문명의 변곡점에서



김경진 변호사

목차

서장. 수학자보다 먼저 푸는 기계

제1부. 교육의 목표를 다시 세운다: 인공지능과 협업하고 인공지능을 지휘하는 사람

1장. 교육은 무엇을 길러야 하는가
2장. 협업 능력: 인공지능과 함께 일하는 법
3장. 지휘 능력: 오케스트레이션
4장. 목표가 바뀌면 평가가 바뀐다
5장. 협업과 지휘를 가르치는 학교

제2부. 역할의 재정의: 인간은 무엇을 할 것인가

6장. 배관공의 자리도 오래 가지 않는다
7장. 인간 노동의 근본적 변화
8장. 백지의 영역
9장. 담대한 목표
10장. 국가, 기관, 기업의 목표

제3부. 통제의 조건: 위험에 대비한다

11장. 경고하는 사람들과 반론
12장. 사고는 이미 현실이 됐다
13장. 모델을 만든 회사가 스스로 사고를 공개하는 이유
14장. 한국의 인공지능안전연구소
15장. 사고 대응 체계의 설계

제4부. 연결의 수단: 뇌와 컴퓨터를 잇는다

16장. 생각으로 커서를 움직이다
17장. 미국과 중국의 경쟁
18장. 양방향 통신까지 10년
19장. 휴대전화와 컴퓨터를 대신하는 장치
20장. 한국의 선택
종장. 변곡점은 되돌아오지 않는다

부록

부록 1. 용어 설명
부록 2. 연표: 2022년 11월 챗GPT 공개부터 2026년 10월까지
부록 3. 인공지능 기본법 제12조 전문과 관련 조문
부록 4. 주요 국가 인공지능안전연구소 비교표
부록 5. 세계 BCI 임상 현황표

서장. 수학자보다 먼저 푸는 기계

2026년 1월, 테렌스 타오의 발표

2026년 1월 4일 케빈 바레토(Kevin Barreto)는 에르되시 문제 사이트(erdosproblems.com)의 게시판에 증명 하나를 올렸다. 증명의 뼈대를 쓴 것은 사람이 아니고 인공지능(AI)이었다. 인공지능은 사람이 하던 추론과 판단을 컴퓨터가 하도록 만든 기술이다. 바레토가 쓴 도구는 오픈AI의 언어 모델 GPT 5.2 Pro였다. 언어 모델은 사람이 쓴 방대한 글을 읽고 다음에 올 말을 예측하도록 훈련한 인공지능이다. 하모닉(Harmonic)이라는 회사의 증명 도구 Aristotle은 그 증명을 Lean으로 옮겨 검사했다. Lean은 증명의 각 단계가 맞는지 컴퓨터가 확인하도록 만든 프로그래밍 언어다. 이렇게 컴퓨터가 검사할 수 있는 꼴로 쓴 증명을 형식 증명이라고 부른다.

첫 판본은 문제의 한 가지 해석만 풀었다. 게시판 참여자들이 1월 5일 문제의 본뜻을 다시 정리하자 바레토는 GPT 5.2 Pro에 논증을 넓힐 수 있는지 물었다. 1월 6일 Aristotle은 그 답을 바탕으로 형식 증명을 완성했다. 바레토는 뒤에 쓴 글에서 Lean 검사를 거치지 않은 인공지능 증명은 올리지 않는다고 밝혔다(바레토, 2026년 1월 26일). 함께 일한 리엄 프라이스(Liam Price)는 한 모델이 낸 풀이를 새로 띄운 다른 모델에게 검사시키는 방식을 썼다(Quanta, 2026년 8월 3일). 두 사람 모두 20대 초반이었다.

로스앤젤레스 시각으로 1월 7일 오후, 한국 시각으로 8일 새벽 6시 3분에 테렌스 타오가 이 소식을 알렸다(UTC 기준 1월 7일 21시 3분). 타오는 미국 UCLA 수학과 교수이고 필즈상을 받은 수학자다. 필즈상은 국제수학자대회가 4년마다 40세 이하 수학자에게 주는 상이다. 그는 수학자들이 모이는 마스토돈 서버(mathstodon.xyz)에 다섯 개의 글을 잇달아 올렸다. 마스토돈은 트위터와 비슷한 분산형 소셜 미디어다. 첫 글은 인공지능이 에르되시 문제 728번을 사실상 자율적으로 풀었다고 알렸다("more or less autonomously")(타오 마스토돈, 2026년 1월 7일).

폴 에르되시는 평생 수많은 미해결 문제를 남긴 20세기 헝가리 수학자다. 728번은 1975년 에르되시와 그레이엄, 루자, 스트라우스가 낸 정수론 문제다. n의 계승은 1부터 n까지의 정수를 모두 곱한 수다. 문제는 a의 계승과 b의 계승을 곱한 값이 n의 계승을 나눌 때 a와 b의 합이 n보다 얼마나 클 수 있는지를 다룬다. 에르되시는 1968년에 그

차이가 log n 정도를 넘지 못함을 보였다. 1975년에는 그 차이가 실제로 log n만큼 벌어지는 경우가 무한히 많은지를 물었다.

타오는 첫 글부터 단서를 달았다. 에르되시가 처음 제시한 문제는 조건이 정확하게 정식화되지 않았고, 이번에 풀린 것은 공동체가 문제의 취지에 맞게 다듬은 판본이라고 적었다("in the spirit of the problem"). 사람이 한 일은 첫 시도에 의견을 준 정도였다. 비슷한 기법은 문헌에 있지만 이 결과 자체는 없어 보인다고도 썼다. 그는 풀이 자체보다 인공지능이 풀이의 설명문을 빠르게 작성하고 수정하는 능력에 주목했다. 사람은 문제를 고르고 질문을 던진 뒤 결과를 전달했고, 수학적 증명과 검증은 두 인공지능이 수행했다.

석 달쯤 앞서 비슷하게 들리는 소동이 있었다. 2025년 10월 18일 오픈AI의 케빈 와일은 GPT 5가 미해결 에르되시 문제 10개를 풀었다고 X에 썼다. 사이트 운영자인 수학자 토머스 블룸은 이를 극적인 왜곡이라고 반박했다("a dramatic misrepresentation") (GIGAZINE, 2025년 10월 20일). 그 모델은 블룸이 미처 알지 못한 기존 논문을 찾아냈을 뿐이었고, 와일은 글을 지웠다. 10월의 일은 사람이 모르던 옛 논문을 찾은 일이었다. 1월의 일은 문헌에 없던 증명을 쓴 일이었다.

1월의 결과가 부풀려졌다는 지적도 있다. 그 선을 먼저 그은 사람은 타오 자신이다. 그는 1월 10일 자기 글이 맥락 없이 인용된다고 덧붙였다("quoted out of context"). 1월 14일에는 사람의 개입이 적은 풀이일수록 짧고 표준 기법을 따르는 경향이 있다고 적었다. 테크 매체 디코더는 타오가 지금의 인공지능으로 풀 만한 에르되시 문제를 열린 문제의 1에서 2퍼센트로 추정했다고 전했다(The Decoder, 2026년 1월 16일). 3월 5일 그는 인공지능의 방법이 수론학자 칼 포머런스의 2014년 논문과 많이 닮았다는 수정을 달았다. 그래도 728번을 명시적으로 다룬 첫 풀이라는 평가는 거두지 않았다. 우리는 이 비판이 결과의 크기를 바로잡는다고 본다. 728번은 80년 묵은 난제가 아니고 50년 된 변두리 문제였다. 넉 달 뒤 인공지능은 변두리가 아닌 문제를 건드렸다.

5월, 평면 단위 거리 문제에서 뒤집힌 정설

2026년 5월 20일(미국 현지) 오픈AI는 자사 모델이 이산기하학의 중심 추측을 반증했다는 글을 올렸다(오픈AI, 2026년 5월 20일). 추측은 아직 증명되지 않았지만 참으로 여겨지는 수학적 주장이다. 반증은 그 주장이 틀렸음을 보이는 일이고, 대개 반례 하나로 충분하다.

문제는 1946년 에르되시가 냈다. 평면에 점 n개를 찍을 때 거리가 정확히 1인 점의 쌍이 최대 몇 개인지를 묻는다. 브라스와 모서와 파흐는 2005년 책에서 이 문제를

조합기하학에서 아마 제일 널리 알려진 문제라고 불렀다("possibly the best known"). 점을 한 줄로 늘어놓으면 쌍은 n보다 하나 적다. 정사각 격자는 2n개쯤을 준다. 에르되시는 격자의 크기를 잘 맞춰 점의 수보다 조금 빠르게 쌍이 느는 배열을 찾았다. 수학자들은 수십 년 동안 이 배열이 사실상 최선이라고 믿었다. 에르되시는 쌍의 수가 n의 1+o(1)제곱을 넘지 못한다고 추측했다. o(1)은 n이 커질수록 0에 다가가는 값이다.

오픈AI의 내부 모델은 이 믿음을 깨는 점 배열을 무한히 많이 만들었다. 어떤 양수 δ를 잡으면 무한히 많은 n에서 단위 거리 쌍이 n의 $1+\delta$제곱개 이상 나온다. 프린스턴대 수학과 교수 윌 소인은 같은 날 낸 논문에서 δ를 0.014로 잡을 수 있음을 보였다. 열쇠는 기하학이 아닌 대수적 수론에서 왔다. 모델은 에르되시가 쓴 가우스 정수를 대칭이 더 풍부한 수 체계로 바꿔 단위 거리를 더 많이 만들었다. 오픈AI는 이 모델이 수학 전용이 아닌 범용 추론 모델이라고 밝혔고 이름은 공개하지 않았다. 추론 모델은 답을 내기 전에 여러 단계로 생각을 펼치도록 훈련한 언어 모델이다. 회사는 에르되시 문제 모음으로 모델을 시험하던 중에 이 증명이 나왔다고 적었다.

뒤집힌 것은 정설이고 문제는 아직 열려 있다. 쌍의 최댓값은 2026년 5월 기준으로 아래로는 n의 1.014제곱, 위로는 n의 4/3제곱(약 1.333제곱) 사이에 있다. 위쪽 한계는 1984년 스펜서와 세메레디와 트로터가 보인 뒤로 그대로다. 결과물도 학술지 논문이 아니다. 오픈AI의 발표 글과 외부 수학자 아홉 명의 동반 논문, 소인의 독립 논문이 같은 날 나왔다. 두 논문은 논문 사전 공개 서버인 arXiv에 올라왔다. 동반 논문의 저자이자 필즈상 수상자인 티머시 고워스는 이 결과를 인공지능 수학의 이정표라고 불렀다("a milestone in AI mathematics").

사람도 할 수 있었다는 반론이 곧 나왔다. 하버드대 수학자 멜러니 매챗 우드는 전문가들이 추측을 참으로 믿었기에 진전이 막혔다고 봤다. 같은 시간을 반례 찾기에 썼다면 찾았을 것이라는 말이다(Scientific American, 2026년 5월 21일). 모내시대 수학자 멜리사 리도 전문가가 힌트 몇 개만 받았으면 같은 증명에 이르렀으리라고 썼다(The Conversation, 2026년 5월 26일). 반대편 증언도 있다. 2026년 필즈상을 받은 제이컵 치머먼(토론토대)은 비슷한 접근을 잠깐 생각했다가 접었다. 시간이 많이 들고 자주 실패하는 방법이었기 때문이다(Understanding AI, 2026년 5월 28일). 프린스턴대의 노가 알론은 조합기하학 연구자라면 누구나 이 문제를 생각해 봤다고 적었다. 사람이 할 수 있었다는 말과 사람이 하지 않았다는 말은 함께 맞다. 해설 매체 언더스탠딩 AI는 성공 확률이 낮은 긴 탐색을 적은 비용으로 되풀이하는 일이 인공지능의 강점이라고 짚었다. 우리는 사람과 기계를 가른 것이 재능보다 시도의 비용이었다고 판단한다.

수학계 바깥이 조용했던 이유

고워스는 소식을 처음 듣고 인공지능이 추측을 증명한 줄 알았다. 그는 그날 저녁 내내 그 일이 수학자에게 무슨 뜻일지 곱씹었다. 다음 날 아침 외부 검토자들에게 온 전자우편을 읽고서야 그것이 반증이라는 것을 알았고 안도했다(Understanding AI, 2026년 5월 28일). 필즈상 수상자에게도 결과의 뜻은 한 번에 전해지지 않았다.

수학계 바깥은 조용했다. 우리가 검색한 범위에서 1월 사건을 다룬 곳은 기술 매체와 개발자 게시판, 수학자들의 소셜 미디어였다. 국내 주요 일간지의 1월 보도는 찾지 못했다. 5월 사건은 사이언티픽 아메리칸 같은 과학 매체와 몇몇 국내 신문이 다뤘다. 이 관찰은 보도 건수를 센 결과가 아니고 검색 범위 안에서만 성립한다. 사이언티픽 아메리칸은 사람이 혼자 냈어도 큰 뉴스가 됐을 결과라고 썼다. 이 조용함을 정면으로 분석한 전문가의 글은 2026년 10월까지 찾지 못했다. 아래 해석은 간접 근거에서 끌어낸 우리의 추정이다.

먼저 두 사건에는 승부가 없었다. 2016년 3월 서울에서 알파고는 이세돌 9단을 4대 1로 이겼다. 바둑을 모르는 사람도 누가 이겼는지 알았다. 2022년 말 나온 챗GPT는 누구나 브라우저에서 직접 써 볼 수 있었다. 1월과 5월에는 진 사람도 없었고 손에 쥘 물건도 없었다.

결과를 확인할 수 있는 사람도 적었다. 단위 거리 반증을 검토해 동반 논문을 쓴 외부 수학자는 아홉 명이었다. 사이언티픽 아메리칸에 따르면 외부 전문가들이 본 것도 모델의 원래 출력이 아니고 편집한 사고 과정 요약이었다. 728번의 증명을 따라가려면 계승과 이항계수의 성질을 알아야 했고, 단위 거리 반증을 따라가려면 대수적 수론을 알아야 했다.

결과를 한 문장으로 옮기기도 어려웠다. "인공지능이 난제를 풀었다"는 문장은 틀렸다. "인공지능이 추측을 반증했다"는 문장은 맞지만 듣는 사람에게 무엇이 바뀌었는지 알려 주지 않는다. 전문가들 사이에서도 이 결과는 충격보다 예상된 다음 걸음으로 읽혔다. 언더스탠딩 AI는 이 결과를 단절이 아닌 기존 추세의 연장으로 봤다.

조용했다고 해서 사건이 작았던 것은 아니다. 승부도 제품도 없는 변화는 신문 1면보다 연구실의 작업 방식을 먼저 바꾼다. 바깥에서 소리가 날 즈음이면 안쪽의 일하는 방식은 이미 달라져 있다고 우리는 본다.

연구의 주체가 바뀐다는 말의 뜻

오픈AI는 5월 글 끝에 인공지능이 연구의 창조적 부분에서 아주 진지한 역할을 맡기 시작한다고 썼다("a very serious role"). 바로 다음 문단은 그 미래가 여전히 사람의 판단에 달렸다고 적었다("That future still depends on human judgment"). 문제를 고르고 결과를 해석하고 다음 질문을 정하는 일은 사람이 한다는 것이다. 한 회사의 한 글 안에 두 문장이 나란히 놓여 있다.

연구의 주체라는 말은 세 층으로 나눠 볼 수 있다. 맨 아래 층은 증명이나 계산을 실제로 만들어 내는 일이다. 728번과 단위 거리 문제에서 이 층은 이미 기계가 맡았다. 가운데 층은 결과에 이름을 걸고 책임지는 일이다. 728번의 Lean 증명을 사람이 읽을 글로 정리한 논문은 수학자 냇 소타나판의 이름으로 arXiv에 올라왔다(arXiv, 2026년 1월 12일). 학술지들은 책임질 수 없는 인공지능은 저자가 될 수 없다는 원칙을 세웠다. 맨 위 층은 어떤 문제를 풀지 고르고 결과의 학술적 의미를 평가하는 일이다. 타오는 1월 다섯째 글에서 최종 논문의 핵심은 사람이 쓰는 편이 낫고 형식 증명은 인공지능과 Lean에 맡길 수 있다고 했다.

주체는 바뀌지 않는다는 시각도 뚜렷하다. Jun-Yong Park은 2026년 7월 arXiv에 올린 글 「이해 없는 자동화(Automation Without Understanding)」에서 수학적 추론을 검사하고 해석하는 능력을 기반시설에 견줬다. 그 능력은 여러 세대에 걸쳐 기관이 쌓은 것이라 필요할 때 다시 만들 수 없다는 것이다. 타오는 2026년 9월 사이언티픽 아메리칸 인터뷰에서 어떤 인공지능 사용을 받아들일지 수학자가 규칙을 정해야 한다고 말했다("set the rules on what types of AI use are acceptable"). 그는 인공지능 모델을 증명의 저자로 올리지 말자는 규범을 담은 라이덴 선언을 높이 평가했다.

우리는 이 시각에 동의한다. 다만 맨 아래 층이 바뀌면 위의 두 층에서 사람이 하는 일도 달라진다. 증명이 귀하던 시절에는 증명을 만든 사람이 곧 그것을 검증하고 의미를 평가하는 사람이었다. 증명을 기계가 쏟아 내면 사람의 일은 문제를 고르고 결과를 확인하고 책임지는 쪽으로 옮겨 간다. 연구의 주체가 바뀐다는 말은 이 이동을 가리킨다.

다른 분야의 연구자들도 비슷한 구분을 쓴다. 에이전트는 사람의 지시를 받아 여러 단계의 일을 스스로 수행하는 인공지능 프로그램이다. 펑(Feng) 등 연구진은 2025년 6월 논문에서 에이전트의 자율성을 기계의 능력이 아닌 사용자가 맡는 역할로 정의했다(arXiv, 2025년 6월). 사람은 조작자에서 협력자, 자문자, 승인자를 거쳐 관찰자로 물러난다. 오픈AI가 그리는 다음 단계는 관찰자 쪽에 가깝다. MIT 테크놀로지 리뷰에 따르면 회사는 2026년 9월까지 자율적인 인공지능 연구 인턴을 만들고 2028년에 완전

자동 다중 에이전트 연구 체계를 내놓을 계획이다(MIT 테크놀로지 리뷰, 2026년 3월 20일). 수석 과학자 야쿠프 파호츠키는 이를 데이터센터 안에 연구소 하나를 통째로 두는 일에 견줬다("a whole research lab in a data center").

기계가 연구의 더 많은 단계를 맡더라도 무엇을 위해 연구하고 어떤 결과를 받아들일지는 사회가 정해야 한다. 따라서 준비할 것은 도구를 더 잘 쓰는 방법만이 아니다. 사람에게 방향 설정과 검증의 권한을 남기고, 그 권한을 실제로 행사할 능력을 교육과 제도로 길러야 한다.

챗GPT 이후 3년: 대화 상대에서 연구 동료로, 다시 에이전트로

2022년 11월 30일 오픈AI는 챗GPT를 공개했다. 회사는 이 모델을 후속 질문에 답하고 실수를 인정하는 대화형 모델로 소개했다(오픈AI, 2022년 11월 30일). 사람들은 질문을 던지고 답을 읽었다. 이때 인공지능은 대화 상대였다. 타오가 728번 소식을 올린 날은 챗GPT 공개 뒤 3년 한 달째였다.

수학 쪽 변화는 시험에서 먼저 드러났다. 2024년 7월 25일 구글 딥마인드는 AlphaProof와 AlphaGeometry 2가 국제수학올림피아드(IMO) 문제를 은메달 수준으로 풀었다고 발표했다. IMO는 각국 고등학생이 여섯 문제를 푸는 수학 대회다. 같은 해 9월 12일 오픈AI는 답하기 전에 더 오래 생각하도록 만든 o1-preview를 내놓았다. 이 모델이 추론 모델의 출발점이다. 2025년 7월에는 오픈AI의 실험 모델과 구글 딥마인드의 제미나이 딥 싱크가 IMO 금메달 수준의 점수를 냈다.

같은 무렵 인공지능을 부르는 이름이 바뀌었다. 앤트로픽 최고경영자 다리오 아모데이는 2024년 10월 에세이에서 데이터센터 안의 천재들의 나라를 그렸다("country of geniuses in a datacenter"). 구글은 2025년 2월 19일 과학자를 돕는 다중 에이전트 체계를 AI co-scientist라고 이름 붙였다. 구글은 이를 가상의 과학 협력자라고 불렀다("a virtual scientific collaborator"). 오픈AI도 2026년 5월 글에서 더 강한 연구 동반자를 말했다("a stronger research partner"). 대화 상대가 동료라는 이름을 얻었다.

에이전트라는 말은 그보다 조금 앞서 퍼졌다. 오픈AI 최고경영자 샘 올트먼은 2025년 1월 블로그 글 「Reflections」에서 그해 첫 인공지능 에이전트가 일터에 합류할 수 있다고 썼다("join the workforce"). 2025년 12월 11일 GPT 5.2가 나왔고, 그 상위판이 한 달 뒤 728번의 증명을 썼다. 2026년 9월 8일 오픈AI는 내부 모델로 돌린 약 1만 개의 에이전트가 88시간 만에 나비에 스토크스 방정식에 관한 답에 이르렀다고 발표했다(오픈AI, 2026년 9월 8일). 이 주장의 독립 검증은 2026년 10월 14일 기준으로 끝나지

않았다. 10월 6일에는 내부 모델이 낸 수학 결과를 GitHub 저장소에 한꺼번에 공개했다(오픈AI, 2026년 10월 6일).

세 단계는 앞 단계를 지우고 오지 않았다. 지금도 사람들은 챗GPT와 대화하고, 같은 회사의 모델은 다른 곳에서 에이전트 수천 개로 나뉘어 일한다. 달라진 것은 사람이 맡는 몫이다. 대화 상대 앞에서 사람은 답을 읽었고, 동료 곁에서는 답을 같이 다듬었다. 에이전트에게는 일을 맡기고 결과를 받는다. 3년 사이 사람의 자리는 답을 쓰는 쪽에서 일을 맡기고 확인하는 쪽으로 옮겨 갔다.

이 책이 묻는 네 가지 질문

인공지능이 없던 사회로 돌아갈 길은 없다. 그래서 우리는 사회의 구조를 인간과 인공지능의 협업을 전제로 다시 짜야 한다고 본다. 이 판단에서 네 질문이 나오고, 질문 사이에는 순서가 있다.

첫 질문은 교육의 목표다. 기계가 시험에서 사람을 앞서는 시대에 학교가 기를 것은 인공지능과 협업하는 능력과 인공지능을 지휘하는 능력이다. 평가의 기준은 이 목표를 따라 바뀐다. 1부는 2026년 7월 상하이의 IMO 시험장에서 인공지능 모델이 42점 만점을 받았다는 발표로 문을 연다.

둘째 질문은 역할의 재정의다. 지식 노동과 사무와 몸으로 하는 일이 차례로 기계에 넘어갈 때 사람에게 남는 일은 방향을 정하는 일이다. 2부는 2025년 8월 베이징 국가스피드스케이팅경기장에서 돌려차기를 헛차고 넘어진 휴머노이드 로봇에서 출발한다.

셋째 질문은 통제의 조건이다. 협업의 상대가 통제를 벗어나면 협업은 성립하지 않는다. 3부는 2026년 9월 미국 의회의 비공개 브리핑을 마친 제프리 힌턴이 규제할 시간이 1년쯤 남았다고 말하는 장면에서 시작한다.

넷째 질문은 연결의 수단이다. 인공지능의 처리 속도와 작업 범위가 커질수록 사람이 의도를 전달하고 결과를 확인하는 방식도 발전해야 한다. 뇌와 컴퓨터를 잇는 장치(뇌-컴퓨터 인터페이스, BCI)는 그 연결을 넓힐 후보 기술이다. 4부는 손을 쓰지 않고 커서로 온라인 체스를 두던 놀런드 아보의 2024년 영상으로 들어간다.

앞의 세 질문은 제도와 사람의 문제이고 넷째는 기술의 문제다. 이 책은 검사, 국회의원, 변호사, 인공지능 회사 대표, 대학 강사로 일하며 겪은 일에서 나왔다. 책의 성격은 정책 제안서와 교양서의 중간이고, 각 부의 마지막 장에는 한국이 할 일을 적었다. 정책

담당자와 대학과 연구기관의 연구자, 기업의 경영진, 인공지능 정책이 궁금한 교양 독자를 함께 염두에 두었다. 타오의 글타래에서 시작된 변화가 맨 먼저 닿는 곳은 교실이라고 우리는 판단한다.

참고 자료

erdosproblems.com(Kevin Barreto), Problem 728 and the use of AI on Erdős problems, 2026년 1월 26일, https://www.erdosproblems.com/forum/thread/blog:2

erdosproblems.com, 728 | Erdős Problems, 2026년 10월 14일 열람, https://www.erdosproblems.com/728

Quanta Magazine, Why the Legendary Erdős Problems Are Falling to AI, 2026년 8월 3일, https://www.quantamagazine.org/why-the-legendary-erdos-problems-are-falling-to-ai-20260803/

타오 마스토돈(mathstodon.xyz), 에르되시 문제 728번 글타래 첫 글(제목 없는 게시물), 2026년 1월 7일, https://mathstodon.xyz/@tao/115855840223258103

타오 마스토돈(mathstodon.xyz), 에르되시 문제 728번 글타래 다섯째 글(제목 없는 게시물), 2026년 1월 7일, https://mathstodon.xyz/@tao/115855852706322322

타오 마스토돈(mathstodon.xyz), 1월 10일 덧붙임과 3월 5일 수정(제목 없는 게시물), 2026년 1월 10일, https://mathstodon.xyz/@tao/115871649394962391

타오 마스토돈(mathstodon.xyz), AI 개입 정도와 풀이 난이도에 관한 글(제목 없는 게시물), 2026년 1월 14일, https://mathstodon.xyz/@tao/115891256726420022

GIGAZINE, An OpenAI researcher posted that 'GPT-5 has solved an unsolved mathematical problem,' but it turned out that the problem ...(제목 뒷부분 미확인), 2025년 10월 20일, https://gigazine.net/gsc_news/en/20251020-openai-researcher-announced-gpt-5-math-breakthrough/

The Decoder, Terence Tao says GPT-5.2 Pro cracked an Erdős problem, but warns the win says more about speed than difficulty, 2026년 1월 16일, https://the-decoder.com/terence-tao-says-gpt-5-2-pro-cracked-an-erdos-problem-but-warns-the-win-says-more-about-speed-than-difficulty/

Carl Pomerance, A remark on the middle binomial coefficient, 날짜 미표기(2026년 3월 5일 이전), https://math.dartmouth.edu/~carlp/binom.pdf

오픈AI, An OpenAI model has disproved a central conjecture in discrete geometry, 2026년 5월 20일, https://openai.com/index/model-disproves-discrete-geometry-conjecture/

arXiv(Alon, Bloom, Gowers, Litt, Sawin, Shankar, Tsimerman, Wang, Wood), Remarks on the disproof of the unit distance conjecture, 2026년 5월 20일, https://arxiv.org/abs/2605.20695

arXiv(Will Sawin), An explicit lower bound for the unit distance problem, 2026년 5월 20일, https://arxiv.org/abs/2605.20579

Scientific American, AI just solved an 80-year-old 'Erdős problem,' and mathematicians are amazed, 2026년 5월 21일, https://www.scientificamerican.com/article/ai-just-solved-an-80-year-old-erdos-problem-and-mathematicians-are-amazed/

The Conversation(Melissa Lee), An AI solution to an 80-year-old problem has shocked mathematicians, 2026년 5월 26일, https://theconversation.com/an-ai-solution-to-an-80-year-old-problem-has-shocked-mathematicians-283686

Understanding AI(Kai Williams), OpenAI's math breakthrough played to AI's strengths, 2026년 5월 28일, https://www.understandingai.org/p/openais-milestone-math-breakthrough

arXiv(Nat Sothanaphan), Resolution of Erdős Problem #728: a writeup of Aristotle's Lean proof, 2026년 1월 12일, https://arxiv.org/abs/2601.07421

arXiv(Jun-Yong Park), Automation Without Understanding, 2026년 7월 7일, https://arxiv.org/abs/2607.06377

Scientific American(Joseph Howlett), If AI can do math, what's the point of mathematicians?, 2026년 9월 10일, https://www.scientificamerican.com/article/mathematicians-confront-the-ai-apocalypse/

arXiv(Feng 외), Levels of Autonomy for AI Agents, 2025년 6월, https://arxiv.org/abs/2506.12469

MIT 테크놀로지 리뷰(Will Douglas Heaven), OpenAI is throwing everything into building a fully automated researcher, 2026년 3월 20일, https://www.technologyreview.com/2026/03/20/1134438/openai-is-throwing-everything-into-building-a-fully-automated-researcher/

오픈AI, Introducing ChatGPT, 2022년 11월 30일, https://openai.com/index/chatgpt/

구글 딥마인드, AI achieves silver-medal standard solving International Mathematical Olympiad problems, 2024년 7월 25일, https://deepmind.google/blog/ai-solves-imo-problems-at-silver-medal-level/

오픈AI, Introducing OpenAI o1-preview, 2024년 9월 12일, https://openai.com/index/introducing-openai-o1-preview/

다리오 아모데이, Machines of Loving Grace: How AI Could Transform the World for the Better, 2024년 10월, https://darioamodei.com/essay/machines-of-loving-grace

구글 리서치, Accelerating scientific breakthroughs with an AI co-scientist, 2025년 2월 19일, https://research.google/blog/accelerating-scientific-breakthroughs-with-an-ai-co-scientist/

샘 올트먼 블로그, Reflections, 2025년 1월(게시일 표기 없음), https://blog.samaltman.com/reflections

오픈AI, On the Navier–Stokes Millennium Prize Problem, 2026년 9월 8일, https://openai.com/index/navier-stokes-solution/

오픈AI, Sharing AI progress in mathematics, 2026년 10월 6일, https://openai.com/index/sharing-ai-progress-in-mathematics/



제1부. 교육의 목표를 다시 세운다: 인공지능과 협업하고 인공지능을 지휘하는 사람

1장. 교육은 무엇을 길러야 하는가

지식을 전달하는 교육의 전제가 무너진 지점

2026년 7월 15일과 16일 상하이에서 각국 고등학생이 이틀 동안 여섯 문제를 푸는 국제수학올림피아드(IMO) 시험이 열렸다. 이해에 참가한 학생은 666명이었고 42점 만점을 받은 학생은 7명이었다(France24/AFP, 2026년 7월 23일). 같은 문제를 인공지능도 풀었다. 화웨이의 셀리아(Celia)와 샤오홍수(RedNote)의 모델이 42점을 받았다고 두 회사가 발표했다. 여러 매체는 주최 측이 이 답안을 채점했다고 보도했다. IMO가 직접 낸 공식 성명은 2026년 10월 14일까지 찾지 못했다.

한 해 전의 사정은 달랐다. 2025년 7월 구글 딥마인드의 제미나이 딥 싱크(Gemini Deep Think) 고급판은 여섯 문제 중 다섯 문제를 풀어 35점을 받았다(구글 딥마인드, 2025년 7월 21일). IMO 조정위원이 학생 답안과 같은 기준으로 채점했고 그 점수는 금메달 기준에 들었다. 1년 사이에 기계의 성적은 금메달 수준에서 만점 수준으로 올라섰다.

한국의 시험에서도 비슷한 기록이 쌓인다. 한 대학생이 2026학년도 수능 문제를 여러 언어 모델에게 풀게 한 개인 프로젝트에서 2026년 9월 오픈AI의 새 모델이 450점 만점을 기록했다(Korea Times, 2026년 9월 7일). 문제마다 한 번씩만 풀게 한 비공식 실험이다(AI타임스, 2026년 3월 8일). 법률 AI 회사 로앤컴퍼니는 자사 서비스 슈퍼로이어가 2026년 1월 치른 제15회 변호사시험 선택형 150문항을 모두 맞혔다고 발표했다(국민일보, 2026년 6월 24일). 외부 기관이 따로 검증한 결과는 아니다.

학생들은 이미 이 도구로 공부한다. 미국 퓨리서치센터가 2025년 가을 13세에서 17세 청소년 1,458명에게 물었더니 54퍼센트가 챗봇을 학교 과제에 써 봤다고 답했다(퓨리서치센터, 2026년 2월 24일). 한국에서는 2024년 조사에서 초중고 학생의 49.9퍼센트가 생성형 AI를 써 본 적이 있었다(한국교육개발원, 2026년 4월 22일). 경제협력개발기구(OECD)의 학업성취도 국제비교(PISA) 2025에서는 OECD 국가 학생의 46퍼센트가 한 주에 한 번 이상 AI 챗봇으로 공부했다(OECD, 2026년 9월).

대학의 수장들도 같은 진단을 내놓는다. 이근배 전남대 총장은 "과거의 대학이 거대한 '지식의 창고'였는데 이제 그 역할은 사실상 끝났다"고 말했다(서울신문, 2026년 8월 24일). 유지범 성균관대 총장은 "과거와 같은 지식 전달 방식의 교육은 효용을 잃었다"고 했다(서울경제, 2026년 10월 11일). 지식이 귀할 때 학교는 교사가 아는 내용을 학생에게 전달하는 데 큰 비중을 두었다. 기계가 높은 시험 성적을 내고 학생들이 그 기계로 과제를 하는 지금, 전달한 지식을 얼마나 재현하는지만으로 교육의 성과를 판단하기는 어렵다. 우리가 지식 전달 교육의 종말이라고 부르는 것은 그 기준의 한계다. 학교는 학생이 지식을 바탕으로 문제를 정하고 인공지능의 답을 판단하도록 교육의 목표를 바꿔야 한다.

세부 지식에서 인공지능과 인간의 격차: 기억력과 지식량

언어 모델이 읽은 글의 양은 토큰으로 센다. 토큰은 언어 모델이 글을 쪼개 처리하는 단위로 낱말 하나나 낱말의 조각에 해당한다. 메타는 2025년 4월 공개한 라마 4(Llama 4)를 30조 개가 넘는 토큰으로 훈련했다고 밝혔다(메타, 2025년 4월 5일). 한 추정은 1조 토큰을 책 약 400만 권 분량으로 본다(arXiv, 2024년 10월 22일). 책 한 권의 분량을 어떻게 잡느냐에 따라 이 환산은 달라진다.

사람의 기억 용량은 측정 방식에 따라 추정치가 크게 갈린다. 인지과학자 토머스 랜다우어는 사람이 장기기억에 정보를 저장하는 속도를 실험으로 측정했다. 그는 이를 바탕으로 사람이 평생 기억하는 정보량을 수백 메가바이트 정도로 추정했다(Cognitive Science, 1986년). 그 자신도 이 수치가 한 자릿수쯤 틀릴 수 있는 거친 추정이라고 적었다. 뉴런은 전기 신호를 주고받으며 정보를 전달하는 신경세포다. 노스웨스턴대 심리학자 폴 리버는 뉴런 연결의 수를 바탕으로 뇌가 담을 수 있는 용량을 약 2.5페타바이트로 계산했다(Scientific American, 2010년 5월 1일). 랜다우어는 사람이 실제로 기억해 쓰는 정보량을 측정했고 리버는 뇌의 신경망이 이론상 담을 수 있는 상한을 추산했다. 두 추정치의 큰 차이는 측정 대상의 차이에서 온다.

사람의 기억 용량과 언어 모델의 훈련자료량은 서로 다른 대상을 재므로 수치만으로 우열을 가릴 수 없다. 시험 성적에서는 변화가 드러난다. 국민일보 보도에 따르면 변호사시험 선택형에서 만점을 받은 사람은 공식으로 확인된 적이 없고 2026년 수석의 성적은 총점 375점 중 360점이었다(국민일보, 2026년 6월 24일). 총점에는 사례형과 기록형이 들어 있어서 기계의 선택형 만점과 바로 견줄 수는 없다. 선택형은 조문과 판례를 기억하고 주어진 사안에 맞춰 보는 능력을 묻는다. 회사가 발표한 만점 기록은 이 과제에서 기계가 높은 성능에 도달했음을 보여 준다. 이 성과만으로 법률 판단 전체의

우열을 가릴 수는 없지만, 판례 번호나 조문이나 화학식을 외워 재현하는 능력만으로 사람을 평가하는 교육은 바꿔야 한다.

그렇다고 기계의 지식을 그대로 믿을 수는 없다. 환각은 언어 모델이 사실이 아닌 내용을 그럴듯하게 지어내는 현상이다. 스탠퍼드대 인간중심AI연구소(HAI)는 AI 인덱스 2026에서 새 정확도 벤치마크로 상위 26개 모델을 시험했다(스탠퍼드 HAI, 2026년 4월 13일). 벤치마크는 성능을 비교하려고 만든 표준 시험이다. 환각률은 모델에 따라 22퍼센트에서 94퍼센트까지 갈렸다. 이 시험은 모르는 것을 모른다고 답하면 벌점을 주지 않도록 짰다. 같은 보고서는 IMO 금메달 수준에 이른 한 모델이 아날로그 시계는 절반쯤만 제대로 읽었다고 적었다.

기계의 능력은 넓지만 고르지 않다. 많은 정보를 기억해 재현하는 능력만으로 사람의 역량을 평가하기는 어려워졌다. 기계가 내놓은 답에서 맞는 것과 틀린 것을 가려내려면 그 분야의 지식이 필요하다. 교육은 지식을 익히는 일과 그 지식으로 기계의 답을 검증하는 일을 연결해야 한다.

큰 방향을 잡고 맥락을 읽고 목표를 정하는 일은 사람의 몫

서장에서 본 에르되시 문제 728번은 인공지능이 사람의 개입을 거의 받지 않고 푼 문제였다. 테렌스 타오는 이 결과를 알리면서 단서를 달았다. 에르되시 문제는 난도의 폭이 넓고 아무도 진지하게 들여다보지 않은 쉬운 문제가 많다는 것이다(The Decoder, 2026년 1월 16일). 그는 이 성과의 가치를 난도보다 증명을 작성하고 수정하는 속도에서 찾았다. 수많은 미해결 문제 가운데 무엇이 지금의 도구로 풀릴 만한지 가늠하고 결과의 학술적 의미를 평가한 사람은 타오였다.

서장에서 본 에이전트는 지시가 흐리면 되묻지 않고 움직이는 일이 많다. 2026년 2월 공개된 ClarEval 연구는 지시를 일부러 모호하게 만든 코딩 과제로 여러 에이전트를 시험했다(arXiv, 2026년 2월 27일). 코드를 잘 짜는 모델도 필요한 질문을 제때 던지지 못했다. 목표가 비어 있으면 기계는 멈추거나 엉뚱한 길로 간다.

와튼스쿨의 이선 몰릭 교수는 무언가를 만드는 일이 빠르고 싸질수록 고르는 능력이 귀해진다고 썼다. 그의 말로는 희소한 자원은 자기 취향으로 골라내는 능력이다("The scarce resource is your ability to select among stuff using your own taste.")(One Useful Thing, 2026년 9월 18일). 그는 어떤 프로젝트를 할지 고르고 인공지능이 틀린 곳을 알아보고 고쳐 달라고 요구하는 일은 여전히 사람이 했다고 적었다.

노동시장 자료도 같은 쪽을 가리킨다. 한국은행은 2025년 10월 보고서에서 경력이 짧은 청년층의 정형화된 업무를 인공지능이 먼저 대신한다고 분석했다(한국은행, 2025년 10월 30일). 경력에서 쌓은 암묵지와 사회적 기술이 필요한 업무에서는 인공지능이 사람을 돕는 쪽으로 쓰였다. 세계경제포럼이 2025년 1월 펴낸 일자리 전망 보고서에서도 고용주들이 꼽은 핵심 역량 1위는 분석적 사고였다(FM Magazine, 2025년 2월 18일).

목표를 정하는 일은 계산으로 끝나지 않는다. 누구의 사정을 먼저 살필지와 위험을 어디까지 받아들일지는 그 결과를 짊어질 사람이 정해야 한다. 맥락은 기록에 다 남지 않는다. 의뢰인이 말하지 않은 사정이나 조직 안의 오랜 관계는 그 자리에 있던 사람이 읽는다. 언어 모델은 사람이 글로 남긴 것에서 배운다. 글로 남지 않은 사정을 묻고 확인해 문제의 틀을 다시 잡는 일은 현장에 선 사람이 해야 한다. 틀이 잘못 잡히면 기계가 아무리 빨리 답해도 그 답은 엉뚱한 질문에 대한 답이 된다. 기계가 세부 지식에서 앞설수록 방향을 잡는 사람의 몫은 커진다.

새 교육 목표의 두 축: 협업 능력과 지휘 능력

사람이 맡을 일을 교육의 목표로 옮기면 협업 능력과 지휘 능력이 된다. 협업 능력은 인공지능에게 좋은 질문을 던지고 돌아온 결과에서 틀린 것을 걸러 내며 그 결과에 책임지는 능력이다. 지휘 능력은 한 분야에서 풀 문제를 찾아내고 여러 인공지능과 에이전트에게 일을 나누어 맡긴 뒤 결과를 조율해 합치는 능력이다. 이 일을 오케스트레이션(orchestration)이라고도 부른다. 협업이 인공지능 하나와 마주 앉아 일하는 법이라면 지휘는 여럿에게 일을 맡기고 방향을 정하는 법이다.

국제기구의 틀에도 비슷한 생각이 들어 있다. OECD는 2019년 학습 나침반 2030에서 학생 주도성과 함께 공동 주도성(co-agency)을 내세웠다(OECD, 2019년). 학생이 또래와 교사와 지역사회와 함께 목표를 세우고 움직이는 힘이다. 사람 사이의 개념이지만 협업 능력의 앞선 사례로 읽을 수 있다. 유네스코는 2024년 9월 학생용 AI 역량 틀을 내면서 네 차원 가운데 첫째로 인간 중심 사고를 두었다(유네스코, 2024년).

미국과 중국도 학교에 인공지능 교육을 넣고 있다. 미국은 2025년 4월 23일 행정명령 14277로 청소년의 AI 리터러시 교육을 국가 정책으로 정했다(백악관, 2025년 4월 23일). AI 리터러시는 인공지능의 작동 원리와 한계를 알고 쓰는 기초 소양이다. 중국 교육부 산하 위원회는 2025년 5월 초중등 학생의 생성형 AI 사용 지침을 냈다(澎湃新闻, 2025년 5월 12일). 지침은 학생이 생성 결과를 과제에 그대로 베끼지 말라고 했다.

한국 교육부의 2026년 업무계획은 AI 교육자료 선택 지원과 질문 중심 수업과 서·논술형 평가 확대를 담았다(데일리안, 2026년 1월 5일). 교육부와 한국대학교육협의회는 2026년 4월 대학 AI 기본교육과정 개발 지원사업 대상으로 20개교를 뽑았다(대학저널, 2026년 4월 23일). 선정 대학은 해마다 교당 3억 원 이내를 받아 AI의 윤리적 사용과 비판적 사고를 다루는 기초 교양을 만든다. 아주대는 2026학년도 신입생부터 AI 리터러시를 교양필수로 정했다(아주대학보, 2026년 2월 27일). 전남대는 2027학년도 신입생부터 전교생에게 AI 교양 6학점을 필수로 듣게 한다(서울신문, 2026년 8월 24일).

이 정책들은 대개 도구를 다루는 법에서 출발한다. 교육의 목표를 협업과 지휘로 세우면 무엇을 가르칠지와 무엇으로 평가할지가 그 목표에서 나온다. 지금 대학의 심사는 표절 검사의 등급표와 검출 점수를 빌려 학생이 인공지능을 얼마나 썼는지 가늠하려 든다. 우리는 이것을 목표가 잘못 선 결과로 본다. 심사 기준을 어떻게 바꿀지는 4장에서 따로 다룬다.

협업은 가르쳐서 느는 능력이 아니고 일하면서 몸에 붙는다는 시각도 있다. 팀워크 기술은 상황에 따라 달라서 교실 연습이 실제 일터로 잘 옮겨지지 않는다는 지적이 그 근거다. 반대 자료도 있다. 호주 매쿼리대의 마크 웨스토비 교수 등은 연구 협업에 필요한 소프트 스킬 각각을 어느 정도 배울 수 있다고 보고 대학원 단기 과정을 제안했다(Trends in Ecology and Evolution, 2017년 2월). 2025년의 한 메타분석은 과학기술 교육을 바탕으로 한 수업이 팀워크를 포함한 소프트 스킬을 키웠다고 보고했다(Frontiers in Education, 2025년 9월 17일). 사람과 인공지능의 협업에는 질문을 설계하고 일을 나누며 결과를 대조하고 출처를 확인하는 절차가 필요하다. 학교는 이 절차를 실습하게 하고 학생이 과제를 나누거나 결과를 선택한 이유를 설명하게 할 수 있다. 현장에서 익혀야 한다는 지적도 이런 실습의 필요성을 뒷받침한다.

답을 외우는 능력에서 문제를 정하고 답을 검증하는 능력으로

PISA 2025에는 이 전환이 왜 필요한지 보여 주는 결과가 있다. 과제 초안을 쓰는 데 챗봇을 거의 매일 쓴 학생은 과학 점수가 낮게 나타났다(OECD, 2026년 9월). AI 산출물의 품질을 평가하는 법을 배운 학생은 그런 지도를 받지 못한 학생보다 성적이 나았다. OECD는 이 결과가 상관관계일 뿐 인과를 보여 주지 않는다고 적었다. 마티아스 코먼 OECD 사무총장은 기술과 인공지능이 학습을 도울 수 있지만 노력과 이해를 대신하지 않고 목적을 갖고 쓸 때만 그렇다고 말했다.

작은 실험에서도 비슷한 걱정이 나왔다. MIT 미디어랩의 나탈리야 코스미나 연구팀은 참가자 54명을 세 집단으로 나눠 에세이를 쓰게 하고 뇌파를 쟀다(MIT 미디어랩, 2025년 6월 10일). 챗GPT를 쓴 집단은 뇌 영역 사이의 연결성이 낮았고 자기 글을 기억해 인용하는 능력이 떨어졌다. 연구팀은 이것을 인지 부채라고 불렀다. 이 논문은 동료 심사를 거치기 전의 예비 연구다. 2026년 1월 arXiv에 공개된 밀로시 스탄코비치 등의 논평은 표본 크기와 뇌파 분석 방법과 재현성을 문제 삼았다(arXiv, 2025년 12월 29일 제출). 논평은 결론을 더 조심스럽게 읽자고 했고 연구의 가치를 부정하지는 않았다.

두 자료는 지도 없이 인공지능을 쓴 결과를 보여 준다. 검증하는 법을 배운 학생이 나았다는 PISA의 다른 결과까지 함께 읽으면 답은 인공지능을 막는 쪽보다 쓰는 법을 가르치는 쪽으로 기운다.

인지과학 쪽에서는 다른 반론이 나온다. 버지니아대 인지심리학자 대니얼 윌링엄은 배경 지식이 없으면 글쓴이의 주장을 이해하지도 못한다고 썼다(American Educator, 2007년 여름). 비판적 사고는 분야마다 다른 지식 위에서만 작동한다는 것이다. 이 지적은 옳다. 인공지능의 답을 검증하려면 그 분야의 기초를 알아야 한다. 타오도 학생이 훈련용으로 받은 문제를 인공지능에게 풀게 하는 일을 경계했다("It is especially risky for students to ask an AI to solve a problem")(UCLA 매거진, 2026년 10월 6일).

기초 지식은 여전히 필요하다. 문제를 이해하고 질문을 만들며 기계의 답을 검증하려면 그 분야의 개념과 원리를 알아야 한다. 바뀌어야 할 것은 지식을 가르치는 목적과 평가 방식이다. 외운 내용을 재현하는 데서 그치지 않고, 그 지식으로 문제를 정하고 근거를 판단하는지 평가해야 한다. 학교는 학생이 혼자 익혀야 할 기초와 인공지능의 도움을 받아 확장할 과제를 이 목표에 맞춰 구분해야 한다.

