 끝났고, 실제 사용법은 커뮤니티가 알아서 정리했다. 이번처럼 레슨·퀴즈·예상 소요 시간을 갖춘 4시간짜리 정규 커리큘럼을 통째로 내놓은 사례는 이 회사 제품군 중에서도 드물다. 그만큼 클로드가 챗봇 하나에서 코드·코워크·태그·플랫폼 네 갈래로 늘어났고, 사용자를 각 갈래로 안내할 지도가 필요해졌다는 뜻으로 읽힌다.

이 크기의 AI 회사가 자체 교육 커리큘럼을 만드는 건 결국 지원 비용을 줄이려는 계산이기도 하다. 신규 이용자가 헬프센터 검색이나 커뮤니티 질문 없이도 정해진 코스만 따라가면 되도록 만들어 두면, 문의 응대에 들어가는 인력을 그만큼 아낄 수 있다.

국내 기업이 클로드 도입을 검토하고 있다면, 계약 전에 이 무료 코스로 담당자 교육을 먼저 마치는 편이 안전하다. 특히 코워크나 태그처럼 아직 손에 익지 않은 협업형 기능은 실제 업무에 붙이기 전에 개념부터 맞춰두는 쪽이 시행착오를 줄인다. 앞으로 몇 주 안에 코워크·태그 강좌가 더 세분화되거나, 기업 온보딩용 패키지로 따로 묶여 나올 가능성이 크다.

커서, 깃허브 장애 날 노려 코드 호스팅 'Origin' 출시

개하지 않았고, 코딩 작업 전반을 지원할 더 넓은 앱 생태계도 구축 중이라고 밝혔다.

커서는 원래 AI 코드 편집기다. 화면 안에서 앤스로픽의 클로드, 오픈AI의 GPT, 구글의 제미나이 같은 다른 회사 모델을 골라 쓸 수 있게 해주는 한편, 컴포저라는 자체 모델도 직접 훈련해 함께 제공해왔다. 2026년 8월 스페이스X가 커서를 인수하면서 커서는 그록을 만드는 xAI와 같은 식구가 됐다. 코드 편집기에서 출발한 회사가 저장소 호스팅까지 손을 뻗은 셈이다.

Origin이라는 이름은 이번이 처음 등장한 게 아니다. 지난 13일 커서는 'Cursor Origin(Cursor Review)'이라는 이름으로 일부 파트너를 대상으로 비공개 베타를 진행 중이었고, 사이드바에 Codebase·Review 탭을 붙여 깃허브 저장소를 동기화하는 기능을 시험하고 있었다. 하루 뒤인 14일에는 벤처캐피털 a16z가 커서 공동창업자 마이클 트루엘의 2025년 인터뷰 발언을 다시 공유했는데, 그는 당시 모델 자체 개발과 멀티프로덕트 전환, 'AI 코딩 번들' 구축을 커서의 전략 축으로 꼽은 바 있다. 비공개 베타로 다지던 기능이 일주일여 만에 정식 출시로 이어진 흐름이다.

Origin을 쓰려면 별도 프로그램 설치 없이 깃허브 계정 연동부터 시작한다. 1) 커서 계정에서 깃허브 계정을 연결한다. 2) 연동할 조직(org)을 선택한다. 3) 동기화 가능한 저장소 목록이 뜨면 원하는 저장소를 고른다. 4) 선택하면 커서가 해당 저장소를 그대로 가져와 Origin 안에서 작업할 수 있게 한다. 요금제나 대기열 같은 이용 조건은 아직 공개되지 않았다.

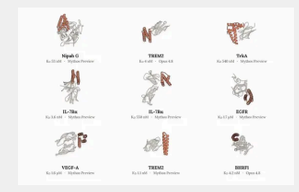
활용 예를 들면, 기존 깃허브 저장소는 그대로 두고 새로 시작하는 프로젝트만 Origin에서 관리해볼 수 있다. 팀 전체를 한 번에 옮기는 대신 일부 저장소만 먼저 동기화해 PR

오늘 지면

1 앤스로픽, 클로드 배우는 무료 교육 사이트 '아카데미' 열었다
2 튜링상 서튼 "합성 데이터로는 AI 확장 못 시킨다"
3 큐원3.8 27B, 뛰어나지만 기본값이 '과잉 사고'
4 구글 리서치, 이동 데이터로 장소 이해하는 AI 프레임워크 공개
5 커서, 깃허브 장애 날 노려 코드 호스팅 'Origin' 출시
6 테렌스 타오 "AI, 괴델 이후 수학 최대 위기 부를 수도"



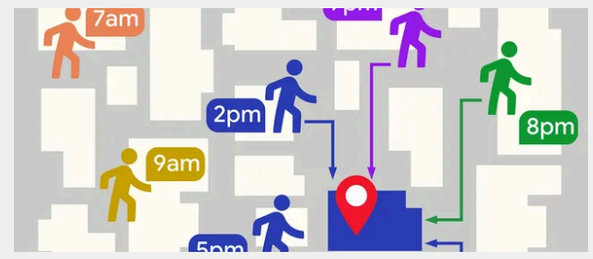
1인 개발자 125M 모델, 아이폰서 피아노 자동완성



클로드, 신약 결합 단백질 설계에서 업계 평균 웃도는 성공률 기록

이 지면은 2026년 8월 23일 일요일 아침 구독자 메일함으로 나간 레터를 신문 판형으로 다시 조판한 것입니다. 기사 전문은 metallab.ai 에서 읽을 수 있습니다.

