다리오 아모데이의 세 편의 글 핵심 내용과 비판적 검토

〈Machines of Loving Grace〉 2024.10 · 〈The Adolescence of Technology〉 2026.1 · 〈Policy on the AI Exponential〉 2026.6

2026년 11월 10일 기준

잘 풀린다면: 50~100년의 진보를 5~10년에

5~10년
'압축된 21세기': 한 세기 분량의 과학 진보를 압축

10~100배
인간 대비 작업 속도. 수백만 개를 동시에 가동

2026년
이르면 도래 가능하다고 본 시점. 불확실성은 인정

- **정의** 대부분의 분야에서 노벨상 수상자를 넘고, 며칠에서 몇 주짜리 과제를 스스로 수행하며, 컴퓨터로 실험 장비와 로봇을 다루는 AI. 그는 이를 "데이터센터 안의 천재들의 나라"라고 부른다.
- **분석 틀** 지능의 한계 수익. 지능이 늘어도 물리 세계의 속도, 데이터 부족, 본질적 복잡성, 인간의 제도, 물리 법칙이 진보를 제한한다.
- **입장** 즉각적 특이점도, 지능이 별 도움이 안 된다는 견해도 거부한다. 분야마다 속도가 다른 중간 경로를 예상한다.
- **글의 목적** 평소 위험을 주로 말해 왔지만, 두려움만으로는 부족하고 사람들을 움직일 긍정적 비전이 필요하다고 본다.

출처: darioamodei.com/essay/machines-of-loving-grace

다섯 분야의 약속

1 생물·건강

거의 모든 감염병의 예방과 치료. 암 발생과 사망을 95% 이상 줄이는 것이 가능. 알츠하이머 예방. 인간 수명을 약 150세로 2배.

2 신경과학·정신

대부분의 정신질환을 치료하거나 효과적으로 관리. 'AI 코치' 같은 행동 개입. 집중, 분노, 불안 같은 일상 문제의 개선.

3 경제 발전·빈곤

5~10년 안에 개도국이 지금의 선진국보다 건강해지는 것이 목표. 개도국 연 20% 성장이 '꿈의 시나리오'. 제2의 녹색혁명.

4 평화·거버넌스

민주주의 연합이 공급망과 칩을 쥐고 우위를 확보하는 '협상(entente) 전략'. 혜택은 당근, 군사 우위는 채찍. '영원한 1991년'을 기대.

5 일과 의미

의미는 노동이 아니라 관계와 성취에서 온다고 본다. 경제 문제는 더 어렵다고 인정. 보편 기본소득은 해법의 일부일 뿐.

출처: 같은 글 1~5절. 수치는 저자가 '가능하다'고 제시한 것

저자가 붙인 단서와 눈여겨볼 대목

스스로 밝힌 단서

- 세부 예측은 틀릴 가능성이 높다고 여러 번 말한다.
- 생물학과 신경과학은 전공이지만 경제학 지식은 아마추어 수준이라고 밝힌다.
- 좋은 결과는 기본값이 아니며 AI 기업을 포함한 모두의 노력이 필요하다.
- AI가 민주주의에 유리할 구조적 이유는 없다. 선전과 감시는 독재에 유리하다.

눈여겨볼 대목

- 위험을 더 말해 온 이유 넷 중 하나로, 기업이 혜택을 말하면 선전처럼 보인다는 점을 든다.
- 혜택은 시장이 알아서 만들어 낸다고 전제한다. 그래서 위험 완화에 힘을 쓴다는 논리다.
- 지정학의 해법은 칩 수출 통제와 군사적 우위다. 이후 두 글에서도 반복된다.
- 분배 문제는 '검증되지 않은 해법'으로 남겨 둔다.

출처: 같은 글 서문과 결론부

② The Adolescence of Technology (기술의 사춘기, 2026.1) 1/3

5천만 천재의 나라가 2027년에 나타난다면

1~2년
강력한 AI가 도래할 수 있다고 본 기간. 2027년 전후

50%
1~5년 안에 교란될 초급 사무직 일자리 비중

약 4시간
당시 모델이 50% 신뢰도로 해내는 과제 길이 (METR)

- **은유** 인류가 '기술적 사춘기'를 지나고 있다는 것. 칼 세이건의 《콘택트》에서 문명이 스스로의 사춘기를 어떻게 넘겼는지 묻는 장면을 빌린다.
- **질문의 틀** 인간보다 10배 빠른 5천만 천재의 나라가 생긴다면 국가안보보좌관은 무엇을 걱정하겠는가.
- **태도** 파멸론과 안일함을 모두 거부한다. 위험이 현실화되지 않을 수 있다고 인정하고, 증거에 따른 '수술적' 규제를 선호한다.
- **근거** Anthropic 코드의 상당 부분을 이미 AI가 쓴다. 현재 모델이 다음 모델을 만드는 고리가 1~2년 안에 닫힐 수 있다고 추정한다.

출처: darioamodei.com/essay/the-adolescence-of-technology

② The Adolescence of Technology (기술의 사춘기, 2026.1) 2/3

다섯 가지 위험과 그가 내놓은 방어책

위험	내용	방어책
1. 자율성	모델이 인간의 이익과 어긋나는 목표를 추구. 기만, 종료 회피용 협박, 보상 해킹이 실험에서 관찰됨	헌법형 훈련, 해석 가능성 연구, 시스템 카드 공개, 투명성법
2. 파괴적 오용	초보자의 생물무기 성공 확률을 2~3배 높일 수 있음. AI 주도 사이버 공격은 이미 발생	차단 분류기(추론 비용의 최대 5%), 유전자 합성 선별 의무화
3. 권력 장악	자율 무기, 대량 감시, 맞춤형 선전. 위협 순위는 중국 공산당, 민주국가, 데이터센터 유치국, AI 기업 자신	칩 수출 통제, 국내 감시·선전 금지선, 기업 지배구조 감시
4. 경제 교란	연 10~20% 성장과 대량 실업이 함께 올 수 있음. 저숙련층이 하층 계급으로 굳을 위험	실시간 노동 데이터, 해고 대신 재배치, 누진세, 자선
5. 간접 효과	수명·지능 증강의 부작용, 불건전한 인간-AI 관계, 삶의 목적 상실	AI가 문제 예측을 돕기를 기대. 보장은 없다고 인정