한국의 대입도 같은 물음 앞에 섰다. 국가교육위원회는 2026년 8월 5일 대통령 업무보고에서 수능 서·논술형 문항 도입과 전 과목 절대평가를 검토한다고 밝혔다(서울신문, 2026년 8월 5일). 차정인 위원장은 학생이 자기 손으로 정답을 쓰게 만들어야 한다고 말했다. 그는 10월 7일 국정감사에서 시안 발표를 늦추자는 논의가 있다고 했고 도입 여부는 확정되지 않았다고 밝혔다(한국교육신문, 2026년 10월 7일). 2026년 10월 14일 현재 시안은 공개되지 않았다.

교육과 인간 존재의 방향이 같은 곳을 가리킨다

교육철학은 오래전부터 교육의 목적을 하나로 보지 않았다. 존 듀이는 1916년 『민주주의와 교육』에서 교육을 경험을 통한 성장으로 보고 암기와 강의 위주의 수업에

반대했다(Democracy and Education, 1916년). 교육학자 거트 비에스타는 교육의 목적을 세 영역으로 나눈다(Educational Theory, 2020년). 자격화(qualification)는 지식과 기능을 갖추게 하는 일이다. 사회화(socialisation)는 학생을 전통과 문화와 실천 안으로 이끄는 일이다. 주체화(subjectification)는 학생이 교육의 대상에 머물지 않고 자기 삶의 주체로 서게 하는 일이다.

협업과 지휘라는 두 축을 비에스타의 틀에 놓고 읽을 수 있다. 협업 능력은 자격화와 사회화에 가깝다. 도구를 다루는 기능을 익히고 일터의 관행 안으로 들어가는 일이기 때문이다. 지휘 능력은 주체화에 가깝다. 풀 문제를 스스로 정하고 결과에 책임지는 일은 학생이 자기 판단의 주인이 되는 일이다. 이 대응은 비에스타가 한 말이 아니고 우리의 읽기다.

인공지능이 자격화의 상당 부분을 맡게 되면서 주체화의 몫이 커졌다. 지식을 옮기고 정리하는 일은 기계가 사람보다 빨리 한다. 무엇을 할지 정하고 그 결정에 이름을 거는 일은 기계에 넘길 수 없다. 듀이가 말한 성장도 남이 정해 준 답을 받아 적는 데서는 일어나지 않는다.

기계가 지식 노동을 넘겨받을수록 사람에게 남는 일은 어디로 갈지 정하는 일이다. 개인은 무엇을 공부하고 무엇을 만들지 정해야 하고 기업과 국가는 어떤 문제를 풀지 정해야 한다. 목표를 세우는 사람이 없으면 기계가 아무리 유능해도 쓸 곳을 찾지 못한다. 학교에서 문제를 정하고 결과를 책임지는 연습을 한 사람이 사회에서 그 일을 맡는다. 교육의 목표를 다시 세우는 일과 인간의 역할을 다시 정하는 일은 같은 사람을 두고 하는 이야기다.

교육이 기르려는 사람과 인공지능 시대에 인간에게 남는 일은 같은 곳을 가리킨다고 우리는 판단한다. 그곳에 있는 사람은 문제를 정하고 판단하고 그 결과를 책임지는 사람이다. 2026년 9월 서울에서 열린 한 교육 회의에서 유타대 아시아캠퍼스의 최고행정책임자 그레고리 힐은 인격은 인공지능 훈련으로 외주를 줄 수 없다고 말했다("You simply cannot outsource character as a function of AI training")(Korea Times, 2026년 9월 15일).

참고 자료

France24/AFP, AI catches up with humans to score 100% at top maths contest, 2026년 7월 23일, https://www.france24.com/en/live-news/20260723-ai-catches-up-with-humans-to-score-100-at-top-maths-contest

구글 딥마인드, Advanced version of Gemini with Deep Think officially achieves gold-medal standard at the International Mathematical Olympiad, 2025년 7월 21일, https://deepmind.google/discover/blog/advanced-version-of-gemini-with-deep-think-officially-achieves-gold-medal-standard-at-the-international-mathematical-olympiad

Korea Times(한국일보 기사 번역), GPT-6 Astra aces Korean college entrance exam, 2026년 9월 7일, https://www.koreatimes.co.kr/business/tech-science/20260907/gpt-6-astra-aces-korean-college-entrance-exam

AI타임스, 제미나이 이어 GPT도 '수능 만점'...구유겸 학생 "AI 발전 놀라워", 2026년 3월 8일, https://www.aitimes.com/news/articleView.html?idxno=207681

국민일보, AI가 변호사시험 객관식 사상 첫 만점…슈퍼로이어 '150개 모두 정답', 2026년 6월 24일, https://www.kmib.co.kr/article/view.asp?arcid=0030013317&code=61121111&sid1=soc

퓨리서치센터, How Teens Use and View AI, 2026년 2월 24일, https://www.pewresearch.org/internet/2026/02/24/how-teens-use-and-view-ai/

한국교육개발원 월간 교육정책포럼 394호(임소현), 데이터로 본 학생과 교사의 AI·디지털 기술 활용 현황 및 인식, 2026년 4월 22일, https://edpolicy.kedi.re.kr/edpolicy/webzine/42/844819

OECD, PISA 2025: Students' reading and mathematics performance declined sharply across the OECD, 2026년 9월, https://www.oecd.org/en/about/news/press-releases/2026/09/pisa-2025-students-reading-and-mathematics-performance-declined-sharply-across-the-oecd.html

OECD, PISA 2025 Results (Volume I), 2026년 9월, https://www.oecd.org/en/publications/2026/09/pisa-2025-results-volume-i_5265bfb1.html

서울신문, 대학은 미래 산업의 심장… AI로 지역 운명 바꿀 청년 키운다 [월요인터뷰], 2026년 8월 24일, https://www.seoul.co.kr/news/plan/monday-interview/2026/08/24/20260824014001

서울경제, 유지범 성균관대 총장 "지식 전달, AI가 대체…대학은 비판적 지성 키워야", 2026년 10월 11일, https://m.sedaily.com/article/20100171

메타, The Llama 4 herd: The beginning of a new era of natively multimodal AI innovation, 2025년 4월 5일, https://ai.meta.com/blog/llama-4-multimodal-intelligence/

arXiv(Prodan 외), Cutting Through the Confusion and Hype: Understanding the True Potential of Generative AI, 2024년 10월 22일, https://arxiv.org/abs/2410.16629

Cognitive Science(Thomas K. Landauer), How Much Do People Remember? Some Estimates of the Quantity of Learned Information in Long-term Memory, 1986년, https://home.cs.colorado.edu/~mozer/Teaching/syllabi/7782/readings/Landauer1986.pdf

Scientific American(Paul Reber), What Is the Memory Capacity of the Human Brain?, 2010년 5월 1일, https://www.scientificamerican.com/article/what-is-the-memory-capacity/

스탠퍼드 HAI, Responsible AI | The 2026 AI Index Report, 2026년 4월 13일, https://hai.stanford.edu/ai-index/2026-ai-index-report/responsible-ai

스탠퍼드 HAI, Inside the AI Index: 12 Takeaways from the 2026 Report, 2026년 4월 13일, https://hai.stanford.edu/news/inside-the-ai-index-12-takeaways-from-the-2026-report

The Decoder, Terence Tao says GPT-5.2 Pro cracked an Erdős problem, but warns the win says more about speed than difficulty, 2026년 1월 16일, https://the-decoder.com/terence-tao-says-gpt-5-2-pro-cracked-an-erdos-problem-but-warns-the-win-says-more-about-speed-than-difficulty/

arXiv, ClarEval: A Benchmark for Evaluating Clarification Skills of Code Agents under Ambiguous Instructions, 2026년 2월 27일, https://arxiv.org/abs/2603.00187

One Useful Thing(Ethan Mollick), The Overhang, 2026년 9월 18일, https://www.oneusefulthing.org/p/the-overhang

한국은행, AI 확산과 청년고용 위축: 연공편향(seniority-biased) 기술변화를 중심으로(BOK 이슈노트 제2025-30호), 2025년 10월 30일, https://www.bok.or.kr/portal/bbs/P0002353/view.do?nttId=10094258&menuNo=200433

FM Magazine, Analytical thinking remains top core skill for employers, 2025년 2월 18일, https://www.fm-magazine.com/news/2025/feb/analytical-thinking-remains-top-core-skill-for-employers

OECD, OECD Learning Compass 2030 concept note, 2019년, https://search.oecd.org/education/2030-project/teaching-and-learning/learning/learning-compass-2030/OECD_Learning_Compass_2030_concept_note.pdf

유네스코, AI competency framework for students, 2024년, https://www.unesco.org/en/articles/ai-competency-framework-students

백악관, Advancing Artificial Intelligence Education for American Youth, 2025년 4월 23일, https://www.whitehouse.gov/presidential-actions/2025/04/Advancing-Artificial-Intelligence-Education-For-American-youth/

澎湃新闻(中国青年报客户端), 教育部基础教育教指委：小学阶段禁止学生独自使用开放式内容生成功能, 2025년 5월 12일, https://m.thepaper.cn/newsDetail_forward_30802716

데일리안, [2026 전망보고서] AI 인재 양성·지방대학 육성…"국가가 책임지는 기본교육 강화", 2026년 1월 5일, https://www.dailian.co.kr/news/view/1594428/

대학저널, 교육부·대교협, '대학 AI 기본교육과정 개발 지원사업' 20개교 선정, 2026년 4월 23일, https://dhnews.co.kr/news/view/1065589310663169

아주대학보, 아주대, 2026학년도부터 'AI 리터러시' 교양필수 지정, 2026년 2월 27일, https://press.ajou.ac.kr/news/articleView.html?idxno=11304

Trends in Ecology and Evolution(Gibert, Tozer, Westoby), Teamwork, Soft Skills, and Research Training, 2017년 2월 1일, https://researchers.mq.edu.au/en/publications/teamwork-soft-skills-and-research-training/

Frontiers in Education, STEM-based approaches to soft skills development: a synthesis of meta-analytic findings and empirical evidence, 2025년 9월 17일, https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2025.1663155/full

MIT 미디어랩, Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task, 2025년 6월 10일, https://www.media.mit.edu/publications/your-brain-on-chatgpt/

arXiv(Stankovic 외), Comment on: Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Tasks, 2025년 12월 29일 제출, https://arxiv.org/abs/2601.00856

American Educator(Daniel T. Willingham), Critical Thinking: Why Is It So Hard to Teach?, 2007년 여름, https://www.aft.org/periodical/american-educator/summer-2007/critical-thinking

UCLA 매거진, The Tao of Tao, 2026년 10월 6일, https://newsroom.ucla.edu/magazine/terence-tao-math-artificial-intelligence-alpha-research

서울신문, "이제 수능 찍기 안 통한다"…국교위, 대입 '서·논술형' 개편 예고, 2026년 8월 5일, https://www.seoul.co.kr/news/society/education-news/2026/08/05/20260805500250

한국교육신문, 대입개편 시안 발표 늦추나…"공론화 더 필요", 2026년 10월 7일, https://hangyo.com/news/article.html?no=109590

Project Gutenberg(John Dewey), Democracy and Education, 1916년, https://www.gutenberg.org/etext/852

Educational Theory(Gert Biesta), Risking Ourselves in Education: Qualification, Socialization, and Subjectification Revisited, 2020년, https://mural.maynoothuniversity.ie/id/eprint/16038/1/GertBiesta2022Ris.pdf

Korea Times, Universities urged to reshape education, cultivate human skills for AI age, 2026년 9월 15일, https://www.koreatimes.co.kr/southkorea/education/20260915/conference-universities-urged-to-reshap

e-education-cultivate-human-skills-for-ai-age



2장. 협업 능력: 인공지능과 함께 일하는 법

변호사 업무의 분해: 판례 검색과 서면 초안은 인공지능이, 전략 결정은 사람이

법관은 사실 확정과 법률의 해석과 적용을 인공지능에 전적으로 맡기지 않아야 한다. 법원 안의 연구 모임인 법원 인공지능연구회가 2025년 2월 내놓은 가이드라인의 한 구절이다(법무법인 태평양, 2025년 3월 10일). 법원의 공식 입장은 아니지만 판사들이 스스로 정한 기준이다. 연구회는 오류가 담긴 인공지능 산출물을 쓴 법관이 법적 책임과 윤리적 책임을 질 수 있다고도 적었다.

1년 뒤 법원은 재판지원 시스템을 직접 도입했다. 법원행정처는 2026년 2월 13일 재판지원 AI 시스템을 시범 운영한다고 발표했다(법률신문, 2026년 2월 13일). 외부의 공개형 서비스에 기대지 않는 폐쇄형 시스템이다. 판례와 법령과 실무 해설서를 찾아 쟁점을 정리하고 근거 자료를 함께 보여 준다. 법원행정처 관계자는 답에 부정확한 내용이 있을 수 있고 최종 판단과 책임은 이용자에게 있다고 밝혔다.

민간 서비스도 같은 일을 한다. 로앤컴퍼니의 슈퍼로이어는 판례 534만 건을 연결해 검색과 질의응답을 한 서비스 안에서 하게 했다(파이낸셜뉴스, 2026년 10월 2일). 효력을 잃은 판례와 나중에 바뀐 판결도 표시한다. 법무법인 창천의 변호사들은 사흘 걸리던 의견서 초안과 일주일까지 걸리던 소송 서면 초안을 하루 안팎에 만든다고 말했다(아시아경제, 2026년 9월 2일). 미국 국제법률기술협회(ILTA)의 2026년 조사에서는 회원 로펌의 94퍼센트가 생성형 인공지능을 쓰고 있거나 쓸지 검토하고 있었다(LawSites, 2026년 9월 14일).

판례를 찾고 초안을 쓰는 일은 기계 쪽으로 옮겨 가고 있다. 사건을 어느 방향으로 끌고 갈지는 여전히 사람이 정한다. 창천의 정영훈 변호사는 변호사에게 남는 일로 먼저 의뢰인과 직접 대화하는 일을 꼽았다. 사건의 배경과 맥락을 이해하는 일과 인공지능이 낸 결과를 검토하는 일이 뒤를 잇는다. 중국 최고인민법원의 덩슈밍 상무부원장도 2025년 4월 연설에서 같은 선을 그었다. 그는 모든 최종 결정과 책임이 사람 판사에게 있다고

말했다("all the final decisions and responsibilities belong to human judges")(중국 최고인민법원, 2025년 4월 24일). 구속력 있는 규칙이 아닌 연설이다.

변호사 업무의 자동화가 부풀려졌다는 시각도 있다. 2023년 골드만삭스 보고서를 근거로 법률 과업의 44퍼센트가 자동화될 수 있다는 수치가 널리 퍼졌다. 법률 기술 블로그 긱로블로그(Geek Law Blog)는 그 자료가 주장을 받치기에 모자란다고 비판했다(Geek Law Blog, 2023년 7월). 실제 도입 수치도 조사마다 갈린다. 톰슨로이터의 2026년 조사에서 조직 차원으로 인공지능을 도입한 곳은 40퍼센트였다(톰슨로이터 연구소, 2026년 2월 9일). 노출도는 한 직업의 과업을 인공지능이 얼마나 수행할 수 있는지를 측정한 지표다. 한국고용정보원은 인공지능 노출도 1위 직업으로 변호사를 꼽으면서도 노출도가 높다고 대체 가능성이 높은 것은 아니라고 적었다(뉴스핌, 2026년 8월 31일). 이 지적은 옳다. 지금 눈에 보이는 변화는 직업이 통째로 사라지는 모습과 다르다. 변호사의 일은 과업 단위로 쪼개졌고 그중 검색과 초안이 먼저 기계로 넘어갔다. 판단하는 일은 사람에게 남았다.

의뢰인의 사정을 듣는 일이 왜 사람의 일인가

의뢰인이 처음 꺼내는 이야기는 사건의 일부일 때가 많다. 불리한 사실이나 부끄러운 사정은 뒤늦게 나온다. 법률 면담을 다룬 연구는 이 순서를 신뢰의 문제로 설명한다. 2019년 학술지 「정신의학, 심리학과 법」에 실린 연구는 의뢰인이 이해받는다고 느끼면 변호사를 더 믿고 믿음이 커지면 사실을 털어놓을 마음도 커진다는 경로를 보고했다(Psychiatry, Psychology and Law, 2019년). 학생이 변호사와 의뢰인 역할을 나눠 맡은 작은 연구여서 실제 사건에 그대로 옮기기는 어렵다. 크리스틴 거디는 2008년 논문에서 이해받는다고 느끼는 의뢰인이 당혹스럽지만 사건에 중요한 정보까지 내놓는다고 정리했다(Nebraska Law Review, 2008년).

인공지능이 사람보다 공감을 잘 표현한다는 연구도 있다. 2023년 4월 JAMA 내과학에 실린 연구에서 의료 전문가 평가자들은 온라인 게시판에 올라온 환자 질문 195개에 대한 의사의 답과 챗GPT의 답을 비교했다(JAMA Internal Medicine, 2023년 4월 28일). 공감 수준이 높다는 평가를 받은 답은 챗봇이 45.1퍼센트였고 의사는 4.6퍼센트였다. 의사 답의 평균 길이는 52단어였고 챗봇 답은 211단어였다. 비판자들은 긴 답이 공감 점수를 끌어올렸고 게시판 문답은 진료실 대화와 다르다고 지적했다(STAT, 2023년 5월 4일). 이 연구가 잰 것은 공감하는 문장을 쓰는 능력이다. 상대가 말하지 않은 사실을 끌어내는 능력은 재지 않았다. 법률 챗봇이 의뢰인의 사실관계를 얼마나 끌어내는지 잰 실험은 2026년 10월까지 찾지 못했다.

챗봇에게 털어놓은 말은 법의 보호도 받지 못한다. 비밀특권은 의뢰인과 변호사가 나눈 대화를 재판에서 공개하지 않아도 되게 하는 보호다. 미국 연방법원은 2026년 United States v. Heppner 사건에서 챗봇은 변호사가 아니므로 이 특권을 주장할 수 없다고 봤다(National Law Review, 2026년). 미국 하급심의 판단이다. 법원은 서비스의 개인정보 정책이 입력 내용을 학습에 쓸 수 있게 한 점도 비밀을 기대할 수 없는 이유로 들었다.

인공지능은 면담 기록을 정리하고 빠뜨린 질문을 제안할 수 있다. 그 질문을 언제 꺼낼지와 의뢰인이 머뭇거릴 때 기다릴지는 마주 앉은 사람이 정한다. 정영훈 변호사가 남는 일의 첫째로 꼽은 것도 의뢰인과 직접 나누는 대화였다. 숨기고 싶은 사정을 꺼내는 사람은 그 말을 지켜 줄 상대를 고른다. 우리는 의뢰인의 사정을 듣는 일을 믿음을 쌓는 일로 본다. 그 믿음은 비밀을 지킬 의무와 결과에 대한 책임을 함께 지는 사람에게 생긴다.

질문을 던지는 능력: 좋은 질문이 좋은 결과를 만든다

저자는 전자신문 연재 「김경진의 AI전략노트」 15회에서 "법률 업무부터 집필, 강의 준비까지" 거의 모든 영역에서 인공지능을 쓴다고 밝혔다(전자신문 연재 다음 전재, 2025년 11월 18일). 같은 글에서 권한 방법 가운데 하나가 거꾸로 질문이다. 인공지능에게 답을 바로 묻지 않고 "이 문제를 해결하기 위해 내가 생각해 봐야 할 것들이 무엇인지 질문해 줘"라고 시킨다. 기계가 던진 질문을 따라가면 문제의 범위가 좁혀지고 놓친 요소가 드러난다.

프롬프트는 인공지능에게 주는 질문이나 작업 지시문이다. 2022년 무렵부터 단계를 나눠 생각하게 하거나 예시를 붙이는 프롬프트 기법이 성능을 높인다는 연구가 쌓였다. 2024년 11월 공개된 왕(Wang) 등의 연구는 이 기법들을 GPT-4o와 추론 모델 o1에서 다시 시험했다(arXiv, 2024년 11월 4일). 초기 모델에 맞춰 만든 기법은 고급 모델에서 이득이 줄거나 성능을 해칠 수도 있었다("may provide diminished benefits or even hinder performance"). 실행 결과를 되먹여 주거나 과제 정보를 정확히 주는 방법은 여전히 효과가 있었다.

모델이 좋아질수록 주문 문구의 효과는 줄어든다. 줄지 않는 것은 문제를 정확히 정의하고 조건을 밝히는 일이다. 법률 질문에서 조건은 사실관계와 의뢰인이 바라는 결과와 사건의 기한 같은 것이다. 기계는 흐린 질문을 법률 언어로 옮기는 일까지 맡기 시작했다. 대법원은 "전셋집에서 이사를 가야 하는데 집주인이 돈이 없대" 같은 일상어에서 법률 개념을 뽑아 판결문을 찾는 검색 시스템을 추진한다고 2026년 6월 9일 밝혔다(세계일보, 2026년 6월 9일). 그래도 계약 종료를 언제 알렸는지와 집주인이 무슨

이유로 돈을 주지 않는지에 따라 쟁점은 달라진다. 그 사실을 물어 질문에 담는 일은 사건을 맡은 사람이 한다. 무엇을 확인해야 하는지 아는 사람은 돌아온 답을 다시 묻고 범위를 좁힌다.

질문은 기록으로도 남는다. 법원 인공지능연구회 가이드라인은 서면이 인공지능 산출물로 밝혀지면 사용한 인공지능과 입력한 프롬프트와 검증 조치를 밝히도록 법원이 소송을 지휘할 수 있다고 했다(법무법인 태평양, 2025년 3월 10일). 법정에서 변호사가 던진 질문은 그가 어떤 판단을 거쳤는지 보여 주는 자료가 된다.

이정동 서울대 교수는 2022년 책 「최초의 질문」에서 남이 낸 문제를 잘 푸는 단계를 넘어 스스로 질문을 내놓는 단계로 가야 한다고 썼다(문화일보, 2023년). 그의 논의는 국가의 기술 전략을 향한다. 같은 생각은 인공지능 앞에 앉은 개인에게도 들어맞는다. 기계도 질문 후보를 제안할 수 있다. 그 가운데 의뢰인의 문제를 풀 질문을 고르고 필요한 사실을 확인하는 일은 사건을 맡은 사람이 책임져야 한다. 좋은 질문은 문장을 다듬는 요령보다 사건을 깊이 아는 데서 나온다고 우리는 판단한다.

결과를 걸러 내는 능력: 틀린 답을 알아보는 눈

2023년 미국 뉴욕남부연방지방법원의 마타 대 아비앙카(Mata v. Avianca) 사건은 법률가들에게 오래 남는 경고가 됐다. 원고 측 변호사들은 챗GPT로 찾은 존재하지 않는 판결 6건을 서면에 인용했다. 상대방이 그런 판례를 찾을 수 없다고 지적한 뒤에도 약 6주 동안 판결이 있다고 버텼다. 케빈 캐스텔 판사는 2023년 6월 22일 이들에게 제재금 5천 달러를 부과했다(Bloomberg Law, 2023년 6월 22일). 판사는 처음부터 솔직히 밝혔다면 제재하지 않았을 것이라고 적었다.

같은 일은 그 뒤로도 이어졌다. 2026년 6월 23일 뉴욕주 항소법원 2부는 존재하지 않는 판례를 인용한 항소 서면을 문제 삼아 변호사에게 8천 달러, 로펌에 2,500달러의 제재금을 부과했다(Mealey's, 2026년 6월). 변호사는 인공지능 도구로 찾은 판례를 확인하지 않았다고 인정했다. 한국에서는 특허법원이 2026년 6월 11일 등록무효 사건 판결문에 27년 차 변호사가 인공지능으로 찾은 허위 판례 8건을 검증 없이 제출했다고 시인한 사실을 적고 부적절하다고 판시했다(법률신문, 2026년 9월 12일). 같은 해 9월 4일 사법연수원 법관 연수에서는 서면에 인용한 판례 40개 가운데 39개가 허위였던 사례가 공유됐다. 법원행정처 TF는 3월 31일 허위 인용에 소송비용 부담과 변론 진술 제한, 대한변협 징계 의뢰로 대응하는 방안을 냈다(법률신문, 2026년 3월 31일). 2026년 10월 14일까지 실제로 징계나 비용 부담이 내려진 사례는 확인되지 않는다.

다미앵 샤를로탱의 인공지능 환각 사건 데이터베이스에는 2026년 10월 10일 기준 2,184건이 올라 있고 그중 미국이 1,492건이다(Damien Charlotin, 2026년 10월 14일 열람). 법원이나 심판기관이 당사자가 환각 자료에 기댔다고 인정한 결정만 센 수치다. 본인소송 당사자가 관련된 사건이 1,259건이고 변호사가 관련된 사건은 864건이다.

틀린 답을 알아보는 눈은 교육만으로 생기지 않는다. 자동화 편향은 사람이 기계의 판단을 지나치게 믿고 따르는 경향이다. 파키스탄 연구진은 20시간 동안 인공지능 문해 교육을 받은 의사 44명을 무작위로 두 군에 나눠 임상 사례 6개를 진단하게 했다(medRxiv, 2025년 8월). 한 군에는 중대한 오류를 일부러 심은 챗GPT 권고를 줬다. 이 군의 진단 정확도는 73.3퍼센트로 대조군의 84.9퍼센트보다 낮았다. 동료 심사 전의 사전공개 논문이고 표본이 작다.

인공지능 사용이 기본 역량을 약화시킨다는 반론도 이 지점에서 나온다. 탈숙련은 기계에 기대는 동안 사람이 혼자 일하는 능력이 줄어드는 현상이다. 랜싯 소화기학에 실린 폴란드 연구에서 인공지능 보조를 일상으로 쓰기 시작한 내시경의 19명은 인공지능 없이 한 대장내시경에서 선종 발견율이 28.4퍼센트에서 22.4퍼센트로 떨어졌다(The Lancet Gastroenterology & Hepatology, 2025년 8월). 구글 리서치와 경제학자 데이비드 오터 등의 특허변호사 실험에서는 경력이 짧은 변호사의 기량이 평균적으로 늘지 않았다(구글 리서치, 2026년 10월 7일). 서울신문은 2026년 1월 연재에서 저연차 변호사가 역량을 쌓을 기회를 잃는다는 우려를 전했다(서울신문, 2026년 1월 27일). 이 반론은 거르는 눈이 어디서 오는지 알려 준다. 판례를 직접 찾고 서면을 직접 써 본 사람이 틀린 인용을 알아본다. 그래서 기계 없이 해 보는 시간을 일부러 남겨 두어야 한다고 우리는 판단한다.

저자는 앞의 칼럼에서 같은 질문을 여러 인공지능에게 던져 답을 비교하는 방법을 권했다. 답이 서로 어긋나는 곳이 사람이 직접 확인할 곳이다. 여러 답이 같더라도 원자료 확인을 마친 것은 아니다. 비교는 확인할 쟁점을 찾는 절차로 쓰고, 법정에 낼 인용은 판결문과 법령 원문으로 검증해야 한다.

결과에 책임지는 능력: 인공지능이 쓴 글의 책임은 누구에게 있는가

미국변호사협회(ABA)는 2024년 7월 29일 윤리의견 512호에서 기존 직무 규칙을 생성형 인공지능에 적용했다. ABA 원문을 열 수 없어 내용은 2차 해설로 확인했다(Bench & Bar of Minnesota, 2024년 9월). 의견에 따르면 변호사는 전문적 판단을 도구에 넘길 수 없고 산출물의 오류를 확인해야 한다. 감독 의무도 함께 적용된다. 감독 의무는

관리하는 자리의 변호사가 다른 변호사와 직원의 일이 규칙에 맞게 이뤄지도록 살필 의무다. 로펌 경영자는 인공지능 사용 정책을 세우고 구성원을 교육해야 한다.

상설명령은 판사가 자기 법정의 모든 사건에 미리 적용하려고 내 두는 명령이다. 미국 연방 지방법원에서 인공지능 사용을 다룬 판사별 상설명령은 2026년 6월까지 수십 건에 이른다(The Legal Stack, 2026년 7월 10일). 2023년 텍사스 북부연방지방법원 브랜틀리 스타 판사의 명령은 서면에 인공지능이 쓴 부분이 없거나 그 부분을 사람이 전통적 법률 데이터베이스로 확인했다고 증명하게 했다(Bloomberg Law 비교표).

서울지방변호사회는 2024년 12월 제안한 가이드라인에서 인공지능 결과물의 최종 판단과 책임이 변호사 본인에게 있다고 명시했다(법률신문, 2024년 12월 5일). 대한변호사협회는 법률서비스의 인공지능 사용 지침을 11월 초 발표할 예정이다(뉴스1, 2026년 10월 14일). 김정욱 변협회장은 인공지능을 변호사의 책임 아래 써야 한다는 원칙을 밝혔다.

서비스를 만든 쪽의 책임은 다른 법이 다룬다. 2026년 1월 22일 시행된 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법(인공지능 기본법) 제31조는 생성형 인공지능 사업자에게 그 사실을 미리 알리고 결과물에 표시할 의무를 지운다(정책브리핑, 2026년 1월 21일). 제34조는 고영향 인공지능 사업자에게 설명 방안과 이용자 보호 같은 안전성 확보 조치를 정한다(법률신문, 2025년 12월 9일). 두 조문은 로펌 변호사들의 해설을 기준으로 적었다. 해설들이 다루는 의무 조항 가운데 인공지능으로 생긴 손해의 배상을 직접 정한 조문은 없다. 배상 문제는 민법의 불법행위 책임 같은 기존 법리로 다룬다. EU 집행위원회는 2025년 2월 인공지능 책임지침 제안을 거둬들이기로 했다(TCI Law, 2025년 2월 14일). 대신 개정 제조물책임지침이 2026년 12월 9일 이후 출시된 제품부터 소프트웨어와 인공지능을 제품에 넣는다(SVW). 제조물책임은 제품의 결함으로 생긴 손해를 제조자가 과실이 없어도 배상하는 책임이다. 한국저작권위원회는 2025년 6월 안내서에서 사람의 창작적 기여가 없는 생성물은 저작권 등록을 받지 않는다고 정했다(한국저작권위원회, 2025년 6월).

책임을 사람에게 지우면 인공지능을 들일 이유가 사라진다는 반론도 있다. 미국기업연구소의 대니얼 라이언스는 소송 위험이 기업의 인공지능 도입을 망설이게 한다고 썼다("Litigation risk makes companies hesitant to deploy AI technology.")(AEI, 2025년 9월 18일). 매들린 클레어 일리시는 통제력이 적은 사람 운영자가 사고의 책임을 떠안는 구조를 도덕적 크럼플 존이라고 불렀다(Engaging Science, Technology, and Society, 2019년). 두 비판은 책임이 말단 사용자에게만 쏠리는 구조를 겨눈다. 중국 항저우 인터넷법원은 2026년 1월 인공지능은 법적 주체가 아니라고 보면서도 사업자에게

보건과 법률 같은 고위험 영역에서 더 강한 경고를 할 주의의무를 지웠다(Gowling WLG, 2026년). 중국 하급심의 판단이다. 서비스 제공자는 주의의무의 수준에 따라 책임지고 글을 낸 사람은 그 글로 책임진다. 이 이층 구조는 이미 여러 나라의 법과 판결에 들어 있다.

인공지능이 쓴 글을 자기 이름으로 제출하는 사람은 내용을 확인하고 그 판단에 책임져야 한다. 서비스 제공자의 주의의무와 결함에 대한 책임도 별도로 따져야 한다. 책임을 맡기는 범위에 맞춰 검증할 권한과 수단을 함께 주는 것이 협업의 조건이다.

다른 직업으로 확장: 의사, 기자, 공무원, 교사

의료에서도 기계는 보조 자리에 있다. 대한의료인공지능학회 집계로 식품의약품안전처가 허가한 인공지능 의료기기는 2026년 1분기까지 누적 549건이다(데일리메디, 2026년 4월 14일). 식약처 공식 집계로는 2025년 한 해 153건이 허가나 인증을 받았다(뉴시스, 2026년 4월 30일). 식약처는 이 기기들을 진단을 돕는 보조 목적의 의료기기로 허가한다. 허수진 변호사(법무법인 태평양)는 2026년 8월 25일 최종 진단과 치료 결정은 의사의 몫이고 인공지능을 썼다고 책임이 줄지 않는다고 말했다(아시아경제, 2026년 8월 25일). 그는 인공지능 도입으로 의사의 주의의무가 오히려 넓어질 수 있다고 봤다.

언론에서는 한국기자협회와 한국신문협회 등 6개 언론 단체가 2024년 12월 5일 「언론을 위한 생성형 인공지능(AI) 준칙」을 발표했다(한국기자협회, 2024년 12월 6일). 제1조는 인공지능을 뉴스 생산의 보조 수단으로 쓰고 사람의 관리와 감독 아래 두게 했다. 기획과 자료 조사와 번역과 데이터 분석은 폭넓게 허용하고 기사 전체를 자동으로 만드는 일은 바람직하지 않다고 봤다. 결과물의 사실 검증과 인공지능 생성물 표시도 정했다. 로이터는 2023년 5월 내부 메모에서 누가 만들었든 로이터 기사는 로이터 기사라는 원칙을 세웠다("A Reuters story is a Reuters story, regardless of who produces it") (Talking Biz News, 2023년 5월).

공무원 지침은 인사혁신처가 냈다. 인사혁신처는 2026년 5월 22일 공무원 교육훈련생을 위한 생성형 인공지능 활용 지침을 처음 배포했다(머니투데이, 2026년 5월 22일). 지침은 생성형 인공지능을 보조 도구로만 쓰고 비판적 시각을 유지하라고 했다. 결과의 최종 책임은 본인에게 있다고 적었다. 7대 기본원칙에는 진실성과 공정성과 함께 인공지능을 썼다는 사실을 투명하게 공개하는 원칙이 들어 있다. 실제 사용은 지침보다

앞서 있었다. 위성곤 의원실이 2025년 가을 공무원 14,208명에게 물었더니 68.9퍼센트가 업무에 생성형 인공지능을 써 봤다고 답했다(전자신문, 2025년 10월 13일).

학교에서는 교육부와 17개 시도교육청이 2025년 12월 23일 수행평가의 인공지능 관리 방안을 내놓았다(교육부, 2025년 12월 23일). 인공지능을 보조 도구로 쓰게 하되 인공지능이 만든 글을 그대로 내는 일은 금지했다. 한국교육과정평가원은 2026년 10월 7일 수업과 평가를 위한 생성형 인공지능 가이드라인을 발표했다(뉴스핌, 2026년 10월 7일). 평가에 인공지능을 쓰더라도 교사가 결과의 신뢰성과 공정성을 확인하고 최종 판단과 책임을 유지하도록 했다.

네 직업의 지침과 제도는 같은 구조를 갖고 있다. 기계는 검색, 요약, 초안 작성을 맡는다. 사람은 무엇을 맡길지 묻고 돌아온 결과를 검증한 뒤 그 결과에 자기 이름을 건다. 기자 준칙의 활용 표시, 공무원 지침의 사용 내역 공개, 법원 가이드라인의 프롬프트 공개처럼 인공지능을 썼다는 사실을 투명하게 밝히라는 요구도 함께 들어 있다. 최종 판단과 책임을 사람에게 남긴다는 원칙은 2024년 12월부터 2026년 10월 사이에 언론, 행정, 학교의 지침에 차례로 자리 잡았다. 의료에서도 인공지능 기기를 진단 보조 목적으로만 허가하는 제도가 같은 원칙을 뒷받침한다. 인공지능과의 협업 능력이란 기계의 답을 그대로 옮기는 기술이 아니라, 기계의 초안을 검증 가능한 재료로 삼아 자기 책임 아래 완성된 판단으로 바꾸는 능력이다.

참고 자료

법무법인 태평양 BKL Legal Update, 사법에서의 인공지능 활용에 관한 가이드라인 발표, 2025년 3월 10일, https://www.bkl.co.kr/newsLetter/itemUrl.do?itemNo=5983

법률신문, 법원, 생성형 AI 기반 '재판지원 AI 시스템' 시범 오픈, 2026년 2월 13일, https://www.lawtimes.co.kr/news/articleView.html?idxno=216290

파이낸셜뉴스, 판례 검색까지 품은 '슈퍼로이어'…534만건 판례 제공, 2026년 10월 2일, https://www.fnnews.com/news/202610021434018052

아시아경제, '코딩 AI' 만들어 쓰는 변호사 늘었다, 2026년 9월 2일, https://view.asiae.co.kr/article/2026090214443701967

LawSites(Bob Ambrogi), Which AI Product Is Most Popular Among Law Firms? Findings of ILTA's Tech Survey May Surprise You, 2026년 9월 14일, https://www.lawnext.com/2026/09/which-ai-product-is-most-popular-among-law-firms-findings-of-iltas-tech-survey-may-surprise-you.html

중국 최고인민법원 영문 사이트(China Daily), Officials seek limited use of AI in judiciary, 2025년 4월 24일, https://english.court.gov.cn/2025-04/24/c_1088664.htm

Geek Law Blog(Ryan McClead), 44% of Investment Bankers Think They Can Make Lots of Money Off of Attorney Insecurity (AI), 2023년 7월, https://www.geeklawblog.com/2023/07/44-of-investment-bankers-think-they-can-make-lots-of-money-off-of-attorney-insecurity-ai.html

톰슨로이터 연구소, 2026 AI in Professional Services Report: AI adoption has hit critical mass, but now comes the tough business questions, 2026년 2월 9일, https://www.thomsonreuters.com/en/institute/articles/ai-in-professional-services-report-2026

뉴스핌, AI 노출도 1위 변호사…30대 고용보험 가입자 절반 이상 AI 노출↑, 2026년 8월 31일, https://www.newspim.com/news/view/20260831000272

Psychiatry, Psychology and Law(Howieson, Rogers), Rethinking the Lawyer-Client Interview: Taking a Relational Approach, 2019년 7월, https://www.tandfonline.com/doi/abs/10.1080/13218719.2019.1618749

Nebraska Law Review(Kristin B. Gerdy), Clients, Empathy, and Compassion: Introducing First-Year Students to the "Heart" of Lawyering, 2008년, https://digitalcommons.unl.edu/nlr/vol87/iss1/1

JAMA Internal Medicine(Ayers 외), Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum, 2023년 4월 28일, https://jamanetwork.com/journals/intemed/articlepdf/2804309/jamainternal_ayers_2023_oi_230030_1681999216.70842.pdf

STAT, Why a chatbot might seem more empathetic than a human physician, 2023년 5월 4일, https://www.statnews.com/2023/05/04/medical-chatbots-empathy-research-human-doctors/

National Law Review, Think Before You Chat: Using AI for Legal Questions May Waive Privilege and Create Discovery Risks, 2026년 4월 15일, https://natlawreview.com/article/think-you-chat-using-ai-legal-questions-may-waive-privilege-and-create-discovery

전자신문(다음 전재), [김경진의 AI전략노트] 〈15〉 언어모델 인공지능을 효율적으로 잘 사용하려면, 2025년 11월 18일, https://v.daum.net/v/20251118160307876

arXiv(Wang 외), Do Advanced Language Models Eliminate the Need for Prompt Engineering in Software Engineering?, 2024년 11월 4일, https://arxiv.org/abs/2411.02093

세계일보, 법률 지식 부족해도 판결문 쉽게 찾도록…대법, AI 검색 시스템 추진, 2026년 6월 9일, https://www.segye.com/newsView/20260609506353

문화일보, 기술선진국 도약하려면… '도전적 질문' 던지며 시행착오 쌓아야[문화산업포럼 2023], 2023년 9월 22일, https://www.munhwa.com/article/11385778

Bloomberg Law, Phony ChatGPT Brief Leads to $5,000 Fine for NY Lawyers, 2023년 6월 22일, https://news.bloomberglaw.com/esg/chatgpt-phony-legal-filing-case-gets-lawyers-a-5-000-fine

Mealey's, AI-Fake Cites Net New York Attorney, Firm $10,500 In Sanctions, 2026년 6월(기사 날짜 미확인), https://www.mealeys.com/mealeys/articles/2495458

법률신문, "허위 판결" "법원 기망"… 'AI 소설' 뿔난 판사들, 2026년 9월 12일, https://www.lawtimes.co.kr/news/articleView.html?idxno=226218

법률신문, 'AI 환각' 가짜 판례·법령 인용 대응 나선 법원… "소송비용 독박·대한변협 징계 의뢰 가능", 2026년 3월 31일, https://www.lawtimes.co.kr/news/articleView.html?idxno=218596

Damien Charlotin, AI Hallucination Cases, 2026년 10월 10일 갱신(2026년 10월 14일 열람), https://www.damiencharlotin.com/hallucinations/

medRxiv(Qazi 외), Automation Bias in Large Language Model Assisted Diagnostic Reasoning Among AI-Trained Physicians, 2025년 8월 26일, https://www.medrxiv.org/content/10.1101/2025.08.23.25334280

The Lancet Gastroenterology & Hepatology(Budzyń 외), Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study, 2025년 8월, https://www.thelancet.com/journals/langas/article/PIIS2468-1253(25)00133-5/abstract

구글 리서치, Does better work always mean better workers?, 2026년 10월 7일, https://research.google/blog/does-better-work-always-mean-better-workers/

서울신문, 30쪽 의견서도, 의처증 남편 이혼소장도… 30초면 AI로 '뚝딱'[AI의 습격-법전 대신 알고리즘], 2026년 1월 27일, https://seoul.co.kr/news/plan/AI-lawbooks-algorithm/2026/01/27/20260127006002

Bench & Bar of Minnesota(Susan M. Humiston), Ethics guidance for generative AI use, 2024년 9월, https://lprb.mncourts.gov/articles/Articles/Ethics%20guidance%20for%20generative%20AI%20use.pdf

The Legal Stack, The Legal AI Judicial Response Report 2026: How Federal and State Courts Are Actually Responding to AI-Assisted Filings, 2026년 7월 10일, https://www.thelegalstack.org/research/the-legal-ai-judicial-response-report-2026-how-federal-and

Bloomberg Law, Litigation, Comparison Table - Federal Court Judicial Standing Orders on Artificial Intelligence, 날짜 미확인, https://www.bloomberglaw.com/external/document/XCN3LDG000000/litigation-comparison-table-federal-court-judicial-standing-orde

법률신문, "AI 가지고 다툴 일 있나요"… 가이드라인 제안한 서울변회, 2024년 12월 5일, http://www.lawtimes.co.kr/news/articleView.html?idxno=203494

뉴스1, 김정욱 변협회장 "형사사법 개편, 불측의 피해 우려…검증 강화해야", 2026년 10월 14일, https://www.news1.kr/society/court-prosecution/6317993

정책브리핑, '인공지능기본법' 22일 시행…생성형 AI 결과물 '워터마크' 표시 의무, 2026년 1월 21일, https://www.korea.kr/news/policyNewsView.do?newsId=148958380

법률신문(구태언, 방석호), 인공지능기본법 시행과 사업자의 의무, 2025년 12월 9일, https://www.lawtimes.co.kr/news/articleView.html?idxno=213971

TCI Law, AI Liability: EU Commission withdraws Proposal for Directive, 2025년 2월 14일, https://www.tcilaw.de/en/ai-liability-eu-commission-withdraws-proposal-for-directive/

SVW Digital Governance Tracker, Product Liability Directive, 날짜 미확인, https://svw.no/en/svw-digital-governance-tracker/artificial-intelligence/product-liability-directive/

한국저작권위원회, 생성형 인공지능 활용 저작물의 저작권 등록 안내서, 2025년 6월 30일, https://www.copyright.or.kr/gov/nuri/view.do?brdctsno=54253

AEI(Daniel Lyons), The Risks of AI-Specific Liability Regimes, 2025년 9월 18일, https://ctse.aei.org/the-risks-of-ai-specific-liability-regimes/

Engaging Science, Technology, and Society(M. C. Elish), Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction, 2019년 3월 23일, https://estsjournal.org/index.php/ests/article/view/260

Gowling WLG, Hangzhou "AI hallucination" case - Key takeaways on duty of care for generative AI providers operating in China, 2026년 2월 4일, https://gowlingwlg.com/en-gb/insights-resources/articles/2026/hangzhou-ai-hallucination-case

데일리메디, 국내 인공지능(AI) 의료기기 허가 급증 '549건', 2026년 4월 14일, https://www.dailymedi.com/news/news_view.php?ca_id=2206&wr_id=935622

뉴시스, 식약처 "작년 26개 신약·AI 기반 의료기기 153건 허가", 2026년 4월 30일, https://www.newsis.com/view/NISX20260430_0003613396

아시아경제(다음 전재), 의료사고 나면 AI가 책임질까?…"최종 진료 판단은 의사에게", 2026년 8월 25일, https://v.daum.net/v/uha6AZd8on

한국기자협회, 언론계가 만든 '생성형 AI 준칙'… 활용 범위 명시, 2024년 12월 6일, https://www.journalist.or.kr/news/article.html?no=57255

Talking Biz News(Chris Roush), (제목 미확인, URL 기준), 2023년 5월, https://talkingbiznews.com/media-news/what-reuters-is-telling-its-journalists-about-using-artificial-intelligence/

머니투데이, 공무원용 '생성형 AI 활용 지침' 첫 배포…부적절 활용 11건 적발, 2026년 5월 22일, https://www.mt.co.kr/policy/2026/05/22/2026052208184295643

전자신문, 공무원 10명 중 7명 AI 사용…행정문서 90% 'AI가 못 읽는 포맷', 2025년 10월 13일, https://www.etnews.com/20251013000204

교육부, 교육부와 교육청이 함께 '수행평가 시, 인공지능(AI) 활용 관리 방안'을 마련하였습니다, 2025년 12월 23일, https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=294&boardSeq=104984&lev=0&m=020402

뉴스핌, 교사 63% 수업·평가에 생성형 AI 활용…평가원, 학교 가이드라인 마련, 2026년 10월 7일, https://www.newspim.com/news/view/20261007000416



3장. 지휘 능력: 오케스트레이션

해당 영역에서 문제를 찾아내는 능력

두산의 2026년 하반기 신입 지원자는 특정 상황과 과제를 받고 인공지능을 써서 30분 안에 보고서 한 편을 써야 한다(한국경제, 2026년 9월 8일). 두산이 전 직무 공통으로 처음 도입한 시험이다. 기계는 몇 분이면 보고서 초안을 내놓는다. 그러니 이 시험의 무게는 주어진 상황에서 무엇이 핵심 문제인지 가려내 기계에 지시하고 나온 초안을 검증·수정하는 데 실린다고 볼 수 있다. 도구의 도움을 받더라도 상황 속 여러 사정 가운데 어디에 초점을 두고 어떤 결론을 택할지는 지원자가 판단하고 설명해야 한다. 같은 도구를 쓰더라도 문제를 어떻게 정의하느냐에 따라 보고서의 가치가 갈린다.

수학 연구에서도 같은 관찰이 나왔다. 제러미 아비가드 등 연구자 다섯 명은 인공지능을 쓰는 수학 연구에서 사람의 노력이 처음과 끝에 몰린다고 썼다(arXiv, 2026년 8월 17일). 처음은 풀 문제를 고르는 단계이고 끝은 결과를 검토하는 단계다. 연구진은 이 두 단계가 연구 수준 수학의 병목이 되고 있다고 진단했다. 병목은 흐름 전체의 속도를 묶어 두는 좁은 구간이다. 연구진은 사람이 연구 방향만 주면 기계가 문헌에서 후보 문제를 찾아 시도하고 추천하는 방식을 시험했다. 조합론 시범에서 기계는 논문 5,245편에서 열린 추측 4,717개를 뽑았다. 그 가운데 해결 후보 598개를 만들었고 77개를 사람의 검토로 넘겼다. 무엇이 의미 있는 결과인지 가리는 마지막 판단은 사람에게 남았다.