이 어 서



구글 리서치, 이동 데이터로 장소 이해하는 AI 프레임워크 공개

지도 앱에 뜨는 어느 카페의 정보는 이름과 주소, 영업시간, '카페'라는 분류표뿐이다. 하지만 그 카페가 실제로 아침 7시에 붐비고 오후 3시에 한산해지는지, 주말엔 완전히 다른 손님층이 오는지는 그 표에 담기지 않는다. 구글 리서치는 8월 21일 블로그에서 이런 '서류상의 정체성'과 '실제 기능적 리듬'의 간극을 메우는 프레임워크 ME-POIs(Mobility-Embedded POIs)를 소개했다. 연구는 마리아 데스포이나 시암포우와 슈시먼

▶ 2면에 계속

튜링상 서튼 "합성 데이터로는 AI 확장 못 시킨다"



리처드 서튼은 강화학습 분야의 표준 교과서를 쓰고, 훗날 알파고를 만든 데이비드 실버를 제자로 길러낸 연구자다. 2019년 쓴 에세이 'The Bitter Lesson'에서 그는 장기적으로는 인간의 사전 지식을 넣은 방법이 아니라 연산량에 따라 그대로 확장되는 탐색·학습 방식이 결국 이긴다고 주장했다. 이 글은 이후 AI 업계에서 스케일링 법칙을 정당화하는 근거로 자주 인용돼 왔다.

그런 서튼이 최근 전 제자 커람 자비드와 함께 세운 스타트업 Oak Lab을 소개하는 자리에서, 정작 지금 업계가 쓰는 확장 전략 하나를 정면으로 비판했다. 합성 데이터, 즉 사람이 만든 데이터가 바닥났을 때 모델이나 시뮬레이터로 데이터를 대량 생성해 학습에 쓰는 방식이다. 서튼은 이를 "그건 그냥 큰 실수다"라고 잘라 말했다.

큐원3.8 27B, 뛰어나지만 기본값이 '과잉 사고'

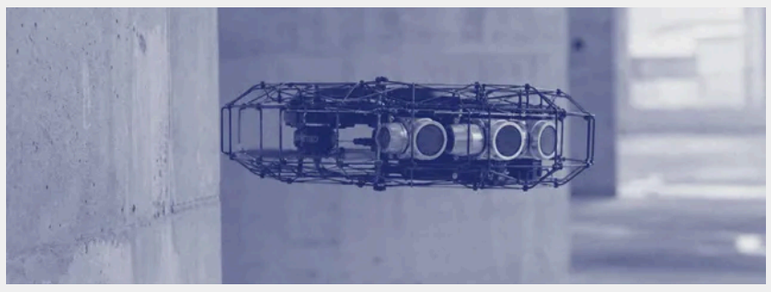


사이먼 윌리슨이 로컬 LLM을 테스트할 때 늘 쓰는 질문이 있다. "자전거를 타는 펠리컨을 SVG로 그려달라"는 것이다. 큐원3.8 27B에게 이 질문을 던졌더니 답이 나오기까지 21분이 걸렸다. 모델은 22,276개의 추론 토큰을 소모한 뒤에야 겨우 3,223개 토큰짜리 SVG를 내놨다. 결과물 자체는 로컬에서 돌아가는 모델 중 가장 정교한 펠리컨 SVG라고 할 만큼 훌륭했지만, 그 완성도를 위해 지불한 시간이 문제였다.

Qwen3.8 27B는 알리바바 큐원(Qwen) 연구소가 아파치 2.0 라이선스로 공개한 270억 파라미터 규모의 비전 지원 언어 모델이다. 27B — 게이밍용 노트북 한 대에서도 무리 없이 돌릴 수 있는 크기 — 급 모델 가운데 전작인 Qwen3.6 27B가 좋은 평가를 받았던 터라 후속작에 관심이 쏠렸다. 알리바바가 공개한 자체 벤치마크에서는 전작은 물론, 한때 큐원 라인업 최상위였던 폐쇄형 모델 Qwen3.7-Plus까지 앞서는 점수가 나왔다고 알려졌다. 다만 이 수치는 알리바바가 직접 발표한 것이라 독립 기관의 검증은 아직 확인되지 않았다.

◀ 1면에서 계속

튜링상 서튼 "합성 데이터로는 AI 확장 못 시킨다"



1 튜링상 수상자 리처드 서튼이 대형언어모델 확장 전략인 합성 데이터를 '큰 실수'라고 비판했다

2 세계는 무한히 복잡해 어떤 시뮬레이션도 '미시적'인 수준에 그치고, 좋은 데이터를 가리는 데 인간 전문가가 필요해 확장에 한계가 있다고 지적했다

3 서튼은 고정된 가중치의 언어모델 대신 스스로 경험을 쌓아 지속적으로 학습하는 에이전트를 대안으로 제시했다

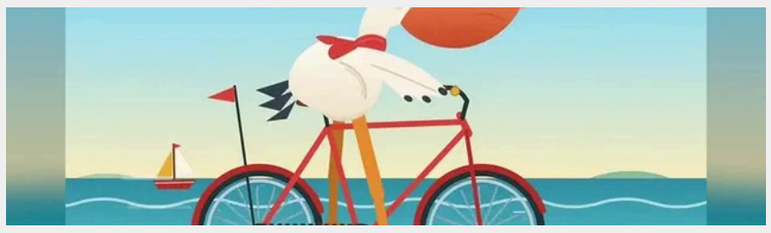
서튼은 대형언어모델을 'Bitter Lesson'의 좋은 사례이자 나쁜 사례로 동시에 꼽는다. 좋은 사례인 이유는 인터넷 전체를 "들이마시듯" 학습해 연산량에 비례해 실제로 성능이 커졌기 때문이다. 나쁜 사례인 이유는 바로 그 지점에서 한계에 부딪혔기 때문이다. 인터넷은 유한하고, 현실 세계는 "우리가 인터넷에 저장한 모든 것보다 훨씬 크다." 인터넷이라는 자원이 소진되자 모델은 결국 인간이 만들어 둔 지식에 더 크게 의존하게 되고, 이것이 오히려 시스템의 발목을 잡는다는 것이 서튼의 진단이다.