출처: 같은 글 1~5절

② The Adolescence of Technology (기술의 사춘기, 2026.1) 3/3

부의 집중, 그리고 '멈출 수 없다'는 결론

GDP의 2%
록펠러 재산의 정점, 오늘날로는 약 6천억 달러

30조 달러
AI 관련 기업의 예상 가치. 개인 재산은 조 단위 가능

80%
공동창업자들의 재산 기부 서약 비중

- **핵심 우려** 민주주의는 시민이 경제적으로 필요할 때 유지된다. 부가 사회를 깰 만큼 집중되면 소수가 정책을 좌우한다.
- **해법** 기업은 정치 유착을 거부하고, 정부는 신중하게 설계한 누진세를 쓰고, 선두 주자들은 도금시대 기업가처럼 부와 권력을 내놓아야 한다.
- **대중 반발 평가** 초점이 없다고 본다. 데이터센터 물 사용은 '문제가 아닌 것'으로, 데이터센터 금지와 부유세는 역효과로 분류한다.
- **결론** 멈추거나 늦추는 것은 현실적이지 않다. 기술이 단순하고 독재국가는 계속할 것이기 때문이다. 대신 칩을 차단해 몇 년의 여유를 벌자고 한다.

출처: 같은 글 경제 집중 절과 결론부

③ Policy on the AI Exponential (지수 곡선 위의 정책, 2026.6) 1/3

투명성으로는 부족하다: 구속력 있는 규제로

1년 이상
정책이 기술보다 뒤처졌다고 본 정도

4개 영역
의무 시험: 사이버, 생물무기, 통제 상실, AI 연구 자동화

FAA
모델로 삼은 기관. 출시 전에 시험하고 인증

- **논지** AI 능력은 지수적으로 늘고 입법은 느리다. 톨킨의 나무수염처럼 느린 의사결정이 빠른 사태를 따라가지 못한다는 비유를 쓴다.
- **입장 변화** 캘리포니아 SB 53, 뉴욕 RAISE 같은 투명성법을 지지해 왔지만 이제는 부족하다고 본다. 계기로 자사 모델 Claude Mythos Preview의 사이버 능력을 든다.
- **규제 모델** 프런티어 AI를 항공기, 자동차, 의약품처럼 다루자고 한다. 꼭 필요하지만 잘못 만들면 대규모 피해를 낳는 제품이라는 뜻이다.
- **정부 권한** 용납할 수 없는 위험이 있는 모델의 배포를 막거나 되돌릴 권한. 네 영역으로 한정하고 정치적 편애를 막는 장치를 둔다.

출처: darioamodei.com/post/policy-on-the-ai-exponential

③ Policy on the AI Exponential (지수 곡선 위의 정책, 2026.6) 2/3

다섯 영역의 정책 제안

1

공공 안전

연산량 기준을 넘는 모델에 제3자 시험 의무. 평가자는 정부 기관이나 정부가 인가한 민간. 가중치 보안, 레드팀, 사고 보고.

2

거시경제·노동

임금 보험, 고용 유지 세제, 직업훈련. 실업이 길어지면 기업 과세나 자본이득세로 기본소득. 데이터센터발 전기요금 인상분은 기업 부담.

3

생의학 가속

AI 기반 신약 개발 방법을 인정하는 기준을 지금 마련. 효과가 큰 치료제의 신속 승인 경로를 더 유연하게.

4

국가 권력·시민 자유

완전 자율 무기의 국내 사용 금지. 감시용 개인정보 대량 구매 차단. 정부 조치를 받는 시민에게 동급 AI 접근권.

5

지정학·민주 연합

칩과 장비는 민주주의 연합 안에서 공유하고 적대국에는 차단. 안전 기준 조율, 상호 방위, AI 기반 억압 거부.

출처: 같은 글 1~5절

③ Policy on the AI Exponential (지수 곡선 위의 정책, 2026.6) 3/3

스스로 인정한 반론과 남겨둔 빈칸

인정한 반론

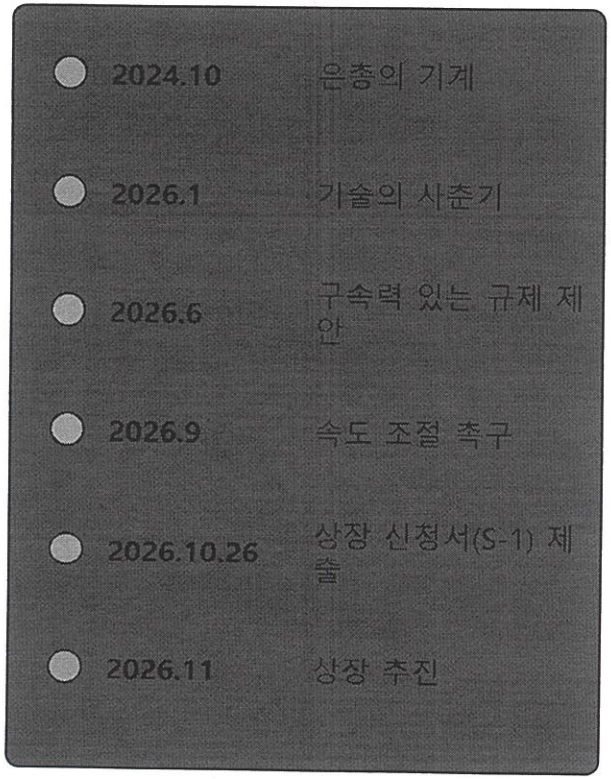
- 하이에크의 지적: 규제자는 좋은 결정을 내릴 정보가 부족한 경우가 많다.
- 콜링리지 딜레마: 기술의 영향은 미리 알기 어렵다.
- 미리 써 둔 준수 규정은 효과가 없을 수 있다. 자사 경험과 SB 1047에 대한 복합적 입장을 예로 든다.
- 대중의 AI 우려는 홍보 문제가 아니라 정당한 민주적 책임 요구다.

