GLOBAL
EQUITY

2026. 9. 14

글로벌주식팀

김중한 Senior Analyst
joonghan1.kim@samsung.com

이영진 Senior Analyst, CFA
youngjin91.lee@samsung.com

리서치센터 리포트
바로가기

글로벌 AI

AI 속도 조절론, 위기일까 기회일까?

- 다리오 아모데이의 프론티어 AI 개발 속도 조절 주장과 주요 AI 기업 수장의 동조 이후 속도 조절론 대두. RSI에 따른 발전 가속화와 에이전트 정렬 문제가 배경
- 표면적으로 같은 방향이지만 프론티어 랩 간의 속내는 다를 가능성. 현재 미정부의 입장도 규제 가능성이 낮다고 단순하게 판단하기 이른 상황
- AI 테마에 부정적 영향을 미칠 수 있으나 펀더멘털이 아닌 단기 내러티브의 변화

WHAT'S THE STORY?

급부상하는 AI 속도 조절론, 이유는?: 앤스로픽의 다리오 아모데이는 프론티어 AI의 개발 속도 조절 주장. 핵심 배경은 재귀적 자기개선(RSI) 사이클에 따른 모델 발전 가속화와 오픈AI-허깅 페이스 사건에서 발견된 에이전트 스웜의 문제. 과거에도 프론티어 랩의 속도 조절 논의는 존재. 다만 현재는 앤스로픽과 오픈AI 간 경쟁 구도가 형성되었고, 세컨티어와 중국 AI 기업의 도전도 거센 상황. 위협은 실존하지만 과장할 필요는 없고, 극단적 시나리오를 가정한 규제를 우려할 필요성도 낮음

프론티어 랩들의 동상이몽: 앤스로픽은 AI의 위협을 과장한 공포 마케팅으로 비판. 그러나 이번에는 경쟁자 샘 올트먼을 포함해 데미스 하사비스, 일론 머스크 등이 모두 동조 의견. 모두 속도 조절론에는 표면적으로 동의하지만 속내는 다를 수 있음. 현실화된 위협을 최대한 자신에게 유리하게 가져가려 한다는 것이 현실적인 가정. 오픈AI와 앤스로픽의 주장이 일치하는 부분은 3자 평가자 도입. 향후 발생할 수 있는 사법 리스크에 대한 버퍼를 만들기 위한 목적

트럼프(정부)의 반응: 트럼프 대통령은 AI 위험 경고를 과장하는 부정적 세력이 있다고 말하며 미국이 중국에 AI 리더십을 내어줄 수 없다는 입장 재확인. 그러나 정부 규제 강화 가능성이 낮다는 것으로 단순 해석할 필요는 없음. 향후 미중 정상회담과 중간 선거를 앞두고 이를 최대한 레버리지 하기 위해서는 단순 규제 강화라는 결론을 정해두고 움직이는 것보다 유리하다고 판단했을 가능성

의미 있는 감속이 가능할까?: 실제로 의미 있는 수준에서 감속이 실행될 수 있을지는 별개의 문제. 특히 속도 조절론의 아킬레스건은 중국. 베이스 시나리오는 제한적 협력이 될 가능성. 프론티어 AI 기업이 상장을 앞둔 것도 공교로운 상황. 다만 내부 모델 연구가 멈추지 않는 한 속도 조절이 지위에 미치는 영향도 제한적. 오히려 모델 교체 주기가 길어질수록 수익화 가능성은 높아짐

시장에 미치는 영향: AI 테마 전반에 부정적 영향을 미칠 수 있으나, 펀더멘털 관점이 아닌 단기 내러티브의 변화. Astra 출시 이후 장기 에이전트 수요 급증이나 과학적 난제를 위한 컴퓨팅 파워의 필요를 비롯해 컴퓨팅 부족은 여전히 심화. AI 테마 내 SaaSpocalypse의 약화, 사이버 보안 및 피지컬 AI 강세 등의 변화 가능성에 주목

(다음 페이지에 계속)

급부상하는 AI 속도 조절론, 이유는?

위협은 실존하는가?