합성 데이터가 이 병목을 뚫을 수 있느냐는 질문에 그는 두 가지 근거를 든다. 첫째는 자비드가 정식화하고 앨버타 연구팀이 오랫동안 다뤄 온 'Big World Hypothesis'다. 세계는 무한히 복잡해서 어떤 마음, 어떤 에이전트보다도 훨씬 크다는 전제다. 이 복잡한 세계를 흉내 낸 시뮬레이션은 아무리 정교해도 결국 "미시적인" 크기에 그친다. 작은 프로그램은 작은 세계밖에 만들 수 없고, 그 세계는 마찰값이 다르거나 로봇 모터의 움직임을 부정확하게 재현하는 식으로 현실과 어긋난다. 게다가 세계에는 다른 행위자들의 내면이 존재하는데, 이건 합성 데이터로 만들 수 있는 대상이 아니다. "다른 사람들의 마음에 대한 합성 데이터는 만들 방법이 없다"는 게 서튼의 지적이다.

두 번째 근거는 인간 병목이다. 어떤 합성 데이터가 좋고 나쁜지는 결국 사람이 판단해야 한다. 자비드는 박쥐처럼 반향정위로 움직이는 드론을 학습시키려면 먼저 그 분야 전문가를 고용해야 한다는 예를 든다. 결국 이 방식은 인간 전문가의 수에 의해 제한되고, 따라서 확장되지 않는다. 시뮬레이션으로 학습시키는 자율주행차도 마찬가지다. 시뮬레이션과 실제 도로 사이의 간극은 결국 사람이 손으로 메워야 한다.

◀ 1면에서 계속

큐원3.8 27B, 뛰어나지만 기본값이 '과잉 사고'



1 알리바바 큐원랩의 270억 파라미터 모델 Qwen3.8 27B가 기본 추론 강도를 xhigh로 설정해 간단한 요청에도 과도하게 사고한다는 리뷰가 나왔다

2 펠리컨 자전거 SVG 테스트에서 추론을 켠 경우 21분과 22,276개 추론 토큰이 들었지만 끈 경우 137초 만에 끝났다

3 128GB 맥북 프로와 엔비디아 DGX Spark에서 실행한 테스트에서 코딩 에이전트 연동과 이미지 바운딩 박스 작업은 준수한 결과를 보였다

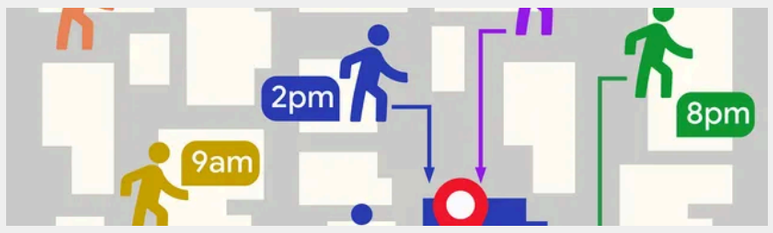
윌리슨은 128GB 메모리를 탑재한 M5 Max 맥북 프로와 엔비디아 DGX Spark 두 기기에서 LM Studio를 돌려 이 모델을 테스트했다. 실제로 실행한 빌드는 디스크 용량 17GB짜리 Q4KM 양자화 버전이었고, DGX Spark에서는 llama-server를 직접 띄워보기도 했다.

문제의 원인은 모델 설정에 있었다. 큐원 공식 문서는 이 모델의 reasoningeffort 추론 강도 기본값을 "xhigh"로 안내하고 있고, LM Studio가 배포한 GGUF 빌드도 이 값을 그대로 물려받았다. 여기에 LM Studio의 기본 컨텍스트 길이 제한인 8,192 토큰이 겹치면서, 모델은 "원을 그려달라"는 아주 단순한 요청에도 사고를 위해 남은 토큰을 전부 소진해버렸다. 윌리슨은 컨텍스트를 모델이 지원하는 최대치인 262,144 토큰으로 늘리고서야 이 문제를 피할 수 있었다.

원 그리기 테스트가 이 현상을 가장 잘 보여준다. xhigh 기본값이 켜진 상태에서 모델의 추론 과정은 "The user is asking for an SVG drawing of a circle"(사용자는 원의 SVG를 요청하고 있다)로 시작해, 동심원과 컴퍼스 눈금, 은은한 애니메이션까지 스스로 덧붙이며 수 분을 흘려보냈다. 결과물은 아름다웠지만 애초에 요청한 '원 하나'는 아니었다.

