AI 뉴스와 용어
월·수·금 오전 7시

METAL
First, AI Magazine

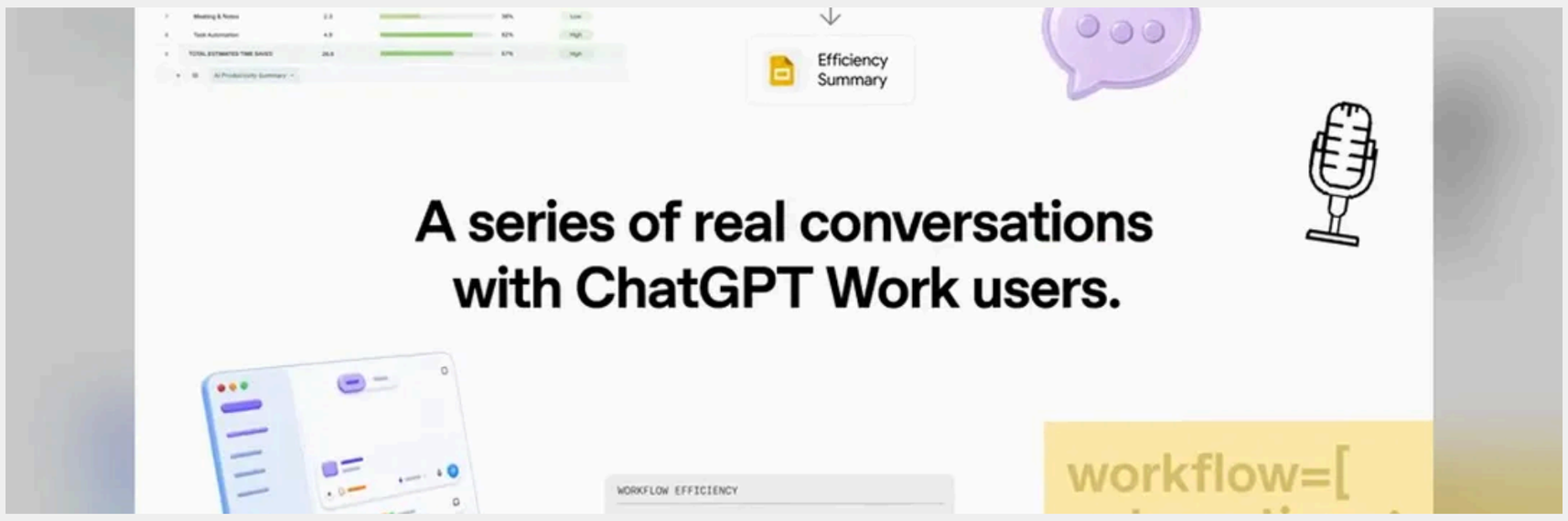
무료 · 언제든 수신거부
metallab.ai

제 41 호 | 2026년 10월 9일 금요일 | 오늘의 AI

AI

오픈AI, 오라클 챗GPT·Codex 도입 사례 공개

오픈AI가 챗GPT 활성 사용자 13만 명, Codex 활성 사용자 9만5000명을 넘긴 오라클의 도입 사례를 공개했다. 채용팀은 2~4일 걸리던 인재 시장 조사를 15~20분으로 줄였고, 업무 사용자는 일상 언어로 사내 데이터를 묻는다.



AI · 오픈AI, 오라클 챗GPT·Codex 도입 사례 공개

오픈AI가 10월 8일 오라클의 챗GPT Work·Codex 도입 사례를 공개했다. 오픈AI에 따르면 오라클에서 챗GPT를 쓰는 활성 사용자는 13만 명, Codex 활성 사용자는 9만5000명을 넘는다. 채용팀은 직접 만든 도구로 인재 시장 조사에 드는 시간을 98% 줄였다. 오픈AI는 채용과 엔지니어링, 운영 전반에서 전문가에게 기대 며칠씩 걸리던 일이 이제 누구나 몇 분 만에 끝내는 일이 됐다고 밝혔다.

가장 큰 숫자는 채용에서 나왔다. 오라클 채용팀은 챗GPT Work로 인재 시장 분석 도구를 만들었다. 직무 기술서를 넣으면 비슷한 직무를 조사하고, 보상 수준을 비교하고, 관련 지역의 인재 풀을 평가한다. 채용 담당자가 채용 관리자와 마주 앉기 전에 갖춰야 할 이 정보는 예전엔 모으는 데 2~4일이 걸렸다. 오라클의 잰 애커먼 수석부사장 겸 글로벌 인재확보 총괄은 "0에서 100으로 뛰었다"며 "이제는 Work로 만든 도구를 써서 15~20분 정도면 준비를 마친다"고 말했다.

