



## 주요 동향(1) : 과학기술

### 1 RAND 연구소, AGI 패권 경쟁에서 라이드아웃 전략 제안

- ⇒ RAND 연구소는 범용인공지능(AGI) 경쟁에서 ‘선점 우위’ 논리를 반박하고 방어·반격 능력으로 전략적 안정성을 높이는 라이드아웃\* 전략을 강조하는 보고서\*\*를 발표(‘26.4)
- \* 라이드아웃(Rideout)은 냉전 시대 핵전략에서 적의 선제 핵 공격을 당한 뒤에도 살아남아 반격할 수 있는 능력을 갖춤으로써 적이 아예 공격하지 못하게 억제하는 교리를 의미
- \*\* The AGI Rideout Strategy for Reducing Strategic Risk and Promoting Stability in the Transition to Artificial General Intelligence
- (배경) 보고서는 국가 간 AGI ‘선점 우위(first-mover advantage)’ 경쟁이 냉전 시대와 유사한 불안과 전쟁 위험을 키운다는 문제의식을 배경으로 작성
    - 그동안 미국은 기술 리더십을 유지·확장함으로써 전략적 우위를 확보할 수 있다는 전제 하에 AGI에 가장 먼저 도달하기 위해 노력
    - 그러나 AGI를 먼저 확보한 국가가 영구적인 패권을 거머쥌 수 있다는 ‘선점 우위’ 논리 속에서 선제 공격이나 무제한적 군비 경쟁 위험은 점점 더 고조
  - (주요 내용) AGI 경쟁을 위한 새로운 전략적 청사진으로서 라이드아웃의 정의, ‘선점 우위’ 문제와 ‘플랜 B’의 중요성, 라이드아웃의 전제·목표·실행법을 설명
    - 핵 억제 전략의 역사적 경험을 바탕으로, 어느 국가가 AGI에 먼저 도달하든 전쟁을 피하고 계속해서 기술·전략 경쟁에 참여할 수 있는 능력을 갖추는 것이 중요함을 강조하면서 AGI 라이드아웃 전략을 제안
    - 이 전략은 AGI 개발 속도를 늦추거나 기술 리더십 추구를 포기하자는 것이 아니라, ‘선점 우위’만을 목표로 내달리기보다는 전쟁 등 지정학적 리스크를 완화하면서 AGI 전환기를 ‘무사히 견뎌내는 것’을 최우선의 목표로 강조
  - (정의) 핵전략의 라이드아웃 용어를 차용해 타국이 AGI에 도달하더라도 계속 경쟁하거나 방어할 수 있는 역량을 갖추는 것을 AGI 라이드아웃으로 정의
    - ‘AGI 경쟁에서 이기는 전략’이 아니라, ‘경쟁 과정에서의 대규모 전쟁과 전략적 붕괴를 막고, 미래 의사결정자의 선택지를 넓게 남겨두는 전략’을 의미
    - AGI 달성 및 선점에 주력하는 기존 목표와 양립 가능하지만 목표를 추구하는 방식에서 지정학적·기타 위험을 헤징하고 안정성을 유지하는 것을 최우선시

- (라이드아웃의 필요성) ‘선점 우위’ 전략의 문제점, AGI 전환기에 해결해야 할 지정학적 리스크, 상호확정 AI 오작동(MAIM\*)의 위험성 등을 설명하며 기존 전략에 대한 ‘플랜B’로서 AGI 라이드아웃 전략의 필요성을 역설

\* Mutual Assured AI Malfunction : AGI가 존재론적 위협이라는 전제 아래 AI를 발전시키되 AGI의 문턱은 넘지 않도록 특정 국가가 AGI 개발을 시도하는 경우 해당 국가의 AI 개발 인프라를 타격하는 방식 등으로 국가 상호 간에 AGI 개발을 막는 역지력을 확보해야 한다는 주장

- (선점 우위의 문제점) ‘선점 우위’ 전략은 4가지 잘못된 낙관적 기대(전제)를 기반으로 하고 있으며, 이러한 전제들은 동시에 완벽하게 충족되지 않을 경우 매우 바람직하지 않은 결과로 귀결될 가능성이 높음

〈 선점 우위 전략의 4가지 잘못된 전제 〉

AGI 선점 우위 전략이 의존하는 4가지 주요 가정

- 미국 AI 기업이 경쟁국보다 먼저 AGI를 개발·운영할 수 있는 능력을 갖추고 있다는 전제
- AGI를 향한 경쟁이 미국의 AI 개발을 저해하거나 고비용의 국제적인 분쟁을 유발하지 않을 것이라는 전제
- AGI가 낙관론자들의 예측처럼 폭발적인 군사적·경제적 이득을 즉각적으로 창출하고 후발국들을 장기간 멀리 따돌릴 수 있다는 전제
- AGI의 편익이 사회·경제적 혼란보다 충분히 커서 가능하면 빨리 이에 도달하는 것이 바람직하다는 전제

- (지정학적 리스크) AGI 도달 이전·이후를 막론하고 AGI 전환기에는 크게 4가지 범주의 지정학적 리스크에 대한 헤징이 필요한데, 라이드아웃 전략은 이러한 리스크로 인한 파국적 결과를 막는데 초점을 설정

