132 이중 용도 위험

Dual-Use Risk

기술이 개발된 합법적 목적과 달리 해로운 방식으로도 사용되는 위험

- 상업적·연구·공공 서비스 등 정상적 목적의 기술이 동일한 능력 때문에 오용될 수 있는 특성
- 기술의 범용성과 접근성을 고려한 예방적 관리가 필요한 영역

이중 용도 위험이란?

이중 용도 위험은 본래 상업적·연구·공공 목적 등 정당한 합법적 목적을 위해 개발된 AI 기술이 의도와 달리 해로운 방식으로도 사용될 수 있는 가능성을 의미합니다. 이는 생명공학·사이버보안 등 다른 분야에서도 오래 논의되어 왔지만, 최근에는 고도화된 AI 기술이 폭넓고 빠르게 확산되면서 그 중요성이 크게 부각되고 있습니다. 고성능 모델은 연구·창작·교육 등 다양한 영역에서 생산성을 높이지만, 동일한 기능이 악용되면 사회적 혼란, 범죄 지원, 보안 위협 등 부정적 결과를 초래할 수 있습니다. 이러한 위험은 기술의 범용성과 접근 용이성에서 비롯되며, 모델 자체의 능력뿐 아니라 배포 방식과 사용 환경에 따라 그 수준이 크게 달라집니다.

AI의 이중 용도 위험

AI의 이중 용도 위험은 능력 강화와 위험 증가가 동시에 나타나는 구조적 위험입니다. 모델이 복잡한 계획이나 분석을 수행할수록 정당한 연구·산업 활용 범위는 넓어지지만, 악의적 사용 시 피해 규모도 기하급수적으로 커질 수 있습니다. 생성형 AI가 허위정보를 대량 생산하는 데 악용되거나, 코드 생성 모델이 취약점 공격 절차와 유사한 정보를 제공하는 경우가 대표적입니다. 또한 사용자의 의도 파악이 어렵고, 모델 출력이 맥락에 따라 쉽게 변하는 특성 때문에 오용을 조기에 탐지하기 어렵다는 문제가 있습니다.

AI의 이중 용도 위험에 대한 대응

기술적 측면에서는 모델이 위험한 요청을 스스로 감지하고 차단하도록 만드는 안전성 튜닝, 민감한 정보를 출력하지 않도록 제어하는 출력 관리, 고위험 기능을 제한된 환경에서만 사용할 수 있도록 하는 접근 통제가 핵심적입니다. 정책적으로는 모델의 용도와 위험 수준을 투명하게 공개하고, 고성능 모델에 대해 더 강한 감독 기준을 적용하는 방향이 논의되고 있습니다.

관련 용어

CBRN (Chemical, Biological, Radiological, Nuclear) 위협

CBRN 위험은 화학·생물·방사능·핵과 관련된 고위험 분야에서 발생할 수 있는 위협을 의미하며, 이 영역은 특히 AI 이중 용도 위험과 밀접하게 연결됩니다. 고도화된 AI가 실험 설계, 위험 물질 정보, 공격 절차 등을 생성하는 데 활용될 경우 심각한 안전 문제를 초래할 수 있어, 국제 보고서에서도 CBRN 관련 정보에 대한 접근 통제와 모델 출력 제한이 중요한 위험 관리 항목으로 다뤄지고 있습니다.