속도만 바뀐 것은 아니다. 예전엔 채용 담당자마다 접수 절차를 다르게 돌렸지만, 도구를 쓰면서 절차가 하나로 맞춰졌다. 오픈AI는 어느 담당자와 일하든 모든 채용 관리자가 같은 품질의 데이터와 분석을 받게 됐다고 설명했다. 메탈이 확인한 77초짜리 인터뷰 영상에서 애커먼은 직무 기술서에서 시작해 직접 만든 GPT로 외부 자료를 뒤져 비슷한 직무와 경쟁사 움직임을 찾는다고 설명했다. 화면에는 AI 커뮤니티 리더 직무를 놓고 수요 지수를 매긴 보고서와 도시별 인재 분포를 거품 크기로 그린 히트맵이 지나간다. 영상은 챗GPT Work 사용자들과 나눈 대화를 묶은 연작의 한 편이다.



메타, 싱가포르 장애인 단체에 AI 안경 기부

오라클 경영진이 꼽은 원칙은 사람이 결과에 책임진다는 쪽에 있다. 램은 "시스템 설계와 아키텍처, 보안, Codex가 코드를 어떻게 짜게 할지에 여전히 매우 책임 있게 임해야 한다"고 말했다. 그는 "Codex와 함께 일하지 않으면 유지보수할 수 없는 코드가 잔뜩 쌓이는 상황에 빠진다"며 코드의 주인이 되라고 덧붙였다. 배리 실모버 CIO 기술 자문 부사장은 "예전엔 아이디어를 종이에 적었지만 이제는 시제품으로 만든다"며 기획서 대신 시제품을 내밀라고 권했다. 그는 다음에 풀 문제가 무엇일지는 몰라도 가장 먼저 Codex를 쓸 것이라고 말했다.

메탈은 더 덴의 챗GPT Work 도입 사례를 보도한 바 있으며, 오픈AI는 사내 데이터에 묻는 챗GPT Work 데이터 에이전트도 내놓았다. 오픈AI는 전 세계 100만 곳이 넘는 기업이 자사 도구로 의미 있는 성과를 내고 있다고 밝혔다. 이번 사례는 그 가운데 직원 10만 명 넘게 챗GPT Work와 Codex를 쓰는 고객을 전면에 세웠다.

마케팅의 눈으로 보면 이 사례의 설계는 숫자 세 개와 직함 세 개다. 13만 명과 9만5000명이라는 규모, 98%라는 단축률로 첫 화면을 채우고, 채용·데이터·엔지니어링 책임자의 육성으로 각 숫자를 증언하게 했다. 다른 기업이 따라 할 수 있는 지점은 도구 자체보다 직무 기술서에서 보고서

값은 속도에 비례해 뛴다. 메탈이 확인한 오픈AI의 요금 비교표에 따르면 GPT-6.1 Sol의 표준 API 단가는 입력 100만 토큰당 2달러, 출력 10달러이고, 2배 빠른 Fast 모드는 4달러와 20달러다. 최대 8배 빠른 Ultrafast는 표준의 정확히 6배다. 같은 표에서 GPT-6 Astra 표준 단가는 입력 10달러, 출력 50달러다. Sol Ultrafast가 Astra 표준보다 1.2배 비싸다. 더 똑똑한 상위 모델을 표준으로 쓰는 것보다 아래 모델을 빠르게 쓰는 쪽이 조금 더 비싼 구도다. 보도에 따르면 캐시 입력은 100만 토큰당 0.60달러, 캐시 쓰기는 15달러이고, 입력이 27만2000토큰을 넘는 긴 프롬프트에는 다른 단가가, 지역 처리를 고르면 10%가 더 붙는다.

메탈이 확인한 오픈AI API 가이드는 Ultrafast를 API에서 가장 빠른 서비스 등급으로 규정하고 "속도가 더 높은 비용을 정당화할 때 쓰라"고 적었다. GPT-6 Astra와 GPT-6.1 Sol에는 정식으로, GPT-5.6 Sol에는 미리보기로 열려 있다. 개발자는 요청의 모델을 gpt-6.1-sol로 두고 servicetier 값을 ultrafast로 지정하면 된다. Astra와 Sol의 Ultrafast는 모든 API 사용자가 쓸 수 있고, 표준·Fast 모드와는 별도의 사용 한도가 걸린다.

한도는 Sol 쪽이 훨씬 넉넉하다. 가이드에 따르면 GPT-6.1 Sol Ultrafast의 기본 한도는 분당 토큰 기준 Build 등급 100만, Launch 400만, Grow 4000만이다. 같은 등급에서 Astra Ultrafast는 50만, 100만, 500만에 그친다. 등급이 오를수록 격차가 2배, 4배, 8배로 벌어진다. 데이터 처리 위치도 다르다. Sol Ultrafast는 미국과 유럽연합(EU) 데이터 상주, 글로벌 처리를 모두 지원하지만 Astra Ultrafast는 미국 상주와 글로벌 처리만 된다.