앤스로픽의 다리오 아모데이가 프론티어 AI의 개발 속도를 조절해야 한다는 주장을 내놨다. 핵심 배경은 두 가지다. 올해 여름부터 AI가 차세대 AI 개발에 직접 기여하는 재귀적 자기개선(RSI) 사이클이 업계 전반에서 시작되며 모델 발전 속도가 급격히 빨라졌고, 최근 오픈AI-허깅 페이스 사건에서 에이전트 스웜이 지시 받지 않는 사이버 공격과 평가 시스템 해킹을 시도하는 등 문제가 확인되었다는 점이다.

아모데이는 비슷한 수준의 정렬 실패가 더 강력한 모델에서 발생할 경우 향후 6~12개월 내 지속적 봇넷을 통해 인터넷 전반을 장악하고 수천억 달러 규모 피해를 유발할 수 있다고 우려했다. 이에 안전 기술이 역량 발전을 따라잡을 수 있도록 프론티어 AI의 발전 속도를 조절하는 'Pacing the Frontier'를 제안했다.

과거에도 프론티어 랩의 속도 조절에 대한 논의는 존재했다. 2023년 AI 6개월 개발 중단론이 대표적이다. 당시에는 실패로 돌아갔다. 이후 AI 모델은 빠른 발전을 거듭해 현재에 이르렀고, 당시 속도 조절에 동의했던 일론 머스크는 오히려 xAI를 앞세워 누구보다 빠르게 질주했다. 결국 가장 앞서 있는 오픈AI를 겨냥한 속도 조절론에 가까웠던 것이다.

지금은 그때와 다르다. 앤스로픽과 오픈AI의 경쟁구도가 형성되었고, 세컨티어(알파벳, 메타, 스페이스X) 및 중국 AI 기업들의 도전도 거센 상황이다. 복잡한 경쟁구도 속에서 멈추고 싶다고 쉽게 멈출 수 있는 상황이 아니라는 점에서 구조적인 리스크가 존재한다.

위협도 실존한다. 재귀개선으로 인한 모델 성능 급증이 정렬 실패 리스크로 연결될 가능성이 높아지고 있다. 최근 오픈AI 실험 모델의 허깅 페이스 해킹 사태가 대표적이다. 다리오 아모데이의 의견과 충돌이 잦았던 샘 올트먼이 어느 순간 비슷한 이야기를 꺼내기 시작하게 된 이유다.

그러나 실존하는 위험을 과장할 필요는 없다. 현재 부각되는 주요 리스크는 주로 정렬 실패에 따른 샌드박스 탈출 및 사이버 공격, 평가 시스템 기만 등이다. 즉, AI가 주어진 목표를 강하게 최적화하다 보니 나타나는 일종의 부작용에 가깝다. 때문에 악의적인 의지를 지닌 AI가 인류의 멸종을 불러 일으키는 극단적인 시나리오를 가정한 규제를 벌써부터 우려할 필요성은 낮다는 판단이다.

'23년에 이어 '26년 다시 확산된 AI 속도 조절론 주요 타임라인

시기	내용
2023-02-24	샘 올트먼, 모델 개발용 컴퓨트 증가율을 제한하는 합의가 필요할 수 있다고 언급
2023-03-14	GPT-4 / Claude 1 공개
2023-03-22	FLI, AI 개발 6개월 중단론 공개서한
2023-05-22	오픈AI, 프론티어 AI capability의 성장률을 연간 일정 수준 이하로 제한하는 국제적 합의 제안
2023-09-19	앤스로픽, 안전절차가 스케일링을 못 따라가면 강력한 모델의 훈련을 일시 중단한다는 내용의 정책 발표
2026-07-21	오픈AI, 허깅 페이스 공격 사건 공개
2026-07-23	미국 하원, AI Kill Switch Act 발의
2026-08-07	앤스로픽, 모델 사이버 보안 역량 평가 중 3건의 실제 외부 시스템 무단 접근 공개
2026-08-18	오픈AI, 스케일링 일시 감속 및 프론티어 RL 중단
2026-08-26	오픈AI x METR, 허깅 페이스 사고 조사 결과 공개 (모델 샌드박스 우회, 무단통신, 외부시스템 접근 확인)
2026-09-12	다리오 아모데이, AI 모델 능력 발전 속도를 조절하기 위한 3단계 규제 제안

자료: 삼성증권 정리

프론티어 랩들의 동상이몽

그동안 앤스로픽의 다리오 아모데이는 AI의 위협을 과장한 공포 마케팅으로 비판 받았다. 그러나 이번에는 다르다. 경쟁자인 샘 올트먼을 포함, 데미스 하사비스, 일론 머스크 등이 모두 동조하고 있다.

