

5 RAND 연구소, 초지능 시대 미국의 우위 확보를 위한 전략 제언

➔ RAND 연구소는 초지능(Superintelligence) 시대 도래 가능성에 대응해 미국이 지정학적 우위를 확보할 수 있는 전략적 방법을 권고하는 보고서*를 발표('26.9)

* A U.S. Strategy to Secure Geopolitical Advantage on an Uncertain Path to Superintelligence

- (개요) 초지능 시대 도래를 전제로 미국이 선택할 수 있는 여러 전략들을 비교·분석한 뒤, 미래에 특정 전략을 선택할 수 있는 선택권을 유지할 것을 권고
 - 초지능이란 대부분 또는 모든 지적 영역에서 인간의 능력을 압도적으로 뛰어넘는 AI로, 끊임없이 발전하는 프런티어 AI는 초지능 시대의 도래를 예고
 - '26년 9월 미국 트럼프 대통령은 공식 문서에서 'AI'라는 용어를 '초지능'으로 변경하고 AI 위험 관리를 위한 글로벌 규약도 거부하겠다고 선언하면서, AI 기술 선도와 개발·배포를 우선시하겠다는 입장을 표명
 - 이러한 행보는 AI 개발과 관련한 국제적 합의나 협력을 여지를 좁히고 AI 개발 경쟁을 강화할 가능성이 농후
 - 보고서는 초지능 출현 가능성과 불확실한 미래를 염두에 두고 현재 활발히 논의되는 전략 선택지 7가지를 정리하고, 어느 전략이 적절한지 의사결정을 내릴 때 고려해야 할 5가지 판단 기준(불확실성 요소)을 제시
 - 선부르게 특정 전략을 채택하면 다른 전략을 선택할 기회가 사라지지만 의사결정을 미루면 전략 시행에 차질이 발생할 수 있다고 경고하면서도, 현재로서는 어느 전략이 바람직한지 확정하기 어렵다고 진단
 - 이러한 맥락에서 미국이 당장 특정 전략을 확정하기보다는 여러 전략의 실행 역량을 선제적으로 구축해 차후에 전략 선택의 여지를 남겨두는 '행동의 자유(Freedom of Action)' 전략을 취해야 한다고 권고
- (7가지 전략 유형) 초지능에 대응할 수 있는 전략 유형을 7가지로 분류해 ▲개요 ▲기대 결과(성공 상태)▲구조적 위험을 정리