오늘 지면

1 AI 오픈AI, 오라클 챗GPT·Codex 도입 사례 공개
2 AI 오픈AI, 래디슨 챗GPT 플러그인 사례 공개
3 AI 오픈AI, 10대 College Planner 발표
4 AI 오픈AI, ChatGPT 전체에 GPT-6 배포
5 AI 오픈AI, GPT-6.1 Sol Ultrafast 출시
6 AI 앤스로픽, Claude Haiku 5.5 공개



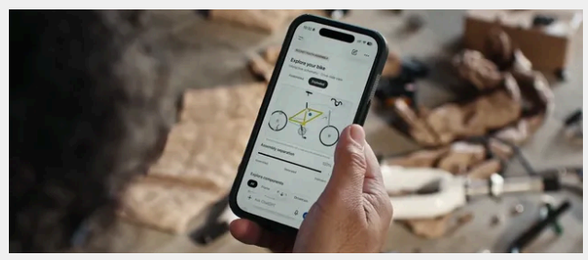
앤스로픽, 제네시스 미션에 1억 5000만 달러 약정



앤스로픽, SDK에 컴퓨터·브라우저 툴셋 탑재

이 지면은 2026년 10월 9일 금요일 아침 구독자 메일함으로 나간 레터를 신문 판형으로 다시 조판한 것입니다. 기사 전문은 metallab.ai 에서 읽을 수 있습니다.

AI



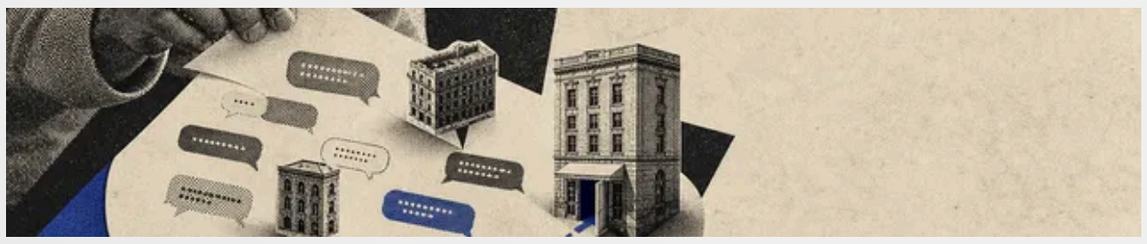
오픈AI, ChatGPT 전체에 GPT-6 배포

오픈AI가 10월 7일 GPT-6를 ChatGPT 전체 사용자에게 배포하기 시작했다. 지난달 유료 고객에게 먼저 내놓은 GPT-6 세대를 매주 ChatGPT를 쓰는 12억 명 이상에게 넓히는 조치다. 이번 배포에는 Intelligent UI라는 새 기능이 함께 들어갔다. 답을 글로만 쓰지 않고 그래픽, 차트, 누를 수 있는 버튼, 입력 폼 같은 화면 요소를 섞어 대화창 안에 바로 띄우는 기능이다. 배포 순서는 요금제로 갈린다. 발표 당일 Plu

▶ 2면에 계속

AI

오픈AI, 래디슨 챗GPT 플러그인 사례 공개

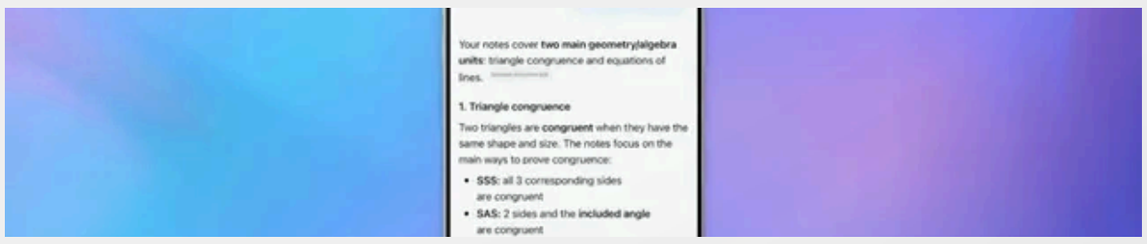


오픈AI가 10월 7일 래디슨 호텔 그룹의 챗GPT 플러그인 사례를 공개했다. 래디슨이 액센추어와 함께 6주 만에 만든 이 플러그인은 챗GPT 대화 안에서 호텔을 찾고 가격을 비교한 뒤 래디슨 자사 사이트로 넘어가 예약하게 한다. 오픈AI에 따르면 2026년 7~8월 플러그인의 방문 대비 예약 전환율은 래디슨 자연 검색의 약 1.5배였다. 래디슨은 플러그인과 함께 챗GPT 스폰서 광고도 돌렸고, 기록된 결제·예약 이벤트의 54%가 광고 노출에서 나온 것으로 집계됐다.