◀ 1면에서 계속

구글 리서치, 이동 데이터로 장소 이해하는 AI 프레임워크 공개



1 구글 리서치가 장소의 텍스트 정보에 이동 데이터를 결합한 ME-POIs 프레임워크를 공개했다

2 방문 의도 예측 최대 81.9%, 가격대 분류 75.1%, 혼잡도 추정 24.7% 향상됐다고 밝혔다

3 방문 정렬·공간 다중스케일 전파·텍스트-이동성 시너지 3단계 파이프라인으로 작동한다

이 프레임워크는 세 단계 파이프라인으로 작동한다. 첫 단계인 '방문 정렬'에서는 특정 장소에 대한 집계 방문 기록을 기본 데이터로 삼아 도착 시간대, 이탈 경향, 평균 체류 시간을 시간 인코더로 벡터 공간에 매핑한다. 이 과정을 거치면 1년 주기와 요일별 패턴을 아우르는 '기능적 중심점'이 만들어진다. 단순 평균이 아니라 다차원 서명을 만드는 방식이라 같은 '카페'라도 아침형 장소와 저녁형 장소가 다른 좌표를 갖게 된다. 이어지는 공간 다중스케일 방문 전파와 텍스트-이동성 시너지 단계에서 이 이동 신호를 텍스트 기반 표현과 결합해 하나의 임베딩 — 뜻을 숫자 좌표로 바꾼 값 — 으로 합친다.

구글 리서치는 이렇게 만든 표현을 언어모델에 결합했을 때 공개 벤치마크에서 다음과 같은 향상을 확인했다고 밝혔다.

세 지표 모두 상대적 개선폭이며, 특히 처음 보는 장소에서도 혼잡도 추정 정확도가 오른 점은 이 표현이 특정 장소를 외운 결과가 아니라 이동 패턴 자체를 일반화해 학습했다는 의미로 읽힌다.

커서, 깃허브 장애 날 노려 코드 호스팅 'Origin' 출시

깃허브가 6시간 넘게 먹통이던 날, 커서가 저장소·PR·협업까지 담은 대체 플랫폼을 냈다



1 커서가 이번 주 코드 호스팅 플랫폼 Origin을 출시해 깃허브가 담당해온 저장소·PR·협업 기능을 그대로 옮겨왔다

2 Origin 출시 당일 깃허브는 6시간 넘게 세계적 장애를 겪었고 오류율이 20%에 달했다. 최근 1년간 장애만 257건이라는 분석도 나왔다

3 Origin은 깃허브 계정을 연결해 저장소를 동기화하는 방식으로 깃허브와 병행 사용이 가능하며, 에이전트 네이티브 기능과 앱 생태계 확장이 예고됐다

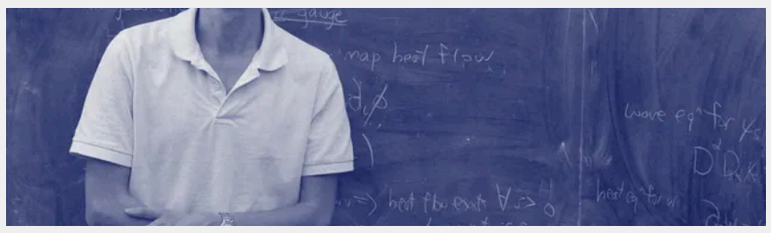
지난 18일(현지시간) 깃허브가 세계적으로 6시간 넘게 접속 장애를 겪었다. 오류율은 한때 20%에 달했다. 공교롭게도 같은 날 커서는 깃허브를 정면으로 겨냥한 코드 호스팅 플랫폼 Origin을 출시했다. 기술매체 LeadDev의 분석에 따르면 깃허브는 최근 1년간 257차례 장애를 겪었고, 이 매체 기자 찰스 험블은 이를 두고 유명 사용자들의 이탈이 눈에 띄게 늘고 있다고 짚었다.

깃허브는 전 세계 개발자가 코드를 올리고 함께 고치는 창고 역할을 해온 곳이다. 2007년 세워져 2012년 마이크로소프트에 인수됐고, 지난해 10월 기준 1억8000만 명이 이곳에서 코드를 관리한다고 스스로 밝히고 있다. 반복되는 장애에도 여전히 세계 최대 코드 호스팅 서비스라는 지위는 흔들리지 않은 셈이다.

Origin은 깃허브가 담당해온 핵심 기능을 그대로 옮겨왔다. 팀원끼리 코드베이스를 같이 작업하고, 코드를 열람·수정하고, 다른 사람이 올린 수정 제안(PR)을 처리하고, 저장소에 코드를 보관하는 일까지 한 화면에서 처리한다. 다만 Origin을 쓴다고 깃허브를 그만 써야 하는 건 아니다. 커서는 자사 블로그에서 "깃허브 저장소는 커서가 호스팅하는 저장소와 나란히 놓일 수 있다"고 소개했다. 깃허브 계정을 커서에 연결하면 조직 단위로 저장소를 골라 동기화할 수 있다는 설명이다. 커서는 이와 함께 '에이전트 네이티브' 기능을 곧 선보이겠다고 예고했지만 구체적인 내용은 아직 공개하지 않았고, 코딩 작업 전반을 지원할 더 넓은 앱 생태계도 구축 중이라고 밝혔다.