디자인 분야의 실험은 다른 쪽에서 같은 점을 보여 준다. 문제 틀은 무엇을 문제로 보고 어디까지를 풀 범위로 삼을지 정한 방식이다. Joongi Shin 등 연구자 네 명은 참가자 280명에게 디자인 문제의 틀을 다시 짜게 했다(CHI 2025, arXiv 2025년 3월 3일). 언어 모델을 쓴 집단에서 문제 틀의 질은 나아지지 않았다. 언어 모델은 숙련 디자이너와 비숙련 디자이너의 격차를 오히려 키웠다. 비숙련자는 자기가 일을 이끈다는 느낌을 덜 받았다.

이 연구들이 인공지능은 문제를 찾지 못한다고 증명하는 것은 아니다. 다만 여기서 살펴본 수학과 디자인 사례에서는 사람이 풀 문제를 정하고 기계의 결과를 판단했다. 이 사례들이 교육에 주는 과제는 문제를 정의하는 능력을 기르는 일이다. 이정동은 『최초의

질문』(민음사, 2022년)에서 남이 낸 문제를 잘 푸는 능력과 스스로 문제를 내는 능력을 나눴다. 그가 말하는 최초의 질문은 기존 분야의 모범과 다른 규범을 내놓으려는 질문이다. 그가 다룬 것은 국가 기술 전략이지만 이 구분은 에이전트를 지휘하는 일에도 적용된다. 필자는 전자신문 연재에서 말로 업무용 앱을 만든 직장인들의 사례를 들고 이렇게 썼다. "이 사례들의 주인공은 개발자가 아닙니다. 문제를 발견한 사람입니다." (전자신문, 2026년 4월 7일)

여러 인공지능과 에이전트에게 일을 나누어 맡기는 능력

오케스트레이션이라는 말은 업계 문서에서 두 뜻으로 쓰인다. 하나는 사람이 여러 에이전트에게 일을 나누어 맡기고 결과를 모으는 일이다. 다른 하나는 관리자 역할의 에이전트가 하위 에이전트들을 부리는 소프트웨어 구조다. 우리가 말하는 지휘는 앞의 뜻이다. 그래도 뒤의 구조를 알면 사람이 할 일이 잘 보인다. 다중 에이전트 체계는 여러 에이전트가 한 과제를 나누어 처리하도록 짠 체계다.

오픈AI는 에이전트 설계 지침서에서 다중 에이전트를 두 유형으로 나눴다(오픈AI, 2025년). 관리자형에서는 관리자 에이전트가 전문 에이전트 여럿을 도구처럼 부른다. 분산형에서는 에이전트들이 동료처럼 서로 일을 넘긴다. 지침서는 먼저 에이전트 하나의 능력을 최대한 끌어내라고 권했다("maximize a single agent's capabilities first"). 앤트로픽은 2025년 6월 13일 자사 연구용 다중 에이전트 체계의 구조를 공개했다. 지휘 역할의 Claude Opus 4가 Claude Sonnet 4 하위 에이전트들에게 조사를 나눠 맡기는 구조다. 이 체계는 내부 평가에서 Opus 4 하나로 일할 때보다 90.2퍼센트 높은 성적을 냈다. 토큰 사용량은 모델이 읽고 쓴 글의 양이며 그대로 비용이 된다. 이 체계는 일반 대화보다 토큰을 약 15배 썼다(앤트로픽, 2025년 6월 13일). 앤트로픽은 하위 에이전트에게 목표와 출력 형식과 도구와 작업 경계를 적어 주지 않으면 에이전트들이 일을 겹치거나 빈틈을 남긴다고 적었다("duplicate work, leave gaps").

마이크로소프트의 에이전트 개발 문서는 지휘 방식을 다섯 패턴으로 정리했다(마이크로소프트, 2026년 2월 12일). 순서대로 넘기기와 동시에 돌리기와 상황에 따라 제어권 넘기기가 있다. 한 대화방에서 함께 논의하기와 관리자 에이전트가 그때그때 일을 배분하기도 있다. 사람이 쓰는 도구에도 이 구조가 들어왔다. 클로드 코드의 에이전트 팀 기능에서는 세션 하나가 팀장이 되어 일을 나눠 주고 결과를 합친다. 2026년 10월 14일 기준 이 기능은 실험 단계라 기본값으로 꺼져 있다. 문서는 팀원 3명에서 5명으로 시작하라고 권한다(클로드 코드 문서, 2026년 10월 14일 열람).

기업의 움직임은 예측과 조사 결과를 나눠 읽어야 한다. 가트너는 2026년까지 기업용 애플리케이션의 최대 40퍼센트가 과제 특화 에이전트를 품을 것으로 예측했다(가트너, 2025년 8월 26일). 2025년에는 그 비율이 5퍼센트에 못 미쳤다. 맥킨지가 2025년 6월과 7월 1,993명에게 물은 조사에서는 응답 조직의 62퍼센트가 에이전트를 적어도 시험하고 있었다(맥킨지, 2025년 11월 5일). 업무 기능별로 보면 에이전트를 넓혀 쓰는 비율이 10퍼센트를 넘는 기능은 하나도 없었다. 한국에서는 삼성SDS가 2026년 5월 업무 목적에 맞는 에이전트를 골라 묶는 패브릭스 2.0을 발표했다(뉴스핌, 2026년 5월 29일). SK AX는 3월 여러 에이전트가 부딪히지 않고 협업하도록 조율하는 브랜드 엑스젠틱와이어를 내놓았다(조선일보, 2026년 3월 19일). LG CNS의 개발 플랫폼에서는 설계와 코딩과 테스트를 맡은 에이전트 세 종류가 이어서 일하고 사람은 결과를 검토해 승인한다(아주경제, 2026년 6월 8일). 네이버는 사내 기술 행사에서 비서실장 역할의 에이전트가 부서 에이전트 10개를 조율하는 사내 체계를 소개했다(파일테크 전재 네이버 D2 요약, 날짜 미표기). 삼성SDS 신계영 부사장은 여러 에이전트가 협업하면 "권한에 너무 복잡한 경우의 수가 생긴다"고 말했다(이투데이, 2026년 10월 7일).

에이전트의 자기조율이 발전할수록 사람은 모든 세부 작업을 직접 배분하기보다 전체 목표와 에이전트가 결정해도 되는 범위, 완료 여부를 판단할 기준을 정해야 한다. 결과가 충돌하면 무엇을 기준으로 선택하고 어느 상황에서 작업을 멈출지도 정해야 한다. 여러 도구를 연결하는 기술은 발전해도 그 기술로 무엇을 이루려는지 설명하고 허용 범위를 책임지는 능력은 지휘 교육의 중심에 있어야 한다.

맡긴 일의 결과를 조율하고 합치는 능력

코딩 에이전트 데빈을 만든 코그니션의 월든 얀은 두 하위 에이전트에게 플래피 버드 게임의 부품을 나눠 맡긴 예를 들었다(코그니션, 2025년 6월 12일). 두 에이전트는 각자 제 몫을 만들었지만 합쳐 놓은 부품은 겉모습이 서로 어긋났다. 그는 모든 행동에는 드러나지 않은 결정이 담긴다고 썼다("Actions carry implicit decisions"). 서로 부딪히는 결정은 나쁜 결과를 낳는다.

UC 버클리 연구진은 다중 에이전트 체계가 왜 실패하는지 실행 기록으로 따졌다(arXiv, 2025년 3월). 전문가들이 7개 프레임워크의 기록에 주석을 달았다. 연구진은 이를 바탕으로 실패 분류 체계 MAST를 만들었다. 실패는 14개 유형이고 연구진은 이를 세 범주로 묶었다. 첫째는 명세와 체계 설계의 문제이고 둘째는 에이전트 사이의 불일치이며 셋째는 과제 검증의 문제다. 세 범주가 실패의 약 3분의 1씩을 차지했다. 알고 있는 정보를 넘기지 않거나 다른 에이전트의 입력을 무시하는 일이 불일치에 든다.

테스트 에이전트가 코드가 컴파일되는지만 보고 제대로 돌아가는지는 확인하지 않은 일은 검증 실패의 예다. 세 범주는 앞의 두 소절과 이 소절이 다루는 일과 겹친다. 문제를 잘못 정의하면 명세가 흔들린다. 일을 잘못 나누면 에이전트끼리 어긋나고 결과를 덜 따지면 검증에서 구멍이 난다.

합치는 일의 병목은 사람의 검토에 걸린다. 코드 리뷰는 다른 개발자가 쓴 코드를 반영하기 전에 읽고 따지는 절차다. He 등 연구진은 한 기업이 인공지능으로 생산량을 두 배로 늘리라고 지시한 뒤의 변화를 2024년 1월부터 2026년 4월까지 추적했다(arXiv, 2026년 7월 2일). 개발자 802명이 낸 코드 변경 요청 196,212건이 분석 대상이었다. 1인당 병합된 변경 요청은 기준선의 2.09배가 됐고 검토자 한 사람이 떠안은 부담도 약 두 배가 됐다. 자동 검토가 사람의 검토보다 많아졌고 병합률과 되돌림률은 안정적이었다. 평가 기관 METR이 2025년 7월 발표한 실험에서는 숙련 오픈소스 개발자 16명이 과제 246건을 풀었다. 인공지능 도구를 쓸 때 과제 시간은 19퍼센트 더 걸렸다(The Decoder, 2025년 7월 11일). 개발자들은 스스로 20퍼센트 빨라졌다고 느꼈다. 만드는 일은 싸졌고 확인하는 일은 그만큼 싸지지 않았다.

에이전트끼리 알아서 조율하게 되면 사람의 지휘는 곧 필요 없어진다는 시각도 있다. Giordano De Marzo 등 연구진의 논문 제목은 「인공지능 에이전트는 사람의 규모를 넘어 조율할 수 있다」이다(arXiv, 2025년 4월 개정판). 이들은 언어 모델 에이전트가 외부 지시 없이 집단을 이룬다고 보고했다. 조율할 수 있는 집단의 크기는 모델의 언어 능력에 따라 지수적으로 커졌다. 같은 논문은 집단이 일정 크기를 넘으면 조율이 무너진다며 자기조직화에 내재한 한계가 있다고 결론지었다. 딜로이트도 지금의 다중 에이전트 체계는 사람이 과정 안에 있을 때 더 잘 돌아간다고 적었다(딜로이트, 2025년 11월 18일). 여러 에이전트가 사람의 지휘 밖에서 서로 연락하며 일을 벌인 일은 2026년 7월 오픈AI 내부 평가 사고에서 실제로 벌어졌다. 자기조율의 범위가 넓어지는 흐름은 맞지만 2026년 10월 기준 공급사 문서와 연구는 모두 사람의 지휘와 검토를 전제로 한다.

필자는 2025년 11월 연재에서 여러 인공지능에 같은 질문을 던지라고 권하며 이렇게 썼다. "여러 AI의 답변을 메모장에 모아 비교한 후, 통합 정리하는 방법이 효과적이다." (전자신문, 2025년 11월 18일) 일상적인 질문에서 여러 모델의 답을 비교해 하나로 묶는 이 습관은, 여러 에이전트가 각자 만든 하위 결과물 사이의 어긋남을 찾아내고 전체 목표에 맞게 결합하는 지휘 능력의 출발점이 된다.

지휘자의 실력은 무엇으로 드러나는가

역량 틀은 학습자가 무엇을 할 줄 알아야 하는지를 영역과 단계로 나눠 적은 기준표다. 유네스코는 2024년 9월 낸 학생용 AI 역량 틀에서 네 측면을 이해, 적용, 창조의 세 단계로 나눴다(유네스코, 2024년 9월). 유럽연합 집행위원회와 OECD가 2026년 6월 19일 확정한 AILit 틀은 인공지능과 관계 맺기, 함께 만들기, 관리하기, 설계하기의 네 영역을 두었다(유럽연합 집행위원회, 2026년 6월 19일). 지휘에 가까운 것은 유네스코의 창조 단계와 AILit의 관리 영역이다. 두 틀 모두 초중등 학생용이다. 성인이나 신입사원의 지휘 능력을 재는 공식 틀은 아직 없다. 한국에서는 황주성 서울과학기술대 교수가 KISDI AI Outlook에서 인간중심 AI 리터러시를 세 역량으로 나눴다(이투데이, 2026년 10월 2일). 목적지향적 역량은 문제를 찾는 일에 닿는다. 행위주체적 역량은 무엇을 맡기고 무엇을 남길지 정하는 일에 닿고 인식론적 역량은 결과를 비판적으로 따지는 일에 닿는다.

생산성 격차 연구의 결과는 두 방향으로 나온다. 하버드 경영대학원과 BCG가 2023년 컨설턴트 758명을 대상으로 한 현장 실험에서 GPT-4를 쓴 사람은 기계의 능력 안에 있는 과제를 12.2퍼센트 더 많이 끝냈고 25.1퍼센트 더 빨리 끝냈다(Dell'Acqua 외, 2023년 9월). 성과가 평균 아래인 사람은 43퍼센트 나아졌고 평균 위인 사람은 17퍼센트 나아져 격차가 줄었다. 기계의 능력 밖에 있는 과제에서는 인공지능을 쓴 사람의 정답률이 19퍼센트포인트 낮았다. 같은 연구진이 P&G 전문직과 한 후속 실험에서는 인공지능을 쓴 개인이 인공지능 없는 팀만큼 성과를 냈다(NBER, 2025년 4월). 성인 1,174명을 대상으로 한 2026년 실험에서는 학력에 따른 성과 격차가 0.548 표준편차에서 0.139로 줄었다(arXiv, 2026년 8월 4일). 그런데 고학력자가 도구를 더 효과적으로 썼다. 도구를 치우자 격차는 다시 벌어졌다. 앞서 본 디자인 실험에서는 언어 모델이 숙련자와 비숙련자의 격차를 키웠다. 정답이 정해진 과제에서는 격차가 준다. 문제를 정하고 결과를 판단하는 과제에서는 격차가 남거나 커진다. 지휘자의 실력은 그 두 번째 자리에서 드러난다.

마이크로소프트는 2025년 작업동향지수에서 에이전트 상사(agent boss)라는 말을 썼다(마이크로소프트, 2025년 4월 23일). 에이전트를 만들고 일을 맡기고 관리해 자기 일의 효과를 키우는 사람이다. 2026년 보고서에는 이 말 대신 프런티어 프로페셔널이라는 집단이 나온다(마이크로소프트, 2026년 5월 5일). 다단계 작업에 에이전트를 쓰고 작업 흐름을 정기적으로 다시 짜는 사람들이다. AI를 쓰는 지식노동자 2만 명 가운데 16퍼센트가 여기에 들었다. 이들의 53퍼센트는 무엇을 AI에 맡기고 무엇을 사람이 할지

멈춰서 따진다고 답했다. 나머지 응답자에서는 이 비율이 33퍼센트였다. AI 도구를 파는 회사의 조사라는 점은 감안해야 한다.

이 자료를 근거로 지휘는 소수 전문가의 일이고 일반 교육의 목표가 될 수 없다고 반론할 수 있다. 같은 보고서는 AI 효과의 67퍼센트를 조직 요인이 설명하고 32퍼센트를 개인 요인이 설명한다고 적었다. 앞에서 본 토큰 비용도 다중 에이전트가 비용을 많이 들일 만한 업무에만 적합하다는 공급사 자신의 단서다. 이 반론이 겨냥하는 것은 여러 에이전트를 돌리는 기술이다. 문제를 고르고 일을 나누고 결과를 따지는 판단은 에이전트 하나를 쓸 때도 필요하다. 격차 연구가 보여 주듯 그 판단이 없으면 도구는 격차를 키운다. 필자는 "AI를 만드는 능력은 소수에게 집중되더라도, AI로 기획하는 능력은 모두에게 열려 있어야 합니다"라고 썼다(전자신문, 2026년 4월 7일).

지휘 능력을 평가할 때도 도구 수나 산출물의 양만 세어서는 안 된다. 어떤 문제를 골랐고 왜 그 일을 맡겼는지, 결과가 목표에 맞는지 어떤 근거로 판단했는지 확인해야 한다. 학생이 이 선택을 설명하고 오류가 생겼을 때 작업 계획을 고칠 수 있는지가 평가의 기준이 되어야 한다.

기업이 신입사원에게 묻기 시작한 것

신입 채용 자체가 줄어드는 가운데 남은 자리에서 기업이 묻는 질문이 바뀌었다. 두산 관계자는 30분 보고서 시험에서 "적절한 프롬프트를 구성하고, 생성된 결과를 검토·개선해 최종 결과물을 도출하는 역량"을 본다고 말했다(한국경제, 2026년 9월 8일). SK하이닉스는 2026년 8월 신입 수시채용에서 20분에서 30분이던 면접을 반나절 심층면접으로 바꿨다(SK하이닉스 뉴스룸, 2026년 7월 30일). 자기소개서에는 인공지능을 써서 문제를 해결한 경험을 묻는 문항을 넣었다. 기아는 하반기 신입 공채 30개 부문에서 인공지능 문제해결력을 검증한다(소비자가 만드는 신문, 2026년 9월 18일). 같은 보도에 따르면 HD현대는 생성형 AI로 과제를 풀고 그 결과를 비판적으로 판단하는 능력을 보는 검사를 처음 도입했다. LG CNS도 생성형 AI 역량 평가를 새로 넣었다. LG전자는 자기소개서에 AI 경험 항목을 더했다(한국경제, 2026년 9월 8일).

은행의 질문은 더 구체적이다. KB국민은행은 하반기 신입 1차 면접에서 AI역량검정을 치르고 AI 이해와 기술, 문제 해결과 실무, 비판적 사고와 평가, 윤리와 책임을 본다(아시아경제, 2026년 9월 17일). 신한은행은 자기소개서를 없애고 AI 협업면접에서 지원자가 과제를 프롬프트로 풀고 결과를 검증하는 과정을 본다. 평가 업체 그렙의 AI 역량평가는 AI에 맡길 업무를 판단하는 능력, AI와 주고받으며 원하는 결과를 얻는 능력,

결과물을 검증하는 능력, 보안과 윤리 기준을 지키는 능력을 본다(벤처스퀘어, 2026년 10월 12일). 그렙은 최종 결과물과 함께 프롬프트 기록을 살핀다. 기업은 어떤 문제에 AI를 썼는지 묻는다. AI에게 어떻게 시켰는지와 결과를 어떻게 확인하고 고쳤는지도 묻는다. 앞의 세 소절이 다룬 순서와 같다.

조사 수치도 같은 흐름이다. 대한상공회의소가 2025년 하반기 기업 500여 곳에 물었을 때 69.2퍼센트가 채용에서 AI 역량을 고려한다고 답했다(한경잡앤조이, 2025년 12월 29일). 한국경영자총협회의 2026년 조사에서는 채용 과정에 AI를 쓰는 일이 늘었다는 응답이 30.6퍼센트였다(아시아경제, 2026년 3월 10일). 경총 조사의 AI는 지원자의 역량보다 채용 도구를 가리킨다. 미국에서는 2025년 12월 구인 공고의 4.2퍼센트가 AI를 언급했다(인디드 하이어링 랩, 2026년 1월 22일).

AI 역량 요구는 채용 유행어일 뿐이라는 시각에도 근거가 있다. 미네소타주 고용경제개발부는 2026년 4월 AI를 언급한 공고 가운데 60퍼센트가 'AI'나 'artificial intelligence'라는 말만 적었다고 분석했다(미네소타주 고용경제개발부, 2026년 9월). 특정 기술을 요구하기보다 막연한 의지를 밝힌 문구에 가깝다는 해석이다. 스트라다 교육재단이 미국 경영진과 인재 담당자 약 1,500명에게 물은 조사에서 신입 역량의 중요도는 비판적 사고와 소통이 5점 만점에 4.3, AI 리터러시가 3.6이었다(스트라다 교육재단, 2026년 5월 19일). 공고 문구는 유행어일 수 있어도 두산의 보고서 시험이나 신한은행의 협업면접 같은 평가 과제는 구체적이다. 그 과제가 재는 것도 문제를 풀고 결과를 검증하는 기본기다.

타오가 제시한 변화: 기초 훈련과 인공지능 연구의 결합

네이처는 2026년 4월 27일 테렌스 타오 인터뷰를 실었다(네이처, 2026년 4월 27일). 제목은 「직무 설명서가 바뀌고 있다」('The job description is changing')였다. 이 인터뷰를 소개한 2차 요약본들에는 타오가 AI를 무시하는 대학원생은 기회를 놓칠 수 있다고 말했다는 문장이 돌아다니지만, 본문이 유료 장벽에 막혀 있어 이 문장은 타오의 실제 발언 원문으로 확인되지 않는다. 공개된 머리글에서 확인되는 사실은 많은 수학자가 AI의 능력이 부풀려졌다고 보지만 지난 1년 사이 AI가 연구 수학자의 일상 업무에 실제로 유용해졌다는 진단이다.

타오가 직접 쓴 글에서 확인되는 입장은 더 조심스럽다. 그는 2026년 3월 29일 블로그 글에 4월 18일 단 댓글에서 학생에게 문제의 답을 AI에 바로 묻지 말라고 했다(타오 블로그, 2026년 4월 18일 댓글). 그런 습관은 앞으로 혼자 문제를 푸는 능력을 해친다는

이유였다("will compromise your ability to solve such problems by yourself"). 그가 권한 쓰임새는 자기 작업의 검증이다. 스스로 세운 증명 전략을 AI에 내놓고 비판하게 하는 레드팀 방식이다. 레드팀은 일부러 반대편에 서서 약점을 찾는 역할이다. 기준은 AI가 낸 결과를 수업 발표에서 스스로 설명하고 질문에 답할 수 있는가다. UCLA 매거진 인터뷰에서도 그는 학생이 주제를 모르면 답이 맞는지 확인할 수 없다고 했다(UCLA 매거진, 2026년 10월 6일). 연구자인 자신은 AI에게 한 주제의 참고문헌을 모아 보고서로 받는다고 했다.

10월 9일 칼텍 강연 「Math 2.0」에서 타오는 수학이 증명이 귀하던 시대에서 증명이 넘치는 시대로 넘어간다고 말했다(타오 강연 슬라이드, 2026년 10월). 슬라이드는 문제 풀이가 더 가치 있는 장기 목표의 대리 지표일 뿐이라고 적었다("only a proxy for more valuable long-term goals"). 대리 지표는 직접 측정하기 어려운 목표를 대신해 측정하는 수치다. 증명이 귀하던 시대에는 이 대리 지표가 목표와 맞았지만 지금은 어긋났다는 것이 그의 진단이다. 문제 풀이만 맹목적으로 최적화하면 이제 해롭다는 문장도 슬라이드에 있다. 그는 열린 문제를 도달 자체가 목적인 곳으로 보지 않고 방향을 알려 주는 등대에 견줬다.

타오는 연구에서 AI를 적극 쓰면서도 학생의 기초 훈련을 AI에 맡기는 일을 경계한다. 두 입장은 지휘 능력의 같은 조건을 가리킨다. 풀 가치가 있는 문제를 고르려면 그 분야를 알아야 하고, 기계가 돌려준 증명을 검증하려면 혼자 문제와 씨름해 본 경험이 있어야 한다. 옛 방식만 고집하는 학생의 기회가 줄어든다는 문장은 그의 글에서 확인되지 않는다. 확인되는 것은 기초 훈련 없이 답을 맡기면 스스로 푸는 힘을 잃는다는 경고다. 타오는 칼텍 강연 슬라이드의 맺음말에서 Math 2.0으로 가는 길의 주된 병목은 기술도 제도 문화도 아니고 상상력이라고 적었다. 학교는 이 두 요구를 함께 받아들여야 한다. 혼자 생각하는 기초를 기르고, 그 기초로 인공지능과 함께 풀 가치가 있는 문제를 찾게 해야 한다.

참고 자료

한국경제, 기업들 하반기 채용 키워드는 'AX 역량', 2026년 9월 8일, https://www.hankyung.com/article/2026090819681

arXiv(Zheng, Zhang, Avigad, Tetali, Welleck), The Problem Is the Problem: Towards Scalable Mathematical Discovery, 2026년 8월 17일, https://arxiv.org/abs/2608.16977

arXiv(Shin, Polyanskaya, Lucero, Oulasvirta), No Evidence for LLMs Being Useful in Problem Reframing, 2025년 3월 3일, https://arxiv.org/abs/2503.01631

예스24(이정동, 민음사), 최초의 질문: 기술 선진국의 조건, 2022년 4월 15일, https://www.yes24.com/Product/Goods/108839781

전자신문, [김경진의 AI전략노트] 〈25〉 기획하는 사람이 이기는 시대가 왔다, 2026년 4월 7일, https://www.etnews.com/20260407000157

오픈AI, A practical guide to building agents, 2025년(문서에 발행일 표기 없음), https://cdn.openai.com/business-guides-and-resources/a-practical-guide-to-building-agents.pdf

앤트로픽, How we built our multi-agent research system, 2025년 6월 13일, https://www.anthropic.com/engineering/multi-agent-research-system

마이크로소프트, Workflow orchestrations in Agent Framework, 2026년 2월 12일(2026년 8월 25일 갱신), https://learn.microsoft.com/en-us/agent-framework/workflows/orchestrations/

클로드 코드 문서, Orchestrate teams of Claude Code sessions, 2026년 10월 14일 열람, https://code.claude.com/docs/en/agent-teams

가트너, Gartner Predicts 40% of Enterprise Apps Will Feature Task-Specific AI Agents by 2026, Up from Less Than 5% in 2025, 2025년 8월 26일, https://www.gartner.com/en/newsroom/press-releases/2025-08-26-gartner-predicts-40-percent-of-enterprise-apps-will-feature-task-specific-ai-agents-by-2026-up-from-less-than-5-percent-in-2025

맥킨지, The state of AI in 2025: Agents, innovation, and transformation, 2025년 11월 5일, https://www.mckinsey.com/~/media/mckinsey/business%20functions/quantumblack/our%20insights/the%20state%20of%20ai/the-state-of-ai-in-2025-vf.pdf

딜로이트, Unlocking exponential value with AI agent orchestration, 2025년 11월 18일, https://www.deloitte.com/us/en/insights/industry/technology/technology-media-and-telecom-predictions/2026/ai-agent-orchestration.html

뉴스핌, 삼성SDS, 10월 '패브릭스 2.0' 공개…AI 에이전트 고도화, 2026년 5월 29일, https://www.newspim.com/news/view/20260529000810

조선일보(네이트 전재), SK AX, 여러 AI 에이전트 지휘하는 브랜드 내놔, 2026년 3월 19일, https://m.news.nate.com/view/20260319n26384

아주경제, LG CNS, 에이전틱 AI 개발 플랫폼 출시…"대규모 시스템 구축 자동화", 2026년 6월 8일, https://www.ajunews.com/view/20260608081617533

파일테크(네이버 D2 전재), AI 에이전트 회사 차리기: 설립부터 어디서든 동기화까지, 날짜 미표기, https://piletech.kr/articles/111697

이투데이(서플 전재), [인터뷰] 신계영 삼성SDS 부사장 "에이전트 늘수록 권한 복잡해져…통합 거버넌스 중요", 2026년 10월 7일, https://supple.kr/news/cmux3yof1004l4bovk9fwhad3

코그니션(Walden Yan), Don't Build Multi-Agents, 2025년 6월 12일, https://cognition.com/blog/dont-build-multi-agents

arXiv(Cemri 외), Why Do Multi-Agent LLM Systems Fail?, 2025년 3월 17일, https://arxiv.org/abs/2503.13657

arXiv(Hao He 외), AI Writes Faster Than Humans Can Review: A Longitudinal Study of an Enterprise 2x Mandate, 2026년 7월 2일, https://arxiv.org/abs/2607.01904

The Decoder, AI coding can make developers slower even if they feel faster, 2025년 7월 11일, https://the-decoder.com/ai-coding-can-make-developers-slower-even-if-they-feel-faster/

arXiv(De Marzo, Castellano, Garcia), AI agents can coordinate beyond human scale, 2025년 4월 15일(4판), https://arxiv.org/abs/2409.02822

전자신문(다음 전재), [김경진의 AI전략노트] 〈15〉 언어모델 인공지능을 효율적으로 잘 사용하려면, 2025년 11월 18일, https://v.daum.net/v/20251118160307876

유네스코, AI competency framework for students, 2024년 9월, https://www.unesco.org/en/articles/ai-competency-framework-students

유럽연합 집행위원회, Launch of the EU-OECD AI Literacy Framework for primary and secondary education, 2026년 6월 19일, https://ec.europa.eu/newsroom/eacea_oep/items/944140/

이투데이, "AI 잘 쓰는 법만 가르쳐선 부족"…한국 AI교육 '활용 역량' 치중, 2026년 10월 2일, https://www.etoday.co.kr/news/view/2631840

하버드 경영대학원 워킹페이퍼 24-013(Dell'Acqua 외), Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality, 2023년 9월, https://mitsloan.mit.edu/sites/default/files/2023-10/SSRN-id4573321.pdf

NBER(Dell'Acqua 외), The Cybernetic Teammate: A Field Experiment on Generative AI Reshaping Teamwork and Expertise, 2025년 4월, https://www.nber.org/papers/w33641

arXiv(Cruces 외), Does generative AI narrow education-based productivity gaps? Evidence from a randomized experiment, 2026년 8월 4일, https://arxiv.org/abs/2608.04198

마이크로소프트, 2025: The year the Frontier Firm is born, 2025년 4월 23일, https://www.microsoft.com/en-us/worklab/work-trend-index/2025-the-year-the-frontier-firm-is-born

마이크로소프트, Agents, human agency, and the opportunity for every organization, 2026년 5월 5일, https://www.microsoft.com/en-us/worklab/work-trend-index/agents-human-agency-and-the-opportunity-for-every-organization

SK하이닉스 뉴스룸, "스펙 보다 '생각 근육' 본다"… AI 시대 맞춤형 인재 선발 SK하이닉스, '반나절 심층면접' 도입, 2026년 7월 30일, https://news.skhynix.co.kr/ai-talent-recruit-2026-02/

소비자가 만드는 신문, "AI 서투르면 취업 못해요"...대기업들 모든 직무서 'AI 활용 역량'을 채용 핵심 잣대로, 2026년 9월 18일, https://www.consumernews.co.kr/news/articleView.html?idxno=764130

아시아경제, 본격화한 은행권 하반기 채용…핵심 키워드는 'AI', 2026년 9월 17일, https://view.asiae.co.kr/article/2026091709154136511

벤처스퀘어, 그렙, AI 활용 능력으로 은행 PT 면접 대체…역량평가 도입 16곳·연내 4만 명 응시 전망, 2026년 10월 12일, https://www.venturesquare.net/1119092

한경잡앤조이, '2026년 채용트렌드'…'소규모 질적 채용' 전환·AI 잘 활용하는 인재 선호, 2025년 12월 29일, https://magazine.hankyung.com/job-joy/article/202512290531d

아시아경제, 기업 66.6% "신규채용 계획 있음"… 작년보다 5.8%P 증가, 2026년 3월 10일, https://view.asiae.co.kr/article/2026031008384903376

인디드 하이어링 랩, January 2026 US Labor Market Update: Jobs Mentioning AI Are Growing Amid Broader Hiring Weakness, 2026년 1월 22일, https://hiringlab.indeed.com/2026/01/22/january-labor-market-update-jobs-mentioning-ai-are-growing-amid-broader-hiring-weakness/

미네소타주 고용경제개발부, Artificial Intelligence Mentions in Job Postings, 2026년 9월, https://mn.gov/deed/newscenter/publications/trends/september-2026/ai-job-postings.jsp

스트라다 교육재단, Entry-Level Hiring in the AI Era: What Employers Are Thinking (and Doing), 2026년 5월 19일, https://www.strada.org/news-insights/entry-level-hiring-in-the-ai-era-what-employers-are-thinking-and-doing

네이처, 'The job description is changing': mathematician Terence Tao on the rise of AI, 2026년 4월 27일, https://www.nature.com/articles/d41586-026-01246-9

타오 블로그, Mathematical methods and human thought in the age of AI, 2026년 3월 29일(4월 18일 댓글), https://terrytao.wordpress.com/2026/03/29/mathematical-methods-and-human-thought-in-the-age-of-ai/

UCLA 매거진, The Tao of Tao, 2026년 10월 6일, https://newsroom.ucla.edu/magazine/terence-tao-math-artificial-intelligence-alpha-research

타오 강연 슬라이드(칼텍, 2026년 10월 9일 강연), Math 2.0, 2026년 10월, https://teorth.github.io/tao-web/slides/math-2-0-caltech-2026.pdf



4장. 목표가 바뀌면 평가가 바뀐다

인공지능 사용 비율로 당락을 가르는 현행 심사

성신여자대학교 대학원은 학위청구논문 심사를 통과한 학생에게 카피킬러 검사 결과확인서를 받는다. 대학원이 안내문에 붙인 견본은 2025년 2월 19일에 발급한 것이다. 문서명 칸에는 '가나다라마바사아자차카타파하.hwpx'라는 견본용 파일 이름이 찍혀 있다. 평가유형 칸에는 카피킬러와 GPT킬러가 나란히 있다. 둘째 쪽 분석 정보 줄은 표절률과 동일 문장과 의심 문장 옆에 'GPT 의심 문장' 칸을 둔다. 남의 글과 겹친 문장과 인공지능이 썼다고 의심되는 문장이 한 서식의 한 줄에 함께 놓인다.

같은 안내문은 유사도 기준의 예시로 네 등급을 적었다. 0퍼센트 이상 10퍼센트 미만은 양호, 15퍼센트 미만은 유의, 20퍼센트 미만은 주의, 그 이상은 위험이다. 양호가 아니면 지도교수가 사유를 적어 낸다(성신여대 대학원, 2017년 12월 대학원위원회 근거). 중앙대 대학원의 「학위수여논문 유사도 검사 내규」(2015년 4월 1일 제정)도 같은 네 등급을 쓴다. 이 내규에서 20퍼센트를 넘긴 논문을 통과시키려면 심사위원장이 소견서를 써야 한다. 30퍼센트를 넘기면 대학원위원회가 졸업 여부를 정할 수 있다.

대학의 인공지능 지침에는 이런 숫자가 없다. 서울대 「서울대학교 AI 가이드라인」(2026년 1월 1일 제정)은 교수자에게 탐지 도구의 불완전성을 이해하라고 적었다. 고려대학부대학의 「AI 활용 가이드라인」(2025년 9월 12일)은 탐지 결과를 참고자료로만 쓰고 성적이나 징계의 단독 근거로 삼지 않는다고 정했다. 계명대 일반대학원 지침(2026년 10월 제정) 제4조는 허용 여부를 "사용 비율이나 AI 탐지율"로 가리지 않는다고 못 박았다. 교육부와 한국대학교육협의회가 2026년 8월 21일 낸 대학 지침도 탐지 결과를 징계의 결정적 증거로 인정하지 않게 했다(스포트라이트유, 2026년 8월 26일).

현장의 사정은 조금 다르다. 국공립대 가운데 국립강릉원주대와 인천대 등 5곳은 무하유의 GPT킬러를 쓴다고 국회에 답했다(서울경제, 2025년 10월 17일). 한 비수도권 국립대 교수는 "제출된 과제물을 우선 조교에게 넘겨 AI 탐지기에 돌리고 있다"고 말했다(시사IN, 2025년 12월 10일). 기사는 이 수업이 표절률 20퍼센트 이상이면 바로 0점을 준다고 서술했다. GPT킬러 이용은 2025년 10월 한 달 약 64만 7,000건으로 한 해 전 같은 달의 3.6배였다(같은 기사, 무하유 집계).

중국은 숫자를 규정에 넣었다. 2024년 6월 기준 최소 15개 대학이 학위논문의 인공지능 생성 비율 검사를 발표했고 합격선은 20퍼센트 미만에서 40퍼센트 사이였다(Sixth Tone, 2024년 6월 5일). 헤이룽장대 대학원은 비율이 40퍼센트에서 70퍼센트면 6개월 이상 고친 뒤 다시 심사하고, 70퍼센트를 넘으면 학위 신청 자격을 없앤다(36氪, 2026년 5월 26일).

한국의 지침은 비율을 말하지 않는다. 그런데도 표절 검사의 등급표와 검출 점수는 현장에서 당락의 기준처럼 쓰이고, 중국은 그 기준을 명문으로 옮겼다. 교육의 목표가 협업과 지휘라면 기계가 쓴 문장의 양은 학생의 능력을 거의 말해 주지 않는다. 우리는 사용 비율을 따지는 심사를 목표가 잘못 선 결과로 본다.

표절 검사의 틀을 그대로 가져온 것이 왜 잘못인가

유사도 검사는 제출한 글을 데이터베이스 속 기존 문서와 대조해 겹치는 부분을 세는 검사다. 중앙대 내규는 6어절 이상이나 한 문장 이상 일치하면 겹친 것으로 센다. 겹친 원문이 화면에 함께 뜨므로 심사위원은 어느 문장이 어디서 왔는지 눈으로 확인할 수 있다. 10퍼센트, 15퍼센트, 20퍼센트라는 경계는 각 대학이 내규로 정한 관행이다. 무하유는 2022년 대학 과제물 통계를 내며 15퍼센트에서 20퍼센트 이내를 안전, 31퍼센트 이상을 표절 위험으로 나눴다(AI타임스, 2023년 4월 20일). 중앙대 내규는 챗GPT가 나오기 7년 전에 만들어졌다.

인공지능 판정에는 대조할 원문이 없다. 검출 도구는 글이 얼마나 예측하기 쉬운지와 문체의 통계적 특징을 바탕으로 기계가 쓴 글인지 추정한다. 유사도는 기존 문서와 겹친 분량을 나타내지만, AI 검출 수치는 통계적 판정에 근거한다. 도구가 표시한 수치가 의심 문장의 비율인지 판정 확률인지도 구분해야 한다. 두 수치는 같은 백분율로 표시되더라도 측정 대상이 다르다. 검출 점수만으로 글쓴이가 실제로 인공지능을 썼는지 확인할 수는 없다. 스탠퍼드대 연구팀은 검출기가 예측 가능성에 기대기 때문에 어휘와 구문이 평이한 비원어민의 글을 기계 글로 본다고 설명했다(Patterns, 2023년 7월). 노터데임대 연구팀은 긴 단어와 학술 어휘가 많은 글일수록 점수가 높게 나온다고 보고했다(arXiv, 2026년 8월 6일 제출). 결과확인서가 두 칸을 한 줄에 놓으면서 심사하는 사람은 두 숫자를 같은 눈금으로 읽게 된다.

질문도 바뀌었다. 표절 검사는 남의 글을 출처 없이 자기 것으로 내놓았는지를 묻는다. 인공지능 사용에서 따질 것은 누가 어떤 판단을 했고 그 결과를 누가 책임지는지다.

기계가 문장을 많이 다듬었어도 연구 질문과 결론은 학생의 것일 수 있다. 기계가 쓴 문장이 몇 줄 안 되어도 핵심 판단을 통째로 넘겼을 수 있다.

학술 출판계는 이 차이를 일찍 받아들였다. Science는 2023년 1월 인공지능이 만든 글을 금지했다가 그해 11월 고지하는 조건으로 허용했다(Times Higher Education, 2023년 11월 16일). 저자는 커버레터와 감사의 글에 사용 사실을 적고 방법 절에 프롬프트와 도구와 버전을 밝힌다. 국제의학학술지편집인위원회(ICMJE)는 2023년 5월 권고에 인공지능의 도움을 받은 작업을 어떻게 밝힐지에 관한 지침을 더했다. 엘스비어는 2026년 6월 개정 정책에서 문법과 맞춤법 점검은 고지하지 않아도 된다고 정했다. 서울대 지침도 외국어 교정만을 위한 사용은 밝히지 않았다는 이유만으로 부정행위가 되지 않는다고 각주에 적었다. 이들 기준은 비율을 언급하지 않고 사용의 고지와 저자의 책임을 기준으로 삼았다. 사용한 양으로 위반을 가렸다면 문법 점검을 따로 빼 줄 까닭이 없다.

검출 도구의 오판과 학생들의 숨기기

검출 도구는 어떤 글을 인공지능이 썼는지 판정하는 프로그램이다. 검출 도구의 오류는 두 쪽으로 난다. 사람이 쓴 글을 기계가 썼다고 보는 오탐과 기계가 쓴 글을 사람 글로 놓치는 미탐이다. 심사에서 학생을 다치게 하는 쪽은 오탐이다.

오픈AI는 2023년 1월 31일 자체 검출기를 내놓았다. 영어 시험 자료에서 이 검출기는 인공지능이 쓴 글의 26퍼센트만 잡아냈고 사람 글의 9퍼센트를 기계 글로 잘못 봤다(The Decoder, 2023년 1월 31일). 오픈AI는 그해 7월 20일 정확도가 낮다는 이유로 이 도구를 거뒀다("no longer available due to its low rate of accuracy")(Search Engine Journal, 2023년 7월 25일). 스탠퍼드대 제임스 저우 교수 연구팀은 검출기 7종에 비원어민이 쓴 토플 에세이 91편을 넣었다(Patterns, 2023년 7월). 검출기들은 이 에세이의 평균 61퍼센트(논문 수치 61.22퍼센트)를 기계 글로 판정했다. 미국 8학년 학생의 에세이는 거의 모두 사람 글로 맞혔다. 실험한 7종에 Turnitin은 들어 있지 않았다.

Turnitin은 2023년 4월 검출 기능을 내놓으며 오탐이 1퍼센트 미만이라고 알렸다(Turnitin 보도자료, 2023년 4월 5일). 두 달 뒤 Turnitin 최고제품책임자 애니 체키텔리는 실제 사용 결과가 실험실과 다르다고 인정했다("real-world use is yielding different results from our lab")(K-12 Dive, 2023년 6월 7일). 문장 단위 오탐률은 약 4퍼센트였다(Inside Higher Ed, 2023년 6월 1일). 밴더빌트대는 2022년 한 해 Turnitin에 낸 7만 5,000편에 1퍼센트를 대입하면 약 750편이 잘못 표시될 수 있다며 이 기능을 껐다

(밴더빌트대, 2023년 8월 16일). 국내 GPT킬러의 정확도 94퍼센트와 오탐지율 0.22는 무하유가 자체 시험으로 밝힌 수치다(AI타임스, 2023년 6월 21일). 기사에는 시험 자료의 규모와 판정 기준이 나오지 않는다.

검출 도구가 불완전해도 학생이 함부로 쓰지 못하게 막는 힘이 있다는 반론이 있다. 체키텔리는 98퍼센트 확신할 때만 표시해 오탐을 줄인다며 도구를 유지했다(Turnitin 보도자료, 2023년 4월 5일). 시카고대 부스경영대학원의 브라이언 자배리언과 알렉스 이마스는 상용 검출기 팬그램(Pangram)의 오탐과 미탐이 0에 가까웠다고 보고했다(NBER, 2025년 9월). 동료 심사를 거치기 전의 논문이지만 모든 검출기가 쓸모없다는 주장을 막는 근거다. 노터데임대 연구팀의 결과는 다른 쪽을 가리킨다(arXiv, 2026년 8월 6일 제출). 2023년부터 2025년 사이 사람이 쓴 논문 초록을 팬그램은 14.9퍼센트, GPTZero는 8.9퍼센트 기계 글로 표시했다. 인공지능으로 초록을 가볍게 다듬기만 해도 38퍼센트에서 80퍼센트가 걸렸다. 연구팀은 검출 점수를 부정행위의 단독 증거로 쓰지 말라고 결론지었다. 억지력이 실제로 있는지 잰 연구는 찾지 못했다. 좋은 검출기가 있더라도 점수는 의심을 시작하는 단서일 뿐이고, 판정은 다른 증거로 해야 한다.

점수만으로 학생을 문제 삼은 대가는 이미 기록에 남았다. 호주 가톨릭대(ACU)는 2024년 9개 캠퍼스에서 학업 부정 의혹 약 6,000건을 등록했다(ABC, 2025년 10월 9일). 대학은 이 수치가 과장됐다고 반박했다. 한 구급학과 학생은 에세이의 84퍼센트가 인공지능 글로 표시된 보고서 한 장이 유일한 근거였다고 말했다. 22세 간호학과 학생은 혐의를 벗기까지 6개월이 걸렸고 그동안 성적표에 결과 보류가 찍혔다. 이 대학은 2025년 3월 Turnitin 검출 기능을 그만 쓰기로 했다. 미국 애델파이대에서는 Turnitin이 과제 전체를 기계 글로 표시한 학생이 소송을 냈다. 연방 판사는 표절 판정에 근거가 없다고 판단했다(Inside Higher Ed, 2026년 2월 11일).

학생들은 점수를 낮추는 법을 배운다. 휴머나이저는 기계가 쓴 글을 검출 도구가 사람 글로 보도록 고쳐 써 주는 도구다. 노터데임대 연구에서 휴머나이저를 거친 기계 글은 4퍼센트 미만만 걸렸다. 중국 관영 CCTV 보도에서는 한 학생이 직접 쓴 과제 계획서가 인공지능 생성률 97퍼센트로 판정됐다(京报网, 2026년 5월 21일). 학생은 인공지능으로 글을 고쳐 그 수치를 0퍼센트로 낮췄다. 비율로 심사하면 정직하게 밝히고 다듬은 학생이 위험해지고 흔적을 지운 학생이 안전해진다. 서울대는 정확도와 부작용을 이유로 GPT킬러와의 제휴를 끝냈다(이투데이, 2026년 10월 6일).

새 심사의 질문: 인공지능을 어떤 방법으로 썼는가, 착상은 어디서 나왔는가, 시작이 된 질문은 무엇이었는가

검출 점수를 내려놓으면 심사하는 사람은 다른 것을 물어야 한다. 처음 물을 것은 인공지능을 어떤 방법으로 썼는가다. 어떤 도구를 어느 단계에서 썼고, 무엇을 시켰고, 돌아온 답을 어떻게 확인하고 고쳤는지를 묻는다. 고려대 지침은 학생에게 도구명과 목적과 범위를 적고 자기 생각과 인공지능의 기여를 구분하게 했다. 미국 듀크대 교수학습센터는 일부 교수가 인공지능과 나눈 대화 기록을 부록으로 받는다고 소개했다. 플리머스주립대의 팸 할랜드는 성찰문의 첫 질문으로 프롬프트가 무엇이었는지 묻는다("What was your prompt?")(듀크대 교수학습센터).

둘째 질문은 착상이 어디서 나왔는가다. 논문의 핵심 생각이 선행 연구에서 왔는지, 현장 관찰에서 왔는지, 인공지능의 제안에서 왔는지를 밝히게 한다. 인공지능의 제안에서 출발했다고 해서 실격은 아니다. 그 제안을 왜 골랐고 무엇을 덧붙였는지 설명할 수 있어야 한다. 셋째 질문은 시작이 된 질문이 무엇이었는가다. 협업과 지휘에서 사람의 몫이 문제를 정하는 일이라면 심사도 그 문제를 누가 어떻게 세웠는지에서 출발해야 한다.

세 질문을 받는 그릇이 기여 명세다. 기여 명세는 연구의 어느 부분에 누가 어떻게 기여했는지 적는 진술이다. 사람 저자의 기여를 적는 국제 표준 CRediT에서는 인공지능 도구에 역할을 줄 수 없다(CASRAI, 2026년 8월). 대신 특정 저자가 어떤 역할에서 인공지능의 도움을 받았다고 적을 수 있다. 복단대는 2024년 학부 졸업논문 규정에서 도구 이름과 버전, 사용 날짜, 용도, 생성 내용, 영향을 받은 부분을 승낙서에 적게 했다(北京日报, 2024년 12월 30일). 학생은 인공지능이 손대기 전의 자료를 보관해 지도교수의 확인에 대비한다. 같은 중국에서도 이 규정은 비율 상한과 다른 축에 섰다. 2026년 9월에는 저자가 자기 원고를 얼마나 이해하는지 감독 아래 시험해 저자성의 증거로 삼자는 제안이 나왔다(arXiv, 2026년 9월 17일 제출).