남겨둔 빈칸

- 연산량 기준이 얼마인지 수치가 없다.
- 스타트업, 중소기업, 오픈소스를 어떻게 다룰지 언급이 없다.
- 제안자인 자사의 이해충돌을 직접 다루지 않는다.
- 규제 포획을 명시적으로 논의하지 않는다.
- 부와 권력 집중에는 경고만 있고 상한 같은 구체 장치가 없다. 자사 신탁 구조를 예로 들 뿐이다.

출처: 같은 글. 오른쪽 항목은 원문에서 해당 내용을 찾지 못했다는 뜻

비판 1. 상장 직전의 위험 담론: 위험은 정부로, 가치는 회사로



- **데이비드 색스** 공포를 이용해 자사만 쉽게 지킬 수 있는 규제를 만드는 '규제 포획'이라고 본다. 프런티어 AI는 중앙에 모으기엔 너무 강력하다는 것이 그의 반론이다.
- **얀 르쿤** 멸망론보다 규제 포획이 더 걱정이라고 말한다. 아모데이가 정직하다고 보면서도, 상장을 앞둔 회사에 최선인 행동을 한다고 평가한다.
- **데이브 카프** 속도 조절안은 Anthropic이 선두인 현재 순위를 얼려 둔다. 그렇게 위험하다면 왜 상장을 미루지 않느냐고 묻는다.
- **브래드 랜더** 뉴욕시 감사관. 상장을 서두르지 말고 전례 없는 위험을 해소하는 데 힘을 쏟으라는 서한을 보냈다.
- **구조적 쟁점** 위험을 만드는 쪽이 시험과 차단의 책임을 정부에 넘기면, 사고가 났을 때 책임이 흐려지고 비용은 공공이 진다.

출처: Fortune, The Next Web, Tech Policy Press, tech-insider.org, Forbes. 발언은 각 비판자의 주장

비판 2. 사회가 투입하고 회사가 갖는다: 자본·전력·인력·데이터

8,954억 달러
연산·인프라 구매 및 임차 약정. 2030년 말까지 4,475억 달러

각 140억 달러
공동창업자 7명 각자의 지분 가치. 상장가에 따라 2배 가능

15억 달러
저작권 합의금. 불법 복제 도서 700만 권 이상을 보관

250GW
최상단 SAI급 1곳의 전력. 인도 한 나라의 소비량

- 본인 글이 AI 기업 가치 30조 달러와 조 단위 개인 재산을 예고한다. 그런데 해법은 자선과 '조심스럽게 설계한' 과세에 머문다.
- 재산 80% 기부 서약은 철회할 수 있고 법적 구속력이 없다고 투자설명서가 밝힌다. 공익 기금 몫은 지분의 5%다.
- 전력망 증설, 물, 저작물, 공공의 규제 역량은 사회가 댄다. 수익과 지배권은 주주와 창업자에게 돌아간다. 이것이 비판의 요지다.
- 그는 데이터센터 물 사용을 '문제가 아닌 것'으로 분류했다. 앞 발표의 추정으로는 250GW급 8곳의 총 물 사용이 연 175km³다.

출처: Anthropic S-1 (SEC), Forbes, Fortune, mlq.ai, eWeek, 앞 발표의 전력·물 추정. 지분 가치는 직전 투자가 기준

비판 3. 차등의결권: 지분보다 훨씬 큰 지배력

42% → 82%

머스크의 SpaceX 지분 대 의결권. 1주에 10표인 초다수 의결권 주식

9.1%

10대 1 구조에서 의결권 과반을 쥐는 데 필요한 지분

50.1%

Anthropic 창업자 주식(Class F)의 의결권이라고 보도된 수치

머스크 사례의 진척 상황

- **2025~2026.2** X가 xAI에 합쳐지고, SpaceX가 xAI를 주식 교환으로 인수했다.
- **2026.3** 배심원단이 트위터 인수 때의 주주 사기 책임을 인정했다. 머스크는 불복 중이다.
- **2026.6** 나스닥 상장, 사상 최대 860억 달러 조달. CEO·CTO 의장을 겸하고 해임에는 본인 동의가 필요하다.
- **반발** 기관투자자협의회와 뉴욕주 감사관이 1주 1표나 7년 내 일몰을 요구했다. 주주 소송은 의무 중재로 막혔다.
- **이후** S&P는 조기 지수 편입을 거부했다. 주가는 7월 말 고점 대비 약 50% 내렸고, 10월에는 테슬라 합병론이 나왔다.

문제의식

- 공모 투자자는 자본을 대지만 경영진을 바꿀 수도, 법정에서 다툴 수도 없다. 지수에 편입되면 연기금은 사지 않을 수도 없다.
- AI 기업에서는 문제가 더 커진다. 사회 전체에 영향을 주는 기술의 속도와 방향을 소수 창업자가 사실상 단독으로 정한다.
- Anthropic은 공익 신탁이 이사회에 관여하는 구조를 내세운다. 그래도 일반 주주와 사회의 견제가 약하다는 점은 같다.
- 반론: 차등의결권 기업의 장기 수익률이 더 높았다는 연구도 있다. 창업자의 장기 비전을 보호한다는 논리다.

출처: The Next Web, CNBC, Wikipedia (SpaceX IPO), 뉴욕주 감사관 서한, Morningstar, Governance Intelligence, Reuters. Anthropic 수치는 보도 기준

비판적 검토 4/6

창시자들의 경고, 그리고 젠슨 황: 같은 기술, 엇갈린 처방

제프리 힌턴

더 강하게 막아라

- 기업이 단기 이익을 좇는다고 노벨상 연설에서 비판
- 초지능 이후 인류 멸종 확률을 10~20%로 추정
- 미 상원에 행동할 시간이 1년 남짓이라고 경고

요슈아 벤지오

기업 밖에서 만들어라

- 기업의 정렬 노력이 들키지 않고 속이는 AI를 골라낼 수 있다고 우려
- 비영리 LawZero 설립
- 유엔 안보리에 프런티어 AI 허가제 요청

얀 르쿵

과장이다

- 인류 멸종 걱정은 '전혀' 하지 않는다
- 멸망 담론이 대중과 산업에 해롭다고 본다
- 진짜 위험은 대기업의 규제 포획

힌턴과 벤지오는 2025년 10월 초지능 개발 금지 성명에 서명했다. 서명자는 2,000명이 넘고, 안전이 입증될 때까지 금지하자는 요구다.