래디슨은 유럽·중동·아프리카와 아시아태평양에서 10개 호텔 브랜드를 운영하는 그룹이다. 운영 중이거나 개발 중인 호텔은 1640곳이 넘는다. 출시 당시 보도에 따르면 앱은 100여 개국 1000곳 이상의 호텔을 검색 대상으로 삼았다. 이용자가 목적지나 편의시설, 여행 취향을 말하면 해당 호텔과 실시간 요금, 주변 정보, 대화형 지도를 보여 주고, 실제 예약은 래디슨 공식 사이트에서 끝내도록 넘긴다.

AI

오픈AI, 10대 College Planner 발표



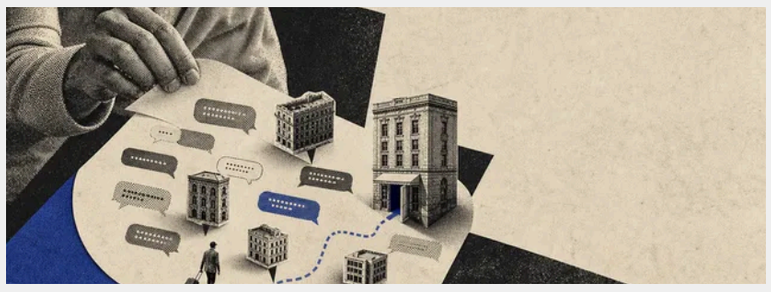
오픈AI가 10월 7일 18세 미만 사용자용 ChatGPT for Teens의 첫 이용 현황을 내놓고, 대학 지원 준비를 돕는 College Planner를 곧 들여오겠다고 발표했다. 회사가 밝힌 10대의 하루 평균 ChatGPT 사용 시간은 15분 미만이다. 같은 날 청소년 보호 비영리단체 커먼센스미디어는 이 10대 모드에 최하 등급인 "용납할 수 없는 위험"을 매겼다. 한 서비스를 두고 회사의 이용 통계와 외부 시험 결과가 같은 날 정반대의 결론을 낸 셈이다.

ChatGPT for Teens는 18세 미만으로 식별된 계정에 자동으로 적용되는 기본 경험이다. 보도에 따르면 오픈AI는 8월부터 이 모드를 배포했고, 나이 정보나 회사의 연령 추정으로 18세 미만이라고 판단되면 켜지며 보호자 통제와 민감한 주제 제한이 함께 작동한다. 메탈은 챗GPT의 13~17세 전용 틴 모드 도입과 오픈AI의 청소년 AI 영향 연구 500만달러 지원을 보도한 바 있다.

◀ 1면에서 계속

AI

오픈AI, 래디슨 챗GPT 플러그인 사례 공개



1 오픈AI가 10월 7일 래디슨 호텔 그룹과 액센추어가 6주 만에 만든 챗GPT 호텔 검색 플러그인 사례를 공개했다.

2 2026년 7~8월 플러그인의 방문 대비 예약 전환율은 래디슨 자연 검색의 약 1.5배였고, 예약은 래디슨 자사 사이트에서 끝난다.

3 7월 시작한 챗GPT 스폰서 광고는 기록된 결제·예약 이벤트의 54%를 광고 노출로 잡았고, 같은 MCP 서버와 API가 광고 경험도 받친다.

플러그인을 만든 쪽은 액센추어의 마케팅·크리에이티브 조직인 액센추어 송이다. 액센추어는 지난 7년 동안 래디슨의 직접 예약 채널을 키우는 일을 맡아 왔고, 오픈AI 파트너 네트워크의 엘리트 파트너로서 챗GPT 안에서의 노출 전략을 오픈AI와 함께 검토했다. 액센추어 송의 보르하 리사라가 로드리게스 데이터·AI 리드는 "거의 하룻밤 사이에 잠재 고객이 영감을 얻고 대안을 비교하고, 호텔 웹사이트에 닿기도 전에 결정을 거의 내리는 새로운 공간이 생겼다"고 설명했다.

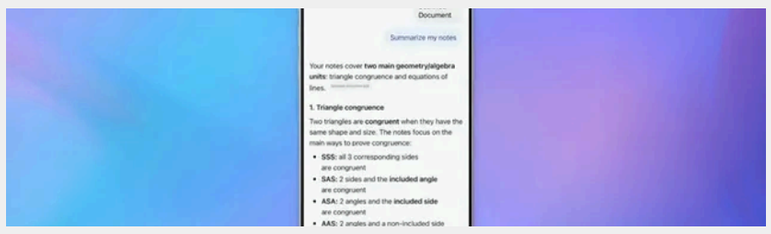
개발에는 에이전트 기반 개발 방식이 쓰였다. 팀은 래디슨 호텔 플러그인을 6주 만에 만들어 내놨고, 이용자는 챗GPT 화면에서 지도 위 추천 호텔과 가격, 편의시설을 바로 확인한다. 보도에 따르면 액센추어는 AI 머천트 센터 안의 MCP 가속기를 써서 래디슨의 콘텐츠와 객실 재고, 가격 신호를 AI 에이전트가 안정적으로 읽을 수 있게 정리했다. 플러그인은 래디슨의 기존 시스템에 직접 연결돼 실시간 정보를 끌어온다.