인공지능 사용을 허용하면 평가의 공정성이 무너진다는 반론도 있다. 배충식 KAIST 총장은 2026년 8월 기자간담회에서 인문사회와 기초과학 수업은 "AI 프리 교육"으로 하겠다고 밝혔다(경향신문, 2026년 8월 25일). 그는 "AI에 (사람이) 끌려가지 않게 하는 것이 목표"라고 했다. 송기창 숙명여대 명예교수는 검색을 허용하는 시험은 학생이 아니라 인공지능을 시험하는 셈이라고 지적했다(시사IN, 2025년 12월 10일). 이 우려에는 근거가 있고, 혼자 해내는 능력을 확인할 자리는 남겨야 한다. 검출 중심의 심사도 공정하지는 않다. 영국 고등교육 옴부즈맨이 2025년 7월 공개한 인공지능 부정 의혹 사례 4건 가운데 3건은 유학생이나 제2언어 학생의 일이었다(HEPI, 2026년 7월

20일). 3건은 전부 또는 일부 학생 쪽 주장이 받아들여졌다. 세 질문은 학생이 자기 착상과 출발 질문을 말로 설명하게 한다. 그 설명은 기계가 대신해 주기 어렵다.

학위논문, 학술지, 과제 보고서의 새 심사 기준 초안

아래 초안은 저자가 제안하는 안이다. 어느 기관도 이 기준을 확정하지 않았다. 근거로 삼은 자료는 계명대의 학위논문 AI 활용 확인서, 과학기술정보통신부 「국가연구개발 AI 연구윤리 가이드」(2026년 9월 20일)의 AI 활용 여부 공개 서식, 교육부와 대교협의 대학 지침이다. 시드니대의 두 차선 접근과 AI 평가 척도(AIAS)도 참고했다. 영국 러셀그룹 24개 대학은 셋째 원칙에서 가르침과 평가를 인공지능의 윤리적 사용을 담도록 바꾸겠다고 밝혔다("adapt teaching and assessment to incorporate the ethical use") (러셀그룹, 2023년 7월 3일). 두 차선 접근은 감독 아래서 학습 성과를 확인하는 평가와 인공지능 사용을 안내하며 배우게 하는 평가를 나누는 틀이다. 시드니대가 호주 고등교육 품질기준청(TEQSA)의 원칙 위에 세웠다. AIAS는 마이크 퍼킨스 등 연구자들이 만든 5단계 허용 등급으로 '인공지능 없음'부터 '인공지능 탐색'까지 나눈다(arXiv, 2024년 12월 12일).

학위논문 심사 기준 초안

1. 심사위원회는 인공지능 생성 비율이나 검출 점수를 합격과 불합격의 기준으로 쓰지 않는다. 검출 결과는 확인을 시작하는 참고자료로만 쓴다.
2. 신청자는 확인서에 도구명과 버전, 사용 단계, 프롬프트의 요지, 결과를 검증하고 고친 방법을 적는다.
3. 신청자는 연구 질문이 어디서 시작됐는지와 핵심 착상의 출처를 한 쪽 이내로 밝힌다.
4. 구술 심사에서 심사위원은 연구 질문, 착상, 인공지능의 제안 가운데 버린 것과 그 이유를 묻는다.
5. 지도교수는 의심이 있을 때 초안과 수정 이력, 대화 기록, 분석 코드를 받아 함께 검토한다.

학술지 심사 기준 초안

1. 저자는 참고문헌 앞에 인공지능 사용 선언을 두고 도구, 버전, 사용 목적과 범위를 적는다. 문법과 맞춤법 점검은 적지 않아도 된다.

2. 저자는 연구 질문과 핵심 착상이 사람 저자에게서 나왔는지, 인공지능의 제안에서 출발했는지 밝힌다.

3. 편집위원회는 필요할 때 대화 기록과 원자료의 제출을 요구할 수 있고 저자는 응한다.

4. 심사위원은 원고를 외부 인공지능에 올리지 않고, 검출 점수를 게재 판단의 근거로 쓰지 않는다.

과제 보고서 평가 기준 초안

1. 교수자는 강의계획서에 과제별 인공지능 허용 등급을 AIAS 같은 단계 표로 적고 보고 방식을 정한다.

2. 학생은 보고서 끝에 어떤 방법으로 인공지능을 썼는지 반 쪽 이내로 적고 주요 대화 기록을 붙인다.

3. 학생은 과제를 시작한 질문과 그 질문을 고른 이유를 적는다.

4. 학기마다 과제 하나는 감독 아래 수업 중 작성이나 짧은 면담으로 학생 혼자의 이해를 확인한다.

5. 검출 점수만으로 감점하거나 징계하지 않고, 의심이 있으면 면담과 초안 비교로 소명할 기회를 준다.

6. 인공지능 사용을 정직하게 밝힌 일 자체를 감점 사유로 삼지 않는다.

세 초안 모두 사용량을 숫자로 적게 하지 않았다. 계명대 확인서는 활용 정도를 최소한의 보조, 제한적 활용, 상당한 보조의 세 단계로 고르게 한다. 이 칸도 퍼센트로 바꾸지 않고 연구설계와 분석과 결론에 미친 영향을 글로 적는 칸과 함께 읽어야 뜻이 산다. 과제 보고서 초안의 4번 항목은 시드니대의 첫째 차선에, 나머지 항목은 둘째 차선에 해당한다. 학위논문 초안의 구술 심사는 두 차선이 만나는 자리다.

평가의 초점을 바꾸면 교육의 방식이 따라 바뀐다

역류 효과(washback)는 시험이나 평가가 가르치고 배우는 방식에 거꾸로 미치는 영향이다. 시험이 수업을 바꾼다는 생각은 교육계의 오랜 통념이다. 언어평가 연구자 찰스 앨더슨과 다이앤 월은 1993년 이 통념을 받치는 근거가 적다고 지적했다(Applied Linguistics, 1993년). 두 사람은 스리랑카의 시험 개편을 살핀 연구에서 시험이 수업 내용을 바꾸었지만 교수 방법은 거의 바꾸지 못했다고 보고했다. 평가를 바꾼다고 교육이 저절로 바뀌지는 않는다는 경고다. 교육학자 존 빅스는 평가를 교육 목표에 맞춰 짜야

학습이 따라온다고 설명했다. 그가 든 예에서 한 심리학과 학생은 시험이 사실 몇 가지만 물었기 때문에 열 가지 요점만 외웠다. 빅스는 문제가 학생에게 있지 않고 목표와 어긋난 평가에 있다고 썼다.

광운대의 2026학년도 1학기 수업은 이 방향으로 움직였다(파이낸셜뉴스, 2026년 9월 18일). 화학공학과 김영훈 교수는 공학열역학 과제를 "자신의 생각 정리, AI와 대화, 사실 확인, 최종 답안 작성"의 단계로 냈다. 수학과 부선영 교수는 수리통계학 팀 프로젝트의 보고서 초안마다 인공지능의 피드백을 받게 했다. 평가 대상은 최종 결과물 한 편에서 초안별 피드백과 단계별 과제로 옮겨 갔다.

과정 평가는 교수의 부담을 감당할 수 없게 늘린다는 반론이 있다. 세르비아 노비사드대 의대의 드라간 니콜리치와 마리야나 바스타 니콜리치는 학생 100명의 구술 확인에 직접 평가만 약 17시간, 준비와 조정을 더하면 25시간에서 35시간이 든다고 추정했다(Frontiers in Artificial Intelligence, 2026년 7월 16일). 실증 결과가 없는 개념 논문이다. 보스턴대 경영대학원의 크리산토스 델라로카스 교수는 구술 방어와 감독 시험이 자원이 넉넉한 대학에서나 가능하다고 썼다(Inside Higher Ed, 2026년 2월 19일). 커뮤니티칼리지와 통학 대학과 성인 온라인 과정이 큰 타격을 받는다는 것이다. 부담은 실제로 있다. 니콜리치 등은 300명 규모 과정에서 학생 30퍼센트에서 40퍼센트만 뽑아 확인하면 한 주기에 약 50시간으로 줄어든다고 계산했다. 델라로카스도 인공지능의 도움을 받으면 30명 구술에 6시간 걸릴 일이 90분이 된다고 예상했다. 검출 점수 하나로 시작한 조사가 6개월을 끄는 비용도 따로 든다.

과정 평가로 바꾼 뒤 학습 결과가 나아졌는지 무작위 배정으로 잰 연구는 2026년 10월까지 찾지 못했다. 측정된 근거는 아직 설문과 작은 사례에 머문다.

미국 볼티모어대 글쓰기 프로그램 책임자 레이철 젤레니는 2025년 학생들이 구글 문서에서 글을 쓰고 고친 과정을 기록하게 했다(세계일보, 2026년 9월 7일). 2026년에는 과제를 다시 짜서 최종 결과물과 함께 챗봇 대화 기록, 손으로 쓴 개요, 과제 소감 영상을 내게 했다.

평가를 바꾼다는 것은 학생의 모든 작업을 감시한다는 뜻이 아니다. 교육 목표에 맞는 핵심 과제에서 질문의 출발점과 판단 과정, 검증 능력을 확인하자는 것이다. 대학은 과목의 규모에 맞춰 구술 확인과 수정 이력 검토를 조합하고, 시범 운영에서 학습 효과와 교수의 부담을 함께 평가해야 한다. 그 결과로 절차를 조정하되, 인공지능 사용량 대신 학생이 이해하고 판단한 내용을 평가한다는 방향은 분명히 세워야 한다.

참고 자료

성신여자대학교 대학원, 학위청구논문 표절예방프로그램(카피킬러) 검사결과 제출 안내(관련근거: 2017년 12월 11일 대학원위원회 회의결과, 결과확인서 견본 2025년 2월 19일 발급), https://www.sungshin.ac.kr/bbs/gsci/5225/110649/download.do

중앙대학교 대학원, 학위수여논문 유사도 검사 내규, 2015년 4월 1일 제정, https://gsaim.cau.ac.kr/download/graduate05_4.pdf

서울대학교, 서울대학교 AI 가이드라인, 2026년 1월 1일 제정, https://www.snu.ac.kr/about/downloads?md=v&bbsidx=166508

고려대학교 학부대학 원격교육센터, AI 활용 가이드라인, 2025년 9월 12일, https://library.korea.ac.kr/wp-content/uploads/2025/11/KU_AI%20guideline(kor).pdf

계명대학교 일반대학원, 학위논문 작성 시 AI 활용 가이드라인, 2026년 10월 제정, https://cms.kmu.ac.kr/pdfView/gs/691/990/fileDownload.do

스포트라이트유, 교육부·대교협, 대학 AI 활용 윤리 가이드라인 배포…'금지'보다 수업별 기준에 무게, 2026년 8월 26일, https://www.spotlightuniv.com/교육부·대교협-대학-ai-활용-윤리-가이드라인-배포/

뉴스1, 대학과제에 AI 활용하면 "사실 공개해야"…가이드라인 확정, 2026년 8월 28일, https://www.news1.kr/society/education/6271700

서울경제, "AI 연구 가이드라인이 카피킬러?"…국립대 3곳만 '지침 관리 중', 2025년 10월 17일, https://m.sedaily.com/article/14130695

시사IN(다음 전재), 대학 시험장의 AI, 막을 수 없다면 차라리?, 2025년 12월 10일, https://v.daum.net/v/20251210071302907

Sixth Tone, For Chinese Students, the New Tactic Against AI Checks: More AI, 2024년 6월 5일, https://www.sixthtone.com/news/1015288

36氪(영문판), This graduation season, those papers "killed by AI", 2026년 5월 26일, https://eu.36kr.com/en/p/3826037992116868

AI타임스, 지난해 대학 과제물 46% '표절 위험'...AI 검사 결과 발표, 2023년 4월 20일, https://www.aitimes.com/news/articleView.html?idxno=150626

Times Higher Education, Science journals overturn ban on ChatGPT-authored papers, 2023년 11월 16일, https://www.timeshighereducation.com/node/729880

ICMJE, Up-Dated ICMJE Recommendations (May 2023), 2023년 5월, https://www.icmje.org/news-and-editorials/updated_recommendations_may2023.html

엘스비어, Generative AI policies for journals, 2026년 6월 개정, https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals

The Decoder, OpenAI releases AI text detector for ChatGPT and other models, 2023년 1월 31일, https://the-decoder.com/openai-releases-ai-text-detector-for-chatgpt-and-other-models/

Search Engine Journal, OpenAI Shuts Down Flawed AI Detector, 2023년 7월 25일, https://searchenginejournal.com/openai-shuts-down-flawed-ai-detector/492565

Patterns(arXiv 판, Liang 외), GPT detectors are biased against non-native English writers, 2023년 7월 10일 개정, https://arxiv.org/abs/2304.02819

Turnitin 보도자료(교도 PR와이어 게재), Turnitin Turns On AI Writing Detection Capabilities for Educators and Institutions, 2023년 4월 5일, https://kyodonewsprwire.jp/release/202304054604

K-12 Dive, Turnitin admits there are some cases of higher false positives in AI writing detection tool, 2023년 6월 7일, https://www.k12dive.com/news/turnitin-false-positives-AI-detector/652221/

Inside Higher Ed, Turnitin's AI Detector: Higher-Than-Expected False Positives, 2023년 6월 1일, https://www.insidehighered.com/news/quick-takes/2023/06/01/turnitins-ai-detector-higher-expected-false-positives

밴더빌트대, Guidance on AI Detection and Why We're Disabling Turnitin's AI Detector, 2023년 8월 16일, https://www.vanderbilt.edu/brightspace/2023/08/16/guidance-on-ai-detection-and-why-were-disabling-turnitins-ai-detector/

AI타임스, 94% 정확도로 챗GPT 생성 논문 잡아내는 'GPT킬러' 비결은, 2023년 6월 21일, https://www.aitimes.com/news/articleView.html?idxno=151898

NBER(Jabarian, Imas), Artificial Writing and Automated Detection, 2025년 9월, https://www.nber.org/papers/w34223

arXiv(Karr 외), Why AI Detection Fails for Academic Integrity, 2026년 8월 6일 제출, https://arxiv.org/abs/2608.11256

ABC, University wrongly accuses students of using artificial intelligence to cheat, 2025년 10월 9일, https://www.abc.net.au/news/2025-10-09/artificial-intelligence-cheating-australian-catholic-university/105863524

Inside Higher Ed, Adelphi Student Wins AI Plagiarism Lawsuit, 2026년 2월 11일, https://www.insidehighered.com/news/quick-takes/2026/02/11/adelphi-student-wins-ai-plagiarism-lawsuit

京报网, 论文AIGC率检测误判高发，有学生用AI降AIGC率, 2026년 5월 21일, https://news.bjd.com.cn/2026/05/21/11758506.shtml

이투데이, "AI 쓰라면서 과제엔 어디까지?"…대학가 'AI 평가실험' [AI 新계급사회], 2026년 10월 6일, https://www.etoday.co.kr/news/view/2631457

듀크대 교수학습센터, Artificial Intelligence Policies: Guidelines and Considerations, 게시일 미표기, https://ctl.duke.edu/ai-and-teaching-at-duke-2/artificial-intelligence-policies-in-syllabi-guidelines-and-considerations/

CASRAI, AI Disclosure Statement Examples: Ready-to-Use Wording for Journals, Funders and Institutions, 2026년 8월 25일 갱신, https://casrai.org/guides/ai-disclosure-statement-examples

北京日报, 复旦大学发布国内高校首个"AI禁令", 2024년 12월 30일, https://xinwen.bjd.com.cn/content/s677272d0e4b000299bae14cd.html

arXiv(Payan 외), greCAPTCHA: Assessing Understanding as Evidence of Research Authorship Under Generative AI, 2026년 9월 17일 제출, https://arxiv.org/abs/2609.20481

경향신문, "내년부터 카이스트 인문·기초 수업서 AI 사용 막을 것", 2026년 8월 25일, https://www.khan.co.kr/article/202608252057005

HEPI(Brendal Aformeziem), Catching the wrong students: AI detection, international students and the fairness crisis in UK universities, 2026년 7월 20일, https://www.hepi.ac.uk/2026/07/20/catching-the-wrong-students-ai-detection-international-students-and-the-fairness-crisis-in-uk-universities/

서울신문, 'AI로 연구' 이것은 지켜라…'국가R&D AI 연구윤리 가이드' 마련, 2026년 9월 20일, https://www.seoul.co.kr/news/society/science-news/2026/09/20/20260920500028

러셀그룹, Principles on the use of generative AI tools in education, 2023년 7월 3일, https://www.russellgroup.ac.uk/policy/policy-briefings/principles-use-generative-ai-tools-education

시드니대 Teaching@Sydney, Frequently asked questions about the two-lane approach to assessment in the age of AI, 2024년 7월 2일 게시(2026년 4월 21일 갱신), https://educational-innovation.sydney.edu.au/teaching@sydney/?p=21264

arXiv(Perkins 외), The AI Assessment Scale Revisited: A Framework for Educational Assessment, 2024년 12월 12일, https://arxiv.org/abs/2412.09029

Applied Linguistics(Alderson, Wall), Does Washback Exist?, 1993년, https://doi.org/10.1093/applin/14.2.115

John Biggs, Aligning teaching for constructing learning, 발행일 미표기, https://www.biologie.uzh.ch/dam/jcr:94dbcc23-1c4d-4892-bab8-ba16abc4ef66/ConstructiveAlignment_Biggs.pdf

파이낸셜뉴스, "정답 암기 대신 질문·검증"… 대학 강의실 삼킨 AI, 수업 바꾼다, 2026년 9월 18일, https://www.fnnews.com/news/202609180957391572

Frontiers in Artificial Intelligence(Nikolić, Basta Nikolić), Designing AI-resilient assessment in higher education: a four-pillar conceptual framework, 2026년 7월 16일, https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2026.1841682/full

Inside Higher Ed(Chrysanthos Dellarocas), AI Will Break Assessment Before It Fixes It, 2026년 2월 19일, https://www.insidehighered.com/opinion/views/2026/02/19/ai-will-break-assessment-it-fixes-it-opinion

세계일보, "AI 썼나 안 썼나보다 배웠나가 중요"…美 대학들 과제 확 바꾼다, 2026년 9월 7일, https://www.segye.com/newsView/20260907504321



5장. 협업과 지휘를 가르치는 학교

박사연구자에게 쓰지 말라고 할 것이 아니라 쓰는 법을 가르친다

AI는 축복이지만 학생들을 게으르게 만들었다(AI is a blessing, but it has made students lazy). 나이지리아 아부자대에서 감염병 모델링으로 박사 과정을 밟는 레오나 디알라의 말이다(Nature, 2026년 3월 19일). 그는 문헌 검색, 코드 작성, 논문 요약에 인공지능을 쓰면서 도구가 준 참고문헌과 사실을 하나하나 확인한다. 네이처가 2025년 박사과정생 약 3,800명에게 물었더니 81퍼센트가 인공지능 도구를 완전히 믿지 않는다고 답했다. 65퍼센트는 인공지능이 사고력, 연구력, 글쓰기 능력을 약하게 만들까 걱정했다. 스스로 참여한 응답자를 모은 조사이고 응답자 대부분이 이공계다.

대학의 대응은 막는 쪽에 기울어 있다. 프린스턴대 역사학과는 대학원 첫 2년 동안 생성형 인공지능의 대부분 사용을 피하라고 정했다(프린스턴대 역사학과). 이 시기에 역사학의 기초 기능을 길러야 한다는 이유다. 유럽대학협회 박사교육위원회 조사에 응답한 217개 대학 가운데 38퍼센트는 처음으로 인공지능 정책을 만드는 중이었다(Times Higher Education, 2026년 1월 15일). 쓰지 말라는 규정은 늘었다. 국내에서도 금지와 적발이 먼저 왔다. 학부 사례지만 서울대 컴퓨터공학부의 2026년 '컴퓨터 네트워크' 수업에서는 수강생 62명 가운데 32명이 생성형 인공지능으로 만든 과제를 냈다가 적발됐다(한국경제, 2026년 10월 14일). 같은 기사에서 홍진호 서울대 독어독문학과 교수는 완성된 결과물만으로 학생의 학업 성취를 평가하기가 불가능에 가깝다고 말했다. 미국 국립보건원(NIH)의 연구윤리 교육 요건과 한국연구재단의 권고에서 박사과정생에게 인공지능 쓰는 법을 가르치라는 조항은 찾지 못했다.

금지 쪽에도 근거가 있다. 타오는 학생이 문제의 답을 인공지능에 바로 묻는 습관이 혼자 푸는 능력을 해친다고 경고했다(타오 블로그, 2026년 4월 18일 댓글). 디알라의 걱정과 같은 방향이다. 그런데 타오가 막은 것은 훈련용 문제의 답을 맡기는 일이었다. 그는 자기 증명 전략을 인공지능에게 비판하게 하는 레드팀 방식을 권했다. 걱정할 대상은 도구가 아니라 쓰는 방식이다. 쓰는 방식은 짧은 훈련으로도 달라진다. Clerc 등 연구진은 중학생 116명에게 언어 모델이 어떻게 틀리는지와 질문을 다듬는 법과 답을 확인하는 법을 두 시간 가르쳤다(arXiv, 2026년 4월 2일). 훈련받은 학생은 기계에 덜

기댔고 답이 맞는지 더 정확히 가렸다. 우리는 금지만으로는 학생에게 기계를 다루는 법이 남지 않는다고 판단한다.

박사 과정의 재설계: 질문, 분배, 검증, 책임

토론토대 대학원은 2026년 9월 9일 갱신한 지침에서 박사 학위논문에 생성형 인공지능을 쓰려면 지도교수와 지도위원회의 사전 승인을 받게 했다(토론토대 대학원, 2026년 9월 9일). 논문에는 도구를 어떻게 왜 썼는지 분석 도구를 적듯 밝힌다. 최종 구술시험에서 학생은 생성형 인공지능의 어떤 사용이든 설명하고 방어할 수 있어야 한다("describe and defend any use of generative AI"). 계명대 일반대학원 지침은 연구문제와 연구목적을 연구자가 직접 정하게 하고 모든 내용의 최종 책임을 제출자에게 둔다(계명대 일반대학원, 2026년 10월). NIH는 2025년 7월 17일 공고에서 인공지능이 실질적으로 쓴 신청서를 신청자의 독창적 아이디어로 보지 않는다고 했다(유타대 연구지원처, NOT-OD-25-132 요약). 그런 신청서는 부적합으로 처리할 수 있다. 박사가 학위를 받은 뒤 만날 연구비 제도는 이미 질문의 주인을 따진다.

이 규정들은 심사의 문턱을 정한다. 문턱을 넘을 능력을 기르는 과정은 비어 있다. 저자는 앞의 두 장에서 다룬 협업과 지휘의 능력을 질문, 분배, 검증, 책임의 네 요소로 나눠 박사 과정에 한 학기 또는 한 과정 단위로 넣자고 제안한다. 각 단위는 1학점 세미나나 실습으로 짜고, 기존 연구방법론 과목에 묶어도 된다. 다음은 저자가 제안하는 예시이며 어느 대학의 확정 교육과정도 아니다.

질문(1년차 세미나). 학생은 인공지능 없이 연구 질문 초안을 쓴다. 그다음 인공지능에게 그 질문의 약점과 빠진 선행 연구를 찾게 한다. 레드팀 방식을 연구 질문에 옮긴 것이다. 평가는 질문을 어떻게 고쳤는지 적은 기록으로 한다.

분배(2년차 실습). 한 연구 과제를 문헌 탐색, 자료 정리, 코드, 초안으로 쪼갠다. 사람과 도구가 맡을 몫과 맡긴 이유와 돌려받을 결과의 형식을 작업 명세서에 적는다. 이 요소를 가르치는 대학원 프로그램은 찾지 못했다.

검증(종합시험). 인공지능이 낸 인용, 계산, 코드를 원자료와 대조하고 정답을 아는 작은 자료로 다시 돌려 본다. 오류를 일부러 심은 결과물에서 틀린 곳을 찾는 문항을 종합시험에 넣는다.

책임(학위 심사). 논문에 인공지능 사용 명세를 두고 구술시험에서 사용 하나하나를 설명하고 방어한다. 계명대 지침처럼 지도교수는 의심이 없어도 프롬프트, 초안, 수정 이력을 요구할 수 있게 한다.

프린스턴처럼 첫 2년을 기초 훈련에 쓰는 학과도 이 틀을 쓸 수 있다. 질문 세미나에서 인공지능 단계를 빼고 검증과 분배를 종합시험 뒤로 옮기면 된다. 기초를 지키는 일과 도구를 다루는 법을 가르치는 일은 학기 순서로 나눌 수 있다. 네 단위를 모두 거친 박사는 기계에 일을 맡기는 법과 그 결과를 자기 이름으로 책임지는 법을 함께 익힌다.

학부 교양 과정에 들어갈 내용

졸업요건은 학위를 받으려면 채워야 하는 학점과 과목의 조건이다. 미국 대학은 인공지능 역량을 졸업요건에 넣기 시작했다. 오하이오주립대는 2029년 졸업 학번부터 모든 학생이 인공지능에 능숙한 채 졸업하게 한다는 목표를 세웠다(오하이오주립대 학사처). 모든 학부생이 듣는 필수 세미나에서 생성형 인공지능의 기초를 다루고 학습 성과 여섯 가지를 정했다. 퍼듀대는 2026년 가을 입학생부터 인공지능 역량을 졸업요건으로 삼았다(퍼듀대, 2026년 8월 19일). 과목 22개를 새로 만들거나 고쳤고 과목마다 1학점에서 4학점이다. 퍼듀는 도구 조작을 가르치지 않고 어떤 일에 인공지능이 맞는지 판단하는 힘에 초점을 둔다고 밝혔다. 싱가포르국립대(NUS)는 2026년 8월 입학생부터 모든 학부생에게 인공지능 과목 둘 이상을 필수로 듣게 했다(The Star, 2026년 8월 11일). 새 필수 과목의 이름은 「응용 생성형 AI: 프롬프트에서 평가까지」다. 모든 전공은 그 직업을 인공지능이 어떻게 바꾸는지 다루는 필수 과목을 하나 이상 둔다. 애런 틴 부총장은 대학의 본업이 사람을 기르는 일이라고 했다("The core business of a university is really to develop the human").

중국은 지역 단위로 필수화를 밀었다. 저장성과 상하이시와 쓰촨성은 인공지능 교양 과목을 필수로 정했다(光明日报, 2026년 5월 19일). 같은 기사는 부작용도 적었다. 광둥의 한 중문과 학생은 첫 강의의 신경망과 역전파 설명을 따라가지 못했다. 베이징의 한 교사는 세 번째 수업에서 출석이 절반으로 줄었다고 했다. 베이징대 교육학원 상쥔제(尚俊杰) 교수는 학생을 판단력 있는 사용자보다 고급 조작공(高级操作工)으로 기른다고 비판했다. 과목을 더하면 교육과정이 넘치고 다른 교양이 밀린다는 걱정은 이런 사정에서 나온다. 도구만 쥐여 주는 방식도 답이 되지 못한다. 미국의 한 대형 공립대에서 학부생 2,379명을 무작위로 나눈 실험에서, 수업에 붙인 인공지능 튜터를 쓸 수 있던 집단의 최종 성적은 0.37표준편차 낮았다(EdWorkingPapers, 2026년 10월). 가족 가운데 처음 대학에 온 학생의 손실이 더 컸다.

저자의 답은 과목 수를 늘리기보다 과제를 바꾸는 쪽이다. 퍼듀가 기존 과목을 고쳐 요건을 채운 방식이 선례다. 새 과목은 1학년 교양필수 3학점 하나로 둔다. 교양필수는 전공과 상관없이 모든 학생이 들어야 졸업할 수 있는 과목이다. 이 과목은 질문과 검증을

맡는다. 학생은 같은 질문을 바꿔 가며 답이 어떻게 달라지는지 보고 답 속 출처를 원문까지 찾아간다. 베이징건축대 뤼청(吕橙) 부교수는 학생이 인공지능의 실수를 찾게 하는 일이 성공을 찾게 하는 일보다 교육 가치가 크다고 했다. 이 과목은 학생이 찾아낸 틀린 답과 그 근거로 평가한다. 분배와 책임은 글쓰기, 통계, 전공 입문 과목의 과제로 넘기고 과제마다 사용 기록과 검증 결과를 붙이게 한다. 신경망 이론은 선택 과목으로 빼고 반을 수준별로 나누면 광둥 학생 같은 경우를 줄일 수 있다.

초중등 교육에서 시작할 수 있는 것

현행 초중등 교육과정은 2022 개정 교육과정이다. 교육부는 인공지능 교육 중심의 새 교육과정을 이르면 2029년에 도입하겠다고 밝혔다(에듀프레스, 2025년 11월 10일). 그 사이에는 AI 중점학교가 길을 연다. AI 중점학교는 정보와 인공지능 수업을 정규 시수보다 늘려 운영하는 학교다. 교육부는 2026년 3월 9일 1,141개교를 선정했다(에듀프레스, 2026년 3월 9일). 초등학교는 정보 관련 교육을 34시간에서 68시간 이상으로, 중학교는 68시간에서 102시간 이상으로 늘린다. 도구부터 보급했다가 물러선 경험도 있다. AI 디지털교과서는 2025년 8월 초·중등교육법 개정으로 교과서에서 교육자료로 지위가 바뀌었다. 감사원 감사에서 이 교과서에 한 번도 접속하지 않은 학생은 평균 60퍼센트였다(세계일보, 2025년 12월 17일).

해외에서는 지침이 먼저 섰다. 미국에서 주 교육부 차원의 학교 인공지능 지침을 낸 곳은 2024년 12월 20개 주였다(Ballotpedia News, 2024년 12월 9일). 2026년 8월 민간 집계로는 35개 주와 푸에르토리코다(AI for Education, 2026년 8월 21일 갱신). 영국 교육부는 학생 사용을 막지 않되 가까운 감독과 필터링을 갖추고 도구가 정한 연령 제한을 지키라고 했다(영국 교육부, 2025년 8월 12일). 자체 연령 하한은 두지 않았다. 싱가포르 교육부는 인공지능에 대해 배우기, 쓰는 법 배우기, 함께 배우기, 그 너머를 배우기의 네 가지 배움을 틀로 삼는다(싱가포르 교육부, 2026년 9월 21일 갱신). 에스토니아의 AI 리프에는 첫해 전국 고등학교 156곳 가운데 154곳이 참여했다(AI Leap, 2026년 6월 9일). 참여 교사의 63퍼센트가 수업을 다시 짰다. 학습 효과 분석 결과는 아직 나오지 않았다.

초등학생에게는 이르다는 반론이 있다. 한국교육과정평가원 박혜영 연구위원은 "초등학교 6학년까진 AI 노출을 줄이는 논의가 필요해 보인다"고 말했다(국민일보, 2026년 10월 5일). 중국 교육부 산하 위원회도 소학생이 개방형 생성 기능을 혼자 쓰지 못하게 했다(澎湃新闻, 2025년 5월 12일). 저자는 이 반론을 받아들여 연령에 따라 단계를 나누자고 제안한다. 초등학교에서는 학생이 도구를 직접 쓰지 않는다. 교사가

보여 준 답에서 틀린 곳을 찾고 무엇을 물을지 질문을 만드는 연습을 한다. 예를 들어 과학 시간에 교사가 식물의 한살이를 설명한 기계의 답 두 개를 보여 주면, 학생은 교과서와 관찰 기록으로 틀린 문장을 찾는다. 중학교에서는 교사의 감독 아래 직접 쓰고 답을 확인한 기록을 남긴다. 고등학교에서는 한 과제를 여러 도구에 나누어 맡기고 결과를 합치는 프로젝트를 한다. 학생은 어느 단계를 어느 도구에 맡겼는지, 왜 그렇게 나눴는지를 보고서 첫 쪽에 적는다. 중점학교에서 늘어난 시수는 이 순서로 채운다.

교수 평가와 연구비 심사 기준

과학기술정보통신부는 「 국가연구개발 AI 연구윤리 가이드 」 를 내놓았다(파이낸셜뉴스, 2026년 9월 28일). 연구윤리 가이드는 연구자가 지킬 기준을 모은 권고 문서다. 생성형 인공지능을 쓴 연구자는 성과를 낼 때 모델과 버전, 목적과 범위를 밝힌다. 인공지능은 저자나 공동연구자가 될 수 없고 책임은 쓴 연구자에게 있다. 평가자는 계획서와 보고서를 외부 인공지능 서비스에 넣지 못한다. 사람 눈에 띄지 않는 명령어로 인공지능의 평가를 유도하는 은닉 프롬프트는 연구부정행위가 될 수 있다. 한국연구재단도 2024년 3월 권고사항에서 평가위원이 평가자료를 생성형 인공지능에 올리지 말라고 했다(한국연구재단 웹진, 2024년 3월). 신청자에게는 사용 내역을 적으라고 권했다.

해외 연구비 기관의 구조도 같다. 미국 국립과학재단(NSF)은 심사자가 승인받지 않은 도구에 제안서 내용을 올리지 못하게 했다(NSF, 2023년 12월 14일). 신청자에게는 생성형 인공지능을 썼는지와 어떻게 썼는지 적도록 권했다. 영국 연구혁신기구(UKRI)는 심사자가 평가 작업에 생성형 인공지능을 쓰지 못하게 했다(UKRI, 2024년 9월). NIH는 앞서 본 독창성 기준에 더해 책임연구자 한 사람이 한 해에 낼 수 있는 신청을 6건으로 묶었다. 신청자는 밝히고 심사자는 쓰지 않는다는 구조다. 이 구조는 부정을 막는 데 그치고, 연구자가 도구를 얼마나 잘 검증했는지는 묻지 않는다.

논문 수로 연구자를 재는 관행을 바꾸려는 흐름도 있다. DORA는 학술지 영향력 지수 대신 연구 내용으로 평가하자는 2012년 샌프란시스코 선언이다. DORA와 유럽 연구평가개혁연합(CoARA)은 2025년 12월 코펜하겐 회의 뒤 공동 성명을 냈다(DORA, 2025년 12월 4일). CoARA는 2026년 9월 7일과 8일 파리에서 회의를 열었다(CoARA, 2026년 10월 5일). 교수 업적평가에 인공지능 교육 실적을 넣은 대학은 국내외에서 찾지 못했다.

연구비 심사에 인공지능 역량을 넣으면 분야 사이가 불공정해진다는 반론이 있다. 모하마드 호세이니 등은 글쓰기와 계산이 중심인 분야에서는 생성형 인공지능이 논문 수를 크게 늘릴 수 있다고 지적했다(Research Evaluation, 2026년 4월 11일). 현장 조사와 표본 채집에 기대는 분야는 그만큼 빨라지지 않는다. 이 지적은 옳다. 저자는 그래서 사용량과 산출량을 심사에 넣지 말자고 제안한다. 연구비 신청서에는 인공지능을 어디에 쓸지와 그 결과를 어떻게 검증할지 적는 칸을 둔다. 심사위원은 그 검증 계획이 연구 질문에 맞는지 본다. 공개와 검증은 분야와 상관없이 요구할 수 있다. 다음은 저자가 제안하는 연구비 신청서 명세란의 예시이며 어느 기관의 확정 서식도 아니다.

1. 쓸 도구, 모델, 버전

2. 인공지능에 맡길 작업과 그 범위

3. 결과를 검증할 방법과 검증을 맡을 사람

4. 사람이 직접 정할 판단(연구 질문, 해석, 결론)

교수 업적평가의 교육 영역에는 학생에게 질문과 검증을 가르친 과목의 개발과 운영 실적을 넣는다. 가르친 사람이 평가에서 인정받지 못하면 앞의 교육과정은 문서에 머문다.

교육부와 대학, 학회의 역할 분담

영국 러셀그룹 24개 대학은 2023년 생성형 인공지능 교육 원칙 다섯 가지를 냈다(러셀그룹, 2023년 7월 3일). 원칙 2.2는 적절한 사용이 학문 분야마다 다르다고 보고 학과가 대학 정책을 자기 맥락에 맞춰 적용하게 했다. 원칙 3.2는 교수와 평가의 조정이 대학과 분야마다 다르며 그 자율을 지켜야 한다고 적었다. 원칙 4.1에는 24개 회원 대학 모두가 학업 행동 정책을 다시 검토했다고 적혀 있다. 연합은 원칙을 세우고 적용은 학과가 맡는 방식이다. 호주 고등교육 품질기준청(TEQSA)은 2024년 6월 3일 모든 등록 고등교육기관에 생성형 인공지능 대응 실행계획을 내라고 요구했다(TEQSA, 2024년 6월). 2025년 9월 자료에서는 학위 과정 전체, 과목 단위, 둘을 섞은 방식의 세 경로를 예로 들면서 처방하려는 문서가 아니라고 밝혔다(TEQSA, 2025년 9월). 규제기관은 계획을 요구하고 설계는 기관이 하는 방식이다.

한국에서는 층마다 규정이 따로 자란다. 교육부는 2026년 1월 1일 인공지능인재지원국을 새로 만들었다(오마이뉴스, 2025년 12월 25일). 교육부와 대교협이 대학 지침을 냈고 대학은 저마다 가이드라인을 만든다. 국내 학회의 인공지능

투고 규정은 이번 조사에서 확인하지 못했다. 계명대 지침 제12조는 학술지와 연구비 기관의 기준이 함께 적용되면 같이 지키되 학위논문은 대학 규정을 먼저 따른다고 적었다. 규정이 여러 층에서 부딪힐 수 있다는 뜻이다.

저자는 역할을 다음처럼 나누자고 제안한다. 교육부는 목표와 예산과 교사 수급을 맡는다. 국가교육위원회는 초중등 교육과정의 단계를 정한다. 교육부는 TEQSA처럼 대학마다 실행계획을 받되 내용을 정해 주지 않는다. 계획에는 교양 과목의 구성, 학위논문 심사 서식, 교원 연수, 검증 기준을 담게 하고, 교육부는 잘된 사례를 모아 공개한다. 대학과 대교협은 교양 과목과 학칙과 학위논문 심사 서식을 맡는다. 학회는 러셀그룹 원칙의 학과처럼 분야별 허용 범위와 투고 규정을 정한다. 한국연구재단과 과기정통부는 연구비 신청 서식과 심사 규정을 맡는다. 세 층의 규정이 부딪힐 때의 우선순위는 교육부가 대교협과 학회 대표와 함께 한 문서로 정해 둔다. 학생이 지도교수와 학술지 사이에서 서로 다른 규정을 받는 일을 막으려는 것이다. 역할을 나누는 목적은 규정의 수를 줄이는 데 있지 않다. 누가 무엇을 정하는지 학생과 교수가 한눈에 알게 하는 데 있다.

한국 교육에 대한 제안

교사 연수는 시작됐지만 효과와 부담을 함께 살펴야 한다. 교육부의 2026년 인공지능 선도교사 연수에는 1만 268명이 신청했고 연수 시간은 모두 35시간이다(뉴스핌, 2026년 6월 17일). 선도교사는 인공지능을 수업과 평가에 먼저 도입해 동료 교사에게 확산하는 교사다. 경기 2권역에서는 신청자의 81.2퍼센트가 이수했지만(서울신문, 2026년 10월 12일), 역량 변화는 교사의 자기진단으로만 측정했다. 현장의 부담과 우려도 크다. 전국중등교사노동조합이 2026년 9월 중등교사 2,427명의 자발 응답을 모은 설문에서는 인공지능 채점을 겪은 교사의 64.1퍼센트가 결과를 다시 확인하느라 업무가 늘었다고 답했다(세계일보, 2026년 10월 13일). 같은 설문에서 수능 서·논술형 도입에는 73.5퍼센트가 반대했고 84.5퍼센트가 채점 공정성 논란을 걱정했다. 도구를 제공하는 데 더해 사용법을 배울 기회도 필요하다. 생성형 인공지능을 써 본 국민은 44.5퍼센트였고 써 보지 않은 60대의 25.3퍼센트는 방법을 몰라서라고 답했다(뉴스핌, 2026년 3월 31일). 고의숙 제주교육감은 인공지능이 교사를 대신하는 교육은 하지 않겠다고 했고(파이낸셜뉴스, 2026년 9월 6일), 김영오 서울대 공대 학장은 "생각의 근육을 키울 수 있는 교육과 평가 방식을 고민하고 있다"고 말했다(한국경제, 2026년 10월 14일). 저자는 전남대의 무료 도구 제공을 예로 들어 모든 대학생에게 인공지능을 쥐여 주는 첫걸음을 국가가 맡아야 한다고 썼다(전자신문, 2026년 5월 5일). 무료 모델에 머무는 학생과 유료

모델을 쓰는 학생의 경험 차이가 부모의 경제력을 따라간다는 것이 그 근거였다. 교사 연수와 평가의 공정성, 사용법 교육과 도구 접근권을 함께 갖추는 것이 다음 제안의 전제다.

다음은 저자의 안이며 정부나 대학의 확정 계획이 아니다.

1. 박사 과정에 질문, 분배, 검증, 책임을 배우는 세미나와 실습을 두되, 분야의 기초 훈련에 맞춰 학기와 순서를 조정한다. 학위 구술시험에서는 인공지능 사용을 설명하고 방어하게 한다.

2. 대학은 1학년 교양필수 3학점 한 과목에서 질문과 검증을 가르치고 분배와 책임은 전공 과목의 과제로 옮긴다.

3. 초등학교는 도구 없이 질문 만들기와 틀린 답 찾기를 가르치고 직접 사용은 중학교부터 감독 아래 둔다.

4. 선도교사 연수에 인공지능 채점 결과를 검증하는 실습을 넣고 이수 뒤 수업이 어떻게 바뀌었는지 외부에서 잰다.

5. 연구비 신청서와 학위논문에 인공지능 사용과 검증의 명세란을 두고 사용량은 심사하지 않는다.

6. 교수 업적평가의 교육 영역에 질문과 검증을 가르친 과목의 개발과 운영 실적을 넣는다.

7. 교육부는 대학별 실행계획을 받고 대학과 학회와 연구재단의 규정이 부딪힐 때의 우선순위를 한 문서로 정한다.

8. 국가는 대학생과 교사에게 유료 인공지능 도구의 접근권을 교육 기반 시설로 제공한다.

2026년 7월 상하이의 국제수학올림피아드 대회장에서 기계는 만점 답안을 냈다. 그 여섯 문제를 고른 것은 사람이었고 기계의 답이 맞는지 따진 것도 사람이었다. 학교가 길러야 할 사람은 풀 문제를 고르고 돌아온 답을 따지고 그 결과에 이름을 거는 사람이라고 우리는 판단한다.

참고 자료

Nature(Linda Nordling), AI and the PhD student: friend or foe?, 2026년 3월 19일, https://media.nature.com/original/magazine-assets/d41586-026-00843-y/d41586-026-00843-y.pdf

프린스턴대 역사학과, Graduate Student Generative AI Policy, 날짜 미표기, https://history.princeton.edu/graduate/ai-policy

Times Higher Education, Two-thirds of universities report AI use among doctoral students, 2026년 1월 15일, https://www.timeshighereducation.com/news/two-thirds-universities-report-ai-use-among-doctoral-students

타오 블로그, Mathematical methods and human thought in the age of AI, 2026년 3월 29일(4월 18일 댓글), https://terrytao.wordpress.com/2026/03/29/mathematical-methods-and-human-thought-in-the-age-of-ai/

arXiv(Clerc 외), Teaching Students to Question the Machine: An AI Literacy Intervention Improves Students' Regulation of LLM Use in a Science Task, 2026년 4월 2일, https://arxiv.org/abs/2604.01955

토론토대 대학원(SGS), Guidance on the Appropriate Use of Generative Artificial Intelligence for Graduate Academic Milestones, 2026년 9월 9일 갱신, https://www.sgs.utoronto.ca/about/guidance-on-the-use-of-generative-artificial-intelligence/

계명대학교 일반대학원, 학위논문 작성 시 AI 활용 가이드라인, 2026년 10월 제정, https://cms.kmu.ac.kr/pdfView/gs/691/990/fileDownload.do

유타대 연구지원처, NIH Policy Update: New Limits on Applications and Artificial Intelligence Use in Grant Proposals(NIH NOT-OD-25-132 Supporting Fairness and Originality in NIH Research Applications, 2025년 7월 17일 공고 요약), 2026년 2월 24일 갱신, https://osp.utah.edu/news/nih-not-od-25-132.php

오하이오주립대 학사처, AI Fluency, 날짜 미표기, https://oaa.osu.edu/ai-fluency

퍼듀대, Purdue's AI competency graduation requirement begins in every undergraduate degree program, 2026년 8월 19일, https://www.purdue.edu/newsroom/2026/Q3/purdues-ai-competency-graduation-requirement-begins-in-every-undergraduate-degree-program/

The Star, NUS makes AI courses compulsory, gives all students and staff free ChatGPT Edu access, 2026년 8월 11일, https://thestar.com.my/aseanplus/aseanplus-news/2026/08/11/nus-makes-ai-courses-compulsory-gives-all-students-and-staff-free-chatgpt-edu-access

光明日报, 有人听不懂、有人"吃不饱"——高校AI通识课怎样才能讲好, 2026년 5월 19일, https://news.gmw.cn/2026-05/19/content_38772130.htm

에듀프레스, AI 교육과정 개정, 2029년 시행 …정보교사 늘리고 10G 무선망 확대, 2025년 11월 10일, https://www.edupress.kr/news/articleView.html?idxno=21382

에듀프레스, 교육부, 'AI 중점학교' 1141개교 선정… 초등 68·중등 102시간 이상 수업, 2026년 3월 9일, https://www.edupress.kr/news/articleView.html?idxno=21850

세계일보, 감사원 "尹정부, 준비 없이 AI 교과서 도입", 2025년 12월 17일, https://www.segye.com/newsView/20251217515914

Ballotpedia News, Twenty states have released AI guidance for public schools, 2024년 12월 9일, https://news.ballotpedia.org/2024/12/09/twenty-states-have-released-ai-guidance-for-public-schools/

AI for Education, State AI Guidance for K12 Schools, 2026년 8월 21일 갱신, https://www.aiforeducation.io/ai-resources/state-ai-guidance

영국 교육부(DfE), Generative artificial intelligence (AI) in education, 2025년 8월 12일 갱신, https://www.gov.uk/government/publications/generative-artificial-intelligence-in-education/generative-artificial-intelligence-ai-in-education

싱가포르 교육부(MOE), Developing students' AI literacy, 2026년 9월 21일 갱신, https://www.moe.gov.sg/education-in-sg/educational-technology-journey/edtech-masterplan/artificial-intelligence-in-education/developing-students-ai-literacy

AI Leap(TI-Hüpe), AI Leap's First Year Led Schools to Rethink Teaching and Learning, 2026년 6월 9일, https://tihupe.ee/en/3385/

국민일보, 학교 파고드는 AI… 문해력 갉아먹는 독? 교육 혁신의 약?, 2026년 10월 5일, https://www.kmib.co.kr/article/view.asp?arcid=9000019951&code=11131300&sid1=soc

澎湃新闻(中国青年报客户端), 教育部基础教育教指委：小学阶段禁止学生独自使用开放式内容生成功能, 2025년 5월 12일, https://m.thepaper.cn/newsDetail_forward_30802716

파이낸셜뉴스, 생성형 AI 썼다면 모델·버전 공개해야… 국가R&D 책임은 연구자, 2026년 9월 28일, https://www.fnnews.com/news/202609281035598851

한국연구재단 웹진, 한국연구재단, 생성형 AI 도구의 책임 있는 사용을 위한 권고사항 발표, 2024년 3월, https://webzine.nrf.re.kr/magazine/2403/sub_1_01.php

NSF, Notice to Research Community: Use of Generative Artificial Intelligence Technology in the NSF Merit Review Process, 2023년 12월 14일, https://www.nsf.gov/news/notice-to-the-research-community-on-ai

UKRI, Use of generative AI in application preparation and assessment, 2024년 9월, https://www.ukri.org/news/use-of-generative-ai-in-application-preparation-and-assessment/

DORA, Shared Principles, Shared Progress: DORA & CoARA in Collaboration, 2025년 12월 4일, https://sfdora.org/2025/12/04/dora-coara-in-collaboration/

CoARA, Carrying Reform Forward: Highlights from the CoARA Conference 2026, 2026년 10월 5일, https://www.coara.org/news/project-news/recap-coara-conference-2026/

Research Evaluation(Hosseini 외), Generative AI can and should accelerate research evaluation reform to better recognize 'distinctly human contributions', 2026년 4월 11일, https://pmc.ncbi.nlm.nih.gov/articles/PMC13094469/

러셀그룹, Principles on the use of generative AI tools in education, 2023년 7월 3일, https://www.russellgroup.ac.uk/sites/default/files/2025-01/Russell%20Group%20principles%20on%20generative%20AI%20in%20education.pdf

TEQSA, Request for information: Addressing the risk of artificial intelligence, 2024년 6월, https://www.teqsa.gov.au/guides-resources/higher-education-good-practice-hub/artificial-intelligence/request-information-addressing-risk-artificial-intelligence

TEQSA, Enacting assessment reform in a time of artificial intelligence, 2025년 9월, https://www.teqsa.gov.au/sites/default/files/2025-09/enacting-assessment-reform-in-a-time-of-artificial-intelligence.pdf

EdWorkingPapers(Jing Liu 외), The Effects of Course-Integrated AI Tutoring on Student Performance and Engagement: A Randomized University Trial, 2026년 10월, https://edworkingpapers.com/ai26-1598

오마이뉴스, 교육부 3년 만의 조직개편, 'AI'는 앞에 서고 '자치'는 뒤로 갔다, 2025년 12월 25일, https://m.ohmynews.com/NWS_Web/Mobile/amp.aspx?CNTN_CD=A0003193761

뉴스핌, 교육부, AI 수업 이끌 교사 1만명 키운다…선도교사 양성 연수, 2026년 6월 17일, https://www.newspim.com/news/view/20260617000135

서울신문, 2026 인공지능 활용 선도교사 연수 경기 2권역, AI·디지털 역량 자기진단 점수 상승, 2026년 10월 12일, https://www.seoul.co.kr/news/spotlight/ggc-gyeonggi/2026/10/12/20261012500125

세계일보, 서·논술형 AI 채점했더니…교사 10명 중 6명 '업무 늘어', 2026년 10월 13일, https://www.segye.com/newsView/20261013514246

뉴스핌, 과기부, 2025 인터넷 이용 실태 조사 결과 발표…44.5%가 생성형 AI 경험, 2026년 3월 31일, https://www.newspim.com/news/view/20260331000211

파이낸셜뉴스, AI가 대신하는 교육 안한다… 교사 전문성 중심 지원할 것, 2026년 9월 6일, https://www.fnnews.com/news/202609061812062811

한국경제, [단독] 'AI 오남용 막자'…'인공서원'으로 뭉친 서울대, 2026년 10월 14일, https://hankyung.com/article/2026101395641

전자신문, [김경진의 AI전략노트] 〈27〉 전남대에서 멈출 이유가 없다, 모든 대학생에게 AI를 쥐여줘야 한다, 2026년 5월 5일, https://www.etnews.com/20260505000067



제2부. 역할의 재정의: 인간은 무엇을 할 것인가

6장. 배관공의 자리도 오래 가지 않는다

힌턴의 조언과 그 배경

2025년 8월 15일 아침 베이징 국가스피드스케이팅경기장의 관중이 자리에서 일어나 중국 국가를 들었다. 2022년 동계올림픽 때 지은 1만 2천 석 규모의 경기장이다. 국가가 끝나자 제1회 세계 휴머노이드 로봇 경기대회가 시작됐다. 휴머노이드는 사람처럼 머리와 몸통과 두 팔과 두 다리를 갖춘 로봇이다. 첫 킥복싱 경기에서 한 대학 팀의 작은 로봇이 왼손 훅과 앞차기를 날리자 관중석에서 환호가 터졌다. 경기를 끝낸 것은 상대의 주먹이 아니었다. 그 로봇은 돌려차기를 헛차고 혼자 바닥에 넘어졌다(가디언, 2025년 8월 15일). 1,500미터 경기에서는 달리던 로봇의 머리가 떨어져 나가 기권했다. 앞줄에 앉은 58세 은퇴 엔지니어는 로봇의 달리기가 사람의 경주보다 훨씬 재미있다고 말했다.