테렌스 타오 "AI, 괴델 이후 수학 최대 위기 부를 수도"

증명 검증할 사람보다 AI가 만드는 증명이 더 빨리 쌓일 수 있다는 경고



1 필즈상 수상자 테렌스 타오가 2026 국제수학자대회(ICM) 에세이에서 AI가 수학계에 1900년대 초 기초 위기에 준하는 혼란을 부를 수 있다고 경고했다

2 이번엔 수학적 진리가 아니라 무엇을 성과로 인정하고 누구의 공로로 볼 것인지를 정하는 암묵적 가치체계가 시험대에 올랐다고 진단했다

3 퍼스트-프루프 프로젝트 실험에서 미공개 연구문제 10개 중 7개가 AI 시스템으로부터 최소 1건의 통과 판정을 받았다고 인용했다

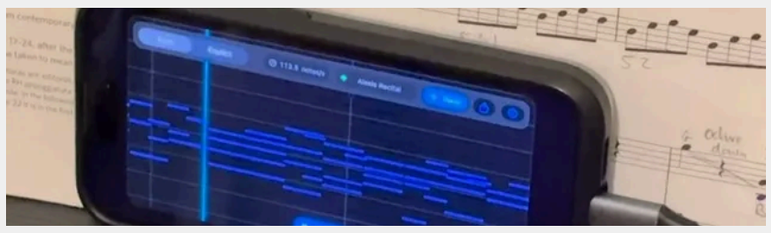
필즈상 수상자 테렌스 타오가 2026 국제수학자대회(ICM)에 제출한 에세이에서 AI가 수학계에 20세기 초 이후 가장 큰 위기를 불러올 수 있다고 경고했다. 그는 수학계가 "AI가 무엇을 할 수 있는가"라는 질문에 매달리는 대신, 그동안 미뤄온 질문과 정면으로 마주해야 한다고 썼다. 무엇을 연구 성과로 인정할지, 누구의 공로로 볼지, 무엇을 이해했다고 말할 수 있는지를 정하는 문제다.

타오는 이 상황을 1900년부터 1930년 사이 수학을 뒤흔든 기초 위기에 비유했다. 당시 러셀의 역설과 괴델의 불완전성 정리가 수학자들이 암묵적으로 넘어갔던 전제를 낱낱이 밝히도록 강제했고, 그 결과 한 세기를 버틴 엄밀한 체계가 만들어졌다. 지금 시험대에 오른 것은 진리의 정합성이 아니라 "무엇을 기여로 볼지, 무엇을 보상할지, 기계가 그 일을 했다고 말할 수 있는지"를 규정하는, 그동안 말로 정리한 적 없는 관행이라고 타오는 짚었다.

타오는 자신의 작업 가설을 이렇게 요약했다. "AI 도구는 머지않아 연구 수준의 수학 과제 상당 부분을, 합리적인 성공률과 품질, 감독 수준, 비용으로 수행할 수 있게 될 것이다." 근거로 든 것이 퍼스트-프루프 프로젝트다. 2차 라운드 실험에서 아직 발표된 적 없는 연구 문제 10개를 AI 시스템 4개에 통제된 조건으로 풀게 했더니, 7개 문제가 적어도 하나의 시스템으로부터 통과 판정을 받았다. 사실상 흠 없는 풀이이거나 소폭 수정만 필요한 수준이었고, 비용은 문제당 수십에서 수백 달러였다.

1인 개발자 125M 모델, 아이폰서 피아노 자동완성

몇 음만 쳐도 나머지를 실시간으로 이어 치는 온디바이스 모델, 아이폰15서 초당 108음



1 1억 2,500만 파라미터 트랜스포머가 아이폰15에서 초당 약 108개 음을 실시간 생성한다.

2 학습 데이터를 5배로 늘리자 오히려 성능이 떨어져, 정제된 수십만 개 MIDI 파일로 최종 모델을 완성했다.

3 DPO 사후학습을 거친 뒤 기존 모델보다 더 나은 이어짐으로 평가된 비율이 69%를 넘었다.

미디(MIDI) 건반에 짧은 멜로디를 몇 마디 치면, 아이폰 화면 뒤에서 작동하는 모델이 나머지 부분을 실시간으로 이어 친다. 해커뉴스 아이디 simedw로 알려진 1인 개발자가 1억 2,500만 파라미터(125M) 규모의 트랜스포머 모델을 훈련해, 아이폰15에서 초당 약 108개 음을 생성하는 피아노 자동완성 앱 'RollTab'을 무료로 공개했다. 사람이 실제로 연주할 수 있는 속도를 웃도는 수치다.

방식은 개발자 사이에서 익숙한 도구에 빗댈 수 있다. 깃허브 코파일럿이 코드 몇 줄이나 주석을 프롬프트 삼아 나머지 코드를 채워주듯, 이 모델은 사용자가 건반으로 친 몇 개 음을 프롬프트로 받아 곡을 이어 붙인다. 다만 코파일럿이 사용자가 고른 타사 모델(GPT, 클로드, 제미나이 등)로 작동하는 창구형 제품인 것과 달리, RollTab은 개발자가 직접 만든 모델을 기기 안에서 통째로 돌린다. 서버 통신 없이 아이폰 안에서 모든 추론이 끝난다는 뜻이다.

개발자는 거의 1년에 걸쳐 열네 차례 실험을 거쳐 이 결과물에 도달했다고 밝혔다. "Think GitHub Copilot, but for piano" — 그가 블로그 글 서두에 남긴 한 줄이 프로젝트의 목표를 압축한다.

오늘 나온 어려운 말

풀이 전문은 metallab.ai/glossary

미토스 프리뷰
Mythos Preview
앤스로픽이 신약 개발용 단백질 설계 같은 실전 과제로 시험한 클로드의 실험용 미리보기 버전.

퍼스트-프루프 프로젝트
First-Proof Project
아직 공개된 적 없는 연구 수준 수학 문제를 AI에게 풀게 해 실전 증명 능력을 검증하는 프로젝트

라이덴 선언
Leiden Declaration
2026년 6월 공개돼 국제수학연맹의 지지를 받은, AI가 만든 수학 증명을 어떻게 검증하고 공로를 인정할지 정한 선언문

코어ML
Core ML
애플이 만든 온디바이스 머신러닝 실행 도구로, 아이폰·맥 안에서 서버 없이 AI 모델을 돌리게 해준다.