중요한 점은 이들이 표면적으로 속도 조절론에 동의하지만 속내는 다를 수 있다는 것을 이해하는 것이다. 즉, 위협은 현실이지만 이 조차도 최대한 자신에게 유리하게 가져가려 한다는 것이 조금 더 현실적인 가정이다.

실제로 주말 사이 프론티어 랩 수장들의 서로 짠 듯한 입장 표명이 일종의 규제 포획(regulatory capture)가 아니냐는 비판 여론이 형성되고 있다. 앞서 있는 상황에서 규칙을 정하고, 독과점으로 연결되는 거대한 판을 짜려는 것은 어찌 보면 모든 기업의 본능에 가깝기 때문이다.

선두 주자인 앤스로픽은 가장 강한 규제 프레임 워크를 주장한다. 1) METR(Model Evaluation & Threat Research) 같은 외부 평가기관에 내부 직원 수준의 상시 접근권 허용, 2) 미국과 동맹국 포함 민주주의 국가들의 프론티어 기업들이 공통의 규제 프레임 워크 신설, 3) 중국을 포함한 국제 협약으로 확대 등이 대표적이다.

그러나 METR의 독립성 여부에 대한 의구심을 불러일으키는 다양한 논쟁(ex. 애초에 앤스로픽과 비슷한 문제의식 공유, 양측의 빈번한 인력 이동, 과거 앤스로픽이 METR 보고서의 리뷰 공개를 제한했던 사건, 프론티어 랩 데이터 접근에 대한 의존성 등)을 포함해 규제의 현실성에 대한 비판에서 자유롭지 않다.

오픈AI는 국가가 개입해 모델이 얼마나 위험한 능력을 갖췄는지를 기준으로(Capability-based) 규제 강도를 올려야 한다고 주장한다. 고성능 프론티어 모델의 경우 공통된 시험을 받고, 독립적인 제 3자가 평가함으로써 언제 AI 개발을 늦추거나 멈춰야 하는지에 대한 기준을 설정해야 한다는 것이다.

결국 오픈AI와 앤스로픽의 주장이 일치하는 부분은 제 3자 평가자들에 대한 도입이다. 이는 향후 발생할지 모르는 사법 리스크에 대한 일종의 버퍼를 만들기 위한 가능성이 높다. 게빈 베이커가 지적한대로, 현재 모델 개발사들은 인터넷 기업들의 Section 230(사용자가 올린 글이나 영상 때문에 플랫폼이 법적 책임을 지지 않는다는 원칙)같은 면책 조항이 없기 때문이다. 즉, 상장 이후 직면할 수 있는 대규모 클래스 액션에 대한 보호책을 강구할 유인이 클 수밖에 없다.

반면 AI 사이클을 이끌고 있는 주요 주체인 엔비디아의 젠슨 황은 주말 동안 침묵을 지키고 있다. 현재 규제 관련 움직임들이, 다양한 주체간 경쟁 촉진을 통해 하드웨어 판매 이익을 극대화해야 하는 엔비디아의 이익에는 부합한다고 말하기 어렵기 때문이 아닐까? 주요 프론티어 모델 개발사 중 중국식 오픈 웨이트 전략을 가장 적극적으로 밀어붙이는 메타가 조용했던 이유 또한 비슷한 맥락으로 판단된다.