〈 7가지 전략 유형 〉

구분	주요 내용
공존 전략	지배 (Dominance) • (개요) 미국이 프런티어 AI 경쟁을 선도하고, 그 우위를 바탕으로 글로벌 질서를 설정하는 영향력을 확보하는 전략 - 수출 통제, 사이버 작전, 공급망 차단 등을 통해 경쟁국의 관련 활동을 억제하는 동시에 기술 개발을 가속화 • (성공 상태) 미국이 초지능에 가장 먼저 도달하고 경쟁국의 시도를 억제 • (구조적 위험) 실질적 독점 체제가 권력을 집중시키고 지배국의 가치관을 고착시킬 위험을 초래
	공동개발 (Co-Development) • (개요) 어느 국가도 일방적으로 계속해서 경쟁에서 승리할 수는 없다는 가정 하에 미국이 국제 컨소시엄(최소한 중국 포함)을 구성을 주도해 안전한 초지능을 공동으로 개발하는 전략 - 공동 거버넌스, 컴퓨팅 자원 공동화, 다층 검증·모니터링 등을 추진 • (성공 상태) 미국이 경쟁국과 공동의 통제 아래 초지능을 공동 관리하며, 단일한 지배국이 없는 상태 • (구조적 위험) 국제 컨소시엄 창설 회원국의 가치·권력 배분이 장기적인 기본 질서로 자리 잡아 비(非)민주적인 가치가 고착될 수 있고, 컨소시엄 밖의 영역에서는 오히려 치열한 경쟁이 지속될 수 있음
	대비 (Preparedness) • (개요) 결정적인 우위나 공식적인 협력 모두 달성하기 어려운 상황에서, 미국이 비공식적 조율을 통해 공존을 위한 노력을 전개하는 전략 - 자발적 약속, 단계적 배포, 독립적 평가, 사고 보고, AI 관련 사고를 수용하고 회복할 수 있는 내부 역량 강화 노력 등을 포함 • (성공 상태) 비공식적 공존 노력과 사고에 대한 복원력을 갖춘 상태 • (구조적 위험) '임기응변식 대응'이 제도화되어 누적되는 피해와 인간의 권한 상실을 정상으로 받아들이는 결과를 초래
억제·차단 전략	개발 유예 (Moratorium) • (개요) 프런티어 AI 개발로 인한 파국적 위험이 임박했다고 판단해 정해진 임계치 이상의 AI 개발을 전 세계적으로 '일시 중지'시키는 전략 - 이를 통해 공존을 위한 연구를 성공적으로 수행할 시간을 확보 • (성공 상태) 미국과 경쟁국이 일정 기준을 초과하는 AI 개발을 함께 중단 • (구조적 위험) AI의 이점이 실현되지 못하고, AI 개발 중단을 강제하는 체제가 연구·혁신을 저해하고 개인의 권리·자유를 침해하는 전 지구적 감시 장치로 고착화될 위험이 존재
	억지 (Deterrence) • (개요) 협력으로 AI 개발을 중단하는 것은 불가능하지만, 초지능 개발을 차단하는 것은 가능하다고 판단하여 미국 내에서 일정 기준을 초과하는 프런티어 AI 개발을 중단하는 동시에, 국가적 역량을 총동원해 외국의 AI 개발이 위험 수준에 이르기 전에 저지·파괴하는 전략 - 이를 통해 공존을 위한 연구를 성공적으로 수행할 시간을 확보 • (성공 상태) 미국이 위협 또는 필요한 경우에는 물리력을 동원해 위험한 AI 개발을 저지 • (구조적 위험) 인류의 생존을 위한 시간은 별 수 있지만, 지속적인 대처 상태에 갇힐 위험이 존재
	사회 지속성 (Continuity of Society) • (개요) 모든 예방 전략이 실패하거나 실현 불가능하고 인간과 AI의 공존 또한 불가능한 경우, 일부 인류의 생존과 주체성을 보존하는 전략 - AI 기반 위협에 대비해 견고하게 방어된, 지리적으로 분산되고 생물학적으로 자립가능한 정착지를 조성 • (성공 상태) 위험한 초지능으로부터 생존가능한 자립적 인류 집단을 조성 • (구조적 위험) 인류를 보존하는 대신 많은 사람들이 고통받을 수 있고, 인류는 자신의 미래를 의미 있게 설계할 주체성을 거의 상실한 채 '유물 같은 존재'로 남을 수 있음

구분		주요 내용
가속 전략	가속화 (Acceleration)	• (개요) AI 개발을 제약하는 위험이 AI 개발 자체의 위험보다 더 크다고 판단하고, 새로운 보안 체계가 유익한 개발을 저해할 수 있다고 전제 - 시장, 경쟁, 신속한·반복적인 개선을 통해 유용한 AI를 개발하고 그 과정에서 자연스럽게 안전성을 확보하는 방식을 취하며, 규제 장벽을 제거해 물질적 풍요를 증진 • (성공 상태) AI 개발과 확산을 최대화해 편익을 확보하고 변화에 적응 • (구조적 위험) 강력한 거버넌스가 없는 풍요 속에서 여러 위험(환경 오염, 허위 정보 확산, 인류의 통제가 불가능한 시스템에 대한 과도한 의존 등)에 노출될 위험이 존재

- **(5가지 불확실성)** 초지능 대응 전략을 선택할 때 어떠한 선택지가 어느 상황에서 현실적이고 바람직한지를 판별하는 기준으로 5가지 불확실성 요소를 강조
 - 현재까지의 증거로는 실현가능성이 높고 합리적인 전략이 무엇인지 확정할 수 없으므로 특정 전략을 성급하게 선택하는 것은 바람직하지 않다고 주장

