미국은 지금

Fable 5 수출 통제와 소버린 AI

키움증권 리서치센터 글로벌리서치팀
US Strategy Analyst 김승혁

Issue Brief

Fable 5 수출 통제 배경과 함의

2026 년 6 월 9 일 출시된 앤트로픽의 최상위 모델 Fable 5 와 Mythos 5 가, 출시 사흘 만인 6 월 12 일 미 상무부의 수출통제 지시에 따라 외국 국적자 접근이 차단됐고, 국가별 선별 집행이 불가능해 전면 비활성화됐다. 앤트로픽이 지시를 받은 것은 6 월 12 일 금요일 오후였다. 이번 조치가 과거 AI 통제와 구분되는 점은 규제 대상이 '칩'에서 '모델'로 이동했다는 점이다.

정부의 통제 명분은 Fable 5 가 안전장치를 우회한 '탈옥(Jailbreak)' 위험이 있다는 것이다. Fable 5 에는 사이버 안보와 같은 위험 분야 질문을 막는 안전장치가 있는데, 일련의 프롬프트로 이를 우회해 모델에게 소프트웨어 코드의 취약점을 발견했다는 것이다. 아마존은 이 방식으로 네 개 이상의 보안 허점을 발견했고, 앤디 재시 CEO 가 그 결과를 정부에 직접 전달했다. 이 결과 베센트 재무장관(베센트), 백악관 사이버 안보 책임자, 수지 와일스 비서실장 등은 긴급 회의를 했고, 그 이후 수출 통제 결정이 내려졌다.

앤트로픽은 두 가지 근거로 정부 통제에 반박했다. 첫째, 이번 사례는 탈옥으로 볼 수 없다는 것이다. 앤트로픽은 아마존 테스트는 보안 담당자가 시스템 취약점을 점검하는 통상적 보안 테스트와 본질적으로 다르지 않았다고 밝혔다. 보안 담당자는 공격당하기 전에 코드 취약점을 먼저 찾아 보완하는데, 아마존의 작업도 모델로 코드 취약점을 탐지하는 정상적 보안 작업일 뿐 **안전장치를 우회·악용한 사례는 아니란 것이다.** 또한 찾아낸 허점이 곧바로 실제 공격에 쓸 수 있는 코드로 만들어졌다고 볼 정황은 확인되지 않았다. 둘째, **이와 같은 취약점은 GPT-5.5 등 다른 모델로도 탈옥 없이 똑같이 찾아낼 수 있다고 주장했다.** 즉 Fable 5 만이 제공하는 고유한 위협은 없다는 의미이다.

다른 관점에서 위 반박과는 별개로, **정부의 "안보"란 명분이 약해지는 지점 역시 확인된다.** 앞서 언급했듯 문제가 된 취약점 탐지 능력은 다른 모델에도 존재하는데, 수출통제를 받은 곳은 앤트로픽뿐이다. 또한, 정부는 앤트로픽의 모델 배포는 막으면서, 같은 시기 그 수준의 모델을 훈련·구동할 수 있는 엔비디아 첨단 칩은 UAE·사우디에 수출하도록 승인하고 있다. 나아가 대중국 경쟁을 안보 강화의 이유로 언급했지만, 정작 미국 기업에서 일하는 외국인 직원의 접근까지 끊었다. 최상위 연구조직의 비미국 국적 인력 비중이 적어도 10%대로 추정되는 만큼, 타격받는 쪽은 중국이 아니라 미국 자신의 개발 역량이기에 대중국 경쟁을 강화 못한다. **안보라는 단일 명분으로 정부의 행보가 설명되지 않기에, 시장은 이를 연초 국방부 갈등으로 정부와 마찰을 빚어온 앤트로픽에 대한 제재로 해석한다. 뒤집어 보면, AI 모델의 경쟁력이 이제 성능만이 아니라 정부와의 관계에서도 갈린다는 의미다.**

롤백 조건을 두고 백악관은 "탈옥을 고치면 풀겠다", 앤트로픽은 "고칠 사안이 아니라 오해이며 곧 복구하겠다"로 맞선다. 시각은 갈리지만 양쪽 다 영구 차단이 아니라 곧 풀릴 사안으로 본다. 현재 두 모델은 이론상 미국 시민만 쓸 수 있으나 실시간 국적 선별이 불가능해 사실상 전원이 막혔고, 미국조차 자국 최고 모델을 못 쓰는 상태라 차단 장기화는 정부에도 부담이다. 앤트로픽은 더 급하다. 6 월 1 일 IPO 를 비공개 신청해 올가을 상장을 목표로 하는데(직전 기업가치 약 9,650 억 달러), 최상위 모델이 막힌 채 시간이 흐를수록 해외 매출과 밸류에이션이 직접 타격을 받기 때문이다. 조건부 단기 복구 시나리오에 무게가 실리는 이유다.

