2 미국 CFDI, 'AI 모델 탈취 방지 법안(DAAMTA)' 개선 의견 제안

미국 데이터혁신센터(CFDI)는 폐쇄형(closed-source) AI 모델 도용을 막기 위해 발의된 'AI 모델 탈취 방지 법안(DAAMTA*)'에 대한 개선 의견을 제안**('26.5)

* Deterring American AI Model Theft Act of 2026

** How to Fix the AI Model Theft Bill Before It Becomes Law

○ **(DAAMTA 개요)** DAAMTA는 적대국들이 미국의 최첨단 폐쇄형 AI 모델을 불법 추출·탈취하는 행위를 막기 위한 법안으로, '26년 4월 하원에서 발의

- **(배경)** 최근 오픈AI, 구글, 앤트로픽 등 미국의 주요 AI 기업들은 중국 정부의 지원을 받는 기업들로부터 일명 '적대적 증류(Adversarial AI Distillation)*' 공격에 노출되었다고 보고

  * 첨단 AI 모델에서 모델 가중치, 아키텍처, 기타 기술적 특성과 같은 핵심 데이터를 체계적으로 추출·탈취하여 저렴하게 자체 모델을 학습시키는 행위

- **(목적)** 중국, 러시아와 같은 적대국들이 미국의 첨단 AI 모델을 무단 추출하거나 학습하는 행위를 국가 안보 및 미국 기업의 경제적 경쟁력에 대한 위협으로 간주하고 이에 대응하기 위해 발의

- **(주요 내용)** 미국 AI 모델에 대한 추출 공격 평가를 의무화하고, 공격을 수행했거나 그러한 행위와 관련된 실체를 식별해 목록으로 관리하는 것이 골자

〈 DAAMTA 주요 내용 〉

구분	주요 내용
AI 모델 추출 공격 (model extraction attack)의 정의	• AI 모델을 복제·개발·훈련·개선하기 위해 폐쇄형 AI 모델(소유자가 복제·재현에 필요한 기본 정보를 공개하지 않는 모델)의 기능을 무단으로 추출하는 행위로 정의하고, 그 예시로 다음을 제시 – 모델 소유자가 구현한 ▲기술적·계약적 제약 및 기타 접근제어 ▲신원확인 요건 ▲지리적 접근 제한 등을 우회하거나, – 사기, 허위진술, 무단 접근을 통해 질의를 수행하거나, – 소유자·제공자가 정한 약관·조건·제한을 위반하는 경우 등
AI 모델 추출 공격 및 사기성 계정 네트워크 제공업체에 대한 평가 의무	• 미국 행정부에 동 법이 제정된 날로부터 180일 이내에 미국의 첨단 AI 모델을 대상으로 정보 추출 공격을 수행했거나 현재 공격을 수행 중인 '우려 기관*'과 '사기 계정 네트워크 제공자**'를 식별·파악할 것을 의무화 * Entity of Concern: 우려 대상 국가에서 활동하거나 본사가 위치하거나, 최종 모회사의 본사가 우려 대상 국가에 있거나, 우려 대상 국가에 본사를 두고 있는 법인의 지시·통제 하에 운영되거나, 미국인 소유의 폐쇄형 AI 모델에 대해 추출 공격을 시도·수행하는 실체(기관이나 개인) ** Fraudulent Account Network Provider: 미국의 폐쇄형 AI 모델에 접근할 수 있는 계정을 고의적·의도적으로 생성·획득·유지·판매·중개 또는 기타 방식으로 제공하는 모든 외국 법인 – 특히, 'AI 모델 추출 공격자 목록' 작성·관리 의무를 도입

구분	주요 내용
	• 평가 활동에 다음 사항들이 포함되도록 규정 – 공격 행위자들의 국적 식별 결과 – 해당 국가 정부의 물질적 지원 내용 – 주요 공격 기법과 사기성 계정 네트워크 제공자의 역할·위치 – 공격 탐지를 위한 다양한 접근방식의 장단점 검토 내용 – 공격의 경제·안보 상의 영향 – 미국 정부가 모델 소유자에게 제공하는 지원 – 공격 탐지·예방을 위해 동맹국·파트너를 활용하는 외교 전략 등
공격 예방·대응을 위한 모범사례 개발 및 정보 공유 메커니즘 구축 의무	• AI 모델 추출 공격의 표적이 되거나 피해를 입은 폐쇄형 AI 모델 소유자(자발적 참여 기반), 기타 기업, 학계 전문가, 산업 포럼 및 단체 등과 협의해 ▲공격 행동 패턴 및 방법 식별 ▲모델 추출 공격 방어를 위한 모범사례 개발 ▲사기 계정 네트워크 제공자의 활동을 식별할 수 있는 모범사례 개발 등을 수행할 것을 요구 – 동 법이 제정된 날로부터 210일 이내에 모델 추출 공격을 탐지·예방·대응하기 위한 모범사례 보고서 발표 의무를 명시 • 폐쇄형 AI 모델 소유자가 모델 추출 공격 및 사기 계정 네트워크 제공자에 대한 정보를 연방정부와 자발적이고 신속하게 기밀로 공유할 수 있도록 정보 공유 메커니즘을 구축할 것을 요구
제재	• 우려 기관에 대해 미국 내 재산 및 재산상 이익에 관련된 모든 거래를 차단·금지할 수 있음을 명시

