

2 미국 정부, 고성능 AI 접근 관리 체계 구축과 제한 공개 확대

→ 앤트로픽 이어 오픈AI까지, 미국 고성능 AI 접근 관리 확산

- 오픈AI, 행정명령 이후 신뢰 파트너 중심 단계적 공개 방식 적용
 - 오픈AI는 '26년 6월 26일 차세대 언어 모델 GPT-5.6 시리즈를 공개하며, 초기 접근 범위를 정부와 사전 공유된 소수의 '신뢰 파트너' 중심으로 제한한다고 발표
 - 오픈AI는 출시 전 모델 역량과 공개 계획을 백악관에 사전 공유했으며, 초기 이용 대상 역시 정부와 협의를 거친 파트너 중심으로 구성한 뒤 향후 일반 공개를 추진하겠다는 입장 제시
 - 이는 '26년 6월 2일 서명된 행정명령에서 제시된 고성능 AI 모델 출시 전 정부와의 사전 공유와 신뢰 파트너 중심 초기 접근 구조가 실제 신모델 공개 과정에 적용된 사례로 평가
- 자율 검토 행정명령 직후 강제 수출 통제 이어지며 접근 관리 흐름 구체화
 - 트럼프 행정부는 행정명령을 통해 자율 공유·사전 검토 체계를 제시했으며, 이후 6월 12일 미국 상무부가 앤트로픽 모델에 대한 수출 통제 지침을 적용하며 강제적 접근제한 조치도 병행하는 양상
 - 상무부는 앤트로픽 최신 모델 미토스(Mythos)·페이블 5(Fable 5)에 대해 해외 국적자의 접근을 제한, 이는 고성능 AI 모델 접근 관리가 자율 협력 차원을 넘어 행정조치로 이어진 선행 사례로 작용
 - 특히 해당 조치는 자국 기업 소속이더라도 해외 국적을 가진 직원까지 이용 제한 대상에 포함한 것으로 알려져, 이용자 개인의 국적과 접근 권한이 주요 관리 기준으로 부상
- 앤트로픽 수출 통제 이후 오픈AI 제한 공개로 정부 개입 방식 확대
 - 미국 고성능 AI 관리는 6월 초 행정명령을 통한 자율 사전 검토 체계 제시, 6월 중순 앤트로픽 대상 수출 통제, 6월 말 오픈AI 신모델 제한 공개로 이어지며 단계적으로 구체화되는 양상
 - 두 기업 모두 정부 요청 또는 조치에 따라 미국 중심 소수 대상에만 모델 접근을 허용하는 방식을 택하며, 기존 일반 공개 중심에서 벗어나 정부 개입이 업계 전반으로 확산되는 흐름 확인

➔ 고성능 AI 모델 공개와 접근 관리 구조의 변화

● 오픈AI, GPT-5.6 3개 모델군 공개와 단계적 출시 전략

- 오픈AI는 GPT-5.6을 솔(Sol)·테라(Terra)·루나(Luna) 3개 등급으로 구성해 공개했으며, 용도와 예산에 따라 선택 가능한 차등형 모델 체계를 제시
- 솔은 코딩·생물학·사이버보안·에이전트형 작업에 특화된 최상위 모델, 테라는 일상 업무와 효율성을 고려한 중간급 모델, 루나는 고속·저비용 대량 처리에 초점을 맞춘 경량 모델로 구성
- 오픈AI는 GPT-5.6 솔이 취약점 탐지·수정 등 방어적 사이버보안 작업에 강점을 갖지만, 악용 가능성과 예상치 못한 위험을 고려해 강화된 안전장치와 단계적 공개 방식을 병행한다고 설명

● 행정명령은 자율 협력을 제시했으나 실제 운영은 제한적 접근 중심으로 전개

- 행정명령은 AI 기업의 정부 사전 평가 참여를 의무적 허가제가 아닌 ‘자발적 절차’로 규정, 고성능 AI 모델 공개 전 정부와 정보를 공유하는 체계를 제시한 것이 핵심
- 다만 실제 운영에서는 초기 접근 대상이 신뢰 파트너 중심으로 제한되고, 정부와의 사전 협의를 거쳐 단계적으로 공개하는 방식이 적용되면서 제한적 접근 구조가 우선 정착되는 양상
- 특히 신뢰 파트너 선정과 초기 공개 범위 설정 과정에서 정부와의 협력 절차가 병행되고 있어, 고성능 AI 모델의 공개 방식이 기존 일반 공개 중심에서 단계적 공개 체계로 변화하는 모습

