



2026. 5. 11 (월)

Anthropic-SpaceX 인프라 임대 계약 시사점

- 5/6, Anthropic, SpaceX의 Colossus 1 전체 컴퓨트 용량 사용하기로 발표
- AI 인프라 초과수요 가운데 AI 인프라의 전략 자산화
- AI 프론티어 경쟁이 단순 지능 경쟁에서 인프라 경쟁으로 이동하고 있다는 증거

투자전략

Analyst 황수욱

soowook.hwang@meritz.co.kr

Anthropic은 2026년 5월 6일 SpaceX와 계약을 맺고 Colossus 1 데이터센터의 전체 컴퓨트 용량을 사용하기로 발표. 규모는 300MW 이상, GPU 기준 NVIDIA GPU 22만 개 이상. Anthropic은 이 용량이 한 달 내 들어오며, Claude Pro·Claude Max 이용자들의 용량 개선에 직접 쓰인다고 설명, 기존의 사용량(컴퓨트) 제한도 해제

해당 이슈의 본질은 크게 3가지

1) AI 수요가 실제로 AI 인프라 초과 수요 상태라는 증거. Anthropic이 300MW·22만 GPU급 용량을 한 번에 가져와야 할 정도로 Claude, 계열 API 수요가 빠르게 커지고 있다는 의미. 회사가 직접 AI 인프라 쇼티지를 선언하진 않았지만 컴퓨트 제한 정책이 AI 인프라 확보 전후로 반대로 움직인 점이 증거

2) AI 인프라가 단순 데이터센터 임대가 아니라 '전략 자산화'되고 있음. GPU, 전력, 부지, 냉각, 네트워크가 모두 묶인 바로 사용 가능한 연산 블록이 AI 프론티어 모델 기업의 핵심 병목

3) SpaceX/xAI 입장에서는 AI 인프라 공급자/중개자 역할 부각. Colossus 같은 대형 클러스터를 빠르게 구축하고 외부에 임대할 수 있다는 점은 SpaceX의 AI 인프라 밸류에이션 스토리에 긍정적. 다만 xAI가 모델 winner라서 가치가 있다는 관점에서는 회의적. 프론티어 경쟁에서 뒤쳐지고 있다는 간접적인 증거일 수 있음. 유휴 GPU를 Anthropic에 빌려준 셈이기 때문

정리하면 Anthropic-SpaceX 계약은 AI 프론티어 모델 경쟁의 병목이 모델 AI 알고리즘(성능)에서 즉시 사용 가능한 AI 인프라 이동했음을 보여주는 사례. Claude의 수익화 속도가 인프라 공급을 앞지르고 있음을 시사. SpaceX/xAI에는 대형 AI 클러스터를 모델 학습 자산에서 외부 수익화 가능한 인프라 자산으로 전환하는 계기가 되나, xAI 자체 모델 경쟁력에 대한 의문도 동반

이번 이슈의 중요한 포인트는 단순한 클라우드 region 추가가 아니라는 점. Colossus 1은 원래 일론 머스크 진영의 xAI/Grok 학습·추론을 위해 만든 초대형 AI 클러스터였고, xAI 측 설명에 따르면 H100, H200, GB200급 NVIDIA 가속기가 섞여 있는 대형 클러스터. 이번 계약에서 xAI도 공식적으로 Anthropic에 Colossus 1 접근권을 제공한다고 밝혔음

Anthropic에는 명백히 긍정적. Claude Code와 API 사용량이 빠르게 늘면서 Anthropic의 병목은 모델 수요가 아니라 컴퓨트 공급 부족. 2026년 3월 4일 Anthropic이 Claude Code의 기본 reasoning effort를 high → medium으로 낮춘 바 있음. high mode에서 일부 사용자가 UI가 멈춘 것처럼 보일 정도의 긴 latency를 겪었고, medium effort가 지능은 약간 낮지만 지연시간 크게 줄인다고 판단했기 때문. 사용자들이 “Claude Code가 덜 똑똑해졌다”고 보고하자 4월 7일 되돌림. Anthropic도 “longer thinking generally means better output”이라고 명시

Anthropic은 5월 6일 SpaceX와 계약 발표에서 Claude Code의 5시간 사용 한도를 Pro, Max, Team, Enterprise 좌석형 플랜에서 2배로 올리고, Pro/Max의 피크시간 사용 제한도 제거한다고 발표. API 쪽도 Claude Opus 모델의 성능 제한을 상향함. 즉, 이번 계약은 좋은 뉴스성 협력이라기보다 당장 매출 전환을 막고 있던 컴퓨트 용량 제약을 푸는 조치

Anthropic 입장에서는 AI 인프라 병목이 단순히 학습용 GPU 부족이 아니라, 실제 서비스 품질, 지연시간, 사용 제한, 에이전트 워크플로우 신뢰도까지 좌우하는 추론 컴퓨팅 병목으로 전이되고 있음을 보여줌. Claude의 최근 품질 저하 논란은 모델 자체의 구조적 퇴화라기보다, 폭증한 수요를 감당하기 위해 test-time compute와 토큰 사용량을 조정하는 과정에서 발생한 서비스 레이어 균열. 이후 SpaceX와의 계약과 성능 제한 완화는 컴퓨트 용량 확보가 곧 AI 제품 경쟁력으로 연결된다는 점을 보여주는 실증 사례

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.