호출 방식도 바뀌었다. 초기에는 질문 앞에 @RadissonHotels를 붙여 불러야 했지만, 최근 챗GPT 업데이트로 플러그인을 설치한 이용자에게는 관련 대화에서 자동으로 뜬다. 액센추어의 브리트니 캔터 에이전틱 커머스 리드는 "마드리드 여행을 계획하고 있고 고려할 만한 새 래디슨 호텔이 있다면, 이제 챗GPT에서는 그냥 자동으로 작동한다"고 말했다. 메탈은 오픈AI가 챗GPT 광고를 동남아와 대만으로 넓혔다고 보도한 바 있다.

◀ 1면에서 계속

AI

오픈AI, 10대 College Planner 발표



1 오픈AI가 10월 7일 ChatGPT for Teens 이용 현황을 공개했다. 10대의 하루 평균 사용 시간은 15분 미만이고, 세 시간 넘게 연달아 쓰는 10대는 2% 미만이다.

2 미국 10~12학년용 College Planner가 곧 들어온다. 지원 요건·마감·할 일·학자금 절차를 하나의 계획으로 묶고, 칼리지 어드바이징 코어와 보스턴 아동병원 학생자문위원회도 지원한다.

3 커먼센스미디어는 4,000개 넘는 프롬프트 시험 끝에 10대 모드에 "용납할 수 없는 위험" 등급을 매겼고, 오픈AI는 시험이 실제 작동을 반영하지 않는다고 반박했다.

오픈AI가 내놓은 수치는 학습 쪽에 몰려 있다. 10대 모드를 쓸 수 있게 된 10대는 아직 쓰지 못한 10대보다 학습 관련 메시지를 평균 약 270만 건 더 보냈다고 회사는 밝혔다. 가장 많이 고른 대화 시작 문구는 "내 생각을 발전시키는 질문을 해 줘", "초안에 피드백을 줘", "근거를 점검하게 도와줘"였다. 한 주 동안 약 120만 명의 10대가 수학·과학 개념을 그림으로 풀어 보는 Learning Visualizations를 썼고, 18만 명 넘게 단계별 학습 도우미인 Study Mode를 썼다.

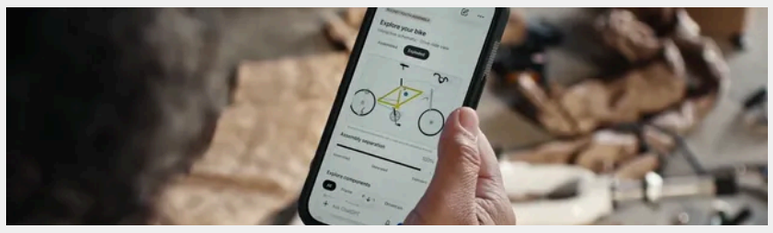
사용 시간 통계는 소셜미디어와의 비교로 짜였다. 세 시간 넘게 연달아 ChatGPT를 쓰는 10대는 2% 미만이고, 휴식 알림이 뜬 10대 대화의 절반 가까이에서 5분 안에 쉬거나 대화를 끝냈다고 오픈AI는 밝혔다. 회사는 공개 보도된 10대의 하루 평균 소셜미디어 사용 시간 5시간을 들어 ChatGPT는 수동적인 스크롤이 아니라 학습과 창작을 위한 도구라고 주장했다. 세 시간이 넘는 세션을 가진 10대 가운데 80% 이상은 학습 관련 프롬프트를 한 번 이상 보냈다는 수치도 함께 내놓았다.

새 기능의 중심은 College Planner다. 오픈AI에 따르면 매주 수십만 명의 미국 10대가 지원서 계획, 마감 관리, 학교 조사, 학자금 이해에 ChatGPT를 쓴다. College Planner는 학생이 고른 학교들의 지원 요건, 마감일, 할 일, 학자금 절차를 하나의 계획으로 묶고, 타임라인으로 남은 일을 보여 준다. 장학금과 전형료 면제를 찾는 일도 돕는다. 첫 버전은 4년제 대학 진학을 준비하는 미국 10~12학년 학생용이며, 다른 나라와 2년제, 기술대학, 직업학교로 넓힐 계획이다. 회사는 어디에 지원할지 정하는 개인적인 일은 대신하지 않고 실무 절차만 돕도록 설계했다고 선을 그었다.