오늘 해볼 것

오늘 지면에 나온 것 중 지금 바로 해볼 수 있는 4가지

1 다음 토큰 예측

Next Token Prediction

언어모델이 지금까지 나온 글 뒤에 올 확률이 가장 높은 다음 조각을 하나씩 골라 이어 붙이며 문장을 만드는 방식

이게 뭔데

다음 토큰 예측은 스마트폰 자판의 다음 단어 추천 기능을 훨씬 정교하게 키운 것이라고 보면 된다. 자판이 지금까지 친 글자를 보고 다음에 올 단어 후보 몇 개를 보여주듯, 언어모델도 지금까지 나온 글을 보고 다음에 올 조각 하나를 확률로 계산해 고른다. 그 조각을 문장에 더하고, 다시 그 문장 전체를 보고 또 다음 조각을 고르는 과정을 계속 반복하면 결국 긴 답변이 완성된다.

재미있는 점은 이 단순한 반복 작업만으로 요약, 번역, 코드 작성처럼 복잡해 보이는 일까지 해낸다는 것이다. 다음에 올 조각을 잘 맞히려면 문법은 물론 맥락과 상식까지 어느 정도 파악해야 하기 때문에, 방대한 글을 보며 이 예측 훈련을 거듭한 모델은 자연스럽게 폭넓은 능력을 갖추게 된다. 반대로 이 방식의 한계도 여기서 나온다. 모델은 사실을 검색해 확인하는 게 아니라 그럴듯하게 이어질 조각을 고르는 것이라서, 확신에 찬 말투로 틀린 내용을 만들어내는 일이 생기기도 한다.

해 보려면

챗봇 창에 다음과 같이 문장을 절반만 던져보자.

다음 문장을 자연스럽게 이어서 완성해줘: 오늘 아침 하늘을 보니

답을 받은 뒤, 왜 그 다음 말을 골랐는지 물어보면 모델이 앞 문맥에 맞춰 확률적으로 다음 조각을 이어간다는 점을 체감할 수 있다.

오늘 지면에서 — 앤스로픽, 클로드 배우는 무료 교육 사이트 '아카데미' 열었다

2 교차엔트로피 손실

cross-entropy loss

모델이 낸 확률 예측이 정답과 얼마나 다른지를 재는 학습용 채점 기준

이게 뭔데

교차엔트로피 손실은 AI 모델이 학습 중에 낸 답이 정답과 얼마나 어긋났는지를 숫자로 매기는 채점 방식이다. 객관식 시험에서 학생이 정답 하나를 확신 있게 고르면 만점에 가깝고, 오답에 확신을 크게 줄수록 감점이 커지는 채점표를 떠올리면 된다. 모델은 다음에 올 단어나 음 하나를 고를 때 여러 후보에 확률을 나눠 부여하는데, 실제 정답에 낮은 확률을 줬다면 벌점을 크게, 높은 확률을 줬다면 벌점을 작게 매겨서 그 점수를 줄이는 방향으로 모델을 조금씩 고쳐나간다.

다만 이 채점은 '다음 한 걸음'만 보고 매기는 근시안적 기준이라는 한계가 있다. 매 순간 정답 하나를 잘 맞히는 습관은 들일 수 있지만, 여러 걸음이 모여 완성된 결과물 전체가 자연스럽고 그럴듯한지는 별도로 따져야 한다. 그래서 기본 문법을 가르치는 초기 학습에는 널리 쓰이지만, 완성본의 전반적인 품질을 다듬는 단계에서는 다른 방식을 덧붙이는 경우가 많다.

해 보려면

챗봇에게 이렇게 물어보면 개념이 잡힌다: "교차엔트로피 손실을 객관식 시험 채점에 비유해서 초등학생도 이해할 수 있게 설명해줘. 모델이 정답에 낮은 확률을 줬을 때와 높은 확률을 줬을 때 점수가 어떻게 달라지는지도 알려줘."

오늘 지면에서 — 1인 개발자 125M 모델, 아이폰서 피아노 자동완성

3 샘플링 방식

Sampling Methods

AI가 다음 단어나 음, 그림 조각을 고를 때 여러 후보 중 하나를 확률에 따라 골라내는 방법

이게 뭔데

샘플링 방식은 AI가 다음에 나올 말이나 음을 고르는 방법이다. 식당에서 오늘의 메뉴판을 받았다고 생각해보자. 셰프는 손님에게 '이 요리가 60% 인기, 이 요리가 25% 인기'라고 확률을 알려줄 수 있는데, 손님이 항상 60% 짜리만 고르면 매번 똑같은 식사가 되고, 가끔은 25%짜리나 그보다 낮은 메뉴도 시도하면 매번 다른 식사가 된다. AI 모델도 다음에 올 단어나 음의 확률을 계산해두고, 이 확률표에서 하나를 뽑는 방식을 정해야 하는데 그 규칙을 샘플링 방식이라 부른다.

가장 단순한 규칙은 확률이 가장 높은 것만 무조건 고르는 것이다. 이렇게 하면 같은 질문에 항상 같은 답이 나오지만, 밋밋하고 반복적인 결과가 나오기 쉽다. 반대로 확률이 낮은 후보도 종종 뽑히도록 무작위성을 섞으면 매번 조금씩 다른, 더 다양하고 자연스러운 결과가 나온다. 아이폰에서 몇 음만 치면 나머지를 이어 치는 피아노 자동완성 앱처럼, 모델이 다음 음 하나를 예측할 때도 이 확률표에서 어떻게 음을 골라내느냐가 결과물의 자연스러움을 크게 좌우한다.

