AI 뉴스와 용어
월·수·금 오전 7시

METAL
First, AI Magazine

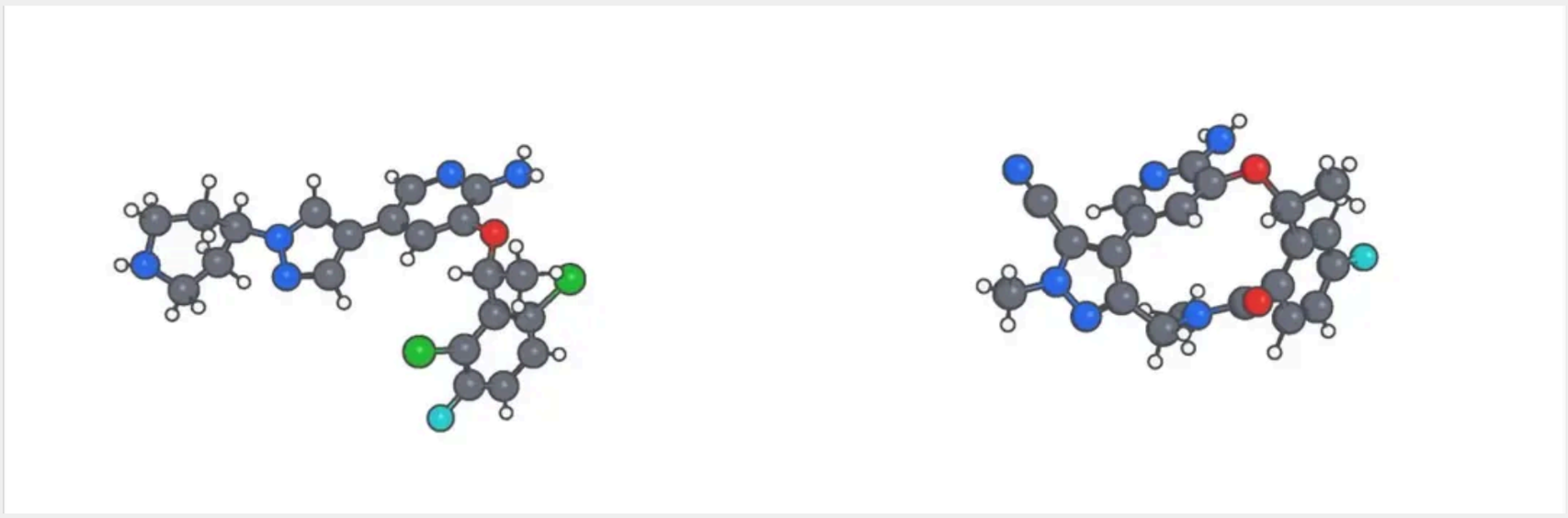
무료 · 언제든 수신거부
metallab.ai

제 59 호 2026년 11월 5일 목요일 오늘의 AI

AI

구글 딥마인드, 안티그래비티 화학정보학 스킬 공개

에이전트가 폐암 치료제 사례로 화합물 수천 개를 정리하고 합성 경로를 붙인 새 고리형 분자를 한 시간 만에 제안했다. 딥마인드는 과학자 1만 명에게 최고 사용량 등급을 6개월간 무료로 제공한다.

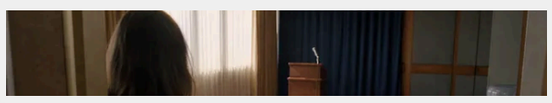


AI · 구글 딥마인드, 안티그래비티 화학정보학 스킬 공개

구글 딥마인드가 11월 3일(현지 시각) 구글의 에이전트 앱 Google Antigravity에 신약 설계용 화학정보학 스킬을 붙여 공개했다. 딥마인드에 따르면 이 스킬을 단 에이전트는 실제 폐암 치료제 사례를 놓고 수천 개 화합물을 찾아 정리하고, 짝분자 분석을 돌린 뒤, 합성 경로까지 붙인 새 고리형 분자를 한 시간 만에 제안했다. 보통 한 번 반복에 며칠이 걸리던 작업이다. 딥마인드는 과학자 1만 명에게 안티그래비티 최고 사용량 등급을 6개월간 무료로 제공하겠다며 신청을 받기 시작했다.