◀ 1면에서 계속

AI

오픈AI, ChatGPT 전체에 GPT-6 배포



1 오픈AI가 10월 7일 GPT-6를 ChatGPT 전체 사용자에게 배포하기 시작했다. 유료 요금제는 당일, Free와 Go는 다음 날부터이며 유료는 GPT-6 Sol, 무료는 GPT-6 Luna가 답한다.

2 새 기능 Intelligent UI는 질문에 따라 그래픽, 차트, 버튼, 입력 폼을 섞어 대화창 안에 화면을 만든다. 네이티브 컴포넌트 라이브러리와 컴파일러로 답이 생성되는 대로 화면이 나타난다.

3 GPT-6는 생각하면서 답을 쓰기 시작하며, 웹 검색 질문에서 GPT-6 Instant는 GPT-5.6 Instant보다 평균 44% 빨리 답을 시작했다고 회사는 밝혔다.

Intelligent UI의 핵심은 답의 형식을 모델이 고른다는 점이다. 오픈AI에 따르면 비교 질문에는 항목을 나란히 놓고, 설명이 필요한 질문에는 눌러 볼 수 있는 도식을 그린다. 글 한 단락이 가장 쓸모 있는 답이면 그냥 글로 답한다. 발표문은 일요일 양고기 로스트 계획을 예로 들었다. GPT-5.6 Instant는 인원별 고기 무게를 표로 적었지만, GPT-6 Instant는 손님 수를 바꾸면 장볼 양이 바로 다시 계산되는 화면과 여덟 단계짜리 요리 체크리스트를 만들었다. 로드트립 경유지를 지도에 찍어 주거나 7단 자전거의 구조를 부품별로 눌러 보게 하는 화면도 시연에 나왔다.

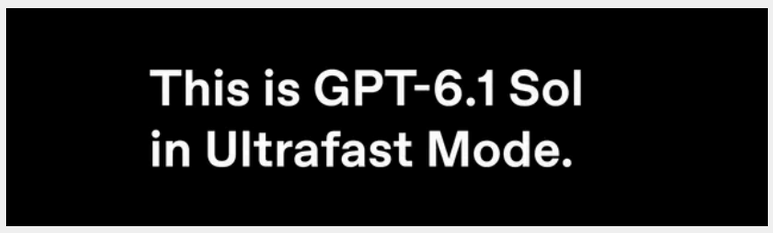
필요한 도구를 그 자리에서 만들어 달라고 할 수도 있다. 오픈AI는 저축액이 어떻게 불어나는지 보여 주는 계산기, 친구들과 저녁값을 나누는 정산기, 대화창 안에서 하는 게임을 예로 들었다. 학습 쪽에서는 입력값을 바꿔 결과가 어떻게 달라지는지 보거나 문제를 한 단계씩 풀어 가는 식의 설명을 내세웠다. 회사는 X 게시물에서 Intelligent UI가 "일상적인 질문을 더 시각적으로 만들고, 복잡한 주제를 이해하기 쉽게 하며, 지금 하는 일에 맞는 도구를 그 자리에서 내놓는다"고 소개했다.

작동 방식은 정해진 부품과 자유로운 조립의 조합이다. 오픈AI는 스트리밍이 되는 네이티브 화면 컴포넌트 라이브러리와, 모델이 생성하는 화면을 그때그때 처리하는 컴파일러를 따로 만들었다. 라이브러리는 모든 답에 익숙한 디자인 바탕을 깔고, 부품을 어떻게 엮을지는 모델이 정한다. 컴파일러 덕분에 화면은 답이 끝나기를 기다리지 않고 생성되는 대로 조금씩 나타난다. 웹과 모바일에서 같은 경험을 내는 것도 이 구조 덕이다. 학습 단계에서는 모델이 만든 화면을 명확성, 유용성, 완결성 기준으로 평가해 언제 상호작용을 넣고 언제 글로 충분한지를 가르쳤다고 회사는 설명했다.

AI

오픈AI, GPT-6.1 Sol Ultrafast 출시

오픈AI가 GPT-6.1 Sol용 Ultrafast를 출시해 API와 Codex, 챗GPT Work에 순차 적용했다. 표준보다 최대 8배 빠른 대신 API 단가는 6배이고, Astra보다 넓은 사용 한도와 EU 데이터 상주를 갖췄다.



1 오픈AI가 10월 8일 GPT-6.1 Sol용 Ultrafast를 출시했고, Sol 표준 모드보다 최대 8배 빠르며 API와 Codex, 챗GPT Work에 순차 적용한다.