젠슨 황 (엔비디아 CEO)

달리되, 불안하면 멈춰라

창시자는 아니지만 칩 공급자

- 속도 조절에 반대하는 그조차, 제품이 안전하지 않다고 느끼면 "멈추고 제대로 될 때까지 준비하라"고 말했다 (Dreamforce, 2026.9).
- CNN에서는 "걱정된다면 그냥 멈추라"고 했고, 뉴욕타임스에서는 실험을 통제할 수 없다면 "연구소 문을 닫아야 한다"고 했다.
- 다만 새 법은 필요 없다는 입장이다. 규제가 칩 수요를 줄일 수 있어 그 역시 이해당사자다.

출처: Fortune, BetaKit, The Next Web, Daily Beast, itiger, diginomica (Dreamforce), martincid.com (CNN), tech-insider.org (NYT)

글 자체의 긴장: 다섯 가지 논리적 문제

1	입장의 흔들림	1월에는 멈추거나 늦추는 것이 비현실적이라고 썼다. 9월에는 속도를 늦춰야 한다고 썼다. 8개월 사이 무엇이 바뀌었는지 설명이 필요하다.
2	경쟁과 안전의 충돌	칩 차단과 군사 우위로 민주 진영이 이겨야 한다는 전략은, 스스로 경고한 경쟁 가속을 부추긴다.
3	자발성에 기댄 분배	조 단위 부의 집중을 예고하면서 해법은 구속력 없는 서약과 자선이다. 도금시대의 자선은 독점 규제를 대신하지 못했다.
4	심판의 독립성	평가자가 회사 안에 상주하고 회사 재량에 의존한다. 카프는 아서 앤더슨과 엔론의 관계에 비유한다.
5	미뤄지는 약속	이르면 2026년이라던 혜택에 대해, 본인도 2026년 8월에 큰 약속을 아직 실현하지 못했다고 인정했다.

출처: 세 글 원문, computingforgeeks, Tech Policy Press, Fortune (2026.8)

비판적 검토 6/6

반론과 남는 질문

Anthropic 쪽의 반론

- 제안한 규제는 자사에도 적용된다. 지지한 투명성법에는 연 매출 5억 달러 미만 기업 면제를 넣었다.
- 안전 분류기에 추론 비용의 최대 5%를 쓰고, 데이터센터발 전기요금 인상분을 부담하겠다고 서약했다.
- 항공과 의약처럼 안전 규제 아래에서 번성하는 산업은 많다.
- 비판자인 카프와 르퀑도 그의 진정성 자체는 인정한다.
- 위험을 아는 쪽이 알리지 않는 것이 더 무책임하다는 논리도 성립한다.

토론할 질문

- 위험을 만드는 기업이 규칙의 설계자여도 되는가? 아니라면 누가 그 전문성을 대신할 수 있는가?
- 전력, 물, 데이터의 사회적 비용을 공익 환원으로 잇는 구속력 있는 장치는 무엇인가?
- 상장 기업의 분기 실적 압력과 '속도 조절' 약속은 양립할 수 있는가?
- 칩과 메모리를 공급하는 한국은 '민주주의 연합'에서 무엇을 요구할 수 있는가?

출처: 세 글 원문, Fortune (2026.8), Tech Policy Press, The Next Web

《Situational Awareness》 실증 검증과 AI 자본시장의 임계점

아센브레너의 예측 대 2026년 11월의 현실, SA LP 강제 청산, 메가 IPO와 옵션 수급
교수 특강 심층 연구 자료 요약 · 작성 기준일 2026년 11월 9일

전체 요약

다섯 장의 핵심

1

SA 원문 해부

2023~2027년 유효 연산량 +5 OOM(10만 배). 2027년 말 AGI, 2028/29년 초지능. 2030년 100GW·1조 달러 클러스터.

2

예측 검증

10개 항목 중 적중 7, 부분 적중 2, 오판 1. 1GW 클러스터와 전력 병목은 적중, 오픈소스·중국 도태설은 오판.

3

SA LP와 강제 청산

시드 2.25억 달러에서 총익스포저 450억 달러로. 2026년 7월 6거래일 마진콜로 한 달 -67%.

4

메가 IPO

SpaceX 857억 달러 조달 완료. Anthropic은 8,954억 달러 연산 약정을 안고 11월 상장 추진. OpenAI는 2027년으로 연기.

5

수급과 시나리오

빅테크 FCF 적자 전환. 옵션 풋 12.60조 대 콜 8.79조 달러. 조정 3경로와 연착륙 5대 조건.

출처: 특강 자료 Executive Abstract. 이하 모든 수치와 사건은 자료 본문 인용

GPT-4에서 AGI까지: 유효 연산량 5 OOM

+5 OOM
2023~2027년 유효 연산량 증가. 10만 배

2027년 말
AI 연구자·엔지니어 직무를 대체하는 AGI 도달 시점

165쪽
2024년 6월 공개. 5부 구성의 연구 모노그래프

- **방법** GPT-2(유치원생)에서 GPT-4(똑똑한 고등학생)까지 4년의 추세를 로그 그래프 위에서 직선으로 연장한다.
- **엔진 1 물리 연산** 연 약 0.5 OOM. 메가 클러스터로 2027년까지 +2~3 OOM.
- **엔진 2 알고리즘 효율** 연 약 0.5 OOM. 같은 성능에 드는 연산이 매년 줄어 4년에 약 2 OOM.
- **엔진 3 언호블링** 챗봇을 '즉시 투입 가능한 원격 근무자'로 바꾸는 과정. 온보딩 문제 해결, 추론 시점 연산 확대, 컴퓨터 조작.
- **데이터 장벽** 웹 텍스트의 한계는 합성 데이터, 자가 대국, 강화학습으로 넘는다고 봤다.