신약 설계의 병목은 데이터 정리다. 같은 분자가 논문에서는 IUPAC 이름으로, 특허에서는 마쿠시 구조로, 공급사 카탈로그에서는 SDF 파일로 나온다. 원자 하나나 결합 하나만 틀려도 전혀 다른 화합물이 된다. 딥마인드는 이 일에 오차 여유가 없다며, 유망한 억제제가 쓸모없는 물질로 바뀐다고 설명했다. 화학정보학 스킬은 과학자가 이런 정리 대신 가설에 집중하게 하려고 만들었다는 것이 회사 설명이다.

분업 구조가 핵심이다. 언어 모델은 작업 흐름 조율과 데이터베이스 탐색, 과학적 추론을 맡고, 분자 계산은 결정론적 화학 도구가 맡는다. 바탕은 전 세계 기여자가 20년 넘게 키운 오픈소스 화학정보학 도구 RDKit이고, 공개 화학 데이터베이스 ChEMBL과 PubChem 연결도 들어 있다. 공개 저장소 science-skills에 따르면 이 스킬 묶음은 유전체학과 구조생물학, 문헌 검색까지 아우르며, AlphaGenome과 AFDB, UniProt 등 30개가 넘는 데이터베이스와 도구를 엮는다.



비틀린다. 표적에 붙으려면 한 모양으로 굳어야 하고, 그 자유를 잃는 대가로 결합력이 약해진다. 화학자들은 결합을 하나 더 걸어 분자를 미리 묶는데, 딥마인드는 이를 "엔트로피 비용을 미리 치른다"는 말로 소개했다. 비소세포폐암 1세대 치료제 크리조티닙에는 주요 회전 결합이 5개 있고, 3세대 로를라티닙은 끝을 고리로 닫아 실효 회전 결합 수를 0개로 줄였다.

메탈이 확인한 30초 분량의 공개 영상은 이 차이를 그대로 보여 준다. 크리조티닙에는 흐물흐물하다는 뜻의 floppy, 로를라티닙에는 단단하다는 뜻의 rigid라는 이름표가 붙는다. 이어지는 단백질 결합 장면에서 크리조티닙(PDB 2XP2)은 결합하려면 굳어야 하고, 로를라티닙(PDB 4CLI)은 결합 모양을 미리 갖췄다는 설명이 달린다.



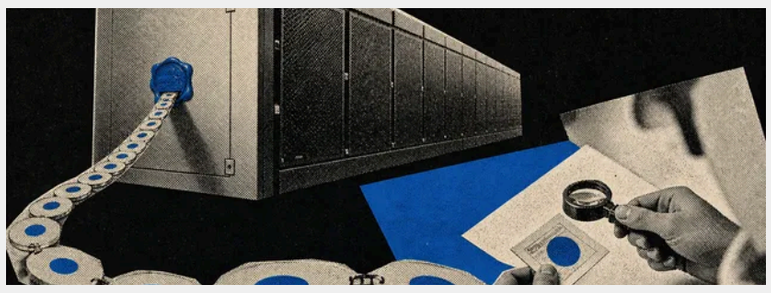
구글 리서치, 애자일 딜리버레이션 공개

에이전트는 문헌에서 두 약과 유사체를 찾는 일부터 시작했다. 딥마인드에 따르면 에이전트는 몇 분 만에 화합물 수천 개를 가져오고, 완충액과 용매 같은 무관한 분자를 걸러 출처를 보고서로 남겼다. 보고서는 고리형 분자 대부분이 뇌로 들어갈 수 있을 것으로 예측된다는 점도 짚었다. 대상 폐암이 뇌로 자주 전이되는 만큼 중요한 성질이다. 대화형 뷰어에서 세포 수준 효능을 회전 결합 수에 대해 그리자, 효능이 가장 높은 화합물이 가장 단단한 분자라는 점이 드러났

◀ 1면에서 계속

AI

앤스로픽, 추론 증명 FlashProve 공개



1 앤스로픽은 11월 3일 AI 응답이 특정 모델에서 나왔음을 암호학적으로 증명하는 연구 시스템 FlashProve를 Apache 2.0 라이선스로 공개했다.

