



글로벌 투자전략-선진국
Analyst 황수욱
02. 6454-4896
soowook.hwang@meritz.co.kr



OpenAI ChatGPT 에이전트 출시

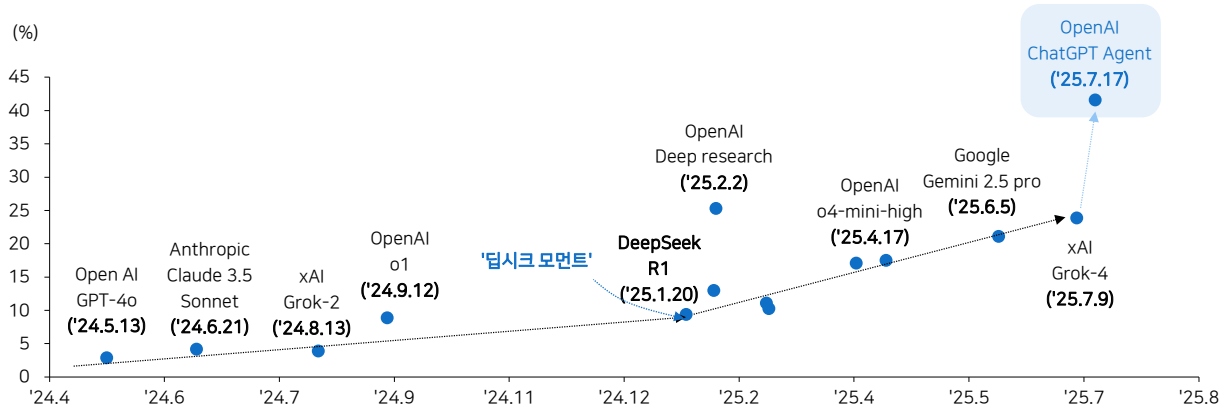
7월 17일 OpenAI는 AI 에이전트 모델 ChatGPT agent 출시. 올해 초 1) 먼저 공개되어 이미 서비스 중인 Deep Research와, 2)리서치 프리뷰 형태로 미국 Pro 요금 사용자에게만 공개되었던 Operator를 결합한 형태의 에이전트 모델. Deep Research의 '심층 분석 능력'과 Operator의 '웹사이트와 상호작용' 및 '사용자 인증 기반 콘텐츠 액세스 능력'을 하나로 합친 것. 이미 있던 GPT에 에이전트 기능을 통합 시작

ChatGPT agent는 현존하는 AI 모델 중 가장 높은 지적 수준과 작업 수행 능력을 지닌 것으로 평가. AI 모델의 지적 역량을 평가하는 벤치마크(시험지) 중 가장 어렵다고 알려진 'Humanity's Last Exam'에서 41.6%의 정답률을 달성. 기존의 LLM들이나 이전까지 최고점이었던 OpenAI Deep Research의 성취도(26.6%) 대비 압도적. 뿐만 아니라 데이터 분석과 모델링 등을 평가하는 DSBench 벤치마크에서 데이터 분석과 데이터 모델링에서 각각 89.9%, 85.5% 성취율을 달성. OpenAI o3(87.9%, 77.1%) 대비 성능 개선, 인간이 직접하는 것(64.1%, 65.0%) 대비해서도 우월한 성능

아쉬운 부분은 '에이전트'에게 기대하는 분야에서 역량이 아직 인간이 직접 수행하는 것에 미치지 못한다는 점. AI 모델의 스프레드시트 활용 능력을 평가하는 벤치마크에서 ChatGPT agent의 성취도(45.5%)는 이전 모델을 압도했으나 인간의 성취도(71.3%)에는 미치지 못함. AI 모델의 웹브라우저 능력을 평가하는 WebArena 벤치마크에서 성취도(65.4%)도 이전 모델 대비 개선되었으나 인간이 직접 하는 것(78.2%) 대비해서는 성취도가 아직 낮음