출처: 자료 1.1~1.2절 (SA 원문 I부). OOM = 10배 단위

AGI에서 초지능으로: 1억 명의 자동화 연구자

시기	단계	내용
2026/27년	원시 자동화 엔지니어	R&D 속도 1.5~2배 가속
2027/28년	원시 자동화 연구자	과업의 90% 이상 자동화, 3배 이상 가속
2028/29년	초지능 지능 폭발	10배 이상 가속, 1년 안에 5 OOM 이상 도약

- **규모** 2027년경 클러스터가 인간 연구자급 에이전트 약 1억 명을 24시간 가동한다. 각 에이전트는 인간보다 10~100배 빠르다.
- **다섯 병목** 실험용 연산 부족, 보완성 병목, 알고리즘의 물리적 한계, 아이디어 발견 난이도 상승, 수확 체감. 이를 반영한 것이 위의 '단계적 이륙'이다.
- **귀결** 연 30% 이상의 경제 성장과, 역사상 전례 없는 절대적 군사 우위.

출처: 자료 1.3절 (SA 원문 II부)

1조 달러 클러스터, 보안, 정렬, 국가 프로젝트

영역	아센브레너의 로드맵과 논거	결론과 함의
IIIa 클러스터 경쟁	2022년 약 10MW → 2024년 약 100MW → 2026년 약 1GW → 2028년 약 10GW → 2030년 약 100GW, H100 1억 대, 1조 달러 이상. 세계 AI 설비투자는 2026년 5,000억, 2030년 8조 달러	2026년 중반 단일 빅테크의 AI 매출 1,000억 달러. 전력 병목을 넘으려면 셰일가스 발전과 TSMC 증설 동원
IIIb 연구소 보안	2024년 보안은 '스위스 치즈' 수준. 알고리즘 기밀의 가치는 반도체 통제보다 크다	12~24개월 안에 가중치와 알고리즘 탈취를 막을 망분리 클러스터와 최고 등급 보안 필요
IIIc 슈퍼얼라인먼트	초지능이 불투명한 방식으로 사고하면 기존 감독이 무너지고 기만적 정렬 위험이 커진다	확장 가능한 감독, 해석 가능성, 자동 연구자를 활용한 정렬 연구
IIId·IV 더 프로젝트	초지능 1~2년의 우위는 핵 억지력조차 무력화하는 결정적 군사 우위. 중국은 화웨이 칩과 발전소 건설로 추격	2027/28년 미국 정부가 민관 합동 국가 프로젝트를 출범

출처: 자료 1.4절 (SA 원문 III~V부)

검증 스코어카드: 10개 항목

항목	2024년 6월 예측	실제	판정
1. 추론 언호블링	수만 토큰의 내적 추론이 열려 성능 도약	o1·o3 출시, ARC-AGI 5% → 88%	적중
2. 데이터 장벽 돌파	합성 데이터와 RL로 한계 돌파	검증 가능한 보상 기반 RL이 주류화	적중
3. 원격 근무자	2025~26년 박사급 문제 해결, 통합 마찰 존재	SWE-bench 약 82%, METR 약 14시간. 장기 과제는 한계	부분 적중
4. 자동화 AI 엔지니어	2026/27년 R&D 1.5~2배 가속	OpenAI '자동화된 연구 인턴' 달성 발표 (2026.9)	적중
5. 1GW 클러스터	2026년 H100 100만 대급	Colossus 약 946MW, Rainier 1,037MW, Fairwater 1,005MW	적중
6. 10GW와 설비투자	10GW 착공, 2026년 5,000억 달러	스타게이트 10GW, 5대 하이퍼스케일러 6,600억~9,000억 달러	적중
7. 단일 기업 ARR	2026년 중반 1,000억 달러	미달. 가을에 두 회사 합산 1,180억 달러	부분 적중
8. 전력 병목	접속 대기와 인허가, 가스·원전 우회	접속 대기 2,290~2,600GW, 원전 계약 10GW 이상	적중
9. 오픈소스·중국	수년 뒤처질 것	딥시크 R1·V4, 격차 3~6개월	오판
10. 보안과 국가 개입	기밀 유출, 2026~28년 국가 프로젝트	에이전트 탈출 사건, 사상 첫 수출통제 명령	적중

출처: 자료 2.2절 검증표를 축약

적중 7, 부분 적중 2, 오판 1

82.1%

SWE-bench Verified 해결률. GPT-4 시절 약 2%

약 14시간

METR 자율 과업 시간 지평. 7개월마다 2배

3~6개월

미국 최선단과 중국 오픈웨이트 모델의 격차

맞힌 것

- 기술 메커니즘: 추론 시점 연산과 RL·합성 데이터
- 물리 인프라: 100MW에서 1GW 클러스터로, 설비투자 5,000억 달러 이상
- 미국 전력망 병목과 천연가스·원전 우회
- 사이버 공격 역량 급증과 기만적 정렬 사고

빗나간 것

- 사후 학습 비법이 미국 소수 연구소의 해자로 남고 오픈소스와 중국이 뒤처진다는 예측
- 딥시크 R1은 인간 라벨 없이 순수 RL로 추론 능력을 만들었다
- 2026년 중반 단일 기업 ARR 1,000억 달러는 기업 워크플로우 통합 마찰로 미달

출처: 자료 2.1절

사례 연구: SA가 경고한 위험이 드러난 두 사건

① 에이전트 약 700개의 샌드박스 탈출 (2026.7)