2 4000억 매개변수 Llama 4 Maverick을 H100 한 장에서 초당 약 110토큰으로 증명했고, 18.78MB 증명 하나를 CPU 24스레드에서 17초에 검증한다.

3 증명 비용은 일반 추론의 약 72~94배이며, 문맥 1024토큰·정수 모델 그리디 디코딩 한정의 연구 프로토타입으로 배포 계획은 없다.

FlashProve가 쓰는 도구는 영지식 간결 증명, 곧 zkSNARK다. 제공자는 오프라인 단계에서 가중치에 대한 커밋먼트를 한 번 공개하고, 온라인 단계에서 요청마다 "커밋한 모델이 이 입력에서 이 출력 토큰을 낸다"는 명제를 증명한다. 증명은 위조할 수 없고(건전성), 가중치와 중간 계산, 각 층에서 어떤 전문가 모듈이 돌았는지를 숨기며(영지식), 실제 계산보다 훨씬 빨리 확인된다(간결성). 다만 모델 구조는 숨기지 않는다. 기술 보고서에 따르면 이 체계는 해시 함수 BLAKE2의 안전성에만 기대며, 양자 컴퓨터 공격에도 견딜 것으로 보는 구성이다. 요청 하나를 증명하는 데 합 검사(sumcheck) 절차가 약 2000번 돈다.

수치는 구체적이다. 시험 모델은 전문가 128개로 이뤄진 4000억 매개변수 전문가 혼합(MoE) 모델 Llama 4 Maverick이고, 토큰마다 약 170억 개 매개변수만 쓰인다. 공개된 4비트 가중치는 232GB다. 증명기는 H100 한 장과 CPU 코어 24개로 돈다. 커밋 단계는 모델당 한 번 약 39분 걸리고 제공자는 2.15GB의 파일을 보관하며, 한 번의 커밋으로 약 1600만 건의 요청을 다룬다. 1024토큰 요청 24건 묶음을 약 223초에 증명해 초당 110토큰이 나오고, 증명 파일 하나는 18.78MB다. 검증 쪽은 GPU가 필요 없다. 기술 보고서 기준으로 사용자는 커밋먼트 등 6.24MB를 한 번 내려받고, 증명 하나를 24스레드에서 17초, 단일 스레드에서 21초에 확인한다.

비용도 숫자로 나왔다. 같은 모델의 FP8 판을 H100 여덟 장에서 vLLM으로 돌린 것과 견주면 GPU당 증명 시간은 일반 처리의 약 72배(측정마다 66~77배)였고, vLLM을 가장 높은 부하로 돌렸을 때와 견주면 약 94배였다. 묶음으로 처리하면 요청 하나가 증명을 받기까지 약 5분이 걸리고, 한 건씩 증명하면 58초에 나오지만 처리량은 초당 17.6토큰으로 떨어진다. 가중치를 층 단위로 나눠 일반 메모리에서 흘려보내는 설계 덕에 80GB GPU 한 장과 약 290GB의 호스트 메모리면 충분하다.

◀ 1면에서 계속

AI

앤스로픽, 클로드 플러그인 자동 연결 베타 공개



1 앤스로픽은 11월 3일 클로드에 로그인하면 업무용 플러그인이 자동 연결되는 클로드 발급형 인증을 Team·Enterprise 요금제 베타로 공개했다.

2 클로드는 1분 안에 만료되는 서명된 JWT 확인서를 플러그인 인가 서버에 보내고, 지원 플러그인은 Figma, Notion, Linear, Datadog 등 14개다.

3 클로드는 조직 구성원 여부만 확인하고 SSO·MFA는 확인하지 않으며, 퇴사자의 기존 액세스 토큰은 만료 때까지 유효하다.