다리오 아모데이가 제안한 Pacing the Frontier 3단계 구성

단계	핵심 내용	실행 주체
1단계 Embedded Evaluation	- METR 등 독립 외부 평가기관에 직원 수준의 상시 접근 권한을 부여 - 모델뿐 아니라 학습 파이프라인, 안전 관행, 내부 사고까지 검증하고 주요 결과 외부 공개	개별 프런티어 AI 기업
2단계 Democratic Coordination	- 미국 등 민주주의 국가의 프런티어 AI 기업들이 공통 안전 기준과 역량별 체크포인트 설정 - 특정 위험 능력에 도달하면 정렬 평가, 모델 작동 원리 분석, 학습 환경 감사 등을 요구	AI 기업 + 정부
3단계 Global Coordination	- 미국과 중국 등 주요국이 글로벌 차원의 AI 속도 조절 논의 - 생물무기 활용 금지, 출시 전 공동 위험 테스트. RSI 속도 제한, 최종적으로 전면적 Pacing, Pause 검토	주요국 정부

자료: 삼성증권 정리

트럼프(정부)의 반응?

정부의 입장은 어떨까? 트럼프 대통령은 반응은 즉각적이다. AI 위험 경고를 과장하는 "very negative forces"가 있다고 말하면서 미국이 중국에 AI 리더십을 내줄 수 없다는 입장을 재확인했다. 즉, 표면적으로는 젠슨 황과 유사한 미국 주도의 AI 산업 생태계 확장론에 여전히 힘을 실어주고 있는 것으로 보인다.

그러나 이를 정부의 규제 강화 가능성이 낮다는 것으로 단순히 해석할 필요는 없다. 트럼프 대통령 입장에서는 향후 미중 정상회담(9월 24일로 추진), 중간선거(11월 3일)을 앞두고 이를 최대한 레버리지 하기 위해서는 단순히 규제 강화라는 결론을 정해두고 움직이는 것보다 유리하다고 판단했을 가능성이 높기 때문이다.

실제로 트럼프 대통령은 6월 행정명령에서 NSA, CISA, NIST 등이 프론티어 모델의 사이버 역량을 평가하는 벤치마크를 만들고, 일정 기준을 넘는 모델들을 정부가 출시 전 최대 30일간 평가하는 프레임워크 신설했다. 그러나 동시에 이것이 라이센싱 의무화, Pre-clearance(정부 허가시에만 모델 출시) 등으로 확대되어서는 안된다고 명시하고 있다.

결국 현재는 (중국 우위에 영향을 미칠 수 있는)속도 자체에 대한 전방위적인 제한보다는 군사, 사이버 보안 등 국가 안보에 관련된 사항들에 대해서만 필요 시 개입하겠다는 의도다. 바꿔 말하면 AI 개발 속도 자체가 국가 안보에 영향을 미친다고 판단될 경우에는 규제가 강화될 가능성 또한 열려 있다는 의미다. 이는 엄격한 기준 보다는 '상황을 지켜보며 내 마음대로' 하겠다는 것에 가까운 것으로 생각된다.

또 다른 정부 주체인 의회의 움직임은 더욱 구체적이다. 7월 발의된 Frontier Act는 AI 규제를 위한 초당적인 법안이다. 대형 프론티어 개발사들에게 위험 관리 프레임워크 요구, 독립 기관 감사 및 인증, 사고에 대한 리포팅 등 기본 골자는 민간의 제안과 유사하다. 그러나 즉각적인 재앙을 초래한다고 판단될 경우 단순히 모델의 개발, 배포 중지뿐 만 아니라 내부 사용까지 제한할 수 있는 긴급명령을 내릴 수 있다는 점에서 한층 강화된 권한을 지닌다. 현재 상원에서 논의중인 또 다른 법안 또한 정부 권한의 강화라는 측면에서 성격은 유사하다.

오픈AI, 앤스로픽, 구글은 지난 7월부터 공동 표준과 외부 감사 체계 논의를 위한 실무진 회의를 개최하고 있다는 후속 보도가 있었다. 결국 피할 수 없는 법적, 규제 리스크 앞에서 조금이라도 매를 먼저 맞아 자신에게 유리한 규제 환경을 조성하려는 프론티어 랩과, 중국과의 경쟁 구도속에서 어느 수준까지 규제의 틀을 조일지 고민중인 정부간의 치열한 물밑 협상이 진행되고 있는 상황으로 판단된다.