○ **(개선 의견)** 미셸 말도나도(Michelle Maldonado) CFDI AI 정책부장은 DAAMTA 제정을 지지하면서도, 실효성을 위해 개선이 필요하다고 제언

- 컴퓨터 사기 및 남용 방지법(CFAA*), 영업비밀 방어법(DTSA**) 등 기존의 법률에는 한계가 있어, DAAMTA가 좋은 입법적 수단이 될 수 있다고 평가

* Computer Fraud and Abuse Act: 권한 없이 컴퓨터 시스템에 접근하거나 정보를 탈취·파괴하는 행위를 처벌하고 민사상 손해배상을 청구할 수 있도록 규정한 미국 연방법

** Defend Trade Secrets Act: 영업비밀 절도 행위에 대해 금전적 손해배상 등을 위한 민사 소송을 제기할 수 있도록 규정한 미국 연방법

- 다만 ▲'모델 추출'에 대한 광범위한 정의 문제 해결 ▲공격자 식별에 관한 증거 기반 확충 ▲오픈소스 AI 개발 및 보안 연구에 대한 보호 추가 ▲조화로운 법적 기반 마련 등을 통해 DAAMTA의 허점을 보완해야 한다고 권고

〈 DAAMTA 개선 방향에 대한 제언 〉

구분	주요 내용
'모델 추출' 정의와 관련된 문제 해결	• DAAMTA는 모델 추출 공격 행위를 'AI 모델 소유자·제공자의 서비스 약관 위반'을 포함해 지나치게 광범위하게 정의 • 약관은 기업과 사용자 간 사적 계약으로 공법이 아니므로, 정부 제재가 기업의 서비스 약관 위반 여부에 좌우되는 것은 바람직하지 않음 • 따라서 고의적인 계정 사기와 체계적인 모델 기능 추출 시도를 결합한 행위를 겨냥하도록 정의를 수정하는 것이 바람직

구분	주요 내용
AI 모델 추출 공격자 목록 작성과 관련해 공격자 식별 증거 확충	• DAAMTA는 AI 모델 추출에 관여한 기관을 식별해 목록을 공개하도록 규정하고 있는데, 이처럼 공개적인 지정이 AI 기업의 신고에만 의존하게 될 가능성이 농후 • 불투명한 증거를 기반으로 한 지정은 절차상 적법성 문제, 오류로 인한 기업 피해 가능성, 국제 분쟁 등을 초래할 수 있으므로 각 지정에 대한 증거 요건을 강화할 수 있는 방안을 검토
오픈소스 AI 개발과 보안 연구에 대한 보호 장치 추가	• DAAMTA는 '폐쇄형 AI 모델'을 주요 대상으로 주요 위험을 적절히 겨냥하고 있으나, 향후 정치적 압력 등으로 인해 규제 범위가 오픈소스 AI 및 연구 활동으로 확대될 위험이 존재 • 오픈소스 AI 개발, 학술연구, 합법적인 보안 테스트 활동을 보호할 수 있는 명확한 법적 안전장치를 추가하는 것이 바람직
정부의 자체 AI 개발 관행에 부합하는 조화로운 법적 기반 마련	• 정부가 적법하게 구매한 모델을 활용해 자체 모델을 개발하는 것은 사기 계정을 만들고 접근통제를 우회해 모델을 추출하는 행위와 본질적으로 다르지만, 이러한 구분을 법률에 명확히 반영해야 할 필요성이 존재

- 또한 다음과 같은 보다 광범위한 노력을 통해 AI 보안을 강화할 것을 촉구

〈 AI 보안 강화 조치에 대한 제언 〉

구분	주요 내용
DAAMTA 보완	• 개선 의견을 반영해 DAAMTA를 발전시킴으로써 법적 실효성을 강화
강력한 기술적 방어 체계에 대한 투자	• 미국 국립표준기술연구소(NIST)와 산하 기관인 인공지능 표준혁신센터(CAISI)를 통해 모델 추출 행위 탐지, 악용가능한 API 사용 제한, 포렌식 분석 등을 지원하는 표준 개발을 추진
동맹국과 협력해 AI 모델 탈취에 대한 공통 규범 수립	• 동맹국들과 협력해 중국 등 특정 국가가 지원하는 AI 모델 탈취 행위에 대한 공통 규범을 확립 • 영국, 일본, 캐나다, 프랑스 등 AI 산업 선도국들과 공동 대응을 추진해 미국의 정책적 영향력을 확대하고 모델 도용에 대한 제재 효과를 확대
기술 자체에 대한 규제 지양	• 유해 행위를 규제하는 데 초점을 맞추되, AI 기술 자체를 규제하지 않도록 주의함으로써 미국의 안보와 혁신을 보호

출처 : 미국 데이터혁신센터(CFDI) (2026.7.28.)

https://datainnovation.org/2026/07/how-to-fix-the-ai-model-theft-bill-before-it

https://www.justsecurity.org/137498/diagnosis-deterrence-us-response-distillation/