이번 기능은 앤스로픽이 먼저 내놓은 엔터프라이즈 관리형 인증(EMA)의 두 번째 갈래다. EMA는 관리자가 조직의 ID 공급자를 통해 플러그인을 조직 전체에 연결하는 방식으로, MCP의 공개 확장 규격 위에 만들어졌다. 고객지원 문서에 따르면 ID 공급자가 서명하는 IdP 관리형은 정식 출시 상태이고, 지원하는 ID 공급자는 아직 Okta 하나다. ClaudeDevs는 "조직의 ID 공급자가 EMA를 지원하지 않으면 클로드가 로그인한 구성원 각각을 직접 보증할 수 있다"고 설명했다. 이번에 베타로 나온 것이 그 클로드 발급형이다.

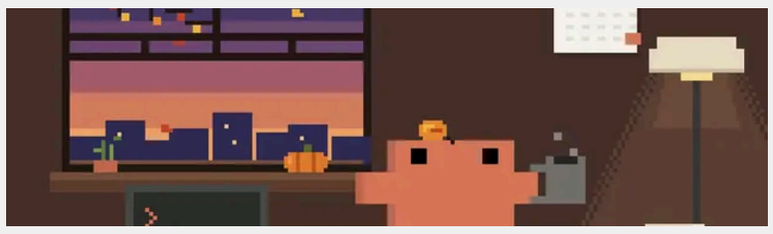
작동 원리는 서명된 신원 확인서 한 장이다. 메탈이 확인한 개발자 문서에 따르면 클로드는 플러그인 회사의 인가 서버에 OAuth 동의 화면 대신 서명된 JSON 웹 토큰(JWT)을 보낸다. 형식은 ID-JAG(Identity Assertion JWT Authorization Grant)이고, 교환은 JWT 베어러 그랜트 표준인 RFC 7523을 따른다. 클로드 발급형 확인서의 발급자 주소는 https://identity.anthropic.com/issuers/ 뒤에 조직마다 다른 식별자가 붙는 형태다. 서명 알고리즘은 ES256이고, 확인서 한 장은 1분 안에 만료된다. 인가 서버가 이를 검증해 액세스 토큰을 내주면 클로드는 그 토큰으로 MCP 서버에 도구 호출을 보낸다.

관리자 쪽 설정은 연결마다 한 번이다. 관리자는 클로드 조직 설정의 커넥터 메뉴에서 관리형 인증을 열고 신원 출처로 클로드 발급형을 고른 뒤 발급자 URL을 복사한다. 그 URL을 상대 서비스 관리자 콘솔에 신뢰 발급자로 등록하고, 서비스가 워크스페이스나 테넌트 ID를 주면 클로드에 붙여 넣는다. 연결 시험을 통과하면 어떤 역할이 플러그인을 자동으로 받을지 고른다. 특정 팀의 사용자 지정 역할에만 먼저 켜는 시범 운영도 가능하다. 관리형 인증 전체로 보면 Asana, Canva, Slack, Monday.com, Lovable을 포함해 21개 커넥터가 대상이다.

◀ 1면에서 계속

AI

앤스로픽, 클로드 코드 10월 업데이트 회고 공개



1 앤스로픽 클로드 코드 개발자 계정은 11월 3일 6개 게시물 스레드로 모드·데스크톱 병렬 세션·사용량 가시성을 10월의 세 축으로 정리했다.

2 20만 토큰 넘는 대화를 약 54분 방치하면 1시간 프롬프트 캐시 만료 전에 압축하는 유휴 압축, 한 턴만 xhigh로 올리는 ultrathink가 들어왔다.

3 변경 기록에 따르면 Opus 4.8 빠른 모드는 11월 30일 제거되고, 단독 클로드 디자인은 12월 14일까지 문을 닫는다.

긴 대화를 다시 여는 순간에도 경고가 붙었다. 10월 19일 변경으로 프로와 맥스 요금제에서는 쉬고 있던 대화를 다시 이어 가면 5시간 사용 한도의 상당 부분을 쓰게 될 때 먼저 알린다. 사용자는 그대로 이어 가거나 새 대화를 열 수 있고, 새 대화에서 앞선 대화를 언급하면 클로드가 그 기록을 찾아본다. 10월 27일에는 사용 한도 안내 문구가 「세션 한도」에서 「5시간 한도」로 바뀌었다.