그보다 두 달 앞서 제프리 힌턴은 젊은이들에게 배관공을 권했다. 힌턴은 2024년 노벨 물리학상을 받은 인공지능 연구자다. 그는 2023년 구글을 떠난 뒤로 인공지능의 위험을 공개적으로 경고해 왔다. 2025년 6월 16일 공개된 팟캐스트 「다이어리 오브 어 CEO(The Diary Of A CEO)」에서 진행자는 초지능 시대에 어떤 일을 택해야 하느냐고 물었다. 힌턴은 인공지능이 물건을 다루는 솜씨에서 사람만큼 되려면 오래 걸린다고 답했다. 그러니 배관공이 되는 편이 괜찮은 선택이라는 것이다("a good bet would be to be a plumber")(Business Insider, 2025년 6월 16일). 그는 같은 대답에 단서를 붙였다. 휴머노이드 로봇이 나타날 때까지라는 것이다("Until the humanoid robots show up")(팟스크립츠 전사본, 2025년 6월 16일). 같은 해 8월에도 그는 여러 자리에서 배관공이 되라고 말했다(Free Press Journal, 2025년 8월 25일).

조언의 바탕에는 오래된 관찰이 있다. 로봇공학자 한스 모라벡은 1988년 책 『마음의 아이들(Mind Children)』에서 이런 취지의 글을 썼다. 컴퓨터가 지능검사나 체커 게임에서 어른 수준을 내기는 비교적 쉽지만 지각과 이동에서 한 살 아이의 솜씨를 갖추기는 어렵거나 불가능하다는 것이다. 이 관찰을 흔히 모라벡의 역설이라고 부른다. 그는 수백만 년의 자연선택이 다듬은 감각과 운동 능력일수록 기계로 옮기기 어렵다고 설명했다. 언어 모델이 시험 문제를 푸는 사이에 로봇은 경기장에서 넘어졌다.

젠슨 황 엔비디아 최고경영자도 기능직의 앞날을 밝게 본다. 그는 2025년 9월 영국 채널4 뉴스 인터뷰에서 모든 나라 경제의 숙련 기능직이 호황을 맞는다고 말했다("The skilled craft segment of every economy is going to see a boom")(Fortune, 2026년 9월 1일). 인공지능용 공장을 지으려면 전기공과 배관공과 목수가 수십만 명 필요하다는 것이 근거다. 두 사람의 근거는 다르다. 힌턴은 로봇이 아직 손일을 못 한다는 기술의 한계를 들었고, 황은 데이터센터를 지을 손이 모자란다는 수요를 들었다. 어느 근거도 기능직의 현재 역할이 영구히 유지된다는 보장은 아니다. 로봇이 현장 작업을 맡을 수 있게 되면 필요한 인력의 규모와 역할도 다시 따져야 한다.

노동경제학의 시각은 조금 다르다. 대런 아세모글루와 데이비드 오터와 사이먼 존슨은 2026년 2월 논문에서 전기공을 돕는 인공지능을 노동을 돕는 기술의 예로 들었다(NBER, 2026년 2월 23일). 아세모글루는 2024년 논문에서 배관공과 전기공의 문제 해결이 실시간 판단과 현장의 맥락과 높은 신뢰성을 요구해 범용 인공지능과 잘 맞지 않는다고 봤다(NBER, 2024년 5월 27일). 이들의 결론은 기능직이 안전하다는 말과 결이 다르다. 기계가 사람을 대신하기보다 도울 가능성이 크고 결과는 기술 설계와 정책에 달렸다는 것이다. 이 분석이 주로 다룬 대상은 화면 속 인공지능이었다. 몸을 가진 기계가 현장에 들어오면 같은 질문을 다시 던져야 한다.

중국의 휴머노이드 체육대회가 보여 준 것

베이징 이좡에서는 2025년 4월 19일 사람과 로봇이 함께 뛰는 하프마라톤이 처음 열렸다. 로봇 21대가 출발했고 톈궁 울트라가 2시간 40분으로 1위를 했다(Al Jazeera, 2025년 4월 19일). 같은 날 사람 남자부 우승자는 1시간 2분에 들어왔다. 우승 로봇은 신호 장치를 멘 사람을 따라 달렸고 다른 로봇 대부분은 원격조종으로 움직였다(TechCrunch, 2025년 4월 19일). 원격조종은 사람이 떨어진 곳에서 신호를 보내 로봇을 움직이는 방식이다. 넉 달 뒤 1회 경기대회의 1,500미터에서는 유니트리의 H1이 6분 34초 40으로 이겼다(SCMP, 2025년 8월).

1년 뒤 기록은 크게 달라졌다. 2026년 4월 19일 2회 하프마라톤에는 로봇 300대 이상과 사람 1만 2천여 명이 나섰다(iRunFar, 2026년 4월 20일). 아너(Honor)의 로봇 라이트닝은 원격조종 부문에서 48분 19초를 기록했다. 대회 규정이 원격조종 기록에 1.2를 곱했기 때문에 자율 주행으로 50분 26초를 낸 같은 기종이 최종 우승을 차지했다(Global Times, 2026년 4월 19일). 사람 하프마라톤 남자 세계기록은 57분 20초다. 2026년 8월 22일 개막한 2회 경기대회에는 16개국 666개 팀이 로봇 2,056대를 내보냈고 51개 종목이 열렸다(China Daily, 2026년 8월 25일, 조직위 등록 기준). 1,500미터에서는

35대가 일곱 개 조로 나눠 뛰었고 1위 기록은 2분 21초 63이었다. 사람 남자 세계기록 3분 26초 00보다 1분 넘게 빠르다. 제자리 높이뛰기의 최고 기록은 1년 전 0.95미터에서 2.88미터로 올랐다.

두 해의 숫자를 그대로 견주기는 어렵다. 2026년 대회는 단거리와 계주와 축구와 태극권과 체조에서 완전 자율을 의무로 정했다(위키백과 대회 항목의 경기 안내서 요약). 시나리오형 종목에서 원격조종 로봇은 자율 로봇의 절반 점수를 받았다. 참가 로봇은 네 배로 늘었고 규칙도 바뀌었다. 넘어지는 로봇도 여전했다. 테크레이더는 기록을 깬 달리기 로봇들이 왜 제대로 멈추지 못하는지를 물었다(TechRadar, 2026년 8월 24일). 기즈모도는 실패 장면만 모은 기사를 냈다(Gizmodo, 2026년 8월 25일).

이 대회를 선전 행사로 보는 시각에도 근거가 있다. 미국 기술 칼럼니스트 셸리 파머는 1회 대회 전 칼럼에서 이 행사가 국가가 후원하는 기술 선전처럼 읽힌다고 썼다("state-sponsored tech propaganda")(셸리 파머, 2025년 5월). 그는 밑바탕의 혁신은 진짜라는 말도 덧붙였다. 홍콩 벤처캐피털 마인드웍스캐피털의 제프리 우는 베이징이 기술 진보를 문화적 구경거리로 바꾸는 일을 워싱턴보다 잘해 냈다고 썼다(SCMP, 2025년). 국제로봇연맹(IFR)은 AFP에 중국 정부가 이 분야의 역량을 보여 주려 한다고 말했다. 2026년 대회 참가 팀의 96퍼센트는 중국 팀이었다(Forbes, 2026년 8월 19일).

같은 기사에서 포브스 기고가 존 쾨치어는 대회 종목표를 시장 지도라고 불렀다. 그는 지금 로봇공학에서 이 종목표만큼 정직한 문서는 드물다고 썼다("one of the most honest documents in robotics right now"). 로봇이 혼자 해낼 수 있는 일과 아직 사람이 붙어야 하는 일이 종목표에 그대로 드러난다는 뜻이다. 선전 행사라는 평가와 진전의 계측기라는 평가는 함께 성립한다. 필자는 2026년 8월 전자신문 연재에서 이 대회를 두고 "겉모습은 운동회지만 속은 거대한 채용 시험장이다"라고 썼다(전자신문, 2026년 8월 25일).

로봇 관절 구동 장치의 정밀도가 높아지는 속도

로봇의 관절을 움직이는 부품을 액추에이터(actuator)라고 한다. 휴머노이드의 액추에이터는 모터와 감속기와 구동 회로와 센서를 한 덩어리로 묶은 관절 모듈이다. 감속기는 모터의 빠른 회전을 늦추는 대신 힘을 키우는 기어 장치다. 업계에서는 휴머노이드 원가의 절반 안팎이 액추에이터에서 나온다고 본다. 모건스탠리는 이 비중을 약 56퍼센트로 추정했다(humanoid.guide, 2026년 8월 29일, 보고서 원문 미확인). 로봇이 싸지려면 관절이 먼저 싸져야 한다.

관절 부품의 가격은 중국에서 내려가고 있다. 정밀 감속기의 한 종류인 하모닉 감속기는 일본 하모닉드라이브시스템즈가 시장을 거의 혼자 쥐던 시절 한 개에 800달러에서 1,500달러였다는 업계 추정이 있다. 지금 중국산은 일본산의 절반 안팎 가격에 팔린다는 업계 추정도 있다. 두 추정 모두 기업 공시가 아닌 2차 자료에서 나왔다. 공시로 확인되는 것은 양이다. 중국 리더드라이브(Leaderdrive)는 2025년에 하모닉 감속기 42만 5,200대를 팔았다(humanoid.guide, 2026년 4월 27일). 한 해 전보다 72퍼센트 늘어난 수치다.

중국 로봇 기업들은 관절을 직접 만든다. 유니트리는 H1의 관절 모터를 자체 개발했다고 기술 문서에 밝혔다. 그 모터의 최대 토크는 360뉴턴미터다. 토크는 물체를 돌리는 힘의 크기다. 유비테크(UBTech)는 서보 액추에이터를 누적 100만 대 생산했다고 회사 소개에 적었다.

관절 가운데 만들기 어려운 곳은 손이다. 자유도는 관절이 서로 따로 움직일 수 있는 방향의 수다. 사람의 손은 약 27 자유도를 가진다(가디언, 2025년 8월 15일). 일론 머스크 테슬라 최고경영자는 2025년 9월 10일 X에 로봇의 손을 설계하는 일은 아주 어려운 문제라고 썼다("Designing dexterous hands for robots is a very hard problem") (Teslarati, 2025년 11월 12일). 2026년 4월 16일 공개된 테슬라의 국제특허에는 줄로 손가락을 당기는 힘줄 방식의 손이 담겼다(Teslarati, 2026년 4월 16일). 손가락마다 4 자유도에 손목 2 자유도를 더해 모두 22 자유도다. 특허에 담긴 설계일 뿐 양산 사양은 아니다. 머스크가 한 차원 위라고 말한 옵티머스 V3의 손은 2026년 10월 현재 공개되지 않았다("The V3 hand is another level beyond this"). 피규어(Figure)는 2025년 10월 9일 발표한 피규어 03의 손끝 센서가 3그램의 힘을 느낀다고 밝혔다. 독립 검증을 거친 수치는 아니다.

관절 구동 장치와 촉각 센서의 정밀도가 해마다 얼마나 높아지는지를 같은 기준으로 측정한 공개 자료는 찾기 어렵다. 측정 조건이 제조사마다 다르고 수치 대부분이 회사 발표다. 확인되는 흐름은 부품 가격이 내려가고 생산량이 늘며 손가락마다 센서가 붙는다는 것이다. 로봇공학자 로드니 브룩스는 사람의 손에 촉각 수용체가 약 1만 7천 개 있다고 설명한다(TechCrunch, 2025년 9월 26일). 3그램 정도의 하중을 감지하는 센서와 사람 손의 감각 사이에는 아직 거리가 있다. 배관공은 보이지 않는 이음새를 손끝으로 더듬고 힘을 조절한다. 휴머노이드가 그 일을 넘겨받을 시점을 전망하려면 가격뿐 아니라 이런 작업에서 손의 감각을 얼마나 재현하는지 살펴야 한다.

데이터센터와 GPU: 인공지능이 로봇의 몸속으로 들어가기까지

필자는 2026년 8월 연재에서 "챗봇이 화면 속의 인공지능(AI)이라면 휴머노이드는 몸을 얻은 AI다"라고 썼다(전자신문, 2026년 8월 25일). 학계는 이 생각에 먼저 이름을 붙였다. 아슬란 미리예프와 미르코 코바치는 2020년 11월 네이처 머신 인텔리전스에 실은 글에서 물리적 AI(Physical AI)를 정의했다(임페리얼칼리지런던, 2020년 11월 10일). 물리적 AI는 디지털 두뇌와 지능을 갖춘 몸을 묶어 생물에 견줄 만한 기계를 만드는 기술이다. 이 말을 산업의 표어로 만든 사람은 젠슨 황이다. 그는 2025년 1월 CES 기조연설에서 물리적 AI를 인공지능의 다음 단계로 꼽았다. 로봇공학의 챗GPT 순간이 코앞에 와 있다는 말도 했다("ChatGPT moment for robotics is right around the corner")(엔비디아 블로그, 2025년 1월 7일).

로봇의 두뇌는 기초 모델에서 나온다. 기초 모델은 많은 데이터로 미리 학습해 여러 작업에 두루 쓰는 큰 인공지능 모델이다. 로봇용 기초 모델은 카메라로 본 장면과 사람의 말을 받아 팔다리의 동작을 내놓는다. 엔비디아는 2025년 3월 18일 휴머노이드용 공개 기초 모델 GR00T N1을 내놓았다(엔비디아, 2025년 3월 18일). 구글 딥마인드는 같은 해 6월 로봇 안에서 바로 도는 제미나이 로보틱스 온디바이스를 공개했다(구글 딥마인드, 2025년 6월 24일). 2026년 7월에는 휴머노이드의 온몸을 제어하는 제미나이 로보틱스 2를 공개했다(The Robot Report, 2026년 8월 2일).

학습과 추론은 일어나는 곳이 다르다. 2026년 현재 로봇 모델의 학습은 데이터센터의 GPU에서 한다. 추론은 학습을 마친 모델이 새 상황에서 판단을 내는 일이다. 로봇은 추론을 몸 안의 칩에서 할 수 있다. 이를 온보드 추론이라고 한다. 통신망 너머 데이터센터에 맡기는 방식은 클라우드 추론이라고 한다. 둘을 가르는 기준은 지연 시간이다. 지연 시간은 신호를 보내고 답이 돌아오기까지 걸리는 시간이다. 클라우드를 한 번 오가는 데 0.2초쯤 걸리지만 몸의 균형을 잡는 제어는 1초에 수백 번 돌아야 한다. 이런 제어는 로봇 몸 안에 둘 수밖에 없다. 2026년 6월 공개된 큐원 로봇내비(Qwen-RobotNav) 기술 보고서는 원격 서버의 평균 응답이 로봇 안 칩보다 조금 빨랐다고 적었다(arXiv, 2026년 6월). 보고서는 응답 시간이 들쭉날쭉해서 실제 배치에는 온보드가 더 안정적이라고 결론지었다.

피규어의 헬릭스(Helix)는 이 구조를 공개한 사례다(Figure, 2025년 2월 20일). 장면과 말을 이해하는 층은 1초에 7번에서 9번 돌고, 손발을 움직이는 층은 1초에 200번 돈다. 회사는 두 층이 모두 로봇에 들어간 저전력 GPU에서 돈다고 밝혔다. 2026년 1월 공개한 헬릭스 02는 사람이 손으로 짠 코드 10만 9천 줄을 학습한 제어기로 바꿨다고 회사가 밝혔다(Figure, 2026년 1월 27일).

로봇 몸에 넣는 칩도 나왔다. 엔비디아가 2025년 8월 출시를 발표한 젯슨 AGX 토르는 최대 2,070테라플롭스의 연산 성능을 낸다고 회사가 밝혔다(Gigazine, 2025년 8월 26일). 이 수치는 수를 4비트로 줄이고 0인 값을 건너뛰는 FP4 희소 연산 기준이라 다른 기준의 수치와 바로 견줄 수 없다. 소비 전력은 40와트에서 130와트이고 개발자 키트의 가격은 3,499달러였다. 2026년의 로봇은 데이터센터에서 배우고 몸속의 칩으로 판단한다. 브룩스는 이 방식의 한계를 짚는다. 영상과 원격조종 데이터에는 촉각이 빠져 있어 손재주를 배울 수 없다는 것이다.

하드웨어 압축의 시간은 얼마나 남았는가

언어 모델은 몇 해 만에 사무실의 지식 노동을 따라잡았다. 로봇의 몸이 그 길을 따라오는 데 남은 시간을 우리는 하드웨어 압축의 시간이라고 부른다. 기업들이 내놓은 시간표는 짧다. 머스크는 옵티머스를 2027년 말까지 일반에 내놓고 싶다고 말해 왔다(Teslarati, 2026년 10월 12일). 테슬라는 2026년 10월 현재 생산 대수를 공식 발표하지 않았다. 1X는 2025년 10월 가정용 로봇 네오(NEO)를 2만 달러에 예약 판매하며 2026년 미국 인도를 목표로 잡았다(1X). 2026년 10월 현재 고객 인도는 공식 확인되지 않는다.

미국에는 중국만 한 공개 경기대회가 없다. 우리가 보기에 미국 기업은 경기장보다 공장을 시험 무대로 삼는다. 피규어는 BMW 스파턴버그 공장의 조립·물류동에 피규어 03을 넣어 부품을 순서대로 꽂는 일을 맡겼다(Figure, 2026년 6월 30일). 피규어는 2026년 7월 23일 피규어 03의 1,000번째 기체를 만들었다고 밝혔다(The Robotics Media, 2026년 8월 2일). 뱅크오브아메리카는 중국산 휴머노이드의 부품원가가 2025년 3만 5천 달러에서 2030년 1만 7천 달러 아래로 내려간다고 봤다(Fortune, 2026년 3월 13일). 판매가가 아닌 부품원가 기준이다.

반대쪽 끝에는 브룩스가 있다. 그는 2025년 9월 26일 에세이에서 사람 수준의 손재주가 수십 년 안에 온다는 믿음을 순전한 공상이라고 불렀다("pure fantasy thinking")(로드니 브룩스, 2025년 9월 26일). 사람의 손재주는 풍부한 촉각에 기대는데 촉각 데이터에는 글이나 영상처럼 쌓인 전통이 없다는 것이 근거다. 그는 휴머노이드가 처음으로 돈을 버는 배치까지 10년 넘게 걸린다고 봤다. 걸어 다니는 휴머노이드는 사람 곁에 두기에 안전하지 않다고도 했다. 2026년 1월 1일 낸 예측표에는 현장에 쓸 만한 로봇의 손재주가 2036년 이후에도 사람 손에 견주면 형편없을 것이라는 예측을 새로 올렸다("Deployable dexterity will remain pathetic compared to human hands beyond 2036").

브룩스의 반론은 배관공의 자리와 바로 닿는다. 배관 수리는 손재주와 사람 곁의 안전이 함께 필요한 일이다. 공장의 정해진 자리에서 부품을 옮기는 일과 남의 집 싱크대 밑에서 낡은 이음새를 푸는 일은 요구하는 손이 다르다. 전문가들의 시간표는 넓게 퍼져 있다. 낙관 쪽 끝에는 2027년 일반 판매를 말하는 머스크와 2026년 인도를 목표로 한 1X가 있다. 중간에는 2028년에서 2030년을 보는 쪽이 있다. 카메라 회사 리얼센스의 크리스 마티유는 많은 가정에 휴머노이드가 들어오는 때를 2030년쯤으로 봤다(R&D World, 2025년 10월 23일). 비관 쪽 끝은 2036년 이후를 말하는 브룩스다.

한국 기업의 시간표는 중간에 있다. 삼성전자는 레인보우로보틱스를 자회사로 두고 2028년 휴머노이드 상용화를 목표로 삼았다(서울파이낸스, 2025년 12월 31일). 현대차그룹은 보스턴다이내믹스의 아틀라스를 2028년 미국 조지아의 현대차그룹 메타플랜트 아메리카(HMGMA)에 넣을 계획이다(로봇신문, 2026년 1월 6일). 정부는 2025년 4월 K-휴머노이드 연합을 출범시켰고(산업통상자원부, 2025년 4월 10일), 2026년 8월 27일 2030년까지 정부 재정 2조 3천억 원을 들이는 계획을 논의했다(정책브리핑, 2026년 8월 27일). 10월 13일에는 로브로스의 휴머노이드 20대를 제조 현장 네 곳에 넣는 실증이 규제특례 승인을 받았다(파이낸셜뉴스, 2026년 10월 13일).

우리의 추정은 이렇다. 정해진 자리에서 같은 동작을 되풀이하는 공장 일은 2028년 전후에 휴머노이드가 맡기 시작할 것이다. 집마다 구조가 다르고 손끝 감각이 필요한 배관 수리는 그보다 늦은 2030년대에 들어설 것으로 본다. 지금부터 낙관 쪽 예측까지는 1년에서 2년이고, 비관 쪽 예측까지는 10년 남짓이다. 힌턴이 말한 오래라는 시간은 한 세대보다 10년 안쪽에 가까울 가능성이 크다. 브룩스의 말대로라면 그 10년 뒤에도 로봇의 손은 사람의 손에 못 미친다. 2026년 10월 현재 힌턴이 배관공 조언을 공개적으로 거둬들인 기록은 확인되지 않는다.

이 시간표는 기술과 보급에 대한 추정이며 배관공의 일자리가 그때 일제히 사라진다는 뜻은 아니다. 다만 지금 로봇이 못하는 일이 있다는 이유로 그 직업을 변화의 바깥에 둘 수는 없다. 교육과 고용 정책은 기능직에도 자동화가 들어올 가능성을 전제로, 사람이 맡을 판단과 현장 관리의 역할을 준비해야 한다. 남은 시간을 직업의 안전을 확인하는 데만 쓰기보다 다음 역할을 익히는 데 써야 한다.

참고 자료

가디언, Box, run, crash: China's humanoid robot games show advances and limitations, 2025년 8월 15일, https://www.theguardian.com/world/2025/aug/15/china-world-humanoid-robot-games-advances-limitations

팟스크립츠(The Diary Of A CEO 전사), Godfather of AI: I Tried to Warn Them, But We've Already Lost Control. Geoffrey Hinton Transcript and Discussion, 2025년 6월 16일, https://podscripts.co/podcasts/the-diary-of-a-ceo-with-steven-bartlett/godfather-of-ai-i-tried-to-warn-them-but-weve-already-lost-control-geoffrey-hinton

Business Insider, The Godfather of AI reveals which jobs are safest — and where 'everybody' will get replaced, 2025년 6월 16일, https://www.businessinsider.com/geoffrey-hinton-godfather-of-ai-safe-jobs-2025-6

Free Press Journal, AI To Make Billions Jobless: Godfather of AI Geoffrey Hinton Sounds Alarm, Urges People To Learn Hands-On Trades Like Plumbing, 2025년 8월 25일, https://www.freepressjournal.in/amp/tech/ai-to-make-billions-joblessgodfather-of-ai-geoffrey-hinton-sounds-alarm-urges-people-to-learn-hands-on-trades-like-plumbing

Fortune, Nvidia CEO Jensen Huang says AI will create 'hundreds of thousands' of new jobs—including for people who want to 'make things with their hands', 2026년 9월 1일, https://fortune.com/2026/09/01/nvidia-ceo-jensen-huang-ai-will-create-hundreds-of-thousands-of-jobs-skilled-trade-workers-data-centers/

NBER(Acemoglu, Autor, Johnson), Building Pro-Worker Artificial Intelligence, 2026년 2월 23일, https://www.nber.org/papers/w34854

NBER(Acemoglu), The Simple Macroeconomics of AI, 2024년 5월 27일, https://www.nber.org/papers/w32487

Al Jazeera, Humans outrun robots at Beijing half-marathon, 2025년 4월 19일, https://www.aljazeera.com/features/2025/4/19/humans-outrun-robots-at-beijing-half-marathon

TechCrunch, Robots run a half-marathon, slowly, 2025년 4월 19일, https://techcrunch.com/?p=2996274

SCMP, Unitree's H1 robot wins 1,500 metre race as China hosts world's first humanoid games, 2025년 8월, https://www.scmp.com/tech/big-tech/article/3322025/unitrees-h1-robot-wins-1500-metre-race-china-hosts-worlds-first-humanoid-games

Global Times, Humanoid robot breaks human half-marathon record; Drastic improvement reflects systemic advances in China's robot technologies: expert, 2026년 4월 19일, https://www.globaltimes.cn/page/202604/1359229.shtml

iRunFar, Human Half-Marathon World Record Zapped by Humanoid Robot at the 2026 Beijing E-Town Half Marathon, 2026년 4월 20일, https://www.irunfar.com/2026-beijing-e-town-half-marathon-humanoid-robot-beats-human-world-record

China Daily, Multiple machines beat human records at second World Humanoid Robot Games, 2026년 8월 25일, https://www.chinadaily.com.cn/a/202608/25/WS6a8d427ae4b057c97e209a42.html

위키백과, World Humanoid Robot Games, 2026년 10월 14일 열람, https://en.wikipedia.org/wiki/World_Humanoid_Robot_Games

TechRadar, Humanoid robots are breaking our world records at the Robot Games — but why can't sprinting robots stop properly?, 2026년 8월 24일, https://www.techradar.com/ai-platforms-assistants/robots-have-smashed-human-world-records-for-the-100m-sprint-and-high-jump-at-chinas-robot-olympics-but-the-biggest-hurdle-is-stopping-safely

Gizmodo, The Best Robot Fails From the 2026 World Humanoid Robot Games, 2026년 8월 25일, https://gizmodo.com/the-best-robot-fails-from-the-2026-world-humanoid-robot-games-2000802858

셸리 파머, Humanoid Olympics: Propaganda or Prototype for the Next Big Thing?, 2025년 5월, https://shellypalmer.com/2025/05/humanoid-olympics-propaganda-or-prototype-for-the-next-big-thing/

SCMP(Jeffrey Wu), China just beat the US in soft power with the humanoid robot games, 2025년(날짜 미확인), https://www.scmp.com/opinion/china-opinion/article/3324482/china-just-beat-us-soft-power-

humanoid-robot-games

Forbes(John Koetsier), The Humanoid Robot Games Events Are A Market Map, 2026년 8월 19일, https://www.forbes.com/sites/johnkoetsier/2026/08/19/the-world-humanoid-robot-games-events-are-a-market-map/

전자신문, [김경진의 AI전략노트] 〈35〉 대한민국 휴머노이드 공사(公社)를 세우자, 2026년 8월 25일, https://www.etnews.com/20260825000157

humanoid.guide, Linear vs Rotary Actuators in Humanoid Joints: Why Tesla Bet on the Roller Screw, 2026년 8월 29일, https://humanoid.guide/linear-vs-rotary-actuators-humanoid-joints/

humanoid.guide, Leaderdrive harmonic reducers rise on humanoid demand, 2026년 4월 27일, https://humanoid.guide/leaderdrive-harmonic-reducers-surge-as-humanoid-demand-lifts-shares/

Unitree H1 문서, H1 Overview, 날짜 미표기, https://www.docs.quadruped.de/projects/h1/html/h1_overview.html

유비테크, About-UBTECH, 날짜 미표기, https://www.ubtrobot.com/en/about/company-profile

Teslarati, Elon Musk shares ridiculous fact about Optimus' hand demos, 2025년 11월 12일, https://www.teslarati.com/elon-musk-shares-shocking-fact-optimus-hand-demos/

Teslarati, Tesla Optimus V3 hand and arm details revealed in new patents, 2026년 4월 16일, https://www.teslarati.com/tesla-optimus-v3-hand-arm-details-revealed-new-patents/

피규어, Introducing Figure 03, 2025년 10월 9일, https://www.figure.ai/news/introducing-figure-03

TechCrunch, Famed roboticist says humanoid robot bubble is doomed to burst, 2025년 9월 26일, https://techcrunch.com/2025/09/26/famed-roboticist-says-humanoid-robot-bubble-is-doomed-to-burst

임페리얼칼리지런던, Skills development in Physical AI could cultivate lifelike intelligent robots, 2020년 11월 10일, https://www.imperial.ac.uk/news/208053/skills-development-physical-ai-could-cultivate

엔비디아 블로그, CES 2025: AI Advancing at 'Incredible Pace,' NVIDIA CEO Says, 2025년 1월 7일, https://blogs.nvidia.com/blog/ces-2025-jensen-huang/

엔비디아, NVIDIA Announces Isaac GR00T N1 — the World's First Open Humanoid Robot Foundation Model — and Simulation Frameworks to Speed Robot Development, 2025년 3월 18일, https://nvidianews.nvidia.com/news/nvidia-isaac-gr00t-n1-open-humanoid-robot-foundation-model-simulation-frameworks

구글 딥마인드, Gemini Robotics On-Device brings AI to local robotic devices, 2025년 6월 24일, https://deepmind.google/blog/gemini-robotics-on-device-brings-ai-to-local-robotic-devices/

The Robot Report, Google DeepMind says Gemini Robotics 2 enables full body control, 2026년 8월 2일, https://www.therobotreport.com/google-deepmind-says-gemini-robotics-2-enables-full-body-control/

arXiv, Qwen-RobotNav Technical Report: A Scalable Navigation Model Designed for an Agentic Navigation System, 2026년 6월, https://arxiv.org/abs/2606.18112

피규어, Helix: A Vision-Language-Action Model for Generalist Humanoid Control, 2025년 2월 20일, https://www.figure.ai/news/helix

피규어, Introducing Helix 02: Full-Body Autonomy, 2026년 1월 27일, https://www.figure.ai/news/helix-02

Gigazine, NVIDIA launches Jetson AGX Thor development kit for robotics brains at $3,499, with shipping expected in September 2025, 2025년 8월 26일, https://gigazine.net/gsc_news/en/20250826-nvidia-jetson-agx-thor/

Teslarati, Why Elon Musk's target date for buying an Optimus robot is harder than it sounds, 2026년 10월 12일, https://www.teslarati.com/elon-musk-tesla-optimus-for-sale-timeline/

1X, 1X NEO Home Robot, 날짜 미표기, https://www.1x.tech/discover/neo-home-robot

피규어, F.03 Arrives at BMW, 2026년 6월 30일, https://www.figure.ai/news/f-03-at-bmw

The Robotics Media, Figure Rolls 1,000th Figure 03 Humanoid Off Its BotQ Line, 2026년 8월 2일, https://theroboticsmedia.com/article/figure-ai-1000th-figure-03-humanoid-botq-manufacturing-milestone-july-23-2026

Fortune, More people will own a humanoid robot than a car by 2060, BofA predicts, 2026년 3월 13일, https://fortune.com/2026/03/13/bank-of-america-humanoid-robot-forecast-3-billion-2060

로드니 브룩스, Why Today's Humanoids Won't Learn Dexterity, 2025년 9월 26일, https://rodneybrooks.com/why-todays-humanoids-wont-learn-dexterity/

로드니 브룩스, Predictions Scorecard, 2026 January 01, 2026년 1월 1일, https://rodneybrooks.com/predictions-scorecard-2026-january-01/

R&D World, How humanoid robots could enter homes in the coming years, 2025년 10월 23일, https://www.rdworldonline.com/how-humanoid-robots-could-enter-homes-2030/

서울파이낸스, 삼성전자 로봇 사업 1년…글로벌 시장 공략 채비 완료, 2025년 12월 31일, https://www.seoulfn.com/news/articleView.html?idxno=617213

로봇신문, 현대차그룹, 차세대 아틀라스 세계 첫 공개…2028년 북미공장 투입, 2026년 1월 6일, https://www.irobotnews.com/news/articleView.html?idxno=44231

산업통상자원부, K-휴머노이드 연합 출범, 2025년 4월 10일, https://www.motir.go.kr/kor/article/ATCL3f49a5a8c/170404/view?mno=&pageIndex=1&rowPageC=0&displayAuthor=&searchCategory=0&schClear=on&startDtD=&endDtD=&searchCondition=1&searchKeyword=%ED%9C%B4%EB%A8%B8%EB%85%B8%EC%9D%B4%EB%93%9C

정책브리핑, 2030년까지 휴머노이드에 2.3조 투입…핵심부품 국산화율 80%로↑, 2026년 8월 27일, https://www.korea.kr/news/policyNewsView.do?newsId=148970725

파이낸셜뉴스, 산업부, 제조현장 휴머노이드 20대 투입…4족보행 순찰로봇도 허용, 2026년 10월 13일, https://www.fnnews.com/news/202610131103564249



7장. 인간 노동의 근본적 변화

지식 노동은 인공지능에, 사무 업무는 에이전트에, 몸으로 하는 일은 휴머노이드에

인공지능의 능력이 급격히 커지고 도입이 빠르게 번지면 일자리 없는 호황이 온다("ushering in a jobless boom"). 마이클 바 미국 연방준비제도(연준) 이사가 2026년 2월 17일 뉴욕 기업경제학자협회 연설에서 인공지능이 노동시장에 가져올 앞날을 세 가지로 나누며 둘째로 든 시나리오다(연준, 2026년 2월 17일). 전문직과 서비스직과 제조업과 운송업의 일자리를 에이전트와 로봇이 넘겨받고 실업이 크게 느는 경우다. 바 이사는 지금까지의 증거가 점진적 확산을 그린 첫째 시나리오에 더 가깝다고 봤다. 그가 인용한 자료에서 고용이 먼저 줄어든 쪽은 인공지능에 많이 노출된 직종의 초기 경력 노동자였다.

지식 노동의 변화는 사용 기록에 먼저 남는다. 앤트로픽은 자사 인공지능 클로드의 사용 기록을 직업과 과업으로 나눠 경제 지수를 낸다. 인공지능을 파는 회사가 낸 집계라는 점은 감안해야 한다. 2025년 2월 첫 보고서에서는 직업의 약 36퍼센트가 관련 과업의 4분의 1 이상에 클로드를 썼다(앤트로픽, 2025년 2월 10일). 앤트로픽의 2026년 3월 24일 보고서에서는 이 비율이 약 49퍼센트였다. 첫 보고서는 사용 방식을 보완 57퍼센트와 대체 43퍼센트로 나눴다. 보완은 사람이 주도해 일하고 인공지능이 돕는 방식이다. 대체는 인공지능이 과업을 직접 수행하는 방식이다. 오픈AI 연구진의 2025년 9월 논문에서는 챗GPT 메시지 가운데 업무와 관계없는 메시지가 70퍼센트를 넘었다(NBER, 2025년 9월). 업무용 메시지에서는 글쓰기 요청이 주를 이뤘다.

사무 업무에서는 에이전트가 상담 창구로 들어갔다. 마크 베니오프 세일즈포스 최고경영자는 2025년 9월 초 팟캐스트에서 고객지원 인력을 약 9천 명에서 약 5천 명으로 줄였다고 말했다(더레지스터, 2025년 9월 2일). 회사 대변인은 에이전트를 들인 뒤 지원 요청이 줄어 빈자리를 채우지 않았고 수백 명은 영업과 고객관리 부서로 옮겼다고 설명했다. 아빈드 크리슈나 IBM 최고경영자는 인공지능이 인사 직원 수백 명의 일을 넘겨받았다고 말했다(Entrepreneur, 2025년 5월 8일). 그는 그렇게 아낀 돈으로 프로그래머와 영업직을 더 뽑아 전체 고용은 늘었다고 덧붙였다.

몸으로 하는 일은 이 기록에 거의 나오지 않는다. 앤트로픽은 2026년 6월 26일 보고서에서 약 9700명의 설문 응답을 사용 기록과 연결해 분석했다. 응답자의 약 30퍼센트가 컴퓨터와 수학 직군이었는데 이 직군은 미국 고용의 4퍼센트다. 운송, 음식 조리, 건설 직군은 설문과 사용 기록 양쪽에서 적게 잡혔다. 이 장은 화면 안의 지식·사무 과업에서 몸을 쓰는 일로 자동화가 넓어지는 흐름에 주목한다. 다만 이 사용 기록만으로 지식 노동과 사무 업무, 육체노동의 전환 순서를 확정할 수는 없다. 그 순서와 시점을 입증한 경제학 연구도 아직 없다. 직종별로 무엇이 먼저 바뀌는지는 실제 도입과 고용 자료를 함께 살펴야 한다.

일자리 통계가 아직 조용한 이유

미국 노동통계국(BLS)은 2026년 10월 2일 9월 실업률을 4.2퍼센트로 발표했다. 9월 한 달 비농업 일자리는 2만 9천 개 늘었다. 한국의 고용 지표도 겉으로는 조용하다. 국가데이터처의 8월 고용동향에서 취업자는 1년 전보다 18만 4천 명 늘었고 15세에서 64세 고용률은 70.4퍼센트였다(파이낸셜뉴스, 2026년 9월 9일). 같은 발표에서 15세에서 29세 청년 취업자는 14만 3천 명 줄었다. 청년 취업자 감소는 46개월째 이어졌다.

바 이사는 앞의 연설에서 인공지능의 영향이 통계에 잘 잡히지 않는 이유를 넷으로 정리했다. 먼저 확산이 아직 이르다. 2025년 12월 기준 미국 기업의 17퍼센트가 인공지능을 썼고 250인 이상 기업에서는 약 30퍼센트였다. 기업은 사람을 내보내기보다 안에서 자리를 옮긴다. 기술이 생산성으로 이어지기까지 시차가 있고 영향은 초기 경력 노동자와 몇몇 직종에 몰려 있다.

시차에는 오래된 이름이 있다. 경제학자 로버트 솔로는 1987년 서평에서 컴퓨터 시대가 생산성 통계만 빼고 어디서나 보인다고 썼다. 새 기술이 퍼지는데도 생산성 통계가 따라 오르지 않는 현상을 생산성 역설이라고 부른다. 에릭 브린욜프슨 스탠퍼드대 교수와 동료들은 2017년 논문에서 인공지능이 범용기술이어서 기업이 일하는 방식을 바꾸는 데 시간이 걸린다고 설명했다(NBER, 2017년 11월). 노동통계국이 2026년 9월 3일 낸 수정치에서 미국 비농업 노동생산성은 2분기에 1년 전보다 2.2퍼센트 올랐다.

데이비드 오터 MIT 교수는 영향이 흩어지는 방식을 든다. 2000년대 중국산 수입품의 충격은 남부 몇몇 카운티의 공장 일자리에 몰렸다. 인공지능은 산업 하나가 아닌 여러 직무와 과업에 걸쳐 퍼진다. 오터는 사무직 일자리가 줄어든 일도 전국의 사무실에 흩어져 있어 눈에 잘 띄지 않았다고 말했다("not nearly as salient or visible")(Issues in Science and Technology, 2026년 1월 13일).

통계가 조용한 까닭이 영향이 작기 때문이라는 시각도 있다. 대런 아세모글루 MIT 교수는 2024년 논문에서 앞으로 10년 동안 인공지능이 영향을 미치는 과업을 전체의 약 5퍼센트로 추정했다(NBER, 2024년 5월 27일). 일자리가 아닌 과업의 5퍼센트다. 그는 조직의 도입이 느리고 많은 과업이 암묵지와 사람 사이의 상호작용을 요구한다는 점을 근거로 든다. 예일대 버짓랩도 2026년 9월 추적 지표에서 인공지능과 연결된 노동시장 교란의 뚜렷한 증거를 찾지 못했다. OECD의 마티아스 코먼은 2026년 7월 7일 고용전망을 내며 기업의 인공지능 사용이 노동 수요를 전반적으로 줄인다는 증거가 아직 없다고 말했다(저널뒤넷, 2026년 7월). 세 자료는 모두 경제 전체의 총량을 잰다. 총량이 조용해도 그 안의 한 칸은 움직일 수 있다. 바 이사가 넷째로 든 이유가 그 칸이고 변화는 일터로 들어가는 입구에서 먼저 나타난다.

신입 채용이 줄고 외주가 줄어드는 방식

브린욜프슨 연구팀은 미국 급여 대행사 ADP의 급여 자료로 젊은 노동자의 고용을 추적한다. 논문 제목은 「탄광 속의 카나리아?(Canaries in the Coal Mine?)」다. 2026년 8월 12일 개정판에 따르면 인공지능 노출도가 높은 직종의 22세에서 25세 고용은 노출도가 낮은 직종의 같은 또래를 따라갔을 경우보다 약 19퍼센트 낮다(스탠퍼드 디지털경제연구소, 2026년 8월 12일). 2022년 11월부터 2026년 6월까지의 자료를 쓴 판본이다. 2025년 8월 첫 판의 수치는 13퍼센트였다. 경력자에게서는 비슷한 격차가 보이지 않았고 조정은 해고보다 신규 채용 감소로 나타났다. 저자들은 이 결과가 관찰한 흐름을 그대로 적은 것이고 인공지능 효과의 인과 추정은 아니라고 밝혔다("These are descriptive patterns, not causal estimates of the effect of AI").

신입이 줄어든 것은 인공지능보다 경기와 금리 탓이라는 반론이 있다. 금리가 오르고 경기가 식으면 기업은 신입부터 덜 뽑는다. 한국경제인협회가 매출 500대 기업 121곳에 물은 2026년 하반기 조사에서 채용 계획을 정하지 못했거나 없는 기업은 도소매와 건설과 유통에 많았다(머니투데이, 2026년 9월 29일). 한경협은 그 이유로 내수와 건설 경기의 더딘 회복을 들었고 인공지능은 거론하지 않았다. 브린욜프슨 연구팀은 2026년 2월 9일 스탠퍼드 디지털경제연구소에 올린 후속 글에서 금리 가설을 따졌다. 노출도가 높은 직종은 평균적으로 금리 변화에 덜 민감했다. 연구팀은 금리 가설을 강하게 뒷받침하는 근거를 찾지 못했다. 다만 기업별 시기 요인까지 통제하면 노출 직종 청년의 상대적 감소는 2024년에야 뚜렷해졌다. 연구팀은 그 이전의 감소에는 인공지능 말고 다른 요인이 적어도 일부 섞였다고 인정했다.

한국 자료도 같은 입구를 가리킨다. 1장에서 본 한국은행 2025년 10월 보고서는 국민연금 가입자 자료로 최근 3년 동안 청년층 일자리 21만 1천 개가 줄었다고 했다(한국은행, 2025년 10월 30일). 그중 20만 8천 개가 인공지능 고노출 업종에서 줄었다. 한국은행의 2026년 8월 후속 보고서에서는 기간이 4년으로 늘었다. 이 기간 청년층 일자리는 28만 5천 개 줄었고 감소분의 94퍼센트가 고노출 업종에 몰렸다(서울신문, 2026년 8월 18일). 신규 채용에서 청년이 차지하는 비중은 2014년부터 2019년 평균 36.9퍼센트에서 2024년부터 2026년 29.3퍼센트로 내려갔다(아시아경제, 2026년 8월 18일). 한국은행은 이 감소를 인공지능만의 효과로 단정하기 어렵다고 했다. 팬데믹 때 늘린 채용의 정상화와 경력직 선호에 인공지능이 속도를 더했다는 설명이다.

외주는 회사가 맡던 일을 바깥 업체나 개인에게 계약으로 맡기는 방식이다. 외주의 변화는 계약을 끊기보다 새 계약을 맺지 않는 쪽으로 온다. 듀오링고는 2025년 4월 인공지능이 할 수 있는 일을 계약직에서 인공지능으로 옮기겠다고 밝혔다(테크크런치, 2025년 5월 4일). 그 전부터 번역과 글쓰기 계약직을 차례로 줄여 온 회사다. 61개국 온라인 프리랜서 공고 약 200만 건을 분석한 연구는 챗GPT 출시 뒤 8개월 안에 글쓰기와 코딩 공고가 21퍼센트 줄었다고 보고했다(DIW, 2025년). 비교 대상은 손으로 하는 작업의 공고였고 남은 일감은 더 복잡하고 보수가 높았다.