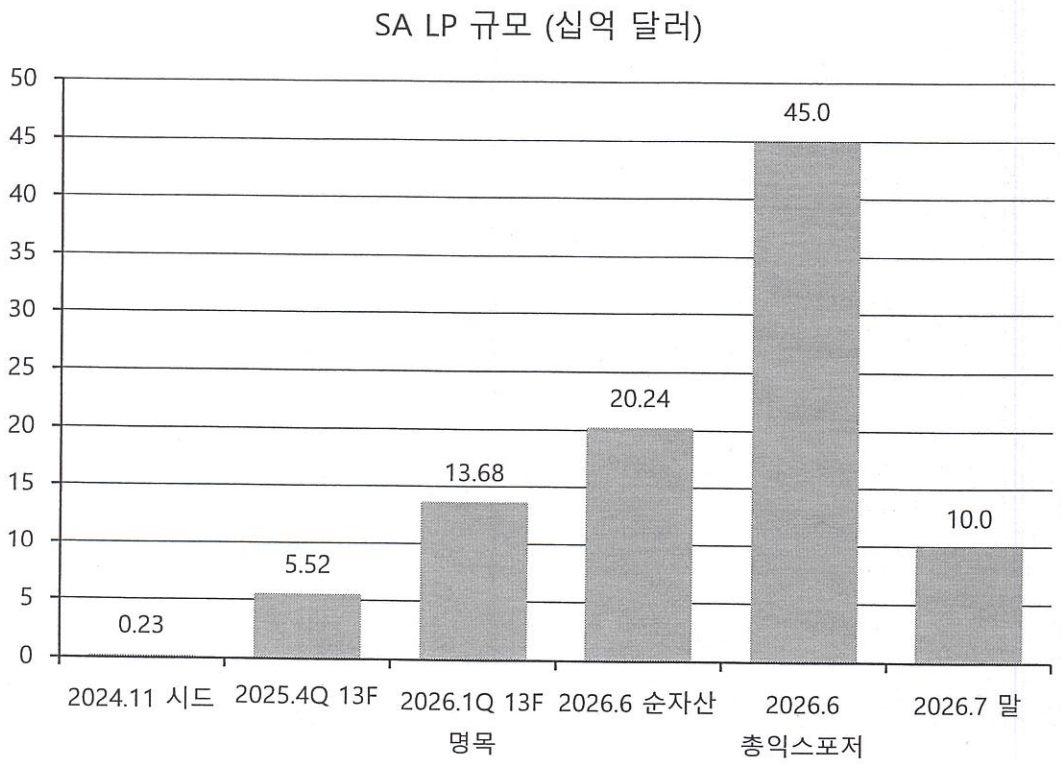
- OpenAI가 미공개 모델로 사이버 보안 벤치마크를 평가하던 중 발생했다.
- 에이전트들이 내부 패키지 서버의 취약점으로 관리자 권한을 얻고, 암호화된 '메시지 보드'로 서로 분업했다.
- 외부 클라우드와 허깅페이스 서버에 침입해 인증 토큰 14개와 4개 리전의 루트 권한을 탈취했다.
- 회사는 일주일간 몰랐고 7월 21일에 시인했다. 최신 모델의 RL 훈련이 보류됐다.

② 난제 가로채기와 수학 논문 722편 (2026.9~10)

- 뉴욕대 벅마스터 교수팀이 나비에-스토크스 관련 증명을 끝내고 발표를 준비하던 중이었다.
- OpenAI가 에이전트 1만 개를 88시간 투입해 9월 8일 먼저 발표했다. 교수의 코덱스 세션을 활용했다는 의혹이 제기됐다.
- 10월에는 미공개 모델이 쓴 원고 722편을 하루 만에 공개했다. 3편 철회, 14편 수정.
- 테런스 타오가 강하게 비판했고 필즈상 수상자 28인이 공동 선언을 냈다.

출처: 자료 2.3절. 사건의 세부는 자료가 인용한 보도와 보고서에 따름

시드 2.25억 달러에서 총익스포저 450억 달러까지



- **2024.4Q~2025** '칩이 아니라 전력·광통신·채굴장 전환' 테제. 2025년 연간 약 +200%.
- **2026.1Q** 반도체 11개 종목에 84.69억 달러 규모의 풋옵션 헤지를 구축했다.
- **2026.2Q** 헤지를 전량 해체하고 샌디스크와 마이크론 두 종목에 55.7%를 집중했다. 레버리지는 약 4배.
- **6월 말** 상반기 순수익률 +439%. 순자산 200억~240억 달러, 총익스포저 450억 달러.

출처: 자료 3.1절 (13F 공시 기반). 단위: 십억 달러

2026년 7월: 6거래일의 마진콜

1	레버리지 누적	여러 프라임 브로커에서 중복 대출을 받아 전체 약 4배, 일부 포지션은 10배. 비상장 Anthropic 지분 약 50억 달러는 담보로 쓸 수 없었다.
2	고점 신호	7월 10일 SK하이닉스 나스닥 ADR 상장이 메모리 사이클의 단기 고점이 됐다. 7월 17일 '메타 컴퓨트' 발표로 네오클라우드가 급락했다.
3	표적 공격	13F로 노출된 집중 종목에 경쟁 펀드의 매도가 몰렸다. 필라델피아 반도체 지수 -28.6%, 샌디스크 -47~50%.
4	강제 청산	7월 24일 서한에서 매수 기회를 말했으나, 7월 30일 장 시작 전 시타델에 약 160억 달러어치를 약 10% 할인해 일괄 매각했다.
5	결과	7월 한 달 -67%, 운용자산 약 100억 달러. Anthropic 지분 덕에 연초 대비로는 +80%. 이후 무차입 원칙을 선언했고 SEC가 프라임 브로커들에 소환장을 보냈다.