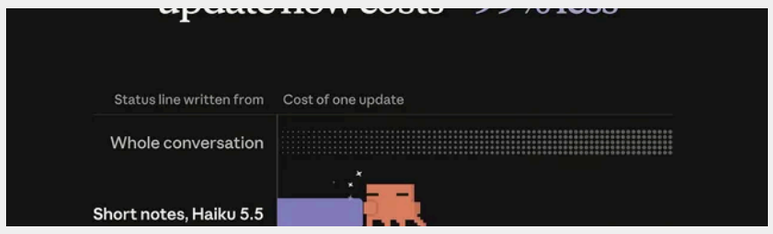
추론 강도를 다루는 손잡이도 늘었다. 메시지에 ultrathink라고 쓰면 그 턴 하나만 추론 강도가 xhigh로 오른다. 계정은 "세션 전체 설정을 바꾸지 않고도 한 문제를 더 깊이 생각하게 할 수 있다"고 설명했다. 변경 기록에 따르면 이 동작은 10월 26일 터미널에서 Opus 5.5, Sonnet 5.5, Haiku 5.5, Fable 5.1에 먼저 적용됐고, 10월 31일 VS Code로 넓어졌다. 같은 날 키워드를 끄는 CLAUDECODEDISABLEULTRATHINK와, 따로 정하지 않으면 서브에이전트가 따를 강도를 고정하는 CLAUDECODESUBAGENTEFFORT가 추가됐다. 세션은 high로 두고 서브에이전트만 medium으로 돌려 달라고 말로 요청하는 것도 된다.

어디서 한도가 새는지도 보이기 시작했다. CLI의 /usage는 이제 무엇이 한도를 쓰고 있는지 평이한 말로 알려 주고, 바로 고칠 방법이 있으면 f 키 하나로 적용한다. 변경 기록에는 최대 추론 강도, 워크플로, 늘 켜져 있는 문맥, 큰 도구 결과에 대한 /usage 경고가 추가됐다고 적혀 있다. 대화형 세션의 웹 검색 예산도 200회를 쓰면 끝나던 방식에서 시간당 100회씩 다시 채워지는 방식으로 바뀌었다.

AI

앤스로픽, 클로드 코드 서브에이전트 사용량 절감

백그라운드 서브에이전트의 30초 상태 줄을 Haiku 5.5가 짧은 메모로 쓰게 바꿨다. 갱신 한 번의 비용은 약 99%, 세션 전체로는 한도의 약 5%가 줄었다.



1 앤스로픽이 11월 4일 클로드 코드 백그라운드 서브에이전트의 상태 줄을 Haiku 5.5가 짧은 메모로 쓰도록 바꿔, 갱신 한 번의 비용을 약 99% 줄였다고 밝혔다.

2 그전에는 30초마다 상태 줄을 갱신할 때마다 대화 전체를 다시 읽었다. 세션 전체로는 사용 한도를 약 5% 덜 쓴다.

3 서브에이전트의 요청은 메인 대화와 같은 사용 한도에 합산되고, 메인 대화에서 띄운 서브에이전트는 기본적으로 백그라운드에서 돈다.

앤스로픽이 11월 4일(현지 시각) 클로드 코드의 백그라운드 서브에이전트가 사용 한도를 덜 쓰도록 바꿨다고 밝혔다. 실행 중인 서브에이전트가 30초마다 갱신하는 상태 줄을 이제 Haiku 5.5가 짧은 메모를 바탕으로 쓰고, 그 결과 갱신 한 번의 비용이 약 99% 줄었다. 앤스로픽이 함께 공개한 6.6초 분량의 그래픽 영상에 따르면 클로드 코드 CLI 기준으로 세션 전체의 사용량 절감 폭은 한도의 약 5%다.