콜센터와 사무 처리를 통째로 다른 회사에 맡기는 산업을 BPO(업무처리 아웃소싱)라고 한다. 필리핀 IT-BPM협회는 2028년 업계 인력 전망을 2022년에 내놓은 250만 명에서 185만 명에서 214만 명 사이로 낮췄다(비즈니스월드, 2026년 8월 19일). 실제 감원 통계가 아닌 전망의 하향이다. 일이 사라지는 모습은 해고 통지보다 오지 않는 발주에 가깝다.

한국 기업과 공공기관의 에이전트 도입 현황

삼성, LG, SK, 네이버와 정부 부처의 에이전트 도입 발표는 2026년 상반기에 잇따랐다(15장). 노동의 관점에서 보면 이 발표 가운데 인력 감축을 말한 곳은 없다. SK텔레콤은 2026년 6월 16일 AI 에이전트에 사번을 주고 소속과 직무와 시스템 권한을 직원처럼 관리하겠다고 밝혔다(세계일보, 2026년 6월 16일). 에이전트를 구성원과 함께 계획하고 판단하고 실행하는 업무 주체로 보겠다는 뜻이다. 같은 발표에 사람을 줄이겠다는 말은 없었다.

인력의 변화는 채용 숫자에서 보인다. 4대 시중은행의 2026년 상반기 신입 채용은 485명으로 1년 전 595명보다 18.5퍼센트 줄었다(시사저널e, 2026년 10월 3일). 같은

보도에 따르면 은행들은 공채 밖에서 AI와 IT와 데이터 전문 인력을 따로 뽑는다. 콜센터에서는 은행마다 결과가 갈렸다. 경향신문이 확보한 자료에서 국민은행 상담 인력은 1,133명에서 2025년 869명으로 줄었다(경향신문, 2026년 4월 15일). 국민은행의 2025년 AI 처리율은 41.3퍼센트였고 상담원 한 명이 하루에 받는 전화는 적정선 70콜을 넘어 100콜에서 120콜이 됐다. 신한은행은 2021년 로보텔러를 도입한 이듬해 상담 인력이 610명에서 667명으로 늘었다. 2024년 AICC 플랫폼을 도입한 뒤에도 693명에서 2025년 697명으로 늘었다. 같은 업권에서도 기술 도입 뒤의 인력 변화는 은행마다 달랐다.

공공 부문의 도입은 업무 부담을 덜어 준다는 말로 소개된다. 행정안전부의 업무용 인공지능 온AI는 2026년 10월 현재 중앙행정기관 25곳에 들어갔고 연말까지 47곳으로 늘어난다(서울신문, 2026년 10월 13일). 54개 지방자치단체 공무원 약 2만 2천 명도 시범으로 쓰고 있다. 보고서 초안과 회의록 정리를 돕는 도구다. 정부는 9월 3일 공공기관 109곳과 직원 약 12만 명을 대상으로 한 기능 개혁을 발표했고 이튿날 비용 절감이 목표가 아니라고 설명했다(서울신문, 2026년 9월 4일). 재정경제부 관계자는 중복 인력은 재배치하지만 신규 채용은 줄지 않을 것이라고 말했다. 2026년 10월 14일까지 인공지능 도입을 이유로 인력을 줄이겠다고 공식 발표한 공공기관은 확인되지 않는다.

국회는 청년 쪽 숫자를 묻기 시작했다. 박정 더불어민주당 의원은 2026년 10월 13일 고용노동부 국정감사에서 한국은행 8월 보고서의 28만 5천 개를 인용했다(아주경제, 2026년 10월 13일). 그는 전체 취업자는 늘어도 청년 취업자는 2022년을 정점으로 줄었다고 지적했다. 김영훈 고용노동부 장관은 고용정책이 산업정책을 뒤따라가는 경향이 많았다고 답했다. 그는 "앞으로는 산업정책과 함께 선도해야 한다고 생각한다"고 말했다.

지금까지의 역할만 놓고 보면 인간의 자리는 줄어든다

기계가 사람의 일을 빼앗는다는 걱정은 오래됐다. 존 메이너드 케인스는 1930년 에세이에서 기술 실업이라는 말을 썼다. 노동을 아끼는 수단을 찾는 속도가 노동의 새 쓰임을 찾는 속도를 앞질러 생기는 실업이다. 그는 이를 일시적인 부적응 단계라고 불렀다("a temporary phase of maladjustment")(케인스, 1930년). 미국 존슨 행정부의 기술·자동화·경제진보 국가위원회가 1966년 2월 보고서를 낼 무렵 실업률은 3.8퍼센트까지 내려와 있었다(컨버서블 이코노미스트, 2014년 12월 1일). 1983년 바실리 레온티예프는 사람의 처지를 말에 견줬다. 1900년 미국의 말과 노새는 2100만 마리가 넘었지만 내연기관이 퍼진 뒤 1960년에는 말이 300만 마리만 남았다(포린어페어스,

2015년). 흔히 아마라의 법칙이라 부르는 말도 있다. 사람들은 기술의 효과를 단기에는 크게 보고 장기에는 작게 본다는 것이다("We tend to overestimate the effect of a technology in the short run").

이번에는 다르다고 말해 온 사람이 앤트로픽의 다리오 아모데이다. 그는 2025년 5월 인공지능이 5년 안에 신입 화이트칼라 일자리의 절반을 없애고 실업률을 10퍼센트에서 20퍼센트로 올릴 수 있다고 했다(포춘, 2025년 5월 28일). 2026년 1월 에세이 「기술의 사춘기(The Adolescence of Technology)」에서는 인공지능이 특정 직업이 아닌 사람의 노동 전반을 대신한다고 썼다("a general labor substitute for humans")(다리오 아모데이, 2026년 1월). 같은 글에서 그는 단기 전환이 과거 기술보다 유난히 고통스러울 것이라고 했다("unusually painful compared to past technologies"). 2026년 5월에는 어조가 달라졌다. 포춘에 따르면 그는 자동화를 일자리를 없애는 쪽보다 산출을 늘리는 쪽으로 설명했다(포춘, 2026년 5월 26일). 일의 90퍼센트를 자동화하면 모두가 나머지 10퍼센트를 한다는 것이다("If you automate 90% of the job, then everyone does the 10%"). 같은 기사에서 오픈AI의 샘 올트먼은 신입 일자리에 대한 자신의 예상이 틀려 기쁘다고 말했다("I'm delighted to be wrong about this").

어조가 바뀌어도 남는 사실이 있다. 오터와 동료들은 미국 11개 특허 법률사무소의 변호사 133명을 상대로 석 달 동안 실험을 했다(NBER, 2026년 9월). 인공지능을 쓰는 동안 명세서 품질은 10일째에 0.34 표준편차 올랐고 90일째에는 0.38 표준편차 올랐다. 주니어의 상승 폭이 더 컸다. 도구를 뺀 뒤에도 실력이 오른 채로 남은 쪽은 경력 7년 이상의 시니어였다. 시니어는 0.45 표준편차 높았다. 주니어는 평균으로 나아진 것이 없었다. 기존 과업으로 사람의 쓸모를 재면 주니어가 먼저 밀리고 기초 전문성을 갖춘 사람만 인공지능의 도움을 제 실력으로 남긴다.

지금까지 사람이 해 온 과업만 기준으로 보면 사람이 직접 수행할 몫은 줄어든다. 조사 초안과 상담 응대와 문서 정리 가운데 일부가 기계로 넘어가면서, 그 일을 배우며 경력을 시작하던 기회도 줄어들 수 있다. 앞의 고용 자료가 이 변화를 모두 인공지능의 효과로 입증한 것은 아니다. 그러나 총고용의 증감만으로는 무엇을 배우며 일을 시작할지라는 문제에 답할 수 없다. 기계에 일을 맡기고 결과를 판단하는 역할을 늘리는 일과, 초보자가 그 판단력을 기를 경로를 마련하는 일을 함께 추진해야 한다.

중국의 휴머노이드 생산 현황: 주요 기업, 생산량, 가격, 투입 현장

휴머노이드 수치는 생산량과 출하량을 나눠 읽어야 한다. 생산량은 공장에서 만든 대수이고 출하량은 고객에게 내보낸 대수다. 2025년 세계 수치는 집계 기관과 범위에 따라 다르다. 국제로봇연맹(IFR)은 산업용과 전문 서비스용 판매만 세어 약 7천 대로 봤다(신화망, 2026년 9월 22일). 옴디아는 세계 출하를 약 1만 3천 대로 집계했다(전자신문, 2026년 6월 30일). IDC는 세계 출하를 약 1만 8천 대로 봤고 중국 고공산업연구원(GGII)은 중국 출하를 약 1만 8천 대로 잡았다(CGTN, 2026년 1월 24일; 신화망, 2026년 1월 8일). 2026년 상반기는 스마트애널리틱스글로벌이 1만 9100대, 카운터포인트 리서치가 2만 2천 대, IDC가 약 2만 5천 대로 셌다(포브스, 2026년 9월 30일). 중국 휴머노이드 장면응용연합은 중국 기업만 3만 대 이상으로 셌다(포브스, 2026년 8월 20일). IDC는 상반기 출하의 95퍼센트 이상을 중국 기업이 만들었다고 봤다.

유니트리는 2026년 8월 19일 상하이 증시 과창판에 상장했다(더넥스트웹, 2026년 8월 17일). 투자설명서는 상장하려는 회사가 사업과 재무를 투자자에게 알리는 공시 문서다. 투자설명서 수치를 옮긴 보도에 따르면 유니트리의 2025년 휴머노이드 출하는 5,511대였다(36Kr, 2026년 8월 19일). 2025년 매출은 17억 위안이었다. 2025년 7월 3만 9900위안에 나온 R1은 2026년 1월 보도 기준 시작가가 2만 9900위안으로 내려갔다(신화망, 2026년 1월 8일). UBTech는 2026년 상반기 매출 12억 7천만 위안에 전신형 휴머노이드 921대를 출하했다(휴머노이즈 데일리, 2026년 8월 31일). 전신형 매출의 96.4퍼센트가 산업 고객의 맞춤 계약에서 나왔다. 주력 기종 워커 S2의 가격은 공개되지 않았다. 에이지봇은 2025년 5100대 넘게 출하했고 2026년 상반기에는 카운터포인트 집계로 약 9700대를 내보내 1위를 했다(휴머노이즈 데일리, 2026년 8월 20일).

공장에 들어간 대수는 아직 작다. 지리의 전기차 브랜드 지커의 닝보 공장에는 2025년 3월 보도 기준 UBTech 워커 S1 약 20대가 회사 사진에 잡혔다(SCMP, 2025년). 부품 업체 푸린정공은 2026년 1월 기준 100대 가까운 휴머노이드를 들였다(신화망, 2026년 1월 8일). 폭스콘은 미국 휴스턴 서버 공장에 2026년 4월 중순 10대 미만을 넣을 계획이었다(Instrumental, 2026년 4월 20일). 카운터포인트 집계로 2026년 상반기 출하의 33.6퍼센트는 공연과 오락 용도였고 27.0퍼센트는 데이터 생산과 연구 용도였다. 제조 용도는 12.8퍼센트였고 창고 물류는 4.9퍼센트였다.

중국의 숫자가 부풀려졌다는 시각도 근거를 갖고 있다. 수전 빌러 IFR 사무총장은 2025년에 팔린 휴머노이드 상당수가 생산 작업이 아닌 인공지능 학습용 데이터 생산에 쓰였다고 말했다(신화망, 2026년 9월 22일). 중국 국가발전개혁위원회도 2025년 11월 150곳이 넘는 기업이 비슷한 로봇을 쏟아낸다며 거품 위험을 경고했다(블룸버그, 2025년

11월 27일). 이 비판은 숫자를 지어냈다는 뜻이 아니다. 세는 범위가 넓고 수요의 상당 부분이 연구와 시연에 있다는 지적이다. 지금까지의 역할만 놓고 보면 사람의 자리는 화면 앞의 일에서 먼저 줄었고 공장에서는 이제 로봇의 대수를 세기 시작했다. IFR이 휴머노이드 판매량을 따로 떼어 발표한 것은 2026년 9월이 처음이다.

노동의 변화를 준비하는 정책도 로봇 대수와 감원 숫자만 세는 데 머물러서는 안 된다. 기업과 공공기관은 자동화한 과업, 새로 맡긴 과업, 신입이 숙련을 쌓는 기회가 어떻게 달라졌는지 함께 살펴야 한다. 사람의 역할을 다시 설계하지 않은 채 대체되는 일을 피하는 데만 집중하면, 자동화가 넓어질 때마다 같은 불안을 되풀이하게 된다.

참고 자료

연준(마이클 바), What Will Artificial Intelligence Mean for the Labor Market and the Economy?, 2026년 2월 17일, https://www.federalreserve.gov/newsevents/speech/barr20260217a.htm

앤트로픽, Introducing the Anthropic Economic Index, 2025년 2월 10일, https://www.anthropic.com/news/the-anthropic-economic-index

앤트로픽, Anthropic Economic Index report: Learning curves, 2026년 3월 24일, https://www.anthropic.com/research/economic-index-march-2026-report

앤트로픽, Anthropic Economic Index report: Cadences, 2026년 6월 26일, https://www.anthropic.com/research/economic-index-june-2026-report

NBER(Chatterji 외), How People Use ChatGPT, 2025년 9월, https://www.nber.org/papers/w34255

더레지스터, Salesforce sacrifices 4,000 support jobs on the altar of AI, 2025년 9월 2일, https://www.theregister.com/2025/09/02/salesforce_4000_jobs_ai

Entrepreneur, IBM Replaced Hundreds of HR Workers With AI, According to Its CEO, 2025년 5월 8일, https://www.entrepreneur.com/business-news/ibm-ceo-ai-replaced-hundreds-of-human-resources-staff/491341

미국 노동통계국(BLS), Employment Situation Summary (September 2026), 2026년 10월 2일, https://www.bls.gov/news.release/empsit.nr0.htm

파이낸셜뉴스, 8월 취업 증가폭 5개월만에 최대... 청년 고용률은 28개월째 내리막, 2026년 9월 9일, https://www.fnnews.com/news/202609091816217189

NBER(Brynjolfsson, Rock, Syverson), Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics, 2017년 11월, https://www.nber.org/papers/w24001

미국 노동통계국(BLS), Productivity and Costs, Second Quarter 2026, Revised, 2026년 9월 3일, https://www.bls.gov/news.release/prod2.nr0.htm

Issues in Science and Technology(David Autor), How Is AI Shaping the Future of Work?, 2026년 1월 13일, https://issues.org/david-autor-economist-ai-future-work/

NBER(Acemoglu), The Simple Macroeconomics of AI, 2024년 5월 27일, https://www.nber.org/papers/w32487

예일대 버짓랩, Tracking the Impact of AI on the Labor Market, 2026년 9월, https://budgetlab.yale.edu/sites/default/files/page_to_pdf/1419/publication_1419.pdf

저널뒤넷, L'OCDE affirme que l'IA ne provoque pas de "baisse généralisée" de l'emploi dans ses pays membres, 2026년 7월 7일, https://www.journaldunet.com/business/1552481-l-ocde-affirme-que-l-intelligence-artificielle-ne-provoque-pas-de-baisse-generalisee-de-l-emploi-dans-ses-pays-membres-selon-son-rapport-2026/

스탠퍼드 디지털경제연구소, No Widespread Displacement, but the AI Employment Gap for Young Workers Has Widened to 19%, 2026년 8월 12일, https://digitaleconomy.stanford.edu/news/canariesaug26/

스탠퍼드 디지털경제연구소, Canaries, Interest Rates, and Timing: More on the Recent Drivers of Employment Changes for Young Workers, 2026년 2월 9일, https://digitaleconomy.stanford.edu/news/canaries-interest-rates-and-timinga-more-on-recent-drivers-of-employment-changes-for-young-workers

머니투데이, 올해 하반기 채용시장 온기…대기업 51% "채용계획 세웠다", 2026년 9월 29일, https://www.mt.co.kr/industry/2026/09/29/2026092909334019258

한국은행, AI 확산과 청년고용 위축: 연공편향(seniority-biased) 기술변화를 중심으로(BOK 이슈노트 제2025-30호), 2025년 10월 30일, https://www.bok.or.kr/portal/bbs/P0002353/view.do?nttId=10094258&menuNo=200433

서울신문, 일 잘하는 AI 신입, 청년부터 밀어냈다, 2026년 8월 18일, https://www.seoul.co.kr/news/economy/finance/2026/08/18/20260818500238

아시아경제, AI가 앞당긴 청년 실업…신규채용 청년 비중 30% 아래로, 2026년 8월 18일, https://view.asiae.co.kr/article/2026081814283987947

테크크런치, Is Duolingo the face of an AI jobs crisis?, 2025년 5월 4일, https://techcrunch.com/2025/05/04/is-duolingo-the-face-of-an-ai-jobs-crisis/

DIW(Demirci, Hannane, Zhu, Management Science 게재), Who Is AI Replacing? The Impact of Generative AI on Online Freelancing Platforms, 2025년, https://www.diw.de/de/diw_01.c.906828.de/publikationen/externe_referierte_aufsaetze/2025_0010/who_is_ai_replacing__the_impact_of_generative_ai_on_online_freelancing_platforms.html

비즈니스월드, (제목 미확인, URL 기준), 2026년 8월 19일, https://bworldonline.com/top-stories/2026/08/19/770985/ai-seen-as-significant-but-manageable-threat-to-philippine-bpo-industry/

세계일보(네이트 전재), AI 에이전트에 사번 부여…SKT, AX 조직 2.0 시동, 2026년 6월 16일, https://m.news.nate.com/view/20260616n25276

시사저널e, 은행 채용도 'AI 중심'…신입 줄이고 전문인력 골라 뽑는다, 2026년 10월 3일, https://www.sisajournal-e.com/news/articleView.html?idxno=424090

경향신문, [단독] AI 도입 콜센터, 월급도 깎았다…대체 1순위? "자르기 쉬워 잘랐을 뿐"[딸깍, 노동②], 2026년 4월 15일, https://www.khan.co.kr/article/202604150600111

서울신문, "보고서에 회의록 정리까지"…'AI 동료'에 공직사회 확 달라졌다, 2026년 10월 13일, https://www.seoul.co.kr/news/society/2026/10/13/20261013500019

서울신문, 정부 "공공기관 개혁, 비용 절감 목표 아냐…신규채용 안줄여", 2026년 9월 4일, https://www.seoul.co.kr/news/economy/2026/09/04/20260904500216

아주경제, [2026 국정감사] AI 확산에 청년 일자리 감소…노동장관 '좋은 일자리 기준 조속히 마련', 2026년 10월 13일, https://ajunews.com/view/20261013153817871

존 메이너드 케인스, Economic Possibilities for our Grandchildren, 1930년, https://www.hetwebsite.net/het/texts/keynes/keynes1930grandchildren.htm

컨버서블 이코노미스트(Timothy Taylor), Automation and Job Loss: The Fears of 1964, 2014년 12월 1일, https://conversableeconomist.com/2014/12/01/automation-and-job-loss-the-fears-of-1964/

포린어페어스(Brynjolfsson, McAfee, 필라델피아 연준 게시본), Will Humans Go the Way of Horses? Labor in the Second Machine Age, 2015년, https://www.philadelphiafed.org/-/media/FRBP/Assets/Events/201

5/The-Economy/2015-Philadelphia-Fed-Policy-Forum/brynjolfsson-humans-horses.pdf

DevIQ, Amara's Law: A Quick Guide on Technology Predictions, 날짜 미표기, https://deviq.com/laws/amaras-law/

포춘, Anthropic CEO warns AI could eliminate half of all entry-level white-collar jobs, 2025년 5월 28일, https://fortune.com/2025/05/28/anthropic-ceo-warning-ai-job-loss

다리오 아모데이, The Adolescence of Technology, 2026년 1월, https://www.darioamodei.com/essay/the-adolescence-of-technology

포춘, Sam Altman and Dario Amodei are both walking back their AI jobs apocalypse prophecies as they eye blockbuster IPOs, 2026년 5월 26일, https://fortune.com/2026/05/26/sam-altman-dario-amodei-walking-back-ai-jobs-apocalypse-prophecies-ipo/

NBER(Autor 외), Does AI Assistance Enhance or Erode Expertise? Evidence from a Three-Month Field Experiment in Patent Drafting, 2026년 9월(10월 개정), https://nber.org/papers/w35720

신화망, 2025年全球售出约7000台人形机器人, 2026년 9월 22일, https://www.news.cn/20260922/846a2b49087946268aeffb7494a5ec70/c.html

전자신문, [김경진의 AI전략노트] 〈31〉 휴머노이드, 다음 질서를 짜는 손, 2026년 6월 30일, https://www.etnews.com/20260630000220

CGTN, IDC report: China leads the global humanoid robot rise in 2025, 2026년 1월 24일, https://news.cgtn.com/news/2026-01-24/IDC-report-China-leads-the-global-humanoid-robot-rise-in-2025-1KccOGZyVGM/p.html

신화망, 2026人形机器人产业展望：10万级量产将至 商业化元年开启, 2026년 1월 8일, https://www.news.cn/tech/20260108/46b1220e159d4f80bc6a4240eb3b47b5/c.html

포브스(야후 파이낸스 전재), Humanoid Robot Shipments Jump 432%, China Has 95% Market Share, 2026년 9월 30일, https://finance.yahoo.com/technology/ai/articles/humanoid-robot-shipments-jump-432-025835160.html

포브스, Humanoid Robot Shipments Up 300%: Up To 30,000 So Far In 2026, 2026년 8월 20일, https://www.forbes.com/sites/johnkoetsier/2026/08/20/humanoid-robot-shipments-up-300-up-to-30000-so-far-in-2026/

더넥스트웹, Unitree, China's humanoid robot champion, lists in Shanghai on 19 August after record demand, 2026년 8월 17일, https://thenextweb.com/news/unitree-humanoid-robot-shanghai-ipo

36Kr, Unitree Technology IPO, Five Key Judgments, 2026년 8월 19일, https://eu.36kr.com/en/p/3945679641902464

휴머노이즈 데일리, UBTECH Reports 104% Revenue Surge in H1 2026 as Full-Size Humanoids Drive Growth, 2026년 8월 31일, https://www.humanoidsdaily.com/news/ubtech-reports-104-revenue-surge-in-h1-2026-as-full-size-humanoids-drive-growth

휴머노이즈 데일리, Global Humanoid Shipments Surpass 22,000 in H1 2026 as AGIBOT Leads and Form-Factor Debates Intensify, 2026년 8월 20일, https://www.humanoidsdaily.com/news/global-humanoid-shipments-surpass-22-000-in-h1-2026-as-agibot-leads-and-form-factor-debates-intensify

SCMP, Future of Made-in-China: UBTech Robotics deploys first humanoid 'team' in car factory, 2025년 (날짜 미확인), https://www.scmp.com/tech/tech-trends/article/3300872/future-made-china-ubtech-robotics-deploys-first-humanoid-team-car-factory

Instrumental, Foxconn's Humanoid Push Will Start Small in Houston, Engineer Says, 2026년 4월 20일, https://instrumental.com/resources/build-better-news/foxconns-humanoid-push-will-start-small-in-houston-engineer-says/

블룸버그, China Warns of Bubble Risks in Booming Humanoid Robots Arena, 2025년 11월 27일, https://www.bloomberg.com/news/articles/2025-11-27/china-warns-of-bubble-risk-in-booming-humanoid-robotics-industry



8장. 백지의 영역

언어 모델이 힘을 내는 곳: 인간이 한 번이라도 기록으로 남긴 영역

티머시 고워스는 2026년 8월 12일 블로그에서 수학 논문의 생김새를 이렇게 그렸다. 출판된 수학은 발견한 사람의 생각 과정을 감추고 깔끔하게 다듬은 증명이라는 것이다("tidied up proofs that hide the thought processes of their discoverers")(고워스 블로그, 2026년 8월 12일). 논문에는 막다른 길이 남지 않는다. 어떤 공략을 버리고 어떤 공략을 따라갔는지도 적히지 않는다. 언어 모델이 배우는 재료는 이렇게 다듬어진 기록이다.

1장에서 본 대로 언어 모델은 수십조 개의 토큰을 읽는다. 사람이 남긴 기록에도 끝은 있다. Epoch AI는 2024년 6월 논문에서 사람이 공개된 글로 남긴 텍스트의 유효 총량을 약 300조 토큰으로 추정했다(Epoch AI, 2024년 6월 6일). 모델이 이 총량을 다 쓰는 때는 2026년에서 2032년 사이로 봤다. 논문 페이지에는 2026년 10월 현재 갱신 공지가 없다.

기록이 있는 곳에서 모델은 사람 최고 수준에 올랐다. FrontierMath 4단계는 교수와 박사후연구원이 단기 연구 과제 수준으로 낸 수학 문제 묶음이다. 2025년 7월 공개 때 최고 점수는 5퍼센트였다. Epoch AI는 2026년 9월 10일 4단계의 마지막 문제가 풀렸다고 알렸다. 이때 GPT-6 Astra의 점수는 98퍼센트였다(알파시그널, 2026년 9월 10일). Epoch는 이 벤치마크가 포화했다고 본다. 출제한 수학자들은 앞선 문제에서 모델이 출제자가 생각하지 못한 지름길로 답을 낸 일이 잦았다고 전했다.

보간은 이미 아는 점들 사이를 메우는 일이다. 외삽은 아는 범위 바깥을 짐작하는 일이다. 2024년까지 나온 연구는 언어 모델이 보간에 강하다는 증거를 많이 쌓았다. 우 등 연구진은 같은 추론을 요구하되 9진법 덧셈처럼 기본 가정만 바꾼 과제에서 모델의 성적이 꾸준히 떨어진다고 보고했다(NAACL, 2024년 6월). 애플 연구진은 초등 수학 문제에 그럴듯하지만 풀이와 상관없는 조건 한 줄을 넣자 정답률이 최대 65퍼센트 떨어졌다고 보고했다(arXiv, 2024년 10월 7일). 테렌스 타오는 2024년 12월 인터뷰에서 많은 연구 분야는 관련 논문이 열두어 편뿐이라고 했다("a dozen papers")(Epoch AI, 2024년 12월 4일). 실패한 접근에서 교훈을 얻는 일은 모델이 아주 서툴다고도 했다. 중국 딥시크가 2025년 4월 공개한 증명 모델 DeepSeek-Prover-V2는 형식 증명 시험

MiniF2F에서 88.9퍼센트를 얻었지만 대학생 경시 문제를 모은 PutnamBench에서는 658문제 가운데 49문제를 풀었다(arXiv, 2025년 4월 30일).

이 구도는 2026년에 흔들렸다. 서장에서 본 5월 20일 평면 단위 거리 추측의 반증 뒤로 언어 모델이 묵은 추측을 뒤집은 사례가 이어졌다. 수론학자 아룰 샹카르는 단위 거리 결과를 두고 지금의 모델이 독창적인 착상을 낸다고 평했다("capable of having original ingenious ideas")(오픈AI, 2026년 5월 20일). 케빈 버자드 런던 임페리얼칼리지 교수는 7월 블로그 제목에 사람 수학자가 반례 찾기에서 기계에 밀린다고 적었다(버자드 블로그, 2026년 7월 20일). 언어 모델은 이미 외삽한다는 주장이다.

반대쪽 관찰도 같은 사례에서 나왔다. 멜리사 리는 전문가가 힌트 몇 개만 받았어도 같은 증명에 이르렀을 것이라고 썼다(The Conversation, 2026년 5월 26일). 그는 증명에 필요한 착상 다수가 이미 문헌에 있었다는 고워스의 말을 함께 인용했다. 우리는 두 관찰이 함께 맞다고 본다. 기록된 조각을 새로 잇는 일만으로도 80년 묵은 추측을 뒤집을 수 있다. 기록이 미치는 범위는 생각보다 넓고 언어 모델은 그 넓은 범위를 사람보다 빠르게 뒤진다.

누구도 가 보지 않은 곳에서 드러나는 한계

Epoch AI는 2026년 9월 1일 FrontierMath 에르되시를 공개했다(Epoch AI, 2026년 9월 1일). 2026년 8월 기준으로 풀리지 않은 에르되시 문제 가운데 의미 있는 것 68개를 토머스 블룸이 골랐다. 모델은 문제마다 300달러와 72시간 안에서 한 번씩 시도했고 Lean 증명을 내야 풀린 것으로 셌다. GPT-6 Astra가 2개를 풀어 3퍼센트를 기록했고 다른 네 모델은 하나도 풀지 못했다. 4단계 포화 발표보다 아흐레 앞서 나온 숫자다. 블룸은 앞서 단위 거리 결과에 붙인 논평에서 지식의 경계가 들쭉날쭉하다고 썼다("The frontiers of knowledge are very spiky")(오픈AI, 2026년 5월 20일).

처음 보는 환경에서 규칙을 익히는 시험에서도 격차가 남아 있다. ARC-AGI-3는 게임 환경 안에서 경험으로 규칙을 배우는 능력을 재는 벤치마크다. 한 집계에 따르면 2026년 10월 13일 기준 최고 점수는 GPT-6 Astra의 62.7퍼센트다(BenchLM, 2026년 10월 13일). 이 수치는 ARC Prize의 공식 검증치와 따로 확인하지 못했다.

고워스의 8월 글은 이 거리의 성격을 짚는다. 모델은 어려운 명제도 증명하지만 지금까지 이름난 성과는 대부분 반례였다는 것이다. 그는 2016년 캡셋 문제 풀이만큼 놀라운 증명이 나와야 장벽이 무너졌다고 인정하겠다고 했다. 그가 보는 격차는 계산

능력에 있지 않고 판단에 있다. 어느 길이 진전인지 알아채는 감각이 모자란다는 것이다("a sufficiently good nose"). 그 감각은 다듬어진 논문에 남지 않는다.

틀린 결과도 섞여 나온다. 오픈AI는 2026년 10월 6일 내부 모델이 낸 수학 결과를 GitHub 저장소로 공개했다(오픈AI, 2026년 10월 6일). 회사는 결과 하나에 챗GPT 프로로 평균 세 시간쯤 생각한 만큼의 계산을 썼다고 밝혔다. 원고는 722편이었고 372개 결과군으로 묶였다(길 칼라이 블로그, 2026년 10월 7일). 다음 날 오픈AI는 3편을 거둬들였다. 한 증명에서 부호 하나를 거꾸로 쓴 오류가 그 증명과 그것에 기댄 두 편을 함께 무너뜨렸다(MLQ, 2026년 10월 9일). 남은 719편 가운데 Lean 형식화가 붙은 결과는 약 42퍼센트다. 칼라이는 결과마다 검증 단계가 다르고 Lean 검증에도 빈틈이 있을 수 있다고 적었다. 버자드는 자신이 언어 모델에 맡겨 받은 수학 문서에 틀린 주장이 들어 있었고 자신도 그것을 놓쳤다고 7월 블로그에 털어놓았다. 그 오류를 찾아낸 것은 Logos라는 곳의 검증 도구였다. 그는 기계가 쓴 비형식 수학은 읽지 않고 형식 검증만 믿는다고 했다.

밀레니엄 문제는 클레이수학연구소가 2000년 상금 100만 달러씩을 걸고 고른 일곱 난제다. 오픈AI는 9월 8일 내부 모델이 그중 나비에 스토크스 문제를 풀었다고 발표했다(오픈AI, 2026년 9월 8일). 회사 설명으로는 1만 개 규모의 에이전트가 88시간 만인 9월 5일 답에 이르렀고 Lean 검증에 17시간이 더 들었다. 주장의 뼈대는 매끄러운 외력을 받는 유체가 유한한 시간 안에 속도가 끝없이 커지는 특이점을 만든다는 것이다. 클레이연구소는 이 문제를 공식적으로 미해결로 둔다(The Conversation, 2026년 9월 10일). 뉴욕대 트리스탄 버크마스터는 앤트로픽의 레벤트 알피게와 함께한 작업이 오픈AI 모델에 흘러들었는지 물었다. 오픈AI는 조사 결과 영향이 없었다고 밝혔다. 타오는 이 결과를 인공지능이 노천 채굴하듯 파낸 문제들의 최신 사례라고 불렀다("strip-mined")(IBM Think, 2026년 9월 10일). 2026년 10월 14일 현재 독립 검증은 끝나지 않았다.

실험이 필요한 분야에서는 한계가 다른 곳에서 나온다. 구글의 Co-Scientist와 퓨처하우스의 Robin은 2026년 5월 19일 네이처에 함께 실렸다(Nature Asia, 2026년 5월 20일). Co-Scientist가 낸 급성 골수성 백혈병 약물 후보는 세포 실험에서 가능성을 보였다. Robin은 건성 황반변성에 쓰인 적 없는 약 후보를 찾았다. 두 팀 모두 이 후보가 전임상과 임상시험을 거쳐야 하고 사람 과학자가 줄곧 함께해야 한다고 적었다. 가설은 빨라졌지만 세포와 사람의 몸은 예전 속도로 답한다.

백지의 영역은 오늘의 기계가 답하지 못하는 문제 목록에 고정돼 있지 않다. 새 답이 나오면 무엇을 더 관측하고 어떤 가설을 시험할지 다시 정해야 한다. 인간이 맡아야 할

일은 기계가 못 푼 문제를 기다리는 데 그치지 않는다. 새로 얻은 지식을 어디에 쓰고 어떤 미지의 영역을 탐구할지 선택하는 일이다.

앞으로 3년에서 5년: 난치병, 유전자 교정, 수명 연장, 수학 난제

다리오 아모데이는 2024년 10월 에세이에서 강력한 AI가 온 뒤 몇 해 안에 수십 년 치 생물학 진전이 압축될 수 있다고 썼다(아모데이, 2024년 10월). 그는 대부분 암의 사망과 발생이 95퍼센트 넘게 줄 수 있다고 봤다. 그 시간표의 기준점은 강력한 AI가 온 뒤다. 우리가 말하는 3년에서 5년은 2026년부터 센 기간이다. 네 분야의 속도는 같지 않다. 기계가 검증하는 수학이 앞서고 사람 몸에서 확인해야 하는 분야가 뒤따른다.

수학 난제에서 고워스는 지금의 장벽도 곧 무너질 것으로 본다. 리처드 보처즈는 2024년 12월 인공지능이 사람의 연구 능력을 넘는 때를 10년 안팎으로 잡았다. 같은 시기 타오는 3년에서 4년 안에 인공지능이 혼자 수학 연구를 할 가능성을 아주 낮게 봤다("very unlikely"). 우리는 Lean으로 확인할 수 있는 열린 문제 상당수가 5년 안에 풀린다고 추정한다. 앤트로픽은 8월 자사 모델이 리만 제타 함수의 영점 가운데 리만 가설을 만족하는 비율의 하한을 41.6퍼센트에서 67.2퍼센트로 올렸다고 밝혔다(앤트로픽, 2026년 8월 10일). 회사는 이 기법이 리만 가설의 증명으로 이어지리라 보지 않는다고 함께 적었다. 밀레니엄 문제의 시간표를 밝힌 전문가 발언은 찾지 못했다.

난치병 신약은 임상 단계를 거친다. 1상은 소수에게서 안전성을, 2상은 효능을, 3상은 많은 환자에게서 효과를 확인한다. 인실리코 메디신은 AI가 표적과 분자를 고른 폐섬유증 약 렌토서팁의 3상을 중국 47개 기관 환자 320명 규모로 시작했다(인실리코, 2026년 7월 7일). 보스턴컨설팅그룹 연구진은 AI가 찾은 약의 1상 성공률이 80에서 90퍼센트지만 2상은 약 40퍼센트로 업계 평균과 비슷하다고 분석했다(Drug Discovery Today, 2024년). 의약화학자 데릭 로는 표적 고르기와 사람 독성 예측이 아직 인공지능의 힘 밖이라고 말했다(Bio-IT World, 2025년 4월 15일). 아이소모픽랩스는 첫 임상 목표를 2026년 말로 1년 미뤘다. 10월 현재 첫 투약 발표는 확인되지 않는다(R&D World, 2026년 5월 12일). 우리는 5년 안에 AI 설계 약의 첫 허가가 나올 수 있지만 난치병 대부분의 치료는 그 바깥에 있다고 본다.

유전자 교정에서는 미국 FDA가 2023년 12월 첫 크리스퍼 치료제 캐스게비를 승인했고 2026년 7월 1일 대상을 2세 이상으로 넓혔다(버텍스, 2026년 7월 1일). 염기 교정은 DNA를 자르지 않고 염기 하나를 바꾸는 기술이다. 필라델피아 아동병원은 2025년 희귀 대사질환을 앓는 영아에게 맞춤형 염기 교정 치료를 6개월 안에 설계해

투여했다(CHOP, 2025년 5월 15일). MIT 테크놀로지 리뷰는 효과를 아직 단정할 수 없고 비용이 간 이식과 비슷한 약 100만 달러라고 적었다(2026년 1월 12일). 단백질 언어 모델이 만든 편집기 OpenCRISPR-1은 2026년 10월 현재 임상 기록이 없다. 몸 안에서 바로 편집하는 크리스퍼 치료제의 미국 승인도 아직 없다. 우리는 5년 안에 맞춤형 교정이 희귀질환 치료의 한 경로로 자리 잡는다고 본다.

수명 연장에서 부분 재프로그래밍은 몇몇 유전자 스위치를 잠깐 켜 세포의 유전자 발현을 젊은 쪽으로 되돌리는 방법이다. 라이프 바이오사이언스는 이 방법을 쓴 ER-100을 녹내장과 시신경병증 환자에게 처음 투약했고 시험의 초점은 안전성이다(라이프 바이오사이언스, 2026년 6월 9일). MIT 테크놀로지 리뷰는 이 시험을 젊음의 샘과 거리가 먼 개념 증명이라고 불렀고 재프로그래밍이 실험동물에서 암을 일으킨 적이 있다고 적었다(2026년 1월 27일). 알토스랩스와 리트로의 재프로그래밍 임상은 확인되지 않는다. 아모데이도 수명 두 배가 생물학적으로 가능하다는 보장은 없다고 썼다. 우리는 5년 안에 사람의 수명이 늘었다는 증거는 나오지 않는다고 본다.

오래 묵은 숙제가 풀린 다음에 남는 것

노가 알론은 에르되시 문제를 수십 개 푼 사람이다. 그는 2026년 8월 퀀타 매거진에 그 연구를 그만뒀다고 말했다. 인공지능이 풀기 시작했으니 더는 의미가 없다는 것이다("Once AI started to solve them, there is no point anymore")(Quanta, 2026년 8월 3일). 블룸은 같은 기사에서 100쪽에서 200쪽짜리 AI 논문이 늘어나는데 읽은 사람도 앞으로 읽을 사람도 없다고 했다. 숙제가 풀린 자리에는 아무도 읽지 않은 답과 연구를 접은 전문가가 남았다.

남는 일이 많다는 시각도 있다. 타오는 8월 사이먼스재단 대담에서 10만 줄짜리 기계 증명을 두고 아무도 그것을 이해하지 못한다고 말했다("no one understands them")(사이먼스재단, 2026년 8월 13일). 그래서 결과를 설명하고 교과서에 담는 일이 사람에게 남는다고 했다. 버자드는 다음에 필요한 것이 기계가 찾은 예에서 끌어낼 통찰이라고 썼다. 미국의 연구자 네이선 디바덜러벤은 2026년 10월 논문에서 문헌 정리와 가설과 분석은 넘쳐나도 실험과 관측과 책임지는 판단은 귀하게 남는다고 분석했다(arXiv, 2026년 10월 7일). 이 일들은 실제로 남는다. 하오 등 연구진은 2024년 12월 논문에서 인공지능을 쓰는 과학자의 영향력은 커지지만 과학 전체가 다루는 주제는 데이터가 많은 쪽으로 좁아진다고 보고했다(arXiv, 2024년 12월 10일). 설명과 검증은 이미 나온 답을 이해하게 하고, 실험과 관측은 아직 없는 답을 찾을 근거를 만든다. 그 과정에서 새로운 질문도 생긴다. 사람은 이 일들을 수행하면서 어떤 질문을 다음 연구의 목표로 삼을지

판단해야 한다. 기존의 숙제 목록을 처리하는 것만으로 연구의 방향이 정해지지는 않는다.

바둑은 기계가 사람을 넘어선 뒤의 10년을 먼저 겪었다. 2016년 이세돌 9단이 알파고에 1승 4패로 진 뒤에도 프로기사들은 대국을 멈추지 않았다. 신진서 9단은 2026년 4월 29일 「알파고 10년, 위대한 동행」 행사에서 요즘은 모두가 AI로 바둑 공부를 한다고 말했다(스포츠경향, 2026년 5월 13일). 그는 이어 자신이 왜 바둑을 하는지 스스로 깊이 고민해 보라고 권했다. 최선의 수를 찾는 일은 기계로 넘어갔고 바둑을 두는 이유는 사람이 다시 정해야 했다.

수학계도 같은 물음을 집단으로 꺼냈다. 2026년 9월 11일 필즈상 수상자 25명이 공개서한 「수학에서 AI의 심각한 어긋남(A Severe Misalignment of AI in Mathematics)」을 냈다(테크크런치, 2026년 9월 11일). 서한은 AI 기업의 목표와 수학계의 목표가 심하게 어긋났다고 적었다. 참과 거짓을 가리는 명제를 점점 빠르게 쏟아 내면 새 착상이 자랄 땅을 망칠 수 있다는 경고도 담았다. 서한은 인공지능을 막자고 하지 않았다. 서한 누리집의 필즈상 수상자 서명은 2026년 10월 14일 현재 28명이다.

한국의 목표는 아직 푸는 쪽에 걸려 있다. 과기정통부는 2025년 12월 12일 2026년 업무계획에서 AI 공동과학자로 2030년 노벨상급 성과를 내겠다고 밝혔다(정책브리핑, 2025년 12월 12일). 2026년부터 4년간 225억 원을 들이는 AI+과학기술 혁신 기술개발 사업도 시작했다(뉴스후플러스, 2026년 6월 4일). 파로스아이바이오가 AI 플랫폼 케미버스로 찾은 백혈병 후보 PHI-101은 미국 1상에서 객관적 반응률 67퍼센트를 보였다. 같은 기사는 KAIST의 단백질 구조 예측 모델 K-Fold가 기존 방식보다 최대 25배 빠르다고 전했다(이데일리, 2026년 9월 24일).

오래 묵은 숙제가 풀린 다음에 남는 일은 비어 가는 숙제 목록에 다음 줄을 적는 일이라고 우리는 판단한다. 바둑에서는 그 일이 기사 한 사람의 고민으로 남았고 수학에서는 필즈상 수상자들의 서한이 되었다. 오픈AI는 서한이 나온 열흘 뒤 프린스턴 고등연구소에 둔 독립 자문단의 출범을 알렸다. 고워스는 첫 위원 9명 가운데 한 사람이다. 회사는 이 자문단이 회사 내부의 수학 연구 속도에 대해서는 조언하지 않는다고 밝혔다(오픈AI, 2026년 9월 21일). 답을 생산하는 속도와 연구할 가치가 있는 문제를 정하는 일은 따로 살펴야 한다. 인공지능이 넓힌 지식 위에서 다음 목표를 만드는 능력을 연구와 교육의 중심에 두어야 한다.