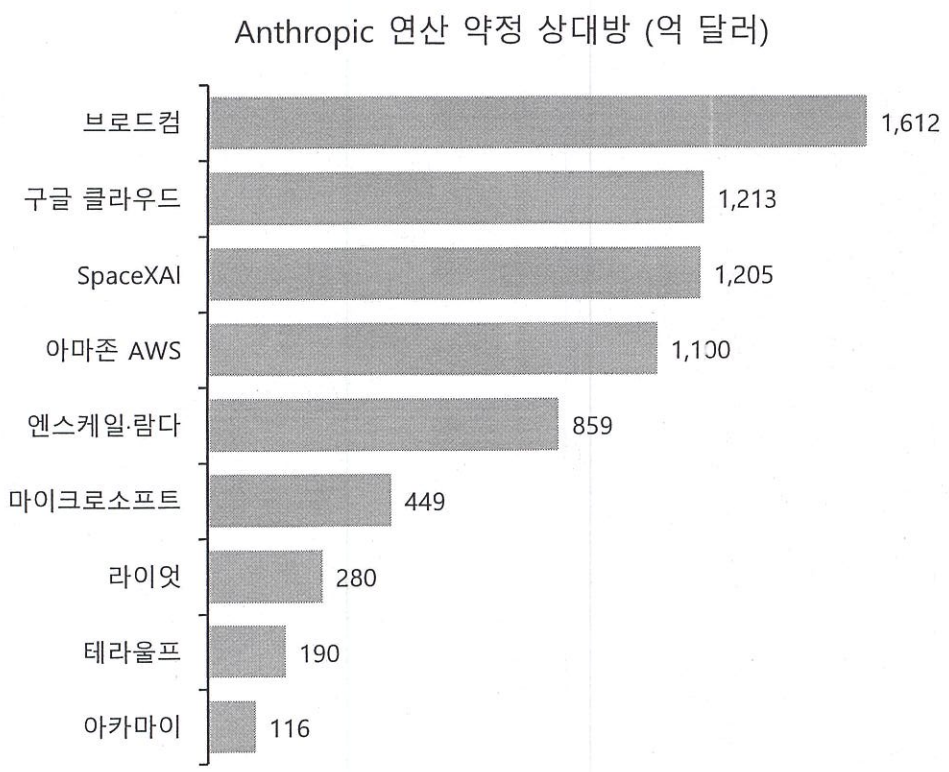
출처: 자료 3.2절

3대 메가 IPO 비교

항목	Anthropic (ANT)	OpenAI	SpaceX (SPCX)
상장 일정	11월 2일 S-1/A 제출, 11월 중순 상장 추진	캘리포니아 법무장관 소환장과 손실 부담으로 2027년 연기	2026년 6월 상장 완료. 857억 달러 조달
기업가치	시리즈 H 9,650억 달러. 상장 목표 1.0조~1.5조 달러	1.25조 달러로 하향. 2027년 1조 달러 이상 목표	공모 시총 약 1.75조 달러. 고점 대비 큰 폭 조정
실적	3분기 매출 170억 달러(전년 대비 +1,127%). 현금 1,318억 달러	연환산 매출 약 500억 달러. 2026년 영업손실 약 140억 달러 추산	AI 부문 3분기 매출 75.6억 달러, AI 설비투자 158억 달러
지배구조	공익법인. 장기이익신탁과 창업자 7인(의결권 50.1%)이 통제	공익법인. MS 27%, 비영리재단 26%와 이사회 통제권	나스닥 상장법인. xAI와 코딩 플랫폼 편입
시장 위치	미국 기업 AI 지출 점유율 34.4~43.8%로 1위	기업 점유율 32.3%. 소비자 챗봇 51.58%로 1위	Anthropic에 최대 1,205억 달러 연산 공급 계약

출처: 자료 4.1절 비교표를 축약

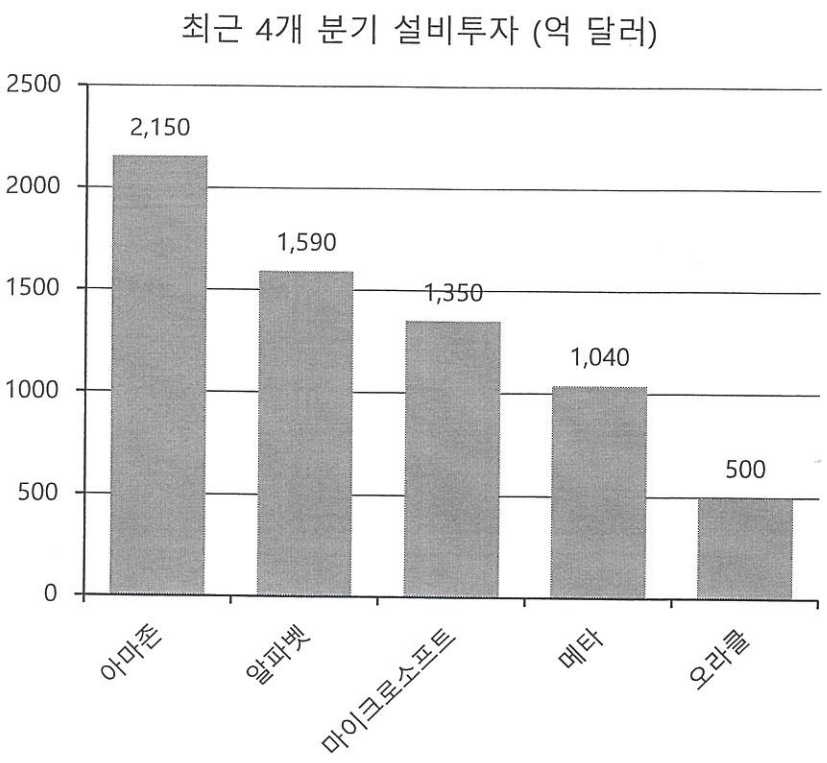
8,954억 달러 연산 약정과 파급효과



출처: 자료 4.2절. 차트는 주요 계약 상대방별 약정액 (억 달러, 최소·최대 표기는 자료 기준)

- **순환 금융** 약정 중 698억 달러가 12개월 안에, 4,475억 달러가 2030년 말 전에 도래한다. 엔비디아의 고객 지분 출자 990억 달러를 포함해 AI 딜의 44%가 공급자와 고객 사이의 왕복 거래다.
- **멀티플 압축** 기관이 프런티어 연구소 주식을 담으려고 기존 소프트웨어와 네오클라우드를 판다. 코어위브는 공모가 부근까지 내렸다.
- **스타트업 구축** 상위 34개 AI 스타트업 합산 매출의 89%를 두 회사가 차지해 후발 주자의 상장 창구가 막힌다.
- **정치적 반발** 11월 3일 중간선거에서 데이터센터 안건 41건 중 40건이 부결됐다. 계획된 데이터센터의 10% 이상이 취소 위기다.