이번 사태의 핵심은 소버린 AI 에 대한 자극이다. 한 번 발생한 차단은 다른 모델·다른 기업으로도 번질 수 있는 선례가 되어, 각국이 모델을 내재화하거나 의존도를 분산할 필요를 키운다. 특히 공공·국방처럼 외부 정책 변수에 좌우되면 안 되는 영역, 그리고 공공데이터·개인정보 보관 부문에서 내재화 압력이 선제적으로 높아질 여지가 있다. 동시에 멀티벤더(OpenAI·제미나이 등)와 오픈웨이트 병용을 통한 모델 의존 분산도 함께 추진될 가능성이 있다. **이러한 자국 AI 역량 확보는 곧 인프라 투자로 직결되므로, 데이터센터·연산·전력 인프라 수요가 늘고 그 하단에서 메모리(HBM·DRAM)와 첨단 공정 수요가 함께 올라간다.** 글로벌 각국 정부가 파운데이션 모델과 공공 클라우드를 직접 키우는 구도 속에서, **그간 미국 중심으로 돌아가던 AI 투자 내러티브가 글로벌로 분산될 가능성이 높다.**

Fable 5 vs Mythos 5

구분	Fable 5 (공개)	Mythos 5 (선별적 공개)
기반 모델	Mythos-class, Opus 상위 등급	동일
차이	안전장치 적용(위험 질문에 대한 자체 차단)	사이버·바이오 부문 안전장치 해제
배포	전 세계 공개	Project Glasswing 통한 검증 기관 한정
안전장치 작동	위험 쿼리 감지 시 Opus 4.8 로 복귀 (세션 5% 미만)	해당 분류기 비활성
핵심 능력	모든 벤치마크 SOTA, 장시간 복합 과업 우위	세계 최강 취약점 탐지·agentic hacking
가격	입력 $10/M · 출력 $50/M	동일

자료: 키움증권 리서치

Fable 5 수출 통제에 대한 정부와 앤트로픽 입장 비교

쟁점	미국 정부	Anthropic
위협의 실체	국가안보 위협, 사이버 무기화	알려진 경미한 취약점 수준, 타사도 동일
탈옥의 성격	안전장치 우회 가능	방어용 프롬프팅(DOP), 보편적 탈옥 아님
조치의 정당성	매출보다 안보 우선	협소한 사유로 상용 모델 회수는 과잉
적용 범위	Anthropic 표적	업계 전체 적용 시 모든 배포 중단 논리
해결 경로	탈옥 패치가 복구 조건	오해 해소를 통한 신속 복구 희망

자료: 키움증권 리서치

Compliance Notice

- 당사는 동 자료를 기관투자자 또는 제 3 자에게 사전 제공한 사실이 없습니다.
- 동 자료에 게시된 내용들은 본인의 의견을 정확하게 반영하고 있으며, 외부의 부당한 압력이나 간섭없이 작성되었음을 확인합니다.

고지사항

- 본 조사분석자료는 당사의 리서치센터가 신뢰할 수 있는 자료 및 정보로부터 얻은 것이나, 당사가 그 정확성이나 완전성을 보장할 수 없고, 통지 없이 의견이 변경될 수 있습니다.
- 본 조사분석자료는 유가증권 투자를 위한 정보제공을 목적으로 당사 고객에게 배포되는 참고자료로서, 유가증권의 종류, 종목, 매매의 구분과 방법 등에 관한 의사결정은 전적으로 투자자 자신의 판단과 책임하에 이루어져야 하며, 당사는 본 자료의 내용에 의거하여 행해진 일제의 투자행위 결과에 대하여 어떠한 책임도 지지 않으며 법적 분쟁에서 증거로 사용 될 수 없습니다.
- 본 조사 분석자료를 무단으로 인용, 복제, 전시, 배포, 전송, 편집, 번역, 출판하는 등의 방법으로 저작권을 침해하는 경우에는 관련법에 의하여 민·형사상 책임을 지게 됩니다.