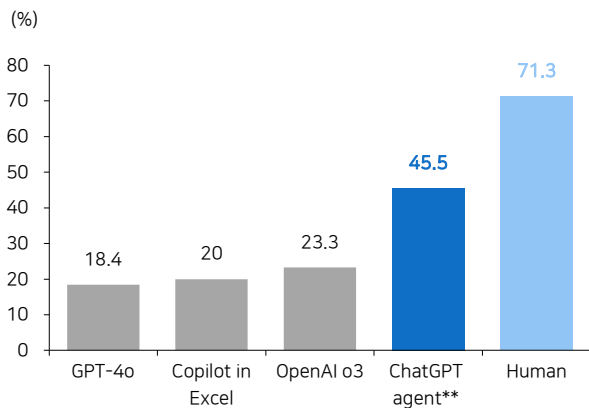
Generative AI에서 '진짜 쓸모 있다'고 사람들이 느끼기 시작한 시점은 AI 모델이 인간의 지적 수준을 넘어서기 시작하면서부터였음. 따라서 B2C AI 에이전트의 수행 능력 기술 단계는 '놀랍지만 아직 쓸모가 있는지 모르겠다'고 느낄 수 있는 구간으로 생각됨. 기술이 쓸모 있는 방향으로 계속 진전하고 있는 것은 분명

그림1 AI 모델 평가하는 벤치마크 중 가장 어렵다고 알려진 Humanity Last Exam에 대한 모델별 성취율



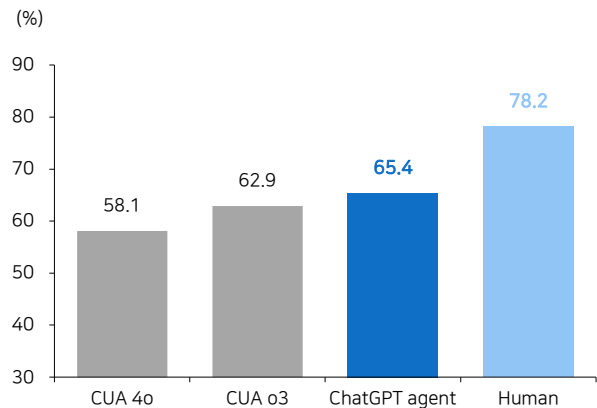
주: ChatGPT agent가 browser+computer+terminal 도구 활용 액세스가 부여되었을 때 점수. 도구 없이 모델만으로는 23.0% 성취도
 자료: Center for AI Safety, 'Humanity's Last Exam(2025)', Artificial Analysis, OpenAI, 메리츠증권 리서치센터

그림2 AI 모델별 SpreadsheetBench* 성취율 비교



주: *스프레드시트를 편집하는 모델의 능력을 평가. **스프레드시트를 직접 편집할 수 있는 권한을 부여했을 때 점수 45.5%, 그렇지 않은 경우 35.3%
 자료: OpenAI, 메리츠증권 리서치센터

그림3 AI 모델별 WebArena* 성취율 비교



주: *웹브라우저 에이전트가 실제 웹 작업을 완료하는 성능을 평가
 자료: OpenAI, 메리츠증권 리서치센터

Compliance Notice

본 조사분석자료는 제3자에게 사전 제공된 사실이 없습니다. 당사는 자료작성일 현재 본 조사분석자료에 언급된 종목의 지분을 1% 이상 보유하고 있지 않습니다.
 본 자료를 작성한 애널리스트는 자료작성일 현재 추천 종목과 재산적 이해관계가 없습니다.
 본 자료에 게재된 내용은 본인의 의견을 정확하게 반영하고 있으며, 외부의 부당한 압력이나 간섭 없이 신의 성실하게 작성되었음을 확인합니다.
 본 자료는 투자자들의 투자판단에 참고가 되는 정보제공을 목적으로 배포되는 자료입니다. 본 자료에 수록된 내용은 당사 리서치센터의 추정치로서 오차가 발생할 수 있으며 정확성이나 완벽성은 보장하지 않습니다. 본 자료를 이용하시는 분은 본 자료와 관련한 투자의 최종 결정은 자신의 판단으로 하시기 바랍니다. 따라서 어떠한 경우에도 본 자료는 투자 결과와 관련한 법적 책임소재의 증빙자료로 사용될 수 없습니다. 본 조사분석자료는 당사 고객에 한하여 배포되는 자료로 당사의 허락 없이 복사, 대여, 배포 될 수 없습니다.