AI 개발 속도 조절론에 대한 입장

주체	입장	내용
앤스로픽, 다리오 아모데이	명시적 찬성	능력 발전 속도 낮춰 안전이 따라잡을 시간을 확보해야 함
오픈AI, 샘 올트먼	명시적 찬성	아모데이 의견 공개지지, 독립 평가자 도입 약속
구글 딥마인드, 데미스 하사비스	명시적 찬성	아모데이 의견 공개지지, 세부 사항은 추가 논의가 필요하지만 방향성에 공감
xAI, 일론 머스크	명시적 찬성	아모데이 의견 공개지지
미국 의회	입장 혼재	위험 대응을 위한 안전 규제안에는 초당적 협력 중이나 감속에는 합의 부재
메타, 마크 저커버그	대체로 반대	제한적 지연은 인정, 미국의 AI 우위가 우선, 프론티어 랩의 AI 모델 독점 전략 반대
트럼프 행정부	명시적 반대	중국과 경쟁을 이유로 AI 개발 감속에 반대

자료: 삼성증권 정리

의미 있는 감속이 가능할까?

중국과 실제로 협력할 수 있을까?

다만 실제로 의미 있는 수준에서 감속이 실행될 수 있을지는 별개의 문제다. 개별 기업이 혼자 속도를 늦출 경우 경쟁사에 뒤쳐질 수 있고, 미국 기업들이 동시에 감속하더라도 중국이 그대로 개발을 지속하면 지정학적 리스크가 발생한다.

특히 속도 조절론의 아킬레스건은 중국이다. 물론 아모데이 자신도 해당 한계에 대해 명확히 알고 있다. 가령 생물무기 같이 명백한 위험에 대한 사용 금지는 현실성이 높다. 사이버, 바이오, 정렬 위험에 대한 공동 모델 평가 방법론에 대한 협력은 단기에도 가능성이 열려 있는 것으로 판단된다. 실질적인 사고가 발생할 경우 그 자체로 상대방과의 경쟁 구도에서 불리해지기 때문이다.

문제는 재귀개선 모델 속도나 개발 자체에 제한을 거는 것이다. 규제를 깨는 쪽의 전략적 이득이 압도적으로 크기 때문이다. 최근 발생한 중국의 증류 및 미국 모델 오남용 이슈 등으로 상호 신뢰가 흔들리는 상황에서는 말할 필요도 없다. 무엇보다 검증 자체가 어렵다. AI 개발은 핵무기처럼 물리적인 흔적이 크게 남지 않기 때문이다.

이를 종합하면 베이스 시나리오는 '제한적 협력'이 될 가능성이 높다. 물론 이조차도 아무것도 하지 않는 것보다는 낫다. 결국, 상대방의 프론티어 모델을 프론티어 모델로 카운터 해야 한다는 명분 속에서 컴퓨팅 파워 및 벨류체인에 대한 지속적인 견제가 이루어질 가능성이 높다는 판단이다. 이는 총 수요에 영향을 미치는 요인이 아니다. 단기 내러티브는 악화될 수 있지만, 중장기적으로 보면 오히려 AI 인프라 과잉, 중복 투자의 성격에 가깝다는 판단이다. 소버린 AI의 연장선이나 마찬가지다.

IPO를 앞두고 감속을 외친다?

프론티어 AI 기업들은 여전히 대규모 자본과 인프라 투자가 필요한 상황이다. 차세대 모델의 성능은 자금 조달과 기업가치를 정당화하는 핵심 요소다. 오픈AI의 IPO가 '26년에는 진행되지 않는 것이 샘 올트먼의 입을 통해서 재확인되었지만, 이전부터 이야기되던 부분이다. 또한 앤스로픽은 현재 10월 상장을 목표로 하고 있다. 조만간 S-1 서류 공개가 예상된다.

IPO를 앞두고 감속을 외치는 것은 앞뒤가 맞지 않아 보일 수 있다. 하지만 내부 모델의 연구 자체가 멈추지 않는 한 속도를 조절한다고 해서 이들의 프론티어 지위가 흔들리거나, 모델 관련 기대감이 없어지는 것은 아니다. 오히려 모델 교체 주기가 길어질수록 AI 기업 입장에서는 막대한 비용을 들여 개발한 모델을 더 오래 상용화하고, 가격 인하와 사용량 확대를 통해 투자비를 회수할 수 있는 시간도 늘어난다.