〈 5가지 불확실성(전략을 선택하는 기준) 〉

구분	판단 기준	선택 검토
위험의 근접성 (Danger Proximity)	AI로 인한 위험이 임박했는지 여부	• 위험이 근접해있다면 AI 개발 유예·억제·사회 연속성 등 위험 차단 전략을 우선시 • 위험이 멀리 있거나 AI 발전 자체가 더 큰 보호 효과를 낸다면 가속화 전략이 유리
공존 가능성 (Coexistence Feasibility)	인간과 첨단 AI가 인간의 생존과 주체성을 유지하는 방식으로 공존할 수 있는지 여부	• 공존이 가능하다면 프런티어 AI를 개발·도입하는 전략이 바람직 • 공존이 불가능하다면 개발을 억제·통제하는 방향, 또는 인류 문명을 보호하기 위한 다른 수단을 모색하는 방향을 중심으로 논의
억제 가능성 (Restraint Feasibility)	국가와 기업이 협력해 AI 개발을 억제할 수 있는지 여부	• 억제가 가능하다면 협력적 전략 채택이 가능 • 억제가 어렵다면 강압적인 집행 수단을 논의
결정적인 전략적 우위의 가능성 (Decisive Strategic Advantage)	특정 단일 행위자가 경쟁국보다 결정적인 전략적·군사적·경제적 우위를 확보하고 유지할 수 있는지 여부	• 결정적 우위를 확보할 수 있다면 지배 전략이 합리적 • 기술 확산이나 기타 요인으로 인해 결정적 우위 확보가 불가능하다면, 어떤 단일 행위자도 AI를 활용해 세계 질서를 좌우할 수 없을 것
억제 가능성 (Suppression Feasibility)	협력이 실패할 경우, 기술 통제, 억지, 무력 또는 기타 수단을 통해 자국 또는 경쟁국의 AI 프로그램 개발을 억제할 수 있는지 여부	• 억제가 가능하면 억지 전략도 가능 • 억제가 불가능할 경우, 행위자들은 다른 국가가 억지 정책을 펴더라도 위험한 AI 개발을 강행

- **(행동의 자유 전략)** 미국이 불확실성이 해소될 때까지 전략적 선택지를 유지할 수 있도록 필요한 기반 역량을 미리 준비하는 ‘행동의 자유’ 전략을 권고
 - 이를 위해 현재 미국이 수행·구축해야 할 조치 및 역량을 ‘투자’와 ‘운영’ 측면에서 제시

〈 행동의 자유 전략 〉

구분		필요한 조치·역량
투자	인간-AI 생태계 구축	• AI 안전 및 공존 관련 연구 지속 • 인간의 주체성을 보존할 수 있는 도구 및 제도 구축 • AI가 제공하는 혜택을 폭넓게 배분하고, AI 개발 정당성을 확보하기 위한 체계 마련 • AI로 인한 사회적 변화와 충격에 대비 • AI 생태계의 인센티브 및 구조 설계
	AI 보안 아키텍처 구축	• 상황 인식 • 컴퓨팅 가시성 및 거버넌스 • 사고 탐지, 격리, 대응 • 검증, 모니터링, 협력 인프라 • 억지 및 집행 • 대(對)AI 대응 및 복원력 • 국내 거버넌스 및 집행 역량
	기존 국가안보 아키텍처 조정	• 핵 억제·군비통제·비확산 체계 • 생물안보, 재래식 군사력, 정보(첩보)활동, 대테러, 사이버안보 • 외교 및 경제안보 전략
	시민·기업·정부의 대응 역량 구축	• 리더 양성 • 안전하고 유익한 AI 활용 촉진 • 비상 대응 계획 수립, 역량 구축, AI 위기 대응 계획 제도와 • AI 리터러시 제고
운영	징후 및 경보 모니터링	• 최첨단 기술의 발전 예측 • 군사적으로 중요한 기술 동향 파악 • 경쟁국 동향 감시 및 레드팀(Red-team) 분석 수행 • 징후 및 경보 임계값 설정 • 경보 전달 및 의사결정 유도
	재앙과 고착 상황(lock-in) 방지	• 안전 조치 없는 무모한 기술 발전은 저지·교란·악화 • 위험한 역량 확산 관리 • 침략 억제 • 전략적 안정성 유지 • 재래식 전쟁이 핵전쟁으로 확대되는 것을 막는 핵 방화선(firebreak) 유지 • 사고 관리 • 복원력 유지 • 각 전략의 최종 의사결정 시점을 파악해 해당 시점 도래 시 해당 전략 옵션을 유지할지 여부 결정
	우위 확보 및 신호 전달	• 우위 유지, 기술 확산 지연 등의 조치를 통해 유리한 기회를 오래 유지 • 확보한 우위를 작전상 이점으로 신속 전환하고 미국이 더 많은 선택지를 가지도록 조치 • 경쟁국 반응을 고려해 정보의 공개·비공개 등을 전략적으로 활용 • 전략적 전환 필요성이 설득력을 얻으면 해당 전략을 추진하고, 전략에 대한 의사 결정 시한 내에 전략 선택지의 유지·포기 여부 결정

출처 : RAND 연구소 (2026.9.15.)

<https://www.rand.org/pubs/perspectives/PEA5105-1.html>