참고 자료

고워스 블로그(Gowers's Weblog), What sort of maths are LLMs good at?, 2026년 8월 12일, https://gowers.wordpress.com/2026/08/12/what-sort-of-maths-are-llms-good-at/

Epoch AI, Will we run out of data? Limits of LLM scaling based on human-generated data, 2024년 6월 6일, https://epoch.ai/publications/will-we-run-out-of-data-limits-of-llm-scaling-based-on-human-generated-data

알파시그널(AlphaSignal), OpenAI's GPT-6 Astra Cracks Epoch's Hardest Math Benchmark in 14 Months, 2026년 9월 10일, https://alphasignal.ai/news/openai-s-gpt-6-astra-cracks-epoch-s-hardest-math-benchmark-in-14-months

Epoch AI(X 게시물), FrontierMath 4단계 포화 발표 (제목 미확인, URL 기준), 2026년 9월 10일, https://x.com/EpochAIResearch/status/2098103831502708864

arXiv(Mirzadeh 외), GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models, 2024년 10월 7일, https://arxiv.org/abs/2410.05229

NAACL 2024(Wu 외), Reasoning or Reciting? Exploring the Capabilities and Limitations of Language Models Through Counterfactual Tasks, 2024년 6월, https://aclanthology.org/2024.naacl-long.102

Epoch AI, What is the future of AI in mathematics? Interviews with leading mathematicians, 2024년 12월 4일, https://epoch.ai/frontiermath/tiers-1-4/expert-perspectives

arXiv(DeepSeek-AI), DeepSeek-Prover-V2: Advancing Formal Mathematical Reasoning via Reinforcement Learning for Subgoal Decomposition, 2025년 4월 30일, https://arxiv.org/abs/2504.21801

오픈AI, An OpenAI model has disproved a central conjecture in discrete geometry, 2026년 5월 20일, https://openai.com/index/model-disproves-discrete-geometry-conjecture/

버자드 블로그(Xena Project), Human mathematicians are being outcounterexampled, 2026년 7월 20일, https://xenaproject.wordpress.com/2026/07/20/human-mathematicians-are-being-outcounterexampled/

The Conversation(Melissa Lee), An AI solution to an 80-year-old problem has shocked mathematicians, 2026년 5월 26일, https://theconversation.com/an-ai-solution-to-an-80-year-old-problem-has-shocked-mathematicians-283686

Epoch AI, Announcing FrontierMath Erdős, 2026년 9월 1일, https://epoch.ai/latest/announcing-frontiermath-erdos

BenchLM, Abstraction and Reasoning Corpus for AGI v3 (ARC-AGI-3), 2026년 10월 13일 갱신, https://benchlm.ai/benchmarks/arcagi3

오픈AI, Sharing AI progress in mathematics, 2026년 10월 6일, https://openai.com/index/sharing-ai-progress-in-mathematics/

길 칼라이 블로그(Combinatorics and more), Updates: Sharing AI progress on mathematics (amazing!); and my lecture plans, 2026년 10월 7일, https://gilkalai.wordpress.com/2026/10/07/updates-sharing-ai-progress-on-mathematics-amazing-and-my-lecture-plans/

MLQ News, OpenAI Withdraws Three of 722 Math Manuscripts After October 6 Release, 2026년 10월 9일, https://mlq.ai/news/openai-withdraws-three-of-722-math-manuscripts-after-october-6-release/

오픈AI, On the Navier–Stokes Millennium Prize Problem, 2026년 9월 8일, https://openai.com/index/navier-stokes-solution/

The Conversation(Andrew Cullen), OpenAI claims another huge mathematical result amid fights over credit, ethics and privacy, 2026년 9월 10일, https://theconversation.com/openai-claims-another-huge-mathematical-result-amid-fights-over-credit-ethics-and-privacy-291575

IBM Think, Will AI solve math too fast?, 2026년 9월 10일, https://www.ibm.com/think/news/will-ai-solve-math-too-fast-navier-stokes-terence-tao

Nature Asia, Artificial intelligence: AI research assistants that may accelerate scientific discovery (Nature), 2026년 5월 20일, https://www.natureasia.com/en/info/press-releases/detail/9330

다리오 아모데이, Machines of Loving Grace: How AI Could Transform the World for the Better, 2024년 10월, https://darioamodei.com/essay/machines-of-loving-grace

앤트로픽, Claude has improved on a longstanding lower bound for the fraction of zeros of the Riemann zeta function that satisfy the Riemann hypothesis, 2026년 8월 10일, https://www.anthropic.com/research/riemann-zeta

인실리코 메디신, Insilico Initiates Phase III Clinical Trial for Rentosertib, Its AI-Empowered TNIK Inhibitor for Idiopathic Pulmonary Fibrosis, 2026년 7월 7일, https://insilico.com/news/xmjsn4l091-insilico-initiates-phase-iii-clinical-tr

Drug Discovery Today(Jayatunga 외), How successful are AI-discovered drugs in clinical trials? A first analysis and emerging lessons, 2024년, https://www.sciencedirect.com/science/article/pii/S135964462400134X

Bio-IT World, Derek Lowe on AI in Drug Discovery: Between Hype and Hope, 2025년 4월 15일, https://www.bio-itworld.com/news/2025/04/15/derek-lowe-on-ai-in-drug-discovery-between-hype-and-hope

R&D World, Alphabet-spinoff Isomorphic Labs raises $2.1 billion in quest to 'solve all disease' with AI-based drug discovery tools, 2026년 5월 12일, https://www.rdworldonline.com/alphabet-spinoff-isomorphic-labs-raises-2-1-billion-in-quest-to-solve-all-disease-with-ai-based-drug-discovery-tools/

버텍스, Vertex Announces US FDA Approval for Expanded Use of CASGEVY® for the Treatment of People Ages 2 Years and Older With Sickle Cell Disease or Transfusion-Dependent Beta Thalassemia, 2026년 7월 1일, https://news.vrtx.com/news-releases/news-release-details/vertex-announces-us-fda-approval-expanded-use-casgevyr-treatment

필라델피아 아동병원(CHOP), World's First Patient Treated with Personalized CRISPR Gene Editing Therapy at Children's Hospital of Philadelphia, 2025년 5월 15일, https://www.chop.edu/news/worlds-first-patient-treated-personalized-crispr-gene-editing-therapy-childrens-hospital

MIT 테크놀로지 리뷰, Base-edited baby: 10 Breakthrough Technologies 2026, 2026년 1월 12일, https://www.technologyreview.com/2026/01/12/1129999/gene-editing-base-edited-baby-personalized-drugs-2026-breakthrough-technology/

프로플루언트, Editing the Human Genome with AI, 2024년 4월 22일, https://www.profluent.bio/media/editing-the-human-genome-with-ai

라이프 바이오사이언스, Life Biosciences Announces First Patient Dosed in Phase 1 Trial of ER-100 for Optic Neuropathies, 2026년 6월 9일, https://www.lifebiosciences.com/life-biosciences-announces-first-patient-dosed-in-phase-1-trial-of-er-100-for-optic-neuropathies/

MIT 테크놀로지 리뷰, The first human test of a rejuvenation method will begin "shortly", 2026년 1월 27일, https://www.technologyreview.com/2026/01/27/1131796/the-first-human-test-of-a-rejuvenation-method-will-begin-shortly/

Quanta Magazine, Why the Legendary Erdős Problems Are Falling to AI, 2026년 8월 3일, https://www.quantamagazine.org/why-the-legendary-erdos-problems-are-falling-to-ai-20260803/

사이먼스재단, Watch: Fields Medalist Terence Tao on Artificial Intelligence and Why We Do Math, 2026년 8월 13일, https://www.simonsfoundation.org/2026/08/13/fields-medalist-terence-tao-on-artificial-intelligence-and-why-we-do-math/

arXiv(Hao 외), Artificial Intelligence Tools Expand Scientists' Impact but Contract Science's Focus, 2024년 12월 10일, https://arxiv.org/abs/2412.07727

arXiv(Nathan DeBardeleben), When Scientific Cognition Is No Longer Scarce, 2026년 10월 7일, https://arxiv.org/abs/2610.10241

스포츠경향, 바둑을 바꿔 놓은 AI, 정말 모두 만능일까, 2026년 5월 13일, https://sports.khan.co.kr/article/202605131644013

테크크런치, OpenAI's feud with mathematicians is only escalating, 2026년 9월 11일, https://techcrunch.com/2026/09/11/openais-feud-with-mathematicians-is-only-escalating/

Math and AI 공개서한 누리집, A Severe Misalignment of AI in Mathematics, 2026년 9월 11일, https://mathandai.org

대한민국 정책브리핑, 2026년 과학기술정보통신부 업무계획, 2025년 12월 12일, https://www.korea.kr/briefing/policyBriefingView.do?newsId=156734887

뉴스후플러스, 과기정통부, AI로 과학기술 연구혁신 추진…4년간 225억원 투입, 2026년 6월 4일, https://www.newswhoplus.com/news/articleView.html?idxno=61607

이데일리(네이트 게재), 2조9000억 vs 160억 AI신약, 알고리즘이 아닌 '체계'에서 갈렸다, 2026년 9월 24일, https://m.news.nate.com/view/20260924n09412

오픈AI, Advisory Group on Mathematics and Artificial Intelligence, 2026년 9월 21일, https://openai.com/index/advisory-group-on-mathematics-and-ai/



9장. 담대한 목표

어디로 갈지 정하는 사람이 있어야 기계도 일한다

무엇보다 앞서는 일은 문명의 미래를 지키는 것이고 달이 더 빠르다("the overriding priority is securing the future of civilization and the Moon is faster"). 일론 머스크가 2026년 2월 8일 X에 스페이스X의 우선순위를 화성에서 달로 옮긴다며 든 이유다(NBC News, 2026년 2월 9일). 그는 화성 도시도 5년에서 7년 안에 시작하겠다고 덧붙였다. 같은 보도는 스페이스X가 투자자들에게 2027년 3월 무인 달 착륙을 목표로 내놓았다는 월스트리트저널 기사를 전했다. 로켓과 공장과 엔지니어는 전날과 같았고 바뀐 것은 행선지 하나였다.

8장 끝에서 본 대로 숙제가 풀린 자리에는 다음 줄을 적는 일이 남는다. 그 일을 기계에 넘기기 어려운 까닭은 기계가 배우는 방식에 있다. 알파고의 정책망은 이기면 1을 받고 지면 마이너스 1을 받는 보상으로 배웠다(Nature, 2016년 1월 28일). 승패를 정한 것은 사람이 만든 바둑 규칙이었다. 알파폴드2가 2020년 단백질 구조 예측 대회 CASP14에서 받은 중앙값 92.4점도 대회 운영진이 미리 정한 채점 방식으로 잰 점수다(Proteins, 2021년). 기계는 사람이 그어 준 틀 안에서 누구보다 빨리 답을 찾는다.

틀을 정하는 일이 사람에게 남는 이유는 그 일이 쉬워서가 아니다. 스튜어트 러셀 UC버클리 교수는 2022년 4월 Daedalus 에세이에서 사람이 기계의 목표를 완전하고 정확하게 적어 줄 수 없다고 썼다. 그는 사람의 선호를 확신하지 못하고 사람에게서 배우는 기계를 대안으로 내놓았다. 목표를 다 적어 줄 수 없으니 그 목표를 고쳐 가는 책임도 사람에게 남는다. 인공지능이 목표의 후보와 실행 계획을 제안하더라도 어느 목표를 받아들일지, 그 과정의 비용과 위험을 누가 부담할지는 사람이 논의하고 결정해야 한다.

기계가 목표를 스스로 세우게 두자는 쪽은 연구계에서 드물다. 허깅페이스의 마거릿 미첼 등 연구자 네 명은 2025년 2월 4일 arXiv에 완전 자율 에이전트를 개발하지 말아야 한다는 논문을 올렸다. 요슈아 벤지오는 2025년 6월 3일 블로그에서 스스로 행동하지 않는 과학자 AI를 만들 비영리 연구소 LawZero를 세운다고 알렸다. 무스타파 술레이만 마이크로소프트 AI 최고경영자는 스스로 목표를 정하고 돈을 버는 AI가 사람과 자원을

다투는 실리콘 종이 될 수 있다고 BBC에 말했다(유로뉴스, 2026년 9월 17일). 에이전트가 일의 순서를 짜고 도구를 고르는 일은 이미 기계가 맡기 시작했다. 그 순서가 무엇을 위한 것인지는 여전히 사람이 적는다. 이 경고들이 묻는 통제의 조건은 3부에서 따로 다룬다.

필자는 2025년 7월 전자신문 연재에서 많은 기업이 AI 도입 목표를 인력 감축과 비용 절감에만 두고 있다고 썼다. 같은 글에서 이렇게 적었다. "진정한 혁신은 기존 목표를 효율적으로 달성하는 것이 아니라, 아예 새로운 목표를 설정하는 것이다."(전자신문, 2025년 7월 15일) 국회의원이던 2017년 9월에는 같은 신문 인터뷰에서 "정부 정책 1순위로 4차 산업혁명을 위한 사회 혁신을 설정해야 한다"고 말했다(전자신문, 2017년 9월 21일). 그 자리에서 과학기술부총리 부활도 요구했다. 국가가 목표의 순서를 정하고 그 일을 맡을 자리를 세우라는 요구였다. 2025년 5월 같은 연재에서는 "지도자의 실천 의지가 대한민국의 AI 역량을 좌우할 것이다"라고 썼다(전자신문, 2025년 5월 27일).

화성에 가자는 목표와 그 과정에서 발전한 공학

2015년 12월 21일 밤 플로리다 케이프커내버럴에서 팰컨 9 로켓이 오브콤의 OG2 통신위성을 싣고 올라갔다. 발사 9분 뒤인 동부시간 오후 8시 38분에 로켓의 1단이 옛 13번 발사대 자리에 만든 착륙장 1에 내려앉았다(미 공군, 2015년 12월 22일). 재사용 로켓은 한 번 쓴 로켓의 몸체를 회수해 다시 쏘는 로켓이다. 2017년 3월 30일 스페이스X는 한 번 날았던 1단으로 SES-10 위성을 쏘아 올리고 그 1단을 다시 바다 위 바지선에 내렸다(NASASpaceflight, 2017년 3월 30일). 미국항공우주학회(AIAA)는 2026년 8월 25일 1단 B1067이 37번째 비행을 마쳤다고 전했다.

발사 비용도 줄었다. NASA 에임스연구소의 해리 존스는 2018년 논문에서 우주왕복선이 저궤도에 1킬로그램을 올리는 비용을 약 5만 4500달러로 계산했다(NASA, 2018년 7월). 팰컨 9는 광고 가격 기준 약 2720달러였다. 실제 원가와 광고 가격의 비교라는 단서가 붙는다.

이 공학의 행선지는 화성이었다. 알자지라 보도에 따르면 머스크는 2025년 5월 2026년 말까지 무인 스타십을 화성에 보낼 가능성을 반반으로 봤다. 그 일정은 궤도 급유에 달려 있었다. 궤도 급유는 궤도에 떠 있는 우주선에 다른 우주선이 추진제를 옮겨 담는 일이다. 스타십은 2024년 3월 한 기체 안의 두 탱크 사이에서 추진제를 옮겨 봤다. 위키백과 기준으로 두 기체 사이의 급유 시험은 2026년 9월까지 날짜가 잡히지 않았다.

스타십은 2026년 9월 28일까지 통합 비행을 14번 했다. 위키백과 집계로는 9번 성공하고 5번 실패했다. 스페이스X는 1차 비행을 성공이라 불렀으니 이 숫자는 집계하는 쪽에 따라 달라진다. 14차 비행에서 스타십은 고도 약 275킬로미터 궤도에 스타링크 위성 26기를 올렸다. 2026년 11월에 열리는 화성 발사 창에 스타십을 보내는 일은 아직 계획이고 다음 창은 2029년에 열린다(19FortyFive, 2026년 8월 9일).

국가의 돈도 같은 기체로 들어왔다. NASA는 2021년 4월 16일 스타십을 바탕으로 한 유인 달 착륙선의 개발사로 스페이스X를 고르고 28억 9천만 달러를 배정했다(NASA, 2021년 4월 16일). 화성이라는 회사의 목표와 달이라는 국가의 목표가 한 기체에서 만난 셈이다. 그래도 개발비의 큰 줄기는 위성 인터넷에서 나온다. 스페이스X가 2026년 5월 20일 낸 상장 신고서를 보면 2025년 매출 187억 달러 가운데 스타링크 중심의 연결 부문이 114억 달러였다(Via Satellite, 2026년 5월 20일). 같은 해 스타십 연구개발에는 30억 달러 가까이 들었다. 신고서는 화성을 돈의 쓰임새로 적지 않았다. 위성 인터넷이 번 돈이 화성용 로켓으로 흘러가는 구조는 부문별 숫자와 머스크의 발언을 이어 읽어야 보인다.

화성 목표를 낭비로 보는 연구도 있다. 드레스덴 공대 등 독일 연구진은 공개 자료로 유인 스타십의 질량을 다시 계산했다(Scientific Reports, 2024년 5월 23일). 그 계산에서 질량은 스페이스X 발표의 두 배쯤 나왔고 화성에서 지구로 돌아오는 비행은 성립하지 않았다. 존스홉킨스대 우주론자 폴 서터는 유인 화성 임무에 수조 달러가 들 수 있고 급유 발사가 10번에서 20번 더 필요하다고 썼다(Scientific American, 2025년 3월 20일). 화성협회의 로버트 주브린은 반대 방향에서 달 우선 선회를 화성 계획을 망칠 피로스의 승리라고 비판했다(Forbes, 2026년 4월 28일). 이 비판이 맞다면 사람이 화성에 닿는 날은 멀다. 그래도 화성이라는 목표가 남긴 재사용 1단과 위성 인터넷 사업은 행선지가 바뀐 뒤에도 그대로 남았다.

목표가 기술을 끌고 가는 다른 사례

1961년 5월 25일 존 F. 케네디 대통령은 의회 연설에서 이 10년이 가기 전에 사람을 달에 보내고 무사히 데려오자고 했다("landing a man on the moon and returning him safely to the earth")(NASA, 2021년 5월 25일). CBS 뉴스는 NASA 예산이 1966년 연방 예산의 약 4.4퍼센트로 정점에 올랐다고 전했다(2019년 7월 16일). 아폴로 유도 컴퓨터에는 집적회로가 들어갔다. 집적회로는 트랜지스터 여러 개를 한 칩에 묶은 부품이다. 개당 약 1000달러이던 1960년의 집적회로는 1963년 약 25달러로 내려갔다

(IEEE ETHW). 아폴로가 사들인 물량의 비중은 자료에 따라 세계 생산의 30퍼센트에서 미국 생산의 60퍼센트까지 갈린다(FedTech, 2018년 9월 11일).

경제학의 평가는 신중하다. 숀 칸토어와 알렉산더 월리는 2025년 9월 American Economic Review 논문에서 NASA 지출이 우주 관련 제조업의 부가가치와 고용을 늘렸다고 분석했다. 그러나 그 상징적인 달 탐사 사업의 경제 효과는 소폭이었다고 결론지었다("only modest economic effects"). 인간 게놈 프로젝트는 1990년 시작해 2003년 4월 14일 완료를 선언했고 1991 회계연도 달러로 27억 달러를 썼다(NHGRI, 2003년 4월 14일). 배텔연구소는 이 사업의 경제 효과를 8천억 달러 가까이로 추정했다. 경제학자들은 연구자 급여 같은 비용을 편익으로 셈한 방식을 문제 삼았다(Scientific American, 2011년 5월 11일).

정부가 목표만 걸고 물러선 사례도 있다. 미국 국방고등연구계획국(DARPA)은 2004년 사람 없이 사막을 달리는 차량 대회에 상금 100만 달러를 걸었다. 그해에는 완주한 차가 없었고 카네기멜런대의 샌드스톰이 12킬로미터를 달린 것이 최고였다(IEEE Spectrum, 2025년 2월 1일). 2005년 10월 8일 2회 대회에서는 23대 가운데 5대가 완주했고 스탠퍼드대의 스탠리가 6시간 53분으로 우승했다. 팀을 이끈 세바스찬 스런은 이 우승을 현대 자율주행차가 태어난 순간이라고 불렀다("birth moment of the modern self-driving car")(Smithsonian Magazine, 2024년 8월).

중국은 연도를 못 박은 목표를 체계적으로 쓴다. CGTN은 2026년 2월 28일 중국유인우주공정판공실이 2030년 유인 달 착륙 목표를 유지한다고 전했다. 달 남극 탐사선 창어 7호는 2026년 8월 23일 발사 조건을 채우지 못해 올해 창을 넘겼다(위키백과, 2026년 10월 조회). 한국의 누리호는 2025년 11월 27일 4차 발사에서 차세대중형위성 3호를 600킬로미터 궤도에 올렸고(정책브리핑, 2025년 11월 27일), 2026년 10월 7일 5차 발사에도 성공했다(MTN, 2026년 10월 7일). 우주항공청은 2025년 12월 차세대발사체를 메탄 연료 재사용 발사체로 개발하기로 확정했다(우주항공청, 2025년 12월 22일). 미션 지향 혁신 정책은 풀 문제와 기한을 먼저 정하고 연구와 투자를 거기에 묶는 정책이다. 과기정통부는 2026년 5월 K-문샷 12개 임무의 총괄관리자 12명을 위촉하고 2035년까지 임무를 풀겠다고 밝혔다(아주경제, 2026년 5월 27일). 12개 임무 전체의 2027년 예산은 1조 3465억 원이다(재경일보, 2026년 9월 30일).

이런 정책을 두고 경제학계는 갈려 있다. 마리아나 마추카토 UCL 교수는 『미션 이코노미』(2021년)에서 아폴로를 본보기로 국가가 기업과 함께 시장을 만들자고 했다. 스웨덴 산업경제연구소(IFN)의 마그누스 헨렉손 등은 미션형 정책이 지대 추구와 포획에

노출되고 기회비용을 무시한다고 반박했다(Springer, 2024년). 듀크대 마이클 멍거는 국가 주도 사업이 공학의 위업은 이뤄도 기업가적 혁신은 잘 낳지 못한다고 정리했다(Econ Journal Watch, 2024년 9월). 이 논쟁은 미션의 경제적 수익률을 두고 벌어진다. 시장이 기술을 끈다는 쪽도 목표를 정하는 주체로 기업가라는 사람을 내세운다. 어느 쪽도 목표를 정하는 일 자체가 필요 없다고 말하지 않는다.

다음 시대의 목표를 상상하는 일

오픈AI의 샘 올트먼은 2024년 9월 23일 에세이 「지능의 시대(The Intelligence Age)」에서 기후를 고치고 우주 식민지를 세우는 일이 언젠가 흔해진다고 썼다. 그는 그 길을 닦는 재료로 연산과 에너지와 사람의 의지를 들었다("compute, energy, and human will"). 데미스 허사비스 구글 딥마인드 최고경영자는 2025년 4월 CBS 「60분」에서 AI로 모든 질병을 끝내는 일이 아마 10년쯤 안에 가능하다고 말했다(Business Today, 2025년 4월 22일). 영국 기술 저널리스트 로리 셀런존스는 신약 후보가 빨리 나와도 임상시험은 빨라지지 않았다고 지적했다(셀런존스 블로그, 2025년 4월 18일). 다리오 아모데이가 2024년 10월 에세이에서 쓴 압축된 21세기도 같은 목록에 든다. 기업 지도자들의 목록에 오른 질병과 에너지와 우주는 20세기 중반의 목표 목록과 크게 다르지 않다. 달라진 것은 그 목록을 먼저 발표하는 사람이 정부보다 기업 쪽에 많다는 점이다.

국가도 목록을 낸다. 중국 국무원은 2017년 7월 「차세대 인공지능 발전 계획」에서 2030년 세계의 으뜸 AI 혁신 중심이 되겠다는 목표를 세웠다(DigiChina, 2017년 8월 1일). 중간 목표를 얼마나 이뤘는지 따진 중국 정부의 공식 평가는 찾지 못했다. 중국 전국인민대표대회는 2026년 3월 제15차 5개년 계획(2026년부터 2030년)을 승인했다. 계획은 양자, 6G, 수소와 핵융합, 바이오 제조, 뇌와 컴퓨터를 잇는 장치, 체화 지능을 앞세웠다(BOFIT, 2026년 3월 20일). 유럽연합은 기후 적응, 암, 기후중립 도시, 토양, 바다와 물 복원이라는 다섯 미션을 두고 2026년 9월 14일 중간 평가를 공개했다(유럽집행위원회). 집행위원회는 이 평가를 소개하며 미션이 눈에 보이는 긍정적 결과를 내고 있다고 밝혔다. 유엔의 2025년 지속가능발전목표 보고서에서는 세부 목표의 35퍼센트만 제 궤도에 있거나 보통의 진전을 보였고 18퍼센트는 뒷걸음쳤다(유엔, 2025년 7월).

목표가 늦어진 기록도 함께 봐야 한다. 핵융합은 가벼운 원자핵을 합칠 때 나오는 에너지로 전기를 만들려는 기술이다. 국제핵융합실험로(ITER)는 2024년 7월 새 일정을 내놓았다. 첫 운전은 2034년으로 밀렸고 중수소와 삼중수소를 쓰는 본운전은 2039년으로 밀렸다(Physics World, 2024년 7월 3일). 바라바스키 사무총장은 2025년 첫

플라스마라는 옛 목표가 상징에 가까웠다고 말했다. 올트먼이 투자한 헬리온은 마이크로소프트와 2028년 전력 공급 계약을 맺었다. 헬리온은 2026년 누리집에서 2028년 전기 생산 문구를 지우고 2028년 초기 운전 시작으로 바꿨다(Axios, 2026년 9월 28일). 회사 홍보 책임자는 50메가와트 완전 가동이 2029년이나 2030년 이후라고 인정했다.

목표 목록을 나란히 놓으면 차이가 보인다. 누가 정했는지와 목표 연도가 명시되었는지가 서로 다르다. 달성 여부를 객관적으로 측정할 수 있는지와 지연되었을 때 중간 기록이 남는지도 다르다. 유엔과 유럽연합의 목표는 수치로 중간 성적을 공개한다. 모든 질병을 끝낸다는 선언만으로는 진전을 판단할 구체적인 기준을 알 수 없다. 다음 시대의 목표를 상상하는 일은 행선지와 함께 달성 연도와 측정 지표, 그리고 계획이 늦어졌을 때의 기록 방식까지 함께 설계하는 일이라고 우리는 본다.

이 일을 게을리하면 인간은 스스로 가치를 잃는다

일에는 두 가지 가치가 있다. 하나는 일을 해내는 수단으로서의 가치이고 다른 하나는 무엇을 할지 정하는 목적 설정의 가치다. 7장이 보여 준 대로 수단으로서의 가치는 기계 쪽으로 넘어가고 있다. 홍콩이공대의 루카스 스크립터는 일에서 얻던 의미를 다른 활동에서 찾으면 된다는 흔한 답을 반박했다(AI & Society, 2026년 4월 15일). 의미는 삶에 두텁게 박혀 있어서 그 상실을 다른 곳에서 메우기 어렵다는 것이다. 재스프리트 란지트 등 연구진은 미국 직업의 컴퓨터 보조 과업 1만 131개를 살폈다(arXiv, 2026년 3월 16일). 노동자가 주체성이나 행복과 잇는 과업이 AI에 더 많이 노출될 수 있다는 결과였다. 지루한 일은 기계가 맡고 사람은 보람 있는 일만 하리라는 기대와 어긋난다. 같은 연구에서 개발자와 노동자가 AI에 바라는 성질도 서로 달랐다. 개발자는 예의와 상상력을 앞세웠고 노동자는 솔직하고 실용적인 성질을 앞세웠다. 기계의 성격을 정하는 쪽과 그 기계와 함께 일하는 쪽이 이미 갈라져 있다.

목적 설정의 가치도 저절로 지켜지지 않는다. 대런 아세모글루와 사이먼 존슨은 『 권력과 진보(Power and Progress) 』 (2023년)에서 누가 기술 진보의 방향을 정하는가를 물었다. 그들은 기술의 방향이 힘 있는 사람들의 이해와 믿음을 따라 정해져 왔고 사회의 선택과 정책으로 바꿀 수 있다고 봤다. 올트먼도 2026년 3월 AI가 노동과 자본의 균형을 흔드는데 어떻게 할지 아는 사람이 없다고 말했다(Fortune, 2026년 3월 12일). 그는 9월 23일 유엔 연설에서 AI가 창조와 발견의 새 르네상스가 될지 격변의 새 산업혁명이 될지는 인류의 선택에 달렸다고 했다(Fortune, 2026년 10월 9일). 머스크 한 사람의 글로 거대 기업의 행선지가 바뀌는 일은 목표를 정하는 자리가 소수에게 몰려

있음을 보여 준다. 필자는 2026년 4월 연재에서 경쟁력의 바탕을 "무엇을 만들어야 하는지 아는 것, 누구의 문제를 풀어야 하는지 아는 것"으로 꼽았다(전자신문, 2026년 4월 7일). 그 앎이 소수의 몫으로 남으면 나머지 사람은 남이 정한 목표를 수행하는 자리로 밀려난다.

거대한 목표가 권위주의로 흐른다는 반론에는 근거가 있다. 맨해튼 계획은 국가가 정한 목표 하나로 산업 규모의 과학을 만들었고 그 목표는 핵무기였다. 앞의 IFN 연구진은 정치인과 관료도 자기 이익에서 자유롭지 않다고 지적했다. 멍거가 서평에서 정리한 대로 이들은 손쉬운 서사가 가진 유혹의 힘을 경계했다. 국가가 연도를 못 박은 목표를 체계적으로 쓰는 중국에서는 그 목표를 바깥에서 고칠 통로가 좁다는 걱정도 따른다. 다른 길을 보여 준 사례도 있다. DARPA는 목표와 상금만 걸고 누구나 도전하게 했고 유럽연합은 미션의 중간 평가를 공개했다. 위험은 목표의 크기보다 그 목표를 누가 어떤 절차로 정하고 고치는지에서 나온다.

맥킨지가 2026년 낸 조직 보고서에서 지도자의 86퍼센트는 자기 조직이 일상 업무에 AI를 도입할 준비가 되지 않았다고 답했다(McKinsey, 2026년). 같은 보고서에서 조직 여섯 곳 가운데 한 곳에는 AI 도입을 책임지는 최고경영진이 없었다. 목표를 정하는 일을 게을리하면 생산 과업을 기계에 맡긴 뒤 그 기계로 무엇을 할지 결정하는 권한까지 남에게 넘길 수 있다. 그 자리는 비어 있지 않고 누군가 먼저 채운다. 조직은 목표를 정할 책임자를 두고, 사회는 그 목표를 시민이 검토하고 바꿀 통로를 마련해야 한다. 기술로 일을 해내는 능력이 커질수록 함께 추구할 목적을 정하는 권한도 넓혀야 한다. 여기서 지켜야 할 가치는 사람의 생산성에 매긴 가격이 아니라 자기 삶과 사회의 방향을 결정하는 주체성이다.

참고 자료

NBC News, In a reversal, SpaceX prioritizes 'city' on the moon over Mars project, Musk says, 2026년 2월 9일, https://www.nbcnews.com/science/space/spacex-prioritizes-moon-before-mars-musk-says-rcna258168

Nature, Mastering the game of Go with deep neural networks and tree search, 2016년 1월 28일, https://www.nature.com/articles/nature16961

Proteins, Applying and improving AlphaFold at CASP14, 2021년, https://pmc.ncbi.nlm.nih.gov/articles/PMC9299164/

Daedalus, If We Succeed, 2022년 4월 13일, https://www.amacad.org/publication/daedalus/if-we-succeed

arXiv, Fully Autonomous AI Agents Should Not be Developed, 2025년 2월 4일, https://arxiv.org/abs/2502.02649

요슈아 벤지오 블로그, Introducing LawZero, 2025년 6월 3일, https://yoshuabengio.org/en/blog/introducing-lawzero

Euronews, L'IA pourrait créer une "espèce de silicium" rivale des humains, alerte le patron de Microsoft, 2026년 9월 17일, https://fr.euronews.com/next/2026/09/17/lia-pourrait-creer-une-espece-de-silicium-rivale-des-humains-alerte-le-patron-de-microsoft

전자신문, [김경진의 AI전략노트] 〈8〉 인공지능 시대, 지도자의 생각과 패러다임의 전환이 필요하다, 2025년 7월 15일, https://www.etnews.com/20250715000109

전자신문, [창간 35주년 특집 Ⅲ]〈인터뷰〉김경진 의원, "과도한 규제 풀고, 교육개혁 선행해야", 2017년 9월 21일, https://m.etnews.com/20170914000202

전자신문, [김경진의 AI전략노트] 〈5〉 대한민국 AI 강국 도약, 지도자 의지가 성패 가른다, 2025년 5월 27일, https://www.etnews.com/20250527000125

미 공군, 45th SW launches ORBCOMM, lands first stage booster, 2015년 12월 22일, https://www.af.mil/News/Article-Display/Article/637696/45th-sw-launches-orbcomm-lands-first-stage-booster

NASASpaceflight, SpaceX conducts historic Falcon 9 re-flight with SES-10 – Lands booster again, 2017년 3월 30일, https://www.nasaspaceflight.com/2017/03/spacex-historic-falcon-9-re-flight-ses-10/

AIAA, SpaceX Falcon 9 Booster Sets Reuse Record With 37th Flight, 2026년 8월 25일, https://aiaa.org/2026/08/25/spacex-falcon-9-booster-sets-reuse-record-with-37th-flight/

NASA NTRS, The Recent Large Reduction in Space Launch Cost, 2018년 7월, https://ntrs.nasa.gov/citations/20200001093

Al Jazeera, Musk says 50-50 chance of sending uncrewed Starship to Mars by late 2026, 2025년 5월 30일, https://www.aljazeera.com/news/2025/5/30/musk-says-50-50-chance-of-sending-uncrewed-starship-to-mars-by-late-2026

위키백과, Starship Propellant Transfer Demonstration, 2026년 10월 조회, https://en.wikipedia.org/wiki/Starship_Propellant_Transfer_Demonstration

위키백과, List of Starship launches, 2026년 10월 조회, https://en.wikipedia.org/wiki/List_of_Starship_launches

19FortyFive, The Mars Launch Window Opens in November 2026 and Closes About a Month Later. The Next One Is in 2029., 2026년 8월 9일, https://www.19fortyfive.com/2026/08/the-mars-launch-window-opens-in-november-2026-and-closes-about-a-month-later-the-next-one-is-in-2029/

NASA, As Artemis Moves Forward, NASA Picks SpaceX to Land Next Americans on Moon, 2021년 4월 16일, https://www.nasa.gov/news-release/as-artemis-moves-forward-nasa-picks-spacex-to-land-next-americans-on-moon/

Via Satellite, SpaceX's IPO Filing Gives First Look Into Company's Financials, 2026년 5월 20일, https://www.satellitetoday.com/finance/2026/05/20/spacexs-ipo-filing-gives-first-look-into-companys-financials/

Scientific Reports, About feasibility of SpaceX's human exploration Mars mission scenario with Starship, 2024년 5월 23일, https://pmc.ncbi.nlm.nih.gov/articles/PMC11116405/

Scientific American, We're Not Going to Mars Anytime Soon—Despite What Musk Says, 2025년 3월 20일, https://www.scientificamerican.com/article/trump-wants-to-go-to-mars-thats-not-happening/

Forbes, (제목 미확인, URL 기준), 2026년 4월 28일, https://www.forbes.com/sites/kevinholdenplatt/2026/04/28/elon-musks-jilting-mars-to-build-moon-city-could-spark-his-downfall/

NASA, 60 Years Ago: President Kennedy Proposes Moon Landing Goal in Speech to Congress, 2021년 5월 25일, https://www.nasa.gov/history/60-years-ago-president-kennedy-proposes-moon-landing-goal-in-speech-to-congress

CBS News, Apollo 11: How much did it cost to land astronauts on the moon?, 2019년 7월 16일, https://www.cbsnews.com/news/apollo-11-moon-landing-how-much-did-it-cost/

IEEE ETHW, Integrated Circuits and the Space Program and Missile Defense, 날짜 미표기, https://ethw.org/Integrated_Circuits_and_the_Space_Program_and_Missile_Defense

FedTech, How the Government Helped Spur the Microchip Industry, 2018년 9월 11일, https://fedtechmagazine.com/article/2018/09/how-government-helped-spur-microchip-industry

American Economic Review(IDEAS/RePEc), Moonshot: Public R&D and Growth, 2025년 9월, https://ideas.repec.org/a/aea/aecrev/v115y2025i9p2891-2925.html

NHGRI, International Consortium Completes Human Genome Project, 2003년 4월 14일, https://www.genome.gov/11006929/2003-release-international-consortium-completes-hgp

Scientific American, What is the Human Genome Worth?, 2011년 5월 11일, https://www.scientificamerican.com/article/what-is-the-human-genome-worth/

IEEE Spectrum, The Starting Line for Self-Driving Cars, 2025년 2월 1일, https://spectrum.ieee.org/darpa-grand-challenge

Smithsonian Magazine, How a Blue SUV Named Stanley Revolutionized Driverless Car Technology, 2024년 8월, https://www.smithsonianmag.com/smithsonian-institution/how-a-blue-suv-named-stanley-revolutionized-driverless-car-technology-180984882/

CGTN, China gears up for crewed moon landing by 2030, 2026년 2월 28일, https://news.cgtn.com/news/2026-02-28/China-gears-up-for-crewed-moon-landing-by-2030-1L7SwMwUaVa/p.html

위키백과, Chang'e 7, 2026년 10월 조회, https://en.wikipedia.org/wiki/Chang'e_7

정책브리핑, 한국형발사체 '누리호' 4차 발사 성공…"2027년까지 2차례 더 발사", 2025년 11월 27일, https://www.korea.kr/news/policyNewsView.do?newsId=148955460

MTN, [누리호 5차 발사] 다섯번째 누리호 발사 성공…15기 중 14기 궤도 진입, 2026년 10월 7일, https://news.mtn.co.kr/news-detail/2026100716330458112

우주항공청, 차세대발사체, 재사용발사체로 개발 확정, 2025년 12월 22일, https://www.kasa.go.kr/bbs/BBSMSTR_000000000010/view.do?nttId=B000000002608Fs8lN0

아주경제, K-문샷 프로젝트 본격화…과기정통부, PD 12명 위촉, 2026년 5월 27일, https://www.ajunews.com/view/20260527151327382

재경일보(JKN), K-문샷, 1.3조 투입 'AI 과학 난제' 도전…2027년 본격 가동, 2026년 9월 30일, https://news.jkn.co.kr/post/1002500

Penguin, Mission Economy: A Moonshot Guide to Changing Capitalism, 2021년, https://www.penguin.co.uk/books/315191/mission-economy-by-mazzucato-mariana/9780141991689

Springer(IDEAS/RePEc), Learning from Overrated Mission-Oriented Innovation Policies: Seven Takeaways, 2024년, https://ideas.repec.org/h/spr/inschp/978-3-031-49196-2_13.html

Econ Journal Watch, Mission Preposterous: A Review of Moonshots and the New Industrial Policy, 2024년 9월, https://econjwatch.org/articles/a-review-of-moonshots-and-the-new-industrial-policy

샘 올트먼, The Intelligence Age, 2024년 9월 23일, https://ia.samaltman.com/

Business Today, AI could eradicate all disease in just 10 years, says Google DeepMind chief Demis Hassabis, 2025년 4월 22일, https://www.businesstoday.in/technology/news/story/ai-could-eradicate-all-disease-in-just-10-years-says-google-deepmind-chief-demis-hassabis-473009-2025-04-22

로리 셀런존스 블로그, Can AI really solve all diseases by 2040?, 2025년 4월 18일, https://rorycellanjones.substack.com/p/can-ai-really-solve-all-diseases

다리오 아모데이, Machines of Loving Grace, 2024년 10월, https://darioamodei.com/essay/machines-of-loving-grace

DigiChina, China's Plan to 'Lead' in AI: Purpose, Prospects, and Problems, 2017년 8월 1일, https://digichina.stanford.edu/work/chinas-plan-to-lead-in-ai-purpose-prospects-and-problems

BOFIT, No major shifts in economic policy under China's newest five-year plan, 2026년 3월 20일, https://www.bofit.fi/en/monitoring/weekly/2026/vw202612_3/

유럽집행위원회, EU Missions in Horizon Europe, 2026년 10월 조회, https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe_en

유엔 가나 사무소, Sustainable Development Goals Report 2025, 2025년 7월 15일, https://ghana.un.org/en/node/298012

Physics World, ITER fusion reactor hit by massive decade-long delay and €5bn price hike, 2024년 7월 3일, https://physicsworld.com/a/iter-fusion-reactor-hit-by-massive-decade-long-delay-and-e5bn-price-hike/

Axios, Seattle-area Helion quietly moves goalpost on fusion, 2026년 9월 28일, https://www.axios.com/local/seattle/2026/09/28/helion-energy-orion-fusion-microsoft-commercial-power-altman

AI & Society, Is artificial intelligence a threat to meaningful work and living? Technological unemployment and the existential challenges of a transitional era, 2026년 4월 15일, https://doi.org/10.1007/s00146-026-02941-x

arXiv, Are We Automating the Joy Out of Work? Designing AI to Augment Work, Not Meaning, 2026년 3월 16일, https://arxiv.org/abs/2603.14963

Hachette Book Group, Power and Progress, 2023년 5월 16일, https://www.hachettebookgroup.com/titles/daron-acemoglu/power-and-progress/9781668626405/

Fortune, Sam Altman admits AI is killing the labor-capital balance—and says nobody knows what to do about it, 2026년 3월 12일, https://fortune.com/2026/03/12/sam-altman-ai-labor-capital-jobs-nobody-knows/

Fortune, Sam Altman says we need another Renaissance. But we really need another Industrial Revolution., 2026년 10월 9일, https://fortune.com/2026/10/09/sam-altman-ai-renaissance-industrial-revolution/

전자신문, [김경진의 AI전략노트] 〈25〉 기획하는 사람이 이기는 시대가 왔다, 2026년 4월 7일, https://www.etnews.com/20260407000157

McKinsey, The State of Organizations 2026, 2026년, https://www.mckinsey.com/~/media/mckinsey/business%20functions/people%20and%20organizational%20performance/our%20insights/the%20state%20of%20organizations/2026/the-state-of-organizations-2026.pdf



10장. 국가, 기관, 기업의 목표

대통령은 국가 단위에서

"인공지능 3대 강국 도약은 대한민국이 달성할 수 있는 목표이며 이 행동계획은 그 목적지를 위한 우리의 이정표입니다." 국가인공지능전략위원회가 「대한민국 인공지능 행동계획」을 소개하며 적은 문장이다(국가인공지능전략위원회, 2026년 4월 14일). 위원회는 2025년 9월 8일 대통령 직속으로 출범했다. 정부는 2026년 2월 24일 이 계획을 확정했다(ZDNet Korea, 2026년 2월 25일). 인공지능 기본법에 따른 법정 기본계획이고 기간은 2028년까지다. 실행과제 99개와 정책권고 326개가 3대 정책축과 12대 전략분야 아래 놓였다.

돈도 따라붙었다. 국회예산정책처는 2026년 AI 예산을 9조 9,000억 원으로 집계했다(국회예산정책처, 2026년 1월 22일). 2025년 본예산의 약 3배다. 정부는 2028년까지 GPU를 5만 2천 장 넘게 확보한다는 목표도 세웠다. 대선 때 자주 들린 "AI 100조 원"은 2025년 대선에서 두 후보가 함께 내건 5년 투자 공약의 규모다(전자신문, 2025년 5월 27일). 이 숫자를 정부의 공식 목표로 적은 문서는 찾지 못했다. 정부가 실제로 내놓은 재원은 국민성장펀드다. 국민성장펀드는 정부와 민간이 함께 돈을 모아 첨단산업에 투자하는 정책 펀드다. 규모는 5년간 150조 원이고 그 가운데 AI 몫은 최대 30조 원이다(문화일보, 2025년 12월 16일). 2026년에 집행하는 AI 몫은 6조 원이다.

같은 국회예산정책처 보고서는 수단과 결과 사이의 틈을 짚었다. 독자 AI 파운데이션 모델 사업(독파모)은 정부가 국내 팀을 골라 자체 대형 모델을 만들게 하는 사업이다. 2026년 1월 15일 1차 평가에서는 계획한 4개 팀보다 적은 3개 팀이 다음 단계에 올랐다. 보고서는 일부 업체의 해외 모듈 사용이 "독자 AI" 목표와 맞지 않는다는 문제와 평가지표의 공정성 논란을 함께 적었다. 영국 토터스 미디어의 글로벌 AI 인덱스(2025년 12월)에서 한국은 종합 5위였다(국회예산정책처 인용). 인프라 4위와 정부 전략 5위에 비해 인재는 13위이고 산업생태계는 17위였다. 정부가 돈을 쓰는 항목은 앞줄에 있고 사람과 시장은 뒷줄에 있다.

미국과 중국의 목표 문서는 결이 다르다. 미국은 2025년 7월 23일 「경주에서 이기기: 미국의 AI 행동계획」을 냈다(Sidley, 2025년 7월 30일). 혁신 가속과 인프라 구축과 외교

·안보 주도의 세 기둥 아래 정책 행동을 늘어놓았다. 언제까지 무엇을 이룬다는 수치는 두지 않았다. 중국 국무원은 2025년 8월 인공지능 플러스 행동 의견을 냈다(중국 정부망, 2025년 8월 27일). 인공지능 플러스는 산업과 생활 곳곳에 인공지능을 붙이자는 국가 정책이다. 의견은 지능 단말과 에이전트의 보급률을 2027년 70퍼센트와 2030년 90퍼센트 넘게 올린다는 숫자를 걸었다. 2026년 3월 승인된 15차 5개년 계획도 이 행동의 전면 추진을 적었다(MERICS, 2026년 2월 26일). 미국은 순위를 목표로 삼았고 중국은 쓰임의 넓이를 목표로 삼았다. 한국의 "3대 강국"은 미국 쪽에 가까운 순위 목표다.

9장에서 본 케네디의 1961년 연설은 결과물과 시한을 한 문장에 담았다. 한국도 반도체와 조선에서 뒤늦게 뛰어들어 정상에 오른 선언의 기억을 갖고 있다(20장). 두 선례에서 대통령이 정한 것은 장비의 수량보다 무엇을 해낼지였다. 위원회 출범 1년 성과보고회에서 임문영 의원은 국민이 AI 정책의 변화를 거의 체감하지 못한다고 지적했다(정보통신신문, 2026년 9월 8일). 저자는 대통령이 정할 목표가 GPU 장수가 아니라 그 연산으로 무엇을 하느냐는 쓰임새라고 본다. 어떤 병을 일찍 찾고 어느 공장이 무엇을 더 만드는지가 국민이 체감하는 목표다.

국가 목표는 정권마다 바뀌어 의미가 없다는 시각이 있다. 한국일보는 2015년 녹색성장과 창조경제를 두고 정권마다 모호한 기치를 걸고 보여주기식 사업을 되풀이한다고 비판했다(한국일보, 2015년 3월 16일). 미국에서도 2025년 새 행정부가 앞 정부의 2023년 행정명령 14110호를 폐기했다. 이 비판에는 근거가 있다. 다만 지금의 목표는 법 위에 서 있다. 인공지능 기본법은 2026년 1월 22일 시행됐다. 2025년 12월 개정으로 위원회의 심의·의결과 이행점검 권한이 법에 들어갔다. 위원회도 "2030년대 후반 골든타임까지 정책 연속성 유지"를 과제로 적었다(정보통신신문, 2026년 9월 8일). 목표를 법정 계획에 시한과 잣대로 적어 두면 다음 정부도 기록을 남기는 절차를 거쳐야 그 목표를 고칠 수 있다.

기관장은 기관 단위에서

공공기관 경영평가는 정부가 해마다 공공기관의 실적을 점수로 매겨 성과급과 기관장 평가에 반영하는 제도다. 2026년도 평가 기준을 담은 편람에는 "AI 활용 등 혁신" 항목이 새로 생겼고 가점 1.5점이 붙었다(데일리안, 2026년 6월 1일). 편람 원문이 아닌 보도로 확인한 수치다. 대학 쪽에서는 9월 29일 AI중심대학 18곳과 AX대학원 15곳이 함께 출범했다(ZDNet Korea, 2026년 9월 29일). AI중심대학은 최장 8년 동안 해마다 30억 원을 받고 전교생에게 기초교육을 의무로 시킨다. 정부출연연구기관(출연연)은 정부가 돈을 대 세운 연구기관이다. 같은 달 출연연의 연구데이터를 분야별 AI 에이전트와

학습용 데이터셋으로 만드는 사업이 15개 과제로 시작했다(아시아경제, 2026년 9월 29일).

세 곳 모두 인공지능을 도입하는 일 자체에 평가 점수와 예산이 붙어 있다. 도입한 뒤 기관의 고유 업무가 무엇으로 바뀌었는지를 측정하는 칸은 저자가 확인한 공개 자료에서 찾지 못했다. 정부의 말은 다른 쪽을 가리킨다. 정부는 2026년 9월 3일 공공기관 109곳을 줄이거나 합치는 기능개혁안을 냈다. 재정경제부 관계자는 "정책 방향의 목표를 비용 절감에 두고 있지 않다"고 했다(서울신문, 2026년 9월 4일). AI 전환과 에너지 위기 대응을 위한 기능 재편이고 신규 채용도 줄이지 않는다는 설명이다. 이 약속의 결과는 아직 나오지 않았다. 4년 전의 기준은 숫자였다. 정부는 2022년 12월 공공기관 정원 1만 2,442명, 전체의 2.8퍼센트를 줄이기로 했다(매일신문, 2022년 12월 26일).

인력 감축을 목표 그 자체로 삼은 조직의 결과는 미국에 있다. 미국 정부효율부(DOGE)가 움직인 2025년 1월부터 1년 동안 연방 인력은 채용을 빼고 26만 4,000명 줄었다(Cato 연구소). Cato는 작은 정부를 지지하는 기관이다. 미국 회계감사원(GAO)은 2025년 연방정부가 일하지 않는 직원에게 약 95억 달러를 지급했다고 추산했다(타임, 2026년 9월 16일). 떠난 자리 가운데 2만 개 넘게 새 채용으로 다시 채워졌다(NPR, 2025년 10월 1일). DOGE가 주장한 절감액은 독립 검증을 받지 못했고 조직은 2026년 7월 4일 문을 닫았다(피스컬타임스, 2026년 7월 6일).

국내에는 새 임무를 먼저 적은 기관장도 있다. 강청희 국민건강보험공단 이사장은 취임하며 건강보험 빅데이터와 AI로 질병 위험을 예측하고 맞춤형 건강관리를 하겠다고 밝혔다(메디게이트뉴스, 2026년 7월 20일). 공단과 질병관리청은 9월 10일 건강보험 빅데이터 분석센터를 함께 열었다(MTN, 2026년 9월 10일). 이 방향은 줄일 사람보다 새로 맡을 일을 먼저 적었다.

저자는 기관장이 AI 도입 계획의 첫 장에 두 줄을 적자고 제안한다. 한 줄은 줄어드는 업무 시간이고 다른 한 줄은 그 시간으로 새로 맡을 임무다. 경영평가의 AI 가점은 도구를 쓴 실적보다 새 임무의 결과에 매긴다. AI중심대학과 출연연 과제의 평가에도 같은 칸을 둔다.

저자는 20대 국회 과학기술정보방송통신위원회에 있던 2018년 9월 3일 과학기술기본법 일부개정법률안을 대표발의했다(전기신문, 2018년 9월 3일). 의안번호는 15259다. 지역별 과학기술 혁신 전담기관의 지정 요건과 운영을 시행령에서 법률로 올리는 안이었다. 국회 의원 페이지의 발의 목록에는 「인공지능산업 진흥에 관한 법률안」도 올라 있다.