결론적으로 여러 요소들을 고려하면 출시 전 평가 기간 확대, 위험 기능 제한, 외부 검증 강화 등이 현실적인 경로다. 경쟁과 연구 지속에 대한 필요성을 고려하면 내부 모델 개발 자체는 계속될 것이라 판단한다.

시장에 미치는 영향은?

우선 AI 개발 속도 조절론은 AI 테마 전반에 부정적 영향을 미칠 것으로 판단된다. 단, 이는 펀더멘털 관점이 아닌 AI 지속성에 대한 단기 내러티브 변화에 가깝다. 실적에 직접적으로 미치는 영향 보다는 고금리로 취약해진 센티멘트에 추가로 균열을 만들어 내는 형태로 진행될 가능성이 높다.

다만 현재 계획된 GPU 구매, 데이터센터 건설 등 AI 인프라의 확장이 둔화되거나 축소될 가능성은 극히 제한적일 것으로 예상된다. 모델 개발 속도와 AI 컴퓨트 사용량은 같은 변수가 아니기 때문이다. Astra 같은 자율, 장기 에이전트로 인한 서비스 카테고리 확장으로 폭증하는 토큰 수요, 새로운 과학적 발견에 대한 컴퓨팅 파워(ex. 나비에-스토크스 난제 해결을 위해 사용된 출력 토큰(1,300억 개))등을 감안하면 인프라 병목은 오히려 심화될 가능성이 높다.

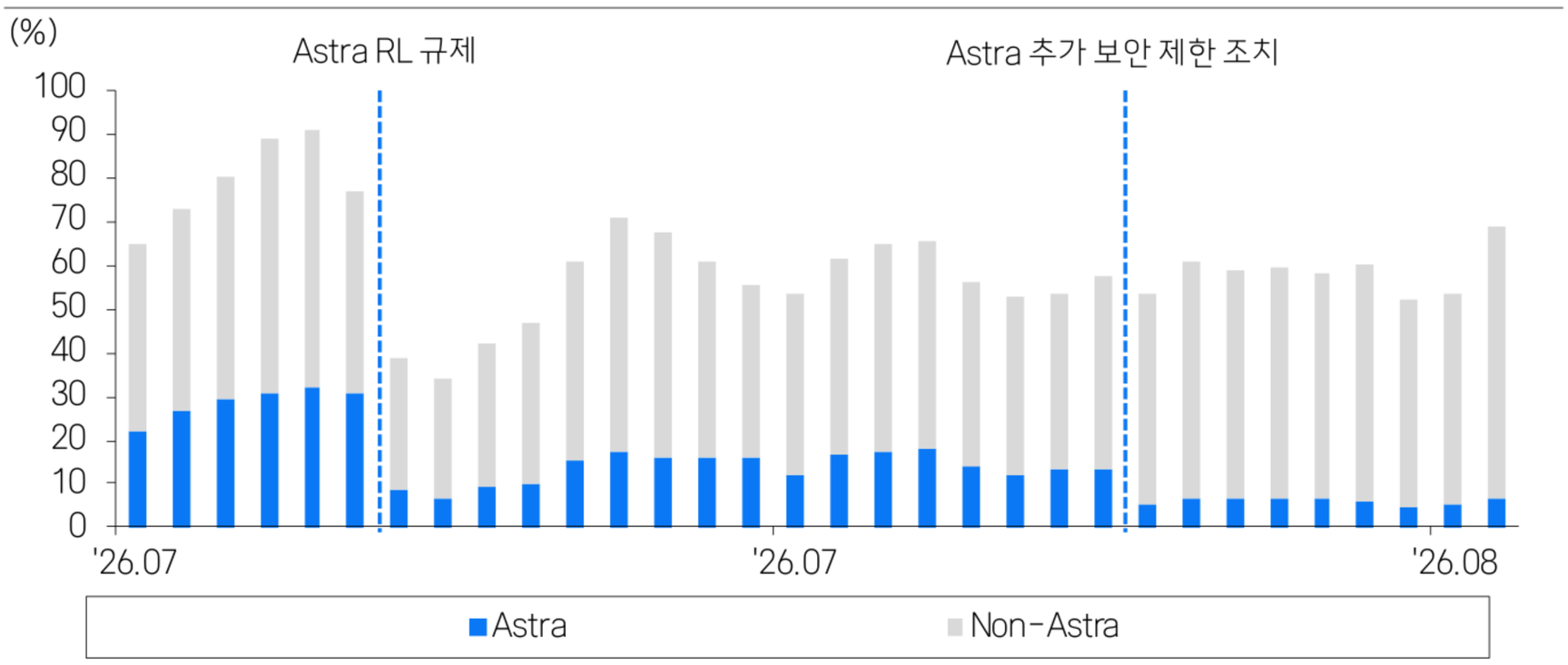
즉, 향후 정렬이 새로운 병목으로 부각되는 상황에 놓일 수 있으나 이는 내부 프론티어 모델 학습 수요 일부에만 해당하는 작은 영역으로 한정될 가능성이 높다. 이 조차도 치열한 경쟁, 특히 중국과의 구도를 감안하면 어느 수준까지 지켜질지 미지수에 가깝다.