같은 모델이라도 샘플링 방식을 어떻게 정하느냐에 따라 결과가 딱딱해 보이거나 창의적으로 보일 수 있다. 그래서 개발자들은 정답이 하나로 정해지지 않는 문제일수록, 확률이 가장 높은 것만 고르는 방식보다 약간의 무작위성을 섞은 샘플링을 선호하는 경우가 많다.

해 보려면

챗봇에게 똑같은 질문을 여러 번 새로 물어보라. 답이 매번 조금씩 다르게 나온다면 그 모델이 확률이 가장 높은 답만 고르지 않고 무작위성을 섞어 다음 단어를 고르고 있다는 뜻이다. 반대로 매번 거의 똑같은 답이 나온다면 가장 확률 높은 후보만 고르는 방식에 가깝다.

오늘 지면에서 — 1인 개발자 125M 모델, 아이폰서 피아노 자동완성

4 INT8 양자화

INT8 quantization

모델 내부 숫자를 8비트 정수로 뭉뚱그려 표현해 크기와 연산량을 줄이는 압축 기법

이게 뭔데

INT8 양자화는 AI 모델 안의 숫자를 더 거친 단위로 반올림해서 저장 용량과 계산량을 줄이는 방법이다. 사진을 원본 그대로 저장하지 않고 압축 파일로 저장하면 용량은 훨씬 줄지만 눈으로 보기엔 큰 차이가 안 나는 것과 비슷하다. 모델은 보통 소수점 아래까지 세밀하게 표현된 숫자 수십억 개로 이루어져 있는데, 이걸 정수 256단계(8비트)로만 뭉뚱그리면 모델 크기가 최대 4분의 1 수준으로 줄고 계산도 훨씬 빨라진다.

즉 미토스 프리뷰는 클로드라는 이름의 완제품이 아니라, '이 정도 자율성과 능력을 실전에서 발휘할 수 있는지' 미리 확인해 보기 위해 잠깐 무대에 올린 시험판에 가깝다.

퍼스트-프루프 프로젝트

First-Proof Project

퍼스트-프루프 프로젝트는 아직 세상에 발표된 적 없는 수학 연구 문제를 AI에게 풀게 해서, 그 실력을 진짜 현장에서 시험해 보는 검증 실험이다. 보통 AI의 수학 실력을 잴 때는 이미 정답이 알려진 문제집을 쓰지만, 이 프로젝트는 다르다. 아무도 답을 내놓은 적 없는 진짜 연구 문제를 새로 준비해서, AI가 어딘가에서 정답을 외워왔을 가능성을 원천적으로 없앤다.

두 번째 라운드 실험에서는 이렇게 준비한 새 문제 10개를 서로 다른 AI 시스템 네 개에게 통제된 조건으로 풀게 했다. 결과는 절반을 훌쩍 넘겼다. 10개 중 7개 문제에서 적어도 하나의 시스템이 거의 흠 없는 풀이를 내놓거나, 약간만 손보면 되는 수준의 답을 만들어냈다. 여기에 든 비용은 문제 하나당 수십에서 수백 달러 정도였다.

이 결과는 필즈상 수상자 테렌스 타오가 자신의 예측—AI가 머지않아 연구 수준 수학 과제 상당 부분을 합리적인 비용과 품질로 해낼 것이라는 가설—을 뒷받침하는 근거로 인용하면서 널리 알려졌다. 다만 타오는 이런 성과가 쌓일수록, 아무도 다 확인하지 못하는 속도로 증명이 쏟아지는 새로운 문제도 함께 생긴다고 경고했다.

라이덴 선언

Leiden Declaration

라이덴 선언은 AI가 만든 수학 증명을 놓고 무엇을 성과로 인정하고, 누구의 공로로 볼지, 어떤 절차로 검증할지를 정리한 문서다. 학회나 단체가 모여 만든 일종의 약속으로, 법적 구속력이 있는 규정이 아니라 업계 관행을 명문화한 선언에 가깝다.

비유하자면 새로운 도구가 등장했을 때 경기 규칙을 다시 쓰는 것과 비슷하다. 예전에는 증명을 사람이 직접 손으로 썼기 때문에 누가 어떤 아이디어를 냈는지, 어디까지 검증됐는지가 비교적 분명했다. 그런데 AI가 순식간에 그럴듯한 증명을 여러 개 만들어내기 시작하자, 이걸 논문으로 인정할 수 있는지, 누구 이름으로 발표해야 하는지, 어느 수준까지 사람이 확인해야 믿을 수 있는지를 두고 기준이 모호해졌다. 라이덴 선언은 이 혼란에 최소한의 원칙을 세우려는 시도다.

2026년 6월 공개됐고 국제수학연맹이 지지를 표명했다는 점에서, 개인이나 특정 회사가 아니라 수학계 전체가 참고할 만한 권고안으로 다뤄지고 있다.

코어ML

Core ML

코어ML은 스마트폰이나 컴퓨터 자체에 내장된 계산기라고 생각하면 쉽다. 보통 AI 기능은 인터넷 너머 큰 서버에 질문을 보내고 답을 받아오는 식으로 작동하지만, 코어ML은 그 계산을 기기 안에서 바로 끝낸다. 애플이 만든 이 도구를 쓰면 개발자는 미리 훈련해 둔 모델을 아이폰이나 맥에 넣어두고, 인터넷 연결 없이도 사진 분류나 소리 인식, 글자 완성 같은 작업을 실시간으로 처리하는 앱을 만들 수 있다.