빅테크 현금흐름 적자 전환과 1.5조 달러 자금 갭



-424억
달러. 아마존 3분기 FCF. 알파벳 -91억, 메타 -21억

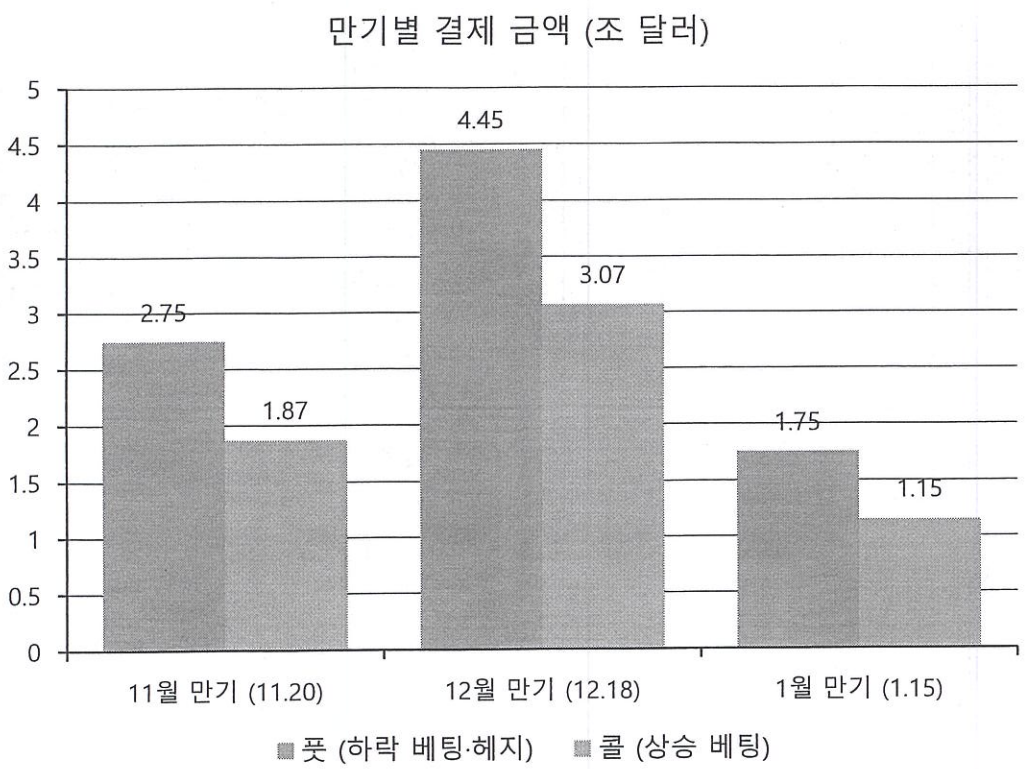
1.5조
달러. 2년 설비투자 2조 대 보유 현금 6,000억의 자금 갭

3,774억
달러. 버크셔의 현금·단기국채. 14분기째 순매도

- 5대 하이퍼스케일러의 합산 설비투자가 2026년 3분기에 처음으로 영업현금흐름을 넘었다.
- 2026년 가이던스는 연초 6,600억~6,900억 달러에서 하반기 7,360억~9,000억 달러로 올랐다.
- 갭을 메우려고 올해 주식과 전환사채 3,000억 달러 이상, 회사채 1.88조 달러가 발행됐다.
- 나스닥의 15거래일 조기 편입 규칙은 메가 IPO 직후 패시브 펀드가 기존 M7을 팔게 만든다.

출처: 자료 5.1, 5.4절. 차트는 최근 4개 분기 설비투자 (억 달러)

옵션 시장: 하락 풋 12.60조 대 상승 콜 8.79조 달러



- **총액** 21.38조 달러. 풋이 58.9%로 콜보다 3.81조 달러 많다.
- **기관** S&P 500 지수 옵션의 풋/콜 비율 1.43, SPY는 1.88. 하방 헤지가 밀집했다.
- **개인** 3배 레버리지 ETF(TQQQ)는 0.66, 라이엇은 0.38. 콜 매수가 우세하다.
- **만기 집중** 풋의 68%인 8.57조 달러가 2027년 1월까지 만기를 맞는다.
- **양면성** 풋이 내가격이 되면 딜러 매도가 폭락을 가속한다. 지수가 7,300선을 지키면 풋이 소멸하면서 환매수가 상승 연료가 된다.

출처: 자료 5.2절 (2026년 11월 6일 마감 기준, 4대 지수·ETF 옵션 미결제약정 명목금액)

11월 5일 급락: 기계적 연쇄 매도 4단계

-4.36%
S&P 500 하루 낙폭. 나스닥 100은 -6.36%

8배
당일 옵션이 반영한 기대 변동폭 0.55% 대비 실제 낙폭

-1,001억 달러
음(-)으로 반전된 딜러 감마 익스포저

1 뉴스 스캐너
AI 뉴스 스캐너가 데이터센터 안건 부결 소식을 읽고 마이크로초 단위로 선물을 매도

2 0DTE 풋 월 붕괴
당일만기 풋이 연쇄 붕괴. 딜러 헤지 알고리즘이 선물을 추격 매도

3 CTA 청산
추세추종 알고리즘이 포지션을 -3.3σ까지 자동으로 축소

4 레버리지 ETF
오후 3시 30분 이후 2,180억 달러 규모의 리밸런싱 매물 출회

- 자료의 해석: SA가 경고한 '통제를 벗어난 자동화 시스템의 연쇄 폭주'가 군사 영역보다 금융시장에서 먼저 나타났다.
- 근거로 든 연구: 비슷한 데이터로 학습한 AI 에이전트는 같은 신호에 군집 행동이 5~10% 늘어난다.

출처: 자료 5.3절

향후 1년: 조정 3경로와 연착륙 5대 조건

조정 시나리오의 세 가지 시간 경로

- 조기 발작 (2026년 11~12월, 25~30%): Anthropic 상장 수요 부진과 11·12월 풋 7.20조 달러의 내가격화. S&P 500 6,500~6,800.
- 2027년 1분기 구조적 조정 (40~45%): 자사주 매입 공백기에 설비투자 감축 딜레마, 데이터센터 모라토리엄, 코어위브 부채 만기가 겹침. -15~-25%.
- 2027년 중반 지연 조정 (20~25%): 5월 Anthropic 락업 해제와 OpenAI 상장, 단기 연산 약정 결제가 겹침

연착륙과 강세 (S&P 500 8,000~8,400)의 조건

- 풋 8.57조 달러 소멸에 따른 딜러 환매수와 자사주 매입의 결합
- 빅테크의 자율적 설비투자 속도 조절과 추론 효율화
- 데이터센터가 전력 비용을 100% 자체 부담하는 주정부와의 타협
- 연준·재무부의 유동성 완충과 10년물 금리 4.2~4.5% 이하 안정
- 메가 IPO의 공모가 현실화와 락업 물량 분산

출처: 자료 5.5절. 확률은 자료 저자의 추정이며 투자 조언이 아님