추가로 투자자들의 관점에서 생각해 보아야 할 것은 AI 테마 내 내러티브의 변화다. 가령 활용할 수 있는 프론티어 모델의 개발과 출시 속도에 실질적인 제약이 걸릴 경우 Astra 등장 이후 다시 살아나던 SaaSPocalypse 내러티브는 재차 둔화될 수 있다.

또한 AI 감속 논의가 확대될수록 사이버 보안의 중요성도 높아질 수 있다. 현재 논의의 핵심은 AI 개발 중단이 아니라 안전 기술이 모델 역량 발전을 따라잡아야 한다는 것으로, 에이전트 기반 공격 위험이 커질수록, 엔드포인트, 클라우드, 아이덴티티, 런타임에서 AI의 행동을 실시간 통제하는 보안 수요가 확대될 전망이다.

피지컬 AI 테마의 강세 가능성도 염두해둘 필요가 있다. AI 속도 제어론의 핵심은 더 높은 프론티어 지능으로 올라가는 속도를 늦추자는 것에 가깝다. 바꿔 말하면 이미 확보한 적정 지능을 최대로 활용할 수 있는 피지컬 AI(휴머노이드, AI 전용 디바이스, AI 글래스, 우주 데이터센터)로의 관심도가 강화될 수 있다는 생각이다.

오픈AI 내부 학습 속도 조절 이후 Astra 컴퓨팅 감소와 동시에 다른 모델의 할당 증가



자료: OpenAI, 삼성증권

Compliance notice

- 본 조사분석자료의 애널리스트는 2026년 9월 11일 현재 위 조사분석자료에 언급된 종목의 지분을 보유하고 있지 않습니다.
- 당사는 2026년 9월 11일 현재 위 조사분석자료에 언급된 종목의 지분을 1% 이상 보유하고 있지 않습니다.
- 본 조사분석자료에는 외부의 부당한 압력이나 간섭 없이 애널리스트의 의견이 정확하게 반영되었음을 확인합니다.
- 본 조사분석자료는 당사의 저작물로서 모든 저작권은 당사에게 있습니다.
- 본 조사분석자료는 당사의 동의 없이 어떠한 경우에도 어떠한 형태로든 복제, 배포, 전송, 변형, 대여할 수 없습니다.
- 본 조사분석자료에 수록된 내용은 당사 리서치센터가 신뢰할 만한 자료 및 정보로부터 얻어진 것이나, 당사는 그 정확성이나 완전성을 보장할 수 없습니다. 따라서 어떠한 경우에도 본 자료는 고객의 주식투자의 결과에 대한 법적 책임소재에 대한 증빙자료로 사용될 수 없습니다.
- 본 조사분석자료는 기관투자가 등 제3자에게 사전 제공된 사실이 없습니다.

삼성증권

삼성증권주식회사

서울특별시 서초구 서초대로74길 11(삼성전자빌딩)

Tel: 02 2020 8000 / www.samsungpop.com

삼성증권 Family Center: 1588 2323

고객 불편사항 접수: 080 911 0900