2 API 요금은 입력 100만 토큰당 12달러, 출력 60달러로 표준 단가의 6배이고, Codex와 챗GPT Work에서는 Pro 500과 일부 Enterprise·Edu 요금제에서만 쓴다.

3 Sol Ultrafast의 분당 토큰 한도는 Grow 등급 4000만으로 Astra Ultrafast의 8배이며, 미국과 EU 데이터 상주를 모두 지원한다.

오픈AI가 10월 8일(현지 시간) GPT-6.1 Sol용 Ultrafast를 출시했다. 오픈AI 개발자 계정은 이날 공식 X에서 Ultrafast가 API와 Codex, 챗GPT Work에 순차 적용되며 Sol 표준 모드보다 최대 8배 빠르다고 밝혔다. API 요금은 입력 100만 토큰당 12달러, 출력 100만 토큰당 60달러다. 9월 29일 데브데이에서 최상위 모델 GPT-6 Astra에 먼저 붙인 속도 등급이 아흐레 만에 한 단계 아래 모델로 내려왔다.

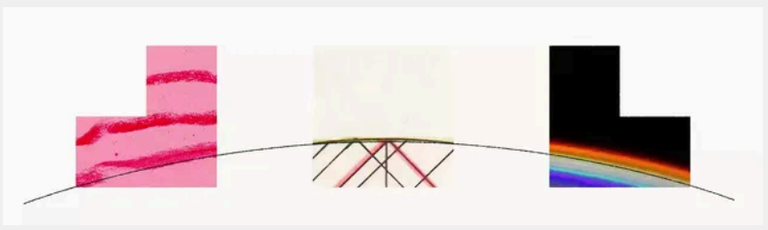
메탈은 오픈AI가 데브데이에서 속도 등급 Ultrafast를 출시했다고 보도한 바 있으며, 당시 회사는 GPT-6.1 Sol용 Ultrafast가 곧 나온다고 예고했다. 같은 날 공개된 GPT-6.1 Sol은 Astra에 가까운 성능을 더 낮은 단가로 쓰게 한다는 모델이다. 메탈은 GPT-6.1 Sol 공개 소식도 전했다. 이번 출시로 오픈AI의 Ultrafast 라인업은 Astra와 Sol 두 모델이 됐다.

오픈AI 개발자 계정은 출시 게시물에서 "Astra에 가까운 지능을 Sol 표준보다 최대 8배 빠른 속도로 낸다"며 "아이디어가 떠오르는 속도만큼 빠르게 만들 수 있다"고 적었다. 이어진 요금 게시물에서는 Ultrafast가 "속도와 지능이 차이를 만드는 일"을 위해 만들어졌다고 설명했다. 회사가 든 예는 장애 원인 추적, 앱을 돌아다니는 에이전트, 1초가 아쉬운 실시간 서비스 세 가지다.

AI

앤스로픽, Claude Haiku 5.5 공개

앤스로픽이 소형 모델 Haiku 5.5를 내놓고 10만 토큰 이하 요청 단가를 전작보다 90% 내렸다. OSWorld 2.1 72.4%로 같은 가격대의 GPT-6 Luna를 앞섰고, Haiku 계열 처음으로 노력 단계를 조절할 수 있다.



1 앤스로픽이 10월 7일 Claude Haiku 5.5를 공개했다. 10만 토큰 이하 요청 단가는 Haiku 4.5보다 90% 낮고, 회사가 계산한 평균 실행 비용은 약 75% 줄었다.

2 발표문 성능표에서 OSWorld 2.1 72.4%, Terminal-Bench 4.0 39.2%, GDPval-AA v2.1 1620으로 GPT-6 Luna(48.9%, 16.4%, 1437)를 앞섰다.

3 같은 날 Sonnet 5.5 캐시 읽기 가격을 절반으로 내렸고 Max·Team 구독자에게 월 API 크레딧을 주기 시작했다.

앤스로픽이 소형 모델 Claude Haiku 5.5를 10월 7일 공개했다. 회사는 이 모델을 "지금까지 내놓은 가장 싸고, 가장 빠르고, 가장 유능한 소형 모델"이라고 소개했다. 10만 토큰 이하 요청의 토큰 단가는 전작 Haiku 4.5보다 90% 낮췄고, 앤스로픽이 계산한 평균 실행 비용은 약 75% 줄었다. 같은 날 Sonnet 5.5의 캐시 읽기 가격을 절반으로 내렸고, Max·Team 구독자에게 매달 API 크레딧을 주기 시작했다.