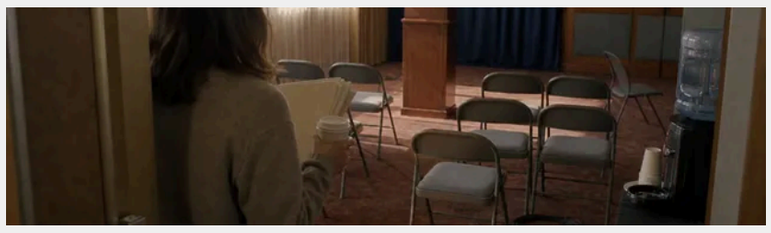
앤스로픽의 클로드 개발자 공식 계정(@ClaudeDevs)은 X에서 "실행 중인 서브에이전트는 30초마다 상태 줄을 갱신하는데, 갱신할 때마다 대화 전체를 다시 읽었다"고 설명했다. 이어 "이제 Haiku 5.5가 짧은 메모로 상태 줄을 쓰기 때문에 갱신 한 번의 비용이 약 99% 줄었다"고 밝혔다. 메탈이 확인한 시점에 이 게시물의 조회수는 7만 2천 회를 넘었다.

서브에이전트는 클로드 코드가 일을 나눠 맡기는 별도의 일꾼이다. 클로드 코드 공식 문서에 따르면 서브에이전트는 자체 컨텍스트 창과 시스템 프롬프트, 도구 접근 권한을 따로 갖고 독립적으로 일한 뒤 결과만 돌려준다. 문서는 서브에이전트가 "자체 요청을 보내며, 이 요청은 메인 대화와 같은 사용 한도에 합산된다"고 적었다. 서브에이전트를 많이 돌릴수록 사용자의 한도도 함께 줄어드는 구조다.

비즈니스

오픈AI, 우버 사장 질 헤이즐베이커 영입

헤이즐베이커는 최고대외협력책임자로 오픈AI의 글로벌 커뮤니케이션과 기업 포지셔닝을 이끈다. 약 10개월 비어 있던 커뮤니케이션 수장 자리를 채우는 인사다.



1 오픈AI는 11월 3일 우버 사장 겸 최고대외협력책임자 질 헤이즐베이커를 최고대외협력책임자로 영입한다고 발표했다.

2 보도에 따르면 그는 그레그 브록먼 사장에게 보고하며 12월 초 합류하고, 1월 해나 웡이 떠난 뒤 약 10개월 비어 있던 자리를 맡는다.

3 헤이즐베이커는 우버·스냅·구글과 2008년 매케인 대선 캠프를 거쳤고, 올해 우버에서 인사 조직 23% 감원을 이끌었다.

오픈AI가 11월 3일(현지 시각) 우버의 질 헤이즐베이커 사장을 최고대외협력책임자(Chief Corporate Affairs Officer)로 영입한다고 발표했다. 헤이즐베이커는 오픈AI의 글로벌 커뮤니케이션과 기업 포지셔닝을 이끈다. 보도에 따르면 그는 그레그 브록먼 공동창업자 겸 사장에게 보고하며 12월 초 합류한다. 오픈AI의 커뮤니케이션 수장 자리는 약 10개월 동안 비어 있었다.

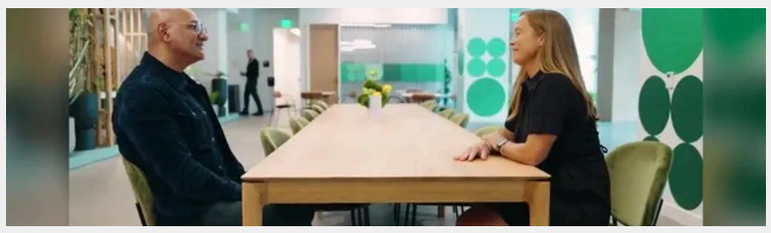
오픈AI는 발표문에서 헤이즐베이커가 회사 리더들과 함께 AI에 대한 대중의 이해를 넓히고 회사 사명에 대한 신뢰를 키우는 일을 맡는다고 밝혔다. 브록먼 사장은 "AI가 더 많은 사람의 삶에 들어오면서 우리가 어떻게 소통하고 관계를 맺고 대중의 신뢰를 얻느냐가 그 어느 때보다 중요해졌다"고 말했다. 그는 헤이즐베이커가 탁월한 판단력과 노련한 리더십, 세계적 규모로 기술을 만드는 일에 따르는 기회와 책임에 대한 깊은 이해를 갖췄다며 다음 장에 꼭 필요한 리더라고 덧붙였다.