서버로 데이터를 보내지 않으니 개인정보가 기기 밖으로 나가지 않고, 응답 속도도 오가는 시간이 없어 빠르다. 대신 큰 서버만큼 강력한 계산은 어렵기 때문에, 코어ML로 돌아가는 모델은 보통 크기를 작게 줄여 기기 성능에 맞춘다. 애플 기기에서 카메라 앱이 사진 속 사물을 즉시 알아보거나, 자판이 다음 단어를 예측하는 기능 뒤에 이런 방식이 흔히 쓰인다.

빅 월드 가설

Big World Hypothesis

빅 월드 가설은 우리가 살아가는 세계가 어떤 사람의 머릿속이나 AI 에이전트가 가진 지식보다 훨씬 크고 복잡하다는 전제다. 아무리 정밀한 지도를 그려도 실제 땅 전체를 지도 안에 다 담을 수는 없다. 세계와 그 세계를 이해하려는 존재 사이에는 항상 이런 크기 차이가 있다는 뜻이다.

이 생각이 중요해지는 지점은 AI를 학습시킬 때다. 사람이 만든 데이터가 부족해지면, 요즘은 모델이나 가상의 시뮬레이터로 데이터를 대량으로 지어내 학습에 쓴다. 그런데 빅 월드 가설에 따르면 이렇게 만든 가짜 세계는 아무리 정교해도 결국 작은 프로그램이 만든 작은 세계일 뿐이다. 실제 마찰력이나 로봇 모터의 미묘한 움직임을 정확히 재현하지 못하고, 무엇보다 다른 사람들의 속마음처럼 애초에 흉내조차 낼 수 없는 부분이 세계에는 존재한다.

그래서 이 가설은 곧 하나의 반론으로 이어진다. 진짜 세계가 그렇게 크고 복잡하다면, AI는 미리 만들어진 가짜 세계로 훈련받는 대신 실제 경험을 통해 끊임없이 스스로 배우고 고쳐나가야 한다는 것이다.

시간 인코더

temporal encoder

시간 인코더는 몇 시에 왔는지, 얼마나 머물렀는지 같은 시간 정보를 AI가 계산할 수 있는 숫자 좌표로 바꿔주는 장치다.

비유하자면 시계 바늘이 가리키는 위치를 읽어서 그걸 지도 위의 한 점으로 옮겨 찍는 일과 비슷하다. 사람은 '아침 7시에 붐빈다'는 말을 그냥 알아듣지만, 컴퓨터는 숫자로 바뀐 좌표가 있어야 그 리듬을 다른 장소와 비교하거나 패턴을 찾을 수 있다. 시간 인코더는 도착 시간대, 머문 시간, 요일별 흐름 같은 정보를 이런 좌표로 정리해주는 역할을 한다.

이 좌표들이 쌓이면 아침형 카페와 저녁형 카페처럼 겉보기엔 같은 업종이라도 서로 다른 위치에 놓이게 된다. 결국 시간 인코더는 '언제'라는 흐르는 개념을 AI가 다룰 수 있는 정적인 숫자로 붙잡아두는 다리 역할을 하는 셈이다.

텍스트-이동성 시너지

text-mobility synergy

텍스트-이동성 시너지란 장소를 설명하는 글자 정보와 사람들이 그 장소를 실제로 오가는 움직임 정보를 하나로 합쳐, 컴퓨터가 다룰 수 있는 하나의 숫자 좌표로 만드는 조립 단계다.

이력서와 실제 근무 태도를 함께 보고 한 사람을 판단하는 상황을 떠올리면 쉽다. 이력서에는 '카페, 종로구, 오전 7시~오후 9시'라고만 적혀 있지만, 그 카페가 아침엔 붐비고 저녁엔 한산한지, 아니면 그 반대인지는 이력서만으론 알 수 없다. 텍스트-이동성 시너지는 이 서류상 정보와 실제 방문 기록을 한 장의 카드로 압축해 담는 마무리 작업이다.

이 작업을 거치면 이름표는 똑같이 '카페'라도 아침형 가게와 저녁형 가게가 서로 다른 좌표를 갖게 된다. 텍스트만 봤을 땐 구분되지 않던 두 장소가, 이동 데이터가 섞이는 순간 비로소 서로 다른 존재로 나뉘는 셈이다.

공간 다중스케일 방문 전파

spatial multi-scale visit propagation

공간 다중스케일 방문 전파는 방문 기록이 거의 없는 장소의 활동 패턴을, 주변 장소들의 이동 데이터로 대신 채워 넣는 기법이다.

동네에 막 생긴 작은 가게를 떠올려보자. 아직 방문객이 적어 '이 가게가 언제 붐비는지' 같은 데이터가 거의 없다. 이럴 때 바로 옆 가게, 한 블록 떨어진 상권, 더 넓은 동네 전체의 방문 패턴을 가까운 범위부터 순서대로 참고하면 '이 동네는 대체로 점심에 붐비고 밤엔 한산하다'는 식의 추정을 세울 수 있다. 이렇게 가까운 규모부터 넓은 규모까지 단계적으로 이웃의 정보를 끌어와 빈틈을 메우는 방식이 바로 이 기법이다.

이 개념은 구글 리서치가 공개한 장소 이해 프레임워크의 한 처리 단계로 등장했다. 유명 랜드마크나 대형 매장은 방문 기록이 풍부해 패턴을 뽑기 쉽지만, 동네 골목의 작은 가게들은 기록이 희박해 신뢰할 만한 패턴을 만들기 어려운 '롱테일' 문제가 있다. 공간