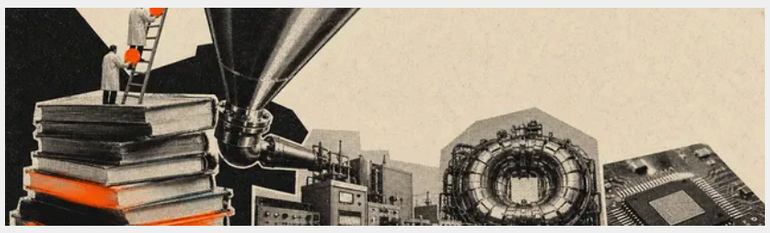
Haiku 5.5는 9월 22일 Opus 5.5, 9월 28일 Sonnet 5.5에 이은 5.5 세대의 세 번째 모델이다. 앤스로픽은 요약과 압축(컴팩션), 데이터베이스 질의, 분류처럼 양이 많고 반복되는 일을 겨냥했다고 밝혔다. 코딩 작업에서는 Opus 5.5나 Sonnet 5.5가 잘게 쪼갠 일을 넘겨받는 하위 에이전트로 쓰라고 권했다. 표준 속도 기준으로 회사에서 가장 빠른 모델이라 실시간 고객 상담과 브라우저 조작에도 맞는다는 설명이다. 다만 Fast Mode로 돌리는 Opus보다는 느리다고 회사는 각주에 적었다.

가격 구조는 요청 길이로 갈린다. 10만 토큰 이하 요청은 입력 100만 토큰당 0.10달러, 출력 0.50달러로 Haiku 4.5의 1달러·5달러에서 90% 내려갔다. 10만 토큰을 넘는 요청은 0.50달러·2.50달러로 50% 인하에 그친다. 앤스로픽에 따르면 Haiku 4.5로 들어온 요청의 약 90%가 10만 토큰 이하였다. 평균 75%라는 수치에는 새 토크나이저가 같은 일에 토큰을 조금 더 쓰는 몫까지 반영했다고 회사는 설명했다. 보도에 따르면 낮은 구간의 입력·출력·캐시 단가는 오픈AI가 지난달 내놓은 저가 모델 GPT-6 Luna와 똑같다.

AI

앤스로픽, 제네시스 미션에 1억5000만 달러 약정

앤스로픽이 미국 연방 AI 과학 프로젝트 제네시스 미션에 3년간 1억5000만 달러를 약정하고 15개 이상 연방 기관에 클로드를 보급한다. 백악관 정상회의에서 11개 기업이 함께 내놓은 24억 달러 가운데 한 몫이다.



1 앤스로픽이 10월 8일 제네시스 미션에 3년간 1억5000만 달러를 약정하고 NASA·NIH·NSF 등 15개 이상 연방 기관에 클로드를 제공한다고 밝혔다.

2 연구 과제 수백 건에 클로드·클로드 코드·API 크레딧을 주고 핵융합·양자컴퓨팅 과제에서 국립연구소와 협업하며 교육·기술 지원을 붙인다.

3 백악관 정상회의에서 엔비디아 10억 달러, AMD 5억 달러, 오픈AI 2억 달러 등 11개 기업이 모두 24억 달러를 약정했다.

AI 기업 앤스로픽이 미국 연방정부의 AI 과학 프로젝트인 제네시스 미션(Genesis Mission)에 3년 동안 1억5000만 달러를 약정했다. 회사는 10월 8일 이 돈으로 미 항공우주국(NASA)과 국립보건원(NIH), 국립과학재단(NSF)을 포함해 제네시스 미션에 참여하는 15개 이상의 연방 기관에서 클로드를 쓸 수 있게 하겠다고 밝혔다. 발표 무대는 백악관 과학기술정책실(OSTP)이 워싱턴 DC에서 연 '과학, 새로운 황금기(Science: A New Golden Age)' 정상회의였다.

약정의 쓰임은 세 갈래다. 앤스로픽은 앞으로 3년 동안 제네시스 미션 연구 과제 수백 건에 클로드와 클로드 코드, API 크레딧을 제공한다. 핵융합 에너지와 양자컴퓨팅처럼 행정부가 최우선으로 꼽은 과학 과제에서는 연방 기관, 국립연구소와 긴밀히 협업한다. 제네시스 미션 과제에 참여하는 과학자에게 교육과 온보딩, 기술 지원을 하고, 미션에 새로 들어온 기관이 첫 과제를 띄우도록 돕는 일도 포함된다.

같은 자리에서 나온 산업계 약정은 앤스로픽 한 곳이 아니었다. 백악관 팩트시트에 따르면 11개 기업이 제네시스 미션 컨소시엄에 과학용 초지능 도구와 컴퓨팅 크레딧으로 모두 24억 달러를 내놓았다. 엔비디아가 10억 달러, AMD가 5억 달러, 오픈AI가 2억 달러를 냈고 앤스로픽과 구글이 각각 1억5000만 달러, AMP와 에메랄드 AI가 각각 1억 달러, AWS와 아마다, 크루소, 마이크론이 각각 5000만 달러를 약정했다. 백악관은 정상회의에서 발표된 연방정부와 산업계, 학계, 자선단체의 투자를 합치면 60억 달러가 넘는다고 집계했다.