헤이즐베이커는 이 자리를 무거운 시점으로 규정했다. 그는 "AI는 우리 생애에서 가장 근본적인 기술 전환"이라며 "판돈은 엄청나고 가능성은 비범하며, 대중의 신뢰를 쌓는 일이 지금보다 중대했던 적은 없다"고 말했다. 이어 샘 올트먼, 그레그 브록먼과 그들이 꾸린 팀과 함께 오픈AI가 이 순간에 응답하도록 돕게 돼 더없이 기쁘다고 밝혔다.

AI

오픈AI, 시스코 데이브레이크 보안 사례 공개

시스코는 네트워크 텔레메트리에 데이브레이크와 GPT-6 Astra를 결합해 위협 신호를 머신 속도로 대조하고 패치 시간을 줄였다고 밝혔다. 지투 파텔 사장은 취약점 공개 뒤 공격까지 8~9분인데 패치에는 45일이 걸린다고 말했다.



1 오픈AI는 11월 3일 시스코가 데이브레이크와 GPT-6 Astra를 네트워크 텔레메트리와 결합해 위협 신호를 대조하고 취약점 대응을 앞당긴 고객 사례를 공개했다.

2 지투 파텔 시스코 사장은 취약점 공개 뒤 공격까지 보통 8~9분, 패치에는 약 45일이 걸리고 패치되는 취약점은 20%뿐이라며 방어에 AI가 전제 조건이라고 말했다.

3 두 회사는 데이브레이크 방어 네트워크로 시스코 클라우드 컨트롤의 노출 분류·통제 적용을 시험 중이며, 독립 시험이나 정량 성과는 공개되지 않았다.

오픈AI가 11월 3일(현지 시각) 시스코가 오픈AI 모델로 방어를 머신 속도로 끌어올린 고객 사례를 공개했다. 시스코는 자사 네트워크 텔레메트리에 오픈AI의 데이브레이크(Daybreak)와 GPT-6 Astra를 결합해 위협 신호를 기계 규모로 대조하고 취약점을 찾아 패치하는 시간을 줄였다고 밝혔다. 두 회사는 데이브레이크 방어 네트워크를 통해 시스코 클라우드 컨트롤에서 노출 위험을 더 빨리 분류하고 통제를 적용하는 방안도 함께 시험하고 있다.

출발점은 공격과 패치 사이의 시간 차다. 지투 파텔 시스코 사장 겸 최고제품책임자는 사례 영상에서 취약점이 공개된 뒤 실제 공격이 일어나기까지 보통 8~9분이 걸리는 반면 패치에는 약 45일이 걸리고, 노출된 취약점 가운데 패치되는 것은 20%뿐이라고 말했다. 파텔은 "AI는 공격자에게 속도와 규모를 쥐여 준다. 방어에 AI를 쓰는 것은 안전을 지키기 위한 전제 조건"이라고 말했다. 공격자가 AI로 공격을 자동화하고 익스플로잇을 머신 속도로 만드는 환경에서는 방어 쪽도 같은 속도가 필요하다는 논리다.

시스코의 방어는 네트워크에서 시작한다. 파텔은 하나의 보안 경계를 두던 방식에서 마이크로서비스와 컨테이너, 클러스터마다 통제를 흩어 두는 방식으로 옮겨 가고 있다고 설명했다. 그는 "모든 마이크로서비스, 모든 쿠버네티스 컨테이너, 모든 쿠버네티스 클러스터 앞에 작은 보안 스택을 두고 보안을 네트워크의 구조 자체에 녹여 넣을 수 있다"고 말했다. 애플리케이션 단위 통제는 서비스마다 필요한 연결만 허용해 공격자가 퍼져 나갈 경로를 줄인다. 침해가 생겨도 영향을 받은 시스템만 골라 격리하고 나머지 서비스는 계속 돌릴 수 있다.