〈 AGI 전환기의 4가지 지정학적 리스크(시나리오) 〉

위험 범주	정의	주요 우려사항
파국적 공격	적대국이 AI·AGI를 활용해 미국의 생존을 위협하거나 막대한 피해를 주는 시나리오	<ul style="list-style-type: none"> <li>• 대규모 사이버 공격</li> <li>• 최신 생물무기</li> <li>• 핵 지휘통제 체계 무력화</li> <li>• 대규모 분쟁 중 발생</li> </ul>
선제적(예방적) 공격	미국의 AGI 개발을 방해하기 위해 적대 세력이 선제적 공격에 나서는 시나리오	<ul style="list-style-type: none"> <li>• 데이터센터, AI 연구소, 반도체 생산시설 등의 무력화 및 핵심 인력에 대한 대규모 공격</li> </ul>
대규모 지역적 공격	AI로 인해 중국·러시아 등 미국의 주요 적대국 전력이 강화되고, 이들 국가와 미국·동맹국 사이에 전쟁 발생 가능성이 커지는 시나리오	<ul style="list-style-type: none"> <li>• 대만 해협 전쟁 역지력 실패</li> <li>• 확산 가능한 기타 지역적 분쟁</li> </ul>
기술적 추월	미국이 글로벌 경제·기술 경쟁에서 뒤처지면서 미국의 전략적 자율성 자체가 훼손되는 시나리오	<ul style="list-style-type: none"> <li>• 미국의 주권 훼손 또는 전략적 지위 상실</li> </ul>

- (MAIM 위험성) MAIM은 ①AGI가 존재론적 위협이라는 정치적 합의를 이루기 어렵고, ②각국의 AGI 개발 임계 시점을 포착하기도 어려우며, ③AI 인프라 타격이 오히려 전쟁 위험을 높인다는 점에서 제대로 작동하지 않을 것

※ 라이드아웃 접근법에 따르면 AGI 개발을 막기 위한 선제적인 AI 인프라 타격은 ‘생존 전략’이 아닌, ‘억제해야 할 위협’에 해당



- (기타 리스크) AGI가 초래할 경제·사회 혼란, 통제 상실(loss-of-control) 사고 등 비지정학적 리스크에 대해서도 라이드아웃 전략의 기본원칙 즉, 조기 위험 탐지, 선제적 대응, AGI 자체에 해결을 맡기지 않는 태도 등은 유효
- (라이드아웃 전략) 라이드아웃 전략의 전제, 목표, 실행법 등을 설명
- (전제) 라이드아웃은 ‘선점 우위’와 다른 전제를 기반으로 한 전략

〈 라이드아웃 전략의 주요 전제 〉

- 미국이 AGI를 먼저 개발할 가능성은 크지만, 중국 등 후발국이 미국이 미처 대비하지 못한 강력한 AI 응용 기술을 찾아낼 수 있다는 전제
- AGI의 편익은 즉각 폭발하기보다 조직 변화와 정치적 조정 속에서 점진적으로 나타날 가능성이 크다는 전제
- AGI를 먼저 확보한다고 해서 기술 독점·독점 유지가 보장되는 것은 아니며, 독점 유지에는 과격한 전쟁까지 요구될 수 있고 그마저도 성공이 불확실하다는 전제
- AGI 경쟁이 심각한 국제 분쟁으로 이어질 수 있으며 그 잠재적 피해가 매우 크다는 전제
- AGI 경쟁에서 최종적으로 승리하는 국가는 누구보다 먼저 AGI를 달성한 국가가 아니라, AGI로 인한 경제·사회적 충격을 가장 잘 관리한 국가가 될 것이라는 전제

- (목표) 미국에 대한 공격 억제, AI 기반 지정학적 리스크 대응, 미국의 AGI 개발·배치에 대한 선택의 여지 확보 등이 주요 목표
- (실행법) 전략 실행의 핵심 요소는 모니터링·적응, 적국 AGI 기반 군사력 상쇄 능력 개발, AGI R&D 인프라 보호, AGI 연구·개발·배포의 연속성 유지 등
- 이를 실행하기 위해 전략AI대응청(SARA)과 국가AI정보센터(NIC-AI)라는 두 개의 새로운 전담기구를 설치하고 국방부 전략능력실(SOC)과 연계해 유기적인 조직 구조를 구축

〈 SARA·NIC-AI·SOC 조직 구조 제안 〉

구분	주요 내용
SARA (Strategic AI Response Agency)	<ul style="list-style-type: none"> <li>• 국방 조직들이 기존 임무를 수행하는 동안 적대국의 AI 기반 공격 저지에 집중</li> <li>• 민간 방위산업기반(DIB)과의 직접 연계를 통해 AI 기반 전력의 신속 양산·확산 담당</li> <li>• 미사일방어청(MDA)과 유사하게 적대국의 AI 응용 기술에 대한 대응력을 개발·통합·조율하고, 억제·거부 전략을 지원하며, AI·AGI 시스템의 생애주기 관리를 유지하면서 방위산업기반 통합을 조율하는 역할 수행</li> </ul>
NIC-AI (National Intelligence Center for AI)	<ul style="list-style-type: none"> <li>• 외국 미사일·우주 활동을 감시하는 우주·미사일사령부(DEFSMAC)를 모델로 AI 기반 첩보 기법을 활용해 안보를 위협하는 글로벌 AI·AGI 발전을 추적하는 전문 정보기구</li> <li>• 신속하고 전문화된 AI 관련 정보를 생산해 SARA 및 국방부 SCO 등과 공유하며 의사결정을 지원</li> </ul>
SOC (Strategic Capabilities Office)	<ul style="list-style-type: none"> <li>• SARA와 군 서비스 간 연결고리로서의 역할 수행</li> <li>• SARA가 개발한 AGI 기반 방어·상쇄 기술을 군 도메인별로 실제 전력으로 통합하고, 중복·비효율을 해소</li> </ul>

출처 : RAND 연구소 (2026.4.27.)

<https://www.rand.org/pubs/perspectives/PEA4347-1.html>