● 모델 성능 외 접근 주체까지 관리 대상으로 확대

- GPT-5.6의 3개 모델군은 성능과 활용 목적에 따라 구분되지만, 초기 접근관리는 특정 모델 등급보다 이용 대상과 활용 목적을 중심으로 적용되는 양상
- 이는 모델 자체의 위험도뿐 아니라 누가 어떤 목적으로 활용하는지를 고려하는 기준으로 허용 여부를 판단하는 접근 주체별 관리 방식으로 이동하고 있음을 시사
- 오픈AI 역시 정부와의 협력을 위한 공식 절차가 마련되는 과정에 있다고 설명한 만큼, 향후 정부가 마련할 세부 기준에 따라 접근 관리 체계도 보다 구체화될 것으로 전망
- 오픈AI가 GPT-5.6을 공개한 날, 엔트로픽도 정부로부터 미토스의 일부 신뢰 파트너를 대상으로 이용 재개를 승인받았다고 발표하며 미국 정부의 두 고성능 AI 모델 접근 관리 사례가 동시에 부각



- 미토스 수출 통제 이후 재개 승인 역시 전면 해제가 아닌 신뢰 파트너로 한정된 조건부 방식, 정부가 최초 진입뿐 아니라 재진입 단계에서도 동일한 개별 승인 방식을 그대로 적용함을 확인
- 오픈AI, 정부 협력 수용하면서도 접근 통제의 상시화는 견제
 - 오픈AI는 정부의 안전관리 체계 구축에 협력한다는 입장을 유지하는 한편, 이번 제한 공개가 장기적인 운영 방식으로 고착되어서는 안 된다는 점을 지속적으로 강조
 - 또한 이번 조치가 개발자, 사이버보안 전문가 등 폭넓은 이용자층의 AI 도구 접근을 제한할 수 있다는 점을 언급하며, 정부와의 협력에는 공감하면서도 과도한 접근제한에는 신중한 입장을 제시
 - 오픈AI는 이번 조치를 절차 수립을 위한 임시 단계로 규정하고 있어, 정부 지시를 수동적으로 따르기보다 향후 업계 표준이 될 심사 절차 설계에 직접 참여하려는 능동적 전략으로 해석 가능

→ 정부 개입의 판단 근거와 절차적 정당성 논쟁

- 사이버보안 위협 판단의 신뢰성 자체에 대한 전문가 반박
 - 스탠퍼드대 알렉스 스타모스(Alex Stamos) 교수는 사이버보안 업계에서는 엔트로픽 수출통제 조치의 근거가 충분하지 않다는 견해가 있다며, 정부 위협 판단 기준에 대한 검토 필요성 제기
 - 스타모스 교수는 자신이 참여한 검토에서 미토스·페이블의 위험 수준이 중국 등에서 개발된 다른 공개 모델과 비교해 유의미하게 높지 않았다고 평가, 위협 판단 근거 자체를 반박
 - 또한 미국이 중국과의 AI 경쟁에서 우위를 확보하려 한다면 이번 조치는 부적절한 선택이라고 평가하며, 예측 불가능한 정부 개입이 오히려 미국 AI 산업 경쟁력을 약화시킬 수 있다고 우려
- 법적 근거 없는 개별 승인 방식, 의회 차원의 절차적 정당성 문제 제기
 - 로리 트라한(Lori Trahan) 미국 민주당 하원의원은 법률, 절차, 감독 체계 없이 임명직 공무원들이 기업별 접근 여부를 결정하고 있다며 절차적 정당성 문제 제기
 - 명문화된 법률이나 공개된 심사 기준 없이 개별 사안별로 접근이 승인·거부되는 현재 방식은 향후 정권이나 담당자 변화에 따라 기준이 흔들릴 수 있다는 우려로 연결

- 예측 불가능한 개입이 오히려 미국 산업 경쟁력을 훼손할 수 있다는 우려
 - 수출통제의 판단 근거와 개별 승인 방식의 절차적 정당성을 둘러싼 논쟁이 지속될 경우, 산업계와 의회 차원의 정책 신뢰가 약화될 가능성
 - 특히 오픈AI·엔트로픽 모두 기업공개(IPO)가 거론되는 시점에서 정부의 개별 승인 방식이 신제품 출시 일정의 불확실성을 높이는 요인으로 지목되며 투자 심리에도 영향 가능성

출처: OpenAI 외(2026.6.)

<https://openai.com/index/previewing-gpt-5-6-sol/>

<https://www.wired.com/story/openai-gpt-5-6-model-release-trump-admin-approval/>

<https://www.goodreturns.in/news/openai-restricts-gpt-5-6-sol-release-after-us-government-review-011-1518437.html>

<https://apnews.com/article/trump-ai-openai-gpt5-6-sol-cybersecurity-mythos-065d5398baac7f16c8265c2cb8ba2baa>

<https://www.businessinsider.com/openai-gpt-5-6-limited-preview-us-government-ai-security-2026-6>

<https://www.techinasia.com/news/trump-admin-asks-openai-to-stagger-gpt-5-6-launch>