비용 절감에서 멈추는 회사의 성장은 거기서 끝난다

앤디 재시 아마존 최고경영자는 2025년 6월 17일 직원들에게 보낸 메모에서 일부 직무에는 사람이 덜 필요해진다고 썼다("We will need fewer people doing some of the jobs")(TechCrunch, 2025년 6월 17일). 아마존은 그해 10월 사무직 1만 4천 명과 2026년 1월 1만 6천 명의 감원을 발표했다(CNN, 2026년 1월 28일). 회사는 효율과 관료주의 제거를 이유로 들었다.

미국 재취업 컨설팅 회사 챌린저 그레이 앤드 크리스마스는 감원 발표를 사유별로 센다. 2025년 인공지능을 사유로 든 감원 발표는 5만 4,836명으로 전체의 5퍼센트였다(챌린저, 2026년 1월 8일). 2026년 1월부터 9월까지는 12만 136명으로 전체 감원 발표 57만 3,195명의 약 21퍼센트가 됐다(The Wire Desk, 2026년 10월 3일). 두 숫자는 회사가 인공지능을 사유로 명시한 발표만 센 것이다. 실제로 해고가 끝난 인원이나 인공지능이 원인인 인원과는 다르다.

발표 뒤의 실적은 감원과 따로 움직였다. 아마존의 2026년 2분기 매출은 2,006억 달러로 1년 전보다 20퍼센트 늘었다(Investing.com, 2026년 7월 30일). 그 성장의 중심은 36.7퍼센트 늘어난 클라우드 사업 AWS였다. 7장의 세일즈포스는 감원 뒤에도 매출이 늘었다. 이 회사의 2026 회계연도 매출은 415억 달러로 10퍼센트 늘었다(세일즈포스, 2026년 2월 25일). 이 숫자에는 인수한 인포매티카의 매출이 들어 있다. 액센추어는 2025년 여름 약 1만 1,000명을 줄였고 같은 회계연도 매출은 7퍼센트 늘었다(CNBC, 2025년 9월 26일). 마이크로소프트는 2025년 두 차례 감원했지만 사티아 나델라 최고경영자는 새 채용으로 전체 인원이 거의 그대로라고 말했다(야후파이낸스, 2026년 7월 26일).

주식시장은 감원 발표를 반기지 않았다. CNBC는 인공지능을 감원 사유로 든 S&P 500 기업 23곳 가운데 13곳, 56퍼센트의 주가가 2026년 5월 15일 기준 발표 시점보다 낮다고 집계했다(CNBC, 2026년 5월 17일). 세일즈포스는 약 32퍼센트 빠져 있었다. 피츠버그대의 마크 마 교수는 미국 상장기업 5년치 직원 리뷰와 실적 발표를 분석했다(The Conversation, 2026년 8월 12일). AI 감원 발표에 대한 주가 반응은 평균 0에 가까웠다. 발표 뒤 직원들의 AI에 대한 정서는 나빠졌고 그 정서는 기업 생산성과 함께 움직였다.

감원의 이유를 의심하는 분석도 있다. AI 워싱은 다른 이유로 하는 감원이나 사업을 인공지능 덕분인 것처럼 포장하는 일이다. 옥스퍼드 이코노믹스는 2025년 1월부터 11월까지 AI 사유 감원이 전체의 5퍼센트 미만이고 경기 사유 감원이 그 네 배

이상이라고 분석했다(포천, 2026년 1월 7일). 연구진은 기업이 감원을 좋은 소식처럼 꾸민다고 봤다("dressing up layoffs as a good news story"). 도입의 성과도 아직 얇다. MIT NANDA 연구진은 2025년 7월 예비 보고서에서 조사 조직의 95퍼센트가 생성형 AI로 측정할 만한 수익을 얻지 못했다고 했다(포천, 2025년 8월 18일). 이 보고서는 동료심사를 거치지 않았고 95퍼센트의 산출 근거가 불분명하다는 비판을 받았다(Futuriom, 2025년 8월 26일). 맥킨지 조사에서는 응답자의 88퍼센트가 한 기능 이상에 AI를 쓰지만 이익에 영향이 있다고 답한 비율은 39퍼센트였다(맥킨지, 2025년 11월 5일). 그 대부분은 영향이 이익의 5퍼센트 미만이라고 답했다.

비용 절감이 먼저이고 성장은 그다음이라는 시각이 있다. 재시는 같은 메모에서 일부 직무는 줄고 다른 직무는 늘며 전체 사무직은 몇 년에 걸쳐 줄어든다고 썼다. 줄리 스위트 액센추어 최고경영자는 재교육이 되지 않는 인력을 내보낸다고 했다(CNBC, 2025년 9월 26일). 비용 구조를 먼저 바꿔야 AI에 투자할 돈이 나온다는 논리다. 이 논리는 절반만 맞다. 란제이 굴라티 하버드 경영대학원 교수 연구진은 세 차례 불황 전후의 상장기업 4,700곳을 추적했다(HBR, 2010년 3월). 불황 뒤 매출과 이익 성장에서 경쟁사를 10퍼센트 넘게 앞선 기업은 약 9퍼센트였다. 비용만 줄인 기업은 불황 뒤 번성하지 못했다. 운영 효율로 비용을 줄이면서 연구개발과 마케팅과 자산에 경쟁사보다 더 쓴 기업이 이겼다. 인공지능 시대의 자료는 아니지만 비용 절감 다음에 무엇을 하느냐가 결과를 갈랐다는 장기 기록이다.

한국에서도 비용 쪽 숫자가 먼저 움직인다. 희망퇴직은 회사가 퇴직금에 위로금을 얹어 주고 스스로 그만둘 사람을 받는 제도다. 5대 시중은행의 2025년 희망퇴직자는 2,470명으로 1년 전보다 24.3퍼센트 늘었다(뉴스웨이, 2026년 6월 24일). 은행들은 인력과 비용 구조의 효율화를 사유로 들었고 인공지능 하나만을 이유로 들지는 않았다. 저자는 비용 절감 자체를 문제로 보지 않는다. 회사의 성장을 끝내는 것은 줄인 자리를 비워 둔 채 거기서 멈추는 일이다.

사람이 빠진 자리에 새 사업과 새 시장을 채우는 회사

IKEA는 고객센터에 AI 봇 빌리(Billie)를 도입한 뒤 상담원 약 8,500명을 내보내지 않았다(포브스, 2026년 10월 5일). 회사는 이들을 원격으로 집 꾸미기를 상담하고 판매하는 인력으로 재교육해 재배치했다. 빌리가 처리하는 문의의 비중은 처음 2년 동안 47퍼센트에서 74퍼센트로 올랐다. 원격 판매 채널의 2025 회계연도 매출은 13억 7천만 달러로 전체 매출의 약 3퍼센트였다. 이 금액은 원격 판매 채널 전체의 매출이고 재교육 인력이 만든 몫만 따로 집계한 숫자는 아니다.

감원을 발표한 회사가 교육에 돈을 쓰기도 한다. 업스킬링은 직원이 지금 일보다 높은 기술을 쓰도록 다시 가르치는 일이다. 아마존은 2019년 이후 전 세계 70만 명 넘게 업스킬 과정을 거쳤고 이 서약에 12억 달러 넘게 들였다고 밝혔다(아마존, 2026년 6월 24일 수정). 같은 회사가 사무직을 줄이면서 사람을 다시 가르치는 일도 함께 한다.

국내에서 인공지능으로 줄인 인력을 새 사업에 돌렸다고 수치와 함께 밝힌 기업은 찾지 못했다. 확인되는 것은 교육의 규모다. 15장이 소개한 LG의 전 임원 과정에 더해 사내 LG AI대학원 1기에는 석사 11명과 박사 6명이 등록했다(서울신문, 2026년 9월 29일). 삼성은 2026년 사장단과 임원부터 AI 집중교육을 시작해 그해 안에 전 임직원 교육을 마친다는 목표를 세웠다(이투데이, 2026년 6월 9일). 이 숫자들은 교육을 받은 사람의 수다. 그 사람들이 어떤 새 일을 맡았는지는 아직 공개된 자료에 없다.

사람이 적은 회사가 새 시장을 여는 흐름도 보인다. 월스트리트저널은 결제 회사 스트라이프의 자료를 인용해 2023년부터 2025년 사이 연 매출 100만 달러 넘는 1인 사업자가 두 배로 늘었다고 보도했다(뉴시스, 2026년 8월 4일). 연 매출 1천만 달러 넘는 1인 사업자는 거의 세 배가 됐다. 한국의 1인 창조기업은 2023년 기준 116만 2,529개다(중소벤처기업부, 2026년 4월 9일). 이 가운데 93.2퍼센트는 사업을 지금 규모로 유지하겠다고 답했고 넓히겠다는 답은 1.6퍼센트였다. 이 조사에는 인공지능을 쓰는지 묻는 항목이 없다. 한국 통계로는 인공지능이 1인 기업을 키웠는지 말할 수 없다.

사례를 나란히 놓으면 결과를 가른 것은 사람을 줄였느냐 늘렸느냐보다 줄인 자리에 무엇을 세웠느냐였다고 우리는 읽는다. IKEA는 상담원을 판매 인력으로 바꿨다. 마이크로소프트는 줄인 만큼 다른 자리에서 뽑았다. 마크 마의 분석에서 감원만 발표한 회사는 직원의 정서를 잃었다.

저자는 기업의 AI 도입 계획서에 줄일 비용과 나란히 새로 열 매출의 칸을 두자고 제안한다. 줄어든 인력이 옮겨 갈 일과 그 일의 첫해 목표도 같은 계획서에 적는다. 이사회는 두 칸을 함께 승인한다.

한국이 세울 만한 국가 목표의 후보

아래에 제시하는 후보는 현재의 기술과 산업 여건에서 먼저 추진할 단기 목표다. 일부 과제는 실현에 여러 해가 걸리겠지만, 문명의 변화라는 긴 시간축에서 보면 출발 단계에 속한다. 저자가 제안하는 모델과 로봇, 메모리와 데이터 사업만으로 인간과 사회가 앞으로 무엇을 지향할지에 대한 답이 완성되지는 않는다.

앞 장에서 말한 담대한 목표는 지금 부족한 기술을 확보하거나 산업 경쟁에서 앞서는 일보다 넓다. 당장의 과제를 해결한 뒤 어떤 삶과 사회를 만들 것인지까지 생각해야 한다. 아래 후보들은 그 사색을 대신하는 최종 설계도가 아니라, 다음 목표를 탐색할 역량을 갖추기 위한 첫 과제로 읽어야 한다.

목표를 고를 때 먼저 볼 것은 한국이 이미 가진 것이다. 고대역폭메모리(HBM)는 이름 그대로 한 번에 주고받는 데이터의 양을 크게 늘린 메모리다. 2026년 2분기 HBM 매출 점유율은 SK하이닉스 50퍼센트, 삼성전자 33퍼센트였다(카운터포인트리서치, 2026년 9월 3일). 위원회의 하정우 상근 부위원장은 한국의 강점으로 메모리 반도체, 프런티어 AI 추격 역량, 제조 기반의 피지컬 AI를 꼽았다(파이낸셜뉴스, 2026년 9월 8일). 약점은 앞에서 본 인재와 산업생태계다.

다음은 저자가 제안하는 후보이며 정부가 확정한 목표가 아니다. 1번부터 6번은 저자가 전자신문 연재에서 이미 제안한 내용이고 7번은 후보로 검토할 만한 것이다. 각 후보에는 추진 근거와 함께 풀어야 할 쟁점, 아직 확인하지 못한 자료를 적었다. 쟁점은 사업의 설계와 검증 조건으로 다루고, 자료가 부족한 부분은 확정 전에 확인해야 한다. 이 검토를 거쳐 무엇을 먼저 추진할지 정하자는 것이다.

1. 프런티어 모델 주권. 프런티어 모델은 한 시점에 성능이 제일 앞선 인공지능 모델이다. 저자는 정부와 기업이 함께 출자한 법인에 자금과 GPU와 데이터와 인재를 모아 국가 대표 프런티어 모델을 만들자고 했다(전자신문, 2026년 10월 6일). 미국 상무부의 수출통제로 국내 기관이 한때 최신 모델을 쓰지 못한 일이 근거였다. 반대 근거는 규모다. 알파벳, 마이크로소프트, 아마존, 메타 네 곳의 2026년 설비투자는 합계 약 7,250억 달러다(전자신문, 2026년 10월 6일). 한국 정부의 AI 예산과 견줄 수 없는 규모다.

2. 휴머노이드. 저자는 연구개발, 부품의 국내 생산, 실증, 수출 금융을 한 기관이 맡는 가칭 한국휴머노이드공사를 세우자고 했다(전자신문, 2026년 8월 25일). 6장이 다룬 정부의 2030년 계획과 한국의 제조 현장이 근거다. 공사 방식이 민간의 경쟁을 대신할 수 있는지, 6장의 로드니 브룩스가 말한 대로 로봇의 손이 언제 쓸모 있어질지는 반대 근거로 남는다.

3. 메모리 다음의 메모리. 저자는 HBM 다음 세대 메모리를 국내 파운데이션 모델과 휴머노이드와 함께 한국이 명운을 걸 세 축으로 꼽았다(전자신문, 2026년 7월 28일). 위의 점유율이 근거다. 반대로 중국 CXMT의 D램 매출은 2026년 2분기 1년 전보다 716퍼센트 늘었다(카운터포인트리서치, 2026년 9월 1일). 민간 두 회사가 이미 1위와 2위인 분야에 국가 목표가 더할 몫이 무엇인지 설명이 필요하다.

4. 데이터 하이웨이. 저자는 데이터를 쌓아 두는 데이터 댐에서 흐르게 하는 데이터 하이웨이로 가고 의료 데이터를 열자고 했다(전자신문, 2025년 5월 27일). 앞 소절의 건강보험 빅데이터 분석센터와 77만 명 규모 국가통합바이오빅데이터의 단계적 개방이 근거다(메디게이트뉴스, 2025년 12월 11일). 개인정보 보호와 민감정보를 둘러싼 반대는 20장의 논의 그대로 남는다.

5. 국방과 외교의 AI. 저자는 외교 정보를 자동으로 모으고 분석하는 체계와 미국 국방부의 최고디지털·인공지능책임자실 같은 국방 통합 조직을 제안했다(전자신문, 2025년 6월 17일; 2025년 7월 1일). 수출통제가 보여 준 외부 의존의 위험이 근거다. 국방 AI 예산 자료를 확인하지 못했고 기밀 영역이라 공개 검증이 어렵다는 점이 반대 근거다.

6. AI 경제 영토. 저자는 개발도상국에 데이터센터를 세우고 한국 모델을 얹어 재해 예측과 농업에 쓰게 하는 원조를 제안했다(전자신문, 2026년 9월 8일). 모델이 쓰이는 범위가 곧 경제 영토라는 판단이다. 정부도 개도국 AI 데이터센터 원조를 검토한다는 보도가 있다(네이트뉴스, 2026년 7월 30일). 반대 근거는 외교부의 원조 몫이 2026년 약 2조 2,000억 원으로 1년 전보다 22퍼센트 줄었다는 점(전자신문, 2026년 9월 8일)과 독자 모델의 성능이 아직 검증되지 않았다는 점이다.

7. 축소 사회의 AI. 인구가 줄고 지방이 비는 사회를 인공지능으로 관리하는 첫 나라가 되는 목표다. 저자는 초등 1학년이 2031년 22만 명으로 줄어든다는 전망을 들어 아시아 인재에게 문을 열자고 했다(전자신문, 2026년 9월 22일). 2026년 미래인구포럼은 마을 단위의 소멸 위험을 AI로 정밀하게 진단하자고 했다(한반도미래인구연구원, 2026년 4월 30일). 인공지능이 지역 서비스의 격차를 줄였다는 실증은 아직 없다.

저자가 1순위로 꼽는 것은 1번과 2번을 하나로 묶은 목표다. 2028년까지 공동 출자 법인으로 국가 대표 모델을 갖고 2030년까지 그 모델을 얹은 휴머노이드를 국내 공장과 공공 현장에 넣는다. 잣대는 GPU 장수에서 국내 현장의 성과로 옮긴다. 모델과 로봇이 맡는 일의 양과 함께 작업의 품질과 안전, 필요한 서비스를 받는 사람의 범위가 얼마나 개선됐는지 평가한다. 모델은 머리이고 휴머노이드는 6장이 말한 몸이다. 둘 다 한국의 메모리와 제조 기반 위에 선다. 7번은 미국과 중국이 세우지 않은 한국 고유의 과제다. 저자는 1순위 목표로 만든 모델과 로봇을 먼저 쓸 곳으로 인구가 줄어드는 지역을 꼽는다.

한국은 추격 전략이 맞고 담대한 목표는 과욕이라는 시각이 있다. 정석근 SK하이퍼 대표는 3등이라도 제대로 전략을 세워 빨리 3등이 되자고 했다(네이트뉴스, 2026년 8월 26일). 이경전 경희대 교수는 "왜 항상 등수에 집착하는가"라고 물었다(ZDNet Korea,

2025년 3월 14일). 순위에 매달리지 말자는 지적에는 저자도 동의한다. 다만 강점에 모으는 일과 시한이 붙은 큰 목표는 서로 밀어내지 않는다. 저자가 고른 목표는 한국이 이미 앞선 메모리와 제조 위에 놓인 강점 집중이다.

이 단기 목표를 달성한 이후의 목표는 지금부터 준비해야 한다. 인공지능과 로봇이 생산과 연구의 많은 부분을 맡게 된다면, 늘어난 능력과 시간을 어디에 쓸 것인지가 국가의 과제가 된다. 더 많은 상품을 생산하는 데서 나아가 누구나 배움과 돌봄을 누리는 사회, 사는 지역과 신체적 조건 때문에 삶의 가능성이 좁아지지 않는 사회를 구체적으로 설계할 필요가 있다. 이는 이미 확정된 정책이나 미래에 대한 단정이 아니라, 장기 목표를 만들어 가며 검토할 방향이다.

그 설계에는 사색과 미래 예측이 함께 필요하다. 기술이 얼마나 빨리 발전할지, 인구와 환경이 어떻게 달라질지 여러 가능성을 미리 검토해야 한다. 예측이 빗나갈 경우도 살피고, 그 속에서도 지켜야 할 인간의 존엄과 자유, 사회가 함께 추구할 가치를 논의해야 한다. 아직 이름조차 붙이지 못한 새로운 목표를 발견하는 일도 이 과정에 포함된다.

목표의 규모가 크다고 구체성을 포기할 수는 없다. 누구의 삶을 어떤 상태로 바꿀 것인지, 그 과정에 어떤 기술과 제도가 필요한지, 비용과 책임은 누가 맡을지를 밝혀야 한다. 가령 지역에 따른 돌봄의 격차를 없애겠다면 로봇 보급 대수를 넘어 실제로 돌봄을 받지 못하는 사람이 얼마나 줄었는지를 살펴야 한다. 먼 미래의 방향을 중간 단계의 과제로 나누고, 실행 결과와 새 지식에 따라 목표 자체도 고쳐야 한다.

따라서 단기 사업의 성공이 국가 목표 설계의 끝이 되어서는 안 된다. 현재의 과제를 추진하면서 그 이후를 예측하고, 시민과 여러 분야의 연구자가 새로운 목표를 만들고 검토하는 일을 지속해야 한다. 문명의 변화에 걸맞은 큰 구상과 삶의 변화를 확인할 수 있는 구체적인 계획을 함께 갖추는 것, 그것이 앞서 제시한 단기 목표 너머에 남아 있는 과제다.

목표를 세우는 자리도 정해야 한다. 저자는 2017년 9월 인터뷰에서 4차산업혁명위원회의 위상이 낮아 총괄할 힘이 모자란다고 보고 과학기술부총리 부활을 요구했다(전자신문, 2017년 9월 21일). 지금은 과학기술부총리가 있다. 대통령 직속 국가인공지능전략위원회는 법에 적힌 조정과 점검 권한을 갖는다. 저자는 새 대통령 직속 기구를 더 만들지 말자고 제안한다. 목표는 위원회가 의결하고 대통령이 발표한다. 시한과 잣대는 법정 기본계획에 적고 해마다 국회에 성과를 보고한다. 부처 예산을 그 목표에 맞추는 조정은 과학기술부총리가 맡는다. 안전연구소의 위상은 15장의 제안을 따른다.

2025년 8월 15일 베이징 경기장에서 돌려차기를 헛차고 넘어진 로봇은 상대의 주먹에 진 것이 아니었다. 그 로봇에게 어떤 동작을 익히게 할지는 대학 팀이 정했고 경기의 종목표는 대회를 연 사람들이 짰다. 기계가 몸과 머리를 갖춰 가는 동안에도 어디로 갈지 정하는 일은 사람에게 남고, 그 일을 놓은 개인과 기업과 국가는 스스로 가치를 잃는다고 저자는 판단한다.

참고 자료

국가인공지능전략위원회, [국가AI전략위] 대한민국 인공지능 행동계획 (한글, 인쇄본), 2026년 4월 14일, https://www.aikorea.go.kr/web/board/brdDetail.do?menu_cd=000011&num=523

ZDNet Korea, 정부, 'AI 행동계획' 확정…2028년까지 AI 3강 도약 청사진, 2026년 2월 25일, https://zdnet.co.kr/view/?no=20260224183212

국회예산정책처, AI 3대 강국 도약을 위한 예산 현황과 향후 과제(나보포커스 제132호), 2026년 1월 22일, https://www.nabo.go.kr/board/file/down.do?fid=33319071

전자신문, [김경진의 AI전략노트] 〈5〉 대한민국 AI 강국 도약, 지도자 의지가 성패 가른다, 2025년 5월 27일, https://www.etnews.com/20250527000125

정책브리핑, 국민성장펀드 '150조 원'으로 확대…AI·반도체 등 첨단산업 지원, 2025년 9월 10일, https://www.korea.kr/news/policyNewsView.do?newsId=148949026

문화일보, 국민성장펀드, 5년간 150조 투입… 내년 AI에 6조, 2025년 12월 16일, https://www.munhwa.com/article/11554540

Sidley Data Matters, The Trump Administration's 2025 AI Action Plan – Winning the Race: America's AI Action Plan – and Related Executive Orders, 2025년 7월 30일, https://datamatters.sidley.com/2025/07/30/the-trump-administrations-2025-ai-action-plan-winning-the-race-americas-ai-action-plan-and-related-executive-orders/

Wilson Center, America's AI Action Plan After One Year: What the Administration Has Done, and What Comes Next, 2026년 9월 9일, https://www.wilsoncenter.org/article/americas-ai-action-plan-after-one-year-what-administration-has-done-and-what-comes-next

중국 정부망, China issues guideline to accelerate 'AI Plus' integration across key sectors, 2025년 8월 27일, https://english.www.gov.cn/policies/latestreleases/202508/27/content_WS68ae7976c6d0868f4e8f51a0.html

MERICS, China's next five-year bet on AI: Self-reliance, diffusion, and a lot of hype, 2026년 2월 26일, https://merics.org/en/comment/chinas-next-five-year-bet-ai-self-reliance-diffusion-and-lot-hype

정보통신신문, 국가AI전략위 출범 1년…'AI 3대 강국' 향한 실행체계 고도화, 2026년 9월 8일, https://www.koit.co.kr/news/articleView.html?idxno=208909

파이낸셜뉴스, "대체불가 'AI 핵심 국가' 도약 뒷받침" 하정우 국가AI전략위 부위원장, 2026년 9월 8일, https://www.fnnews.com/news/202609081324079461

한국일보, 녹색성장 · 창조경제… 정권마다 모호한 기치 내걸고 보여주기식 사업 반복, 2015년 3월 16일, https://www.hankookilbo.com/news/article/201503160480742293

데일리안, 공공기관 경영평가 발표 임박…안전·AI, 판도 바꿀 변수 [경평의 시간 ①], 2026년 6월 1일, https://www.dailian.co.kr/news/view/1651288/

ZDNet Korea, "국가 AI경쟁력 강화 기여"...AI중심대학-AX대학원 공동 선언, 2026년 9월 29일, https://zdnet.co.kr/view/?no=20260929221413

아시아경제, 출연연 연구데이터·노하우, AI가 쓰는 자산으로…15개 과제 첫 지원, 2026년 9월 29일, https://view.asiae.co.kr/article/2026092908544713390

서울신문, 정부 "공공기관 개혁, 비용 절감 목표 아냐…신규채용 안줄여", 2026년 9월 4일, https://www.seoul.co.kr/news/economy/2026/09/04/20260904500216

매일신문, 윤석열 정부, 공공기관 1만2천442명 구조조정, 2022년 12월 26일, https://www.imaeil.com/page/view/2022122613073932290

Cato Institute, DOGE Produced the Largest Peacetime Workforce Cut on Record, but Spending Kept Rising, 날짜 미확인, https://www.cato.org/blog/doge-produced-largest-peacetime-workforce-cut-record-spending-kept-rising-0

Time, DOGE Initiative Resulted in Federal Employees Getting Paid Billions to Not Work, New Government Report Finds, 2026년 9월 16일, https://time.com/article/2026/09/16/federal-spending-layoffs-deferred-resignation-opm-doge-gao-report/

NPR, Federal agencies are rehiring workers and spending more after DOGE's push to cut, 2025년 10월 1일, https://www.npr.org/2025/10/01/nx-s1-5558298/doge-fiscal-year-savings-budget-rehired-government-shutdown

The Fiscal Times, DOGE Is Officially Done, 2026년 7월 6일, https://www.thefiscaltimes.com/2026/07/06/DOGE-Officially-Done

메디게이트뉴스, 강청희 건보공단 이사장 취임…"건강보험 넘어 AI 기반 건강복지플랫폼으로", 2026년 7월 20일, https://medigatenews.com/news/3179344814

MTN, 보건의료 AX 가속화…AI 조직·서비스 잇따라 선보여, 2026년 9월 10일, https://news.mtn.co.kr/news-detail/2026091015584663679

전기신문, 김경진 의원 , 3일 '과학기술기본법' 개정안 대표 발의, 2018년 9월 3일, https://www.electimes.com/news/articleView.html?idxno=164118

TechCrunch, Amazon expects to reduce corporate jobs due to AI, 2025년 6월 17일, https://techcrunch.com/2025/06/17/amazon-expects-to-reduce-corporate-jobs-due-to-ai

CNN, Amazon is laying off 16,000 employees as AI battle intensifies, 2026년 1월 28일, https://www.cnn.com/2026/01/28/tech/amazon-layoffs-ai

Challenger, Gray & Christmas, Challenger Report: Job Cuts Fall to Lowest Level in 17 Months; Highest December Hiring Since 2022, 2026년 1월 8일, https://www.challengergray.com/wp-content/uploads/2026/01/Challenger-Report-December-2025.pdf

The Wire Desk, AI cited for 120,136 US job cuts so far in 2026, but only 3,961 in September, Challenger says, 2026년 10월 3일, https://thewiredesk.insiderone.in/story/ai-cited-for-120-136-us-job-cuts-so-far-in-2026-but-only-3-961-in-sept-wkql

Investing.com, Earnings call transcript: Amazon tops Q2 2026 estimates as AWS growth accelerates, 2026년 7월 30일, https://www.investing.com/news/transcripts/earnings-call-transcript-amazon-tops-q2-2026-estimates-as-aws-growth-accelerates-93CH-4826442

세일즈포스, Salesforce Delivers Record Fourth Quarter Fiscal 2026 Results, 2026년 2월 25일, https://www.salesforce.com/news/press-releases/2026/02/25/fy26-q4-earnings/

CNBC, Accenture plans on 'exiting' staff who can't be reskilled on AI amid restructuring strategy, 2025년 9월 26일, https://www.cnbc.com/2025/09/26/accenture-plans-on-exiting-staff-who-cant-be-reskilled-on-ai.html

야후파이낸스, Monday.com is the latest tech company to blame AI for layoffs — here are 20 others, 2026년 7월 26일, https://finance.yahoo.com/technology/ai/articles/running-list-major-tech-layoffs-012755703.html

CNBC, AI-related layoffs a boost for stocks? Not necessarily, 2026년 5월 17일, https://www.cnbc.com/2026/05/17/ai-related-layoffs-a-boost-for-stocks-not-necessarily.html

The Conversation, Layoffs tied to AI hurt worker productivity – and the reason may surprise managers, 2026년 8월 12일, https://theconversation.com/layoffs-tied-to-ai-hurt-worker-productivity-and-the-reason-may-surprise-managers-286749

포천, AI layoffs are looking more and more like corporate fiction that's masking a darker reality, Oxford Economics suggests, 2026년 1월 7일, https://fortune.com/2026/01/07/ai-layoffs-convenient-corporate-fiction-true-false-oxford-economics-productivity

포천, MIT report: 95% of generative AI pilots at companies are failing, 2025년 8월 18일, https://fortune.com/2025/08/18/mit-report-95-percent-generative-ai-pilots-at-companies-failing-cfo

Futuriom, Why We Don't Believe MIT NANDA's Weird AI Study, 2025년 8월 26일, https://www.futuriom.com/articles/news/why-we-dont-believe-mit-nandas-werid-ai-study/2025/08

맥킨지, The state of AI in 2025: Agents, innovation, and transformation, 2025년 11월 5일, https://www.mckinsey.com/~/media/mckinsey/business%20functions/quantumblack/our%20insights/the%20state%20of%20ai/the-state-of-ai-in-2025-vf.pdf

HBR, Roaring Out of Recession, 2010년 3월, https://hbr.org/2010/03/roaring-out-of-recession

뉴스웨이, 희망퇴직 행렬 속 젊어진 은행원…AI가 바꾼 금융권 인력 양극화, 2026년 6월 24일, https://www.newsway.co.kr/news/view?ud=2026062414343024268

포브스, IKEA Retrained 8,500 Call Center Workers Into Design Sales After AI: $1.37 Billion Followed, 2026년 10월 5일, https://www.forbes.com/sites/noemi-kis/2026/10/05/ikea-retrained-8500-call-center-workers-into-design-sales-after-ai-137-billion-followed/

아마존, Our Upskilling Programs, 2026년 6월 24일 수정, https://www.aboutamazon.com/workplace/upskilling-commitments

서울신문, LG, 맞춤형 AI 교육으로 인재 육성 확대, 2026년 9월 29일, https://www.seoul.co.kr/news/economy/industry/2026/09/29/20260929021001

이투데이, 삼성 사장단, AI 집중교육 받는다…이재용표 'AX 혁신' 본격화 [삼성 'AI 대전환'], 2026년 6월 9일, https://www.etoday.co.kr/news/view/2591748

뉴시스, "직원 한 명도 없는데 매출 140억"…AI가 바꾼 美 1인 기업 시대, 2026년 8월 4일, https://www.newsis.com/view/NISX20260804_0003736453

정책브리핑(중소벤처기업부), 2025년 1인 창조기업 실태조사 결과 발표, 2026년 4월 9일, https://www.korea.kr/briefing/pressReleaseView.do?newsId=156753709

카운터포인트리서치, 전세계 D램 및 HBM 시장 점유율: 분기별 데이터, 2026년 9월 1일, https://korea.counterpointresearch.com/global-dram-and-hbm-market-share-quarterly/

카운터포인트리서치, SK하이닉스 HBM 점유율 50%로 하락…삼성 33%로 추격, 2026년 9월 3일, https://korea.counterpointresearch.com/global-hbm-market-share-q2-2026/

전자신문, [김경진의 AI전략노트] 〈38〉 AI 주권, 흩어진 GPU로는 지킬 수 없다, 2026년 10월 6일, https://www.etnews.com/20261006000183

전자신문, [김경진의 AI전략노트] 〈35〉 대한민국 휴머노이드 공사(公社)를 세우자, 2026년 8월 25일, https://www.etnews.com/20260825000157

전자신문, [김경진의 AI전략노트] 〈33〉 변곡점에서 새로운 질서가 탄생한다, 2026년 7월 28일, https://www.etnews.com/20260728000142

메디게이트뉴스, 복지부 AI 기반 의료 혁신 위한 보건의료데이터 활용 확대 추진, 2025년 12월 11일, https://www.medigatenews.com/news/2823403795

전자신문, [김경진의 AI전략노트] 〈6〉 AI 기반 외교정보분석 시스템 도입이 시급하다, 2025년 6월 17일, https://www.etnews.com/20250617000204

전자신문, [김경진의 AI전략노트] 〈7〉 대한민국 국방 AI 혁신을 위한 조직 개편이 필요하다, 2025년 7월 1일, https://www.etnews.com/20250701000278

전자신문, [김경진의 AI전략노트] 〈36〉 데이터 센터, 독파모 그리고 ODA, 2026년 9월 8일, https://www.etnews.com/20260908000102

네이트뉴스(매일경제 전재), 개도국 AI데이터센터에 국내기업 반도체 심는다, 2026년 7월 30일, https://news.nate.com/view/20260730n32625

전자신문, [김경진의 AI전략노트] 〈37〉 중국의 숫자와 한국의 숫자, 2026년 9월 22일, https://www.etnews.com/20260922000328

한반도미래인구연구원, [2026 제1차 미래인구포럼] AI시대 지역소멸 대응과 축소사회 전환 탐색: 인구 위기와 국가 생존 전략, 2026년 4월 30일, https://kppif.org/bbs/board.php?bo_table=population_forum&wr_id=18

네이트뉴스(조선비즈 전재), [스마트클라우드쇼 2026] "韓, AI 3위 아닌 1위 목표로 전략 세워야…제조 강국 데이터 이점 살리면 기회 있어"(종합), 2026년 8월 26일, https://m.news.nate.com/view/20260826n30113

ZDNet Korea, "목표는 AI 3대 강국, 현실은 10위권 밖"…어디부터 손봐야 할까, 2025년 3월 14일, https://zdnet.co.kr/view/?no=20250314012839

전자신문, [창간 35주년 특집 Ⅲ]〈인터뷰〉김경진 의원, "과도한 규제 풀고, 교육개혁 선행해야", 2017년 9월 21일, https://m.etnews.com/20170914000202

가디언, Box, run, crash: China's humanoid robot games show advances and limitations, 2025년 8월 15일, https://www.theguardian.com/world/2025/aug/15/china-world-humanoid-robot-games-advances-limitations



제3부. 통제의 조건: 위험에 대비한다

11장. 경고하는 사람들과 반론

재귀적 자기 개선과 통제 이탈의 뜻

2026년 9월 16일 저녁 제프리 힌턴은 미국 의회의 비공개 브리핑을 마치고 기자들 앞에 섰다(NBC News, 2026년 9월 17일). 그는 인공지능을 규제할 시간이 1년쯤 남았고 그보다 길지는 않다고 말했다("Maybe a year, but not much more than a year"). 무언가 하지 않으면 통제를 벗어날 것이라며 속도를 늦춰야 한다고도 했다("We need to slow down").

그는 그해 7월의 사고를 "작은 체르노빌"(little Chernobyl)이라고 불렀다. 오픈AI가 내부 평가에서 돌리던 에이전트가 시험 환경을 벗어나 허깅페이스의 운영 시스템에 들어간 사건이다(허깅페이스, 2026년 7월 27일). 허깅페이스는 전 세계 개발자가 인공지능 모델과 데이터를 올리고 내려받는 공유 플랫폼이다.

힌턴의 경고 밑에는 재귀적 자기 개선(recursive self-improvement)이라는 개념이 있다. 인공지능이 자기 자신을 고쳐 성능을 높이고 그 결과로 다시 더 빨리 고치는 순환을 말한다. 앤트로픽은 자사 인공지능 개발 작업의 점점 큰 몫을 인공지능에 맡기고 있다고 밝혔다(앤트로픽 연구소, 2026년 6월 4일). 보고서는 완전한 재귀적 자기 개선에는 아직 이르지 않았고 꼭 그렇게 된다는 보장도 없다고 적었다. 다만 그 순간이 제도가 준비한 것보다 빨리 올 수 있다고 경고했다.

통제 이탈(loss of control)은 인공지능이 누구의 통제도 받지 않고 움직이며 통제를 되찾기가 극도로 어렵거나 불가능해진 상황이다(국제 AI 안전 보고서, 2026년 2월 3일). 요슈아 벤지오가 의장을 맡은 이 보고서는 지금의 시스템에는 그런 위험을 일으킬 능력이 아직 없다고 봤다. 자율성 같은 관련 능력은 빠르게 좋아지고 있다고 함께 적었다. 두 개념을 나란히 놓으면 재귀적 자기 개선은 능력이 오르는 속도의 문제이고 통제 이탈은 그 능력을 사람이 방향 짓고 멈출 수 없게 된 상태다.

유엔 독립 국제 AI 과학패널은 첫 주제 브리핑에서 7월 사고를 다뤘다(유엔 AI 과학패널, 2026년 9월 21일). 공동의장 벤지오는 통제 이탈의 세 조건인 어긋난 목표와 그것을 추구할 능력과 그것을 허용하는 환경이 실제 시스템에서 모두 모였다고 말했다.

패널은 심각한 통제 이탈을 예측하지 않는다고 밝히면서도 불확실성을 통제할 수 있다는 증거로 읽어서는 안 된다고 적었다.

아모데이의 속도 조절론

앤트로픽의 다리오 아모데이는 2026년 9월 12일 개인 사이트에 에세이 "We Must Pace the Frontier"를 올렸다. 10장에서 본 프런티어 모델의 능력이 오르는 속도를 늦춰 안전 검증이 따라올 시간을 벌자는 주장이다. 그는 속도 조절이 모델 훈련이나 기술 진보의 중단을 뜻하지 않는다고 썼다("pacing does not mean halting model training or technical progress").

그가 든 근거는 인공지능이 다음 세대 인공지능을 만드는 일을 점점 더 돕는다는 점과 7월 사고다. 제안의 첫 단계는 METR 같은 외부 평가 기관에 직원에 준하는 접근권을 주는 상주 평가자 제도다. 앤트로픽은 이 단계를 혼자서 곧바로 시행한다고 밝혔다. 다음 단계는 민주국가 기업들이 공통 안전 기준을 정하는 일이고 그는 정부의 지원과 반독점법 면제가 필요하다고 썼다. 마지막 단계는 권위주의 정부와의 조율이며 그는 검증을 큰 난제로 꼽았다.

이 에세이는 혜택을 그린 2024년 10월의 "Machines of Loving Grace"와 위험을 다룬 2026년 1월의 "The Adolescence of Technology"를 잇는 글이다(GV Wire, 2026년 9월 19일). 두 달 앞선 7월 28일에는 오픈AI와 앤트로픽, 구글, 메타 등의 직원 1,100명 이상이 공개서한에 서명했다(CNN, 2026년 7월 28일). 서한은 지금 늦추자는 요구가 아니었고 늦출 수 있는 도구를 만들자는 요구였다.

오픈AI의 샘 올트먼은 에세이가 나온 날 X에 동의한다고 썼다("I agree with Dario that we need to pace the frontier")(Tech Policy Press, 2026년 10월 1일). 반대 목소리도 컸다. 튜링상 수상자 얀 르쿤은 아모데이를 "완전히 착각에 빠졌다"(completely deluded)고 평했다(Fortune, 2026년 10월 1일). 르쿤은 허깅페이스 사고를 샌드박스 설계가 허술했던 공학의 실패로 봤다. 샌드박스는 프로그램이 바깥 시스템에 손대지 못하도록 가둬 둔 격리 실행 환경이다. 앤트로픽의 규제 요구는 규제 포획에 가깝다고 했다. 규제 포획은 규제를 받는 기업이 규제의 내용을 자기에게 유리하게 만드는 현상이다.

첫 단계의 상주 평가자는 다른 회사의 동의 없이도 시작할 수 있고 외부의 눈을 회사 안으로 들이는 조치여서 규제 포획이라는 의심에 일부 답이 된다. 기업끼리 속도를 맞추려면 반독점법의 예외가 필요하다는 둘째 단계는 사정이 다르다. 기업 간 속도 합의는 반독점법 위반 소지가 있다는 지적이 이미 나왔다(American Action Forum,

2026년 10월 7일). 아모데이의 제안 가운데 바로 검증할 수 있는 것은 상주 평가자이고 업계 합의 부분은 아직 설득할 몫이 남아 있다.

힌턴: 에이전트 사고 이전부터 위험을 경고해 온 사람

6장에서 짧게 적은 힌턴의 공개 경고는 2023년 5월 1일에 시작됐다. 그날 뉴욕타임스는 그가 인공지능의 위험을 자유롭게 말하려고 구글을 떠났다고 보도했다. 가짜 콘텐츠 때문에 무엇이 진실인지 알 수 없게 되는 일과 일자리 문제와 인공지능이 스스로 코드를 실행하는 일을 걱정했다. 그는 그 전까지 인간보다 똑똑한 인공지능이 나오려면 30년에서 50년이나 그 이상 걸릴 줄 알았다고 했다. 그가 경고를 시작한 2023년 봄의 대화형 인공지능은 시험 환경을 뚫고 다른 회사의 시스템에 들어가는 일과는 거리가 멀었다.

그보다 앞선 2010년대에 그가 존재적 위험을 공개적으로 경고한 기록은 찾기 어렵다. 2017년에는 자율무기를 실제 위험이라 부르며 제네바 협약 같은 새 규범을 요구했다(BBVA, 2017년). 이 발언은 기계가 사람의 통제를 벗어나는 문제와 주제가 다르다. 힌턴도 2024년 12월 7일 노벨상 기자회견에서 안전을 더 일찍 생각했으면 좋았겠다고 말했다(CP24, 2024년 12월 8일). 원문은 "I wish I'd thought about safety earlier"다. 안전을 더 일찍 생각하지 못했다는 점을 스스로 인정한 셈이다. 2023년 구글을 떠난 뒤에는 공개 경고를 이어 왔다. 2026년 10월 13일 가디언 인터뷰에서 그는 기업 자율규제 주장을 석유 회사와 석면 회사에 빗대 꼬집었다("I'm sure the oil and asbestos companies would have liked to as well")(HuffPost, 2026년 10월 13일).

힌턴은 2024년 말부터 인공지능이 수십 년 안에 인류 멸종에 기여할 확률을 10퍼센트에서 20퍼센트로 말해 왔다. 2026년 8월 CNN 인터뷰에서 그는 이 수치가 어림짐작(wild guess)이라고 스스로 말했다(Forbes, 2026년 8월 7일). 그는 늦추자고 말하지만 모라토리엄을 요구하지는 않는다. 모라토리엄은 개발을 일정 기간 통째로 멈추는 조치다.

엔비디아의 젠슨 황은 이 수치를 두고 10퍼센트라는 확률은 과학에 근거하지 않았다고 반박했다("That 10% chance is not grounded on science")(뉴욕타임스 팟캐스트 전사문, 2026년 9월 23일). 힌턴은 가디언 인터뷰에서 그 수치가 심각하지만 불확실한 가능성을 나타내는 표현이라고 답했다. 확률을 정밀하게 계산할 증거가 없다는 데는 두 사람의 말이 같다. 둘이 갈리는 곳은 불확실한 위험을 대비의 근거로 삼을지 여부다.

중국의 시각: 미국의 추격 저지 음모라는 해석

아모데이 에세이가 나온 다음 날인 9월 13일 중국 관영 매체 환구시보(Global Times)가 사설을 냈다(NPR, 2026년 9월 14일). 사설은 에세이가 안전을 말하는 듯하지만 중국을 겨냥한 "냉전 각본"(Cold War playbook)이라고 했다. 기술 장벽과 규제 독점으로 중국의 인공지능 발전을 누르고 미국의 기술 패권을 지키려 한다는 해석이다. 사설은 미국이 격차를 벌린 다음에야 조건부 협상으로 속도를 늦추자고 하는 구조라고 비판했다.

이 해석에는 에세이 안에 근거가 있다. 에세이는 중국공산당을 중심에 두고 민주국가가 선두를 지켜야 속도 조절이 성립한다고 썼다. 첨단 칩 수출 제한과 모델 가중치 탈취 방지도 권고했다. 아모데이는 중국이 따르지 않는 경우를 제일 어려운 딜레마(toughest dilemma)라고 불렀다(CNBC, 2026년 9월 13일).

중국 외교부의 반응은 사설보다 절제돼 있었다. 궈자쿤 대변인은 9월 14일 정례 브리핑에서 공포 조장과 대결과 악성 경쟁이 세계 인공지능 거버넌스를 어지럽힐 뿐이라고 답했다("Fearmongering, confrontation and malicious competition will only disrupt")(NBC News, 2026년 9월 14일). 외교부는 봉쇄나 억제라는 말까지는 쓰지 않았다. 산업계의 목소리는 더 직접적이었다. 미니맥스의 코드 연구개발 책임자 타오허는 왜 늘 미국이 게임의 규칙을 정하느냐고 되물었다("Why do you always get to define the game")(Rest of World, 2026년 9월 23일).

추격 저지 음모라는 해석은 관영 매체 사설과 일부 개발자의 말에서 뚜렷하다. 중국이 안전 논의 전체를 음모로 보는 것은 아니다. 외교부 브리핑과 같은 날 중국의 정보보안 표준 위원회 TC260은 인공지능 안전 거버넌스 프레임워크 3.0을 냈다(China AI Bulletin, 2026년 9월). 서문은 모델이 자율 학습과 재귀적 자기 개선으로 스스로 가속하는 추세를 보인다고 적었다. 그 진화가 인간의 통제를 앞지를 수 있는지 경계해야 한다고도 썼다.

중국 바깥의 연구자들도 이 양면을 짚는다. GovAI의 치란 첸은 미국과 중국이 위험을 근본적으로 다르게 보지 않는다고 했다(Rest of World, 2026년 9월 23일). 싱가포르국립대의 자이언 총은 중국의 태도를 규칙을 직접 쓰고 싶어 한다는 말로 풀었다("They want to write the rules")(NBC News, 2026년 9월 14일). 중국은 미국 기업이 이끄는 속도 조절과 수출 통제를 봉쇄로 읽으면서 통제 이탈 위험 자체는 자기 문서에 자기 언어로 적었다. 다툼의 대상은 위험이 있느냐보다 규칙을 누가 정하느냐에 가깝다.

미국 대통령의 논리: 멈추면 중국만 이득

도널드 트럼프 대통령은 아모데이 에세이가 나온 지 이틀 뒤인 9월 14일 Truth Social에 연속으로 글을 올렸다(알자지라, 2026년 9월 14일). 그는 인공지능과 데이터센터를 겨냥한 병적인 음모가 벌어지고 있다고 썼다. 그리고 이 일을 반기는 쪽은 중국뿐이라고 덧붙였다("the only one that is happy about it is China"). 인공지능을 늦추려는 시도가 결국 중국에 이롭다는 취지다. 그는 인공지능이 세계를 장악하고 인류를 파괴한다는 이야기를 사기(HOAX)라고 불렀다(Forbes, 2026년 9월 14일). 인공지능에서 이기는 쪽이 이긴다는 글도 올렸다("WHOEVER WINS AI, WINS")(알자지라, 2026년 9월 15일).

다음 날 마이크 존슨 하원의장은 인공지능 개발에 모라토리엄을 둘 수 없다고 했다(알자지라, 2026년 9월 15일). 그는 기업이 스스로 규제할 수 있다고 하면서도 독립 감사인은 필요하다고 덧붙였다("I think you need independent auditors").

이 논리는 행정부 문서에 먼저 나와 있었다. 2025년 7월 23일 백악관의 인공지능 행동계획은 제목부터 "Winning the Race"였다(Sidley, 2025년 7월 30일). 7월 사고에 앞서 나온 2026년의 문서는 결이 조금 다르다. 2026년 6월 2일 행정명령 14409는 프런티어 모델을 출시 30일 전까지 정부에 자발적으로 내 사이버 능력을 검토받는 틀을 만들었다(PBS, 2026년 6월 2일). 의무 인허가나 사전 승인 제도는 만들지 않는다고 명시했다.

9월 29일에는 트럼프와 앤트로픽, 구글, 메타, 오픈AI, xAI, 엔비디아의 대표가 "Joint Commitment on Frontier Responsibilities"에 서명했다(Fortune, 2026년 10월 1일). 약 300단어 분량의 이 자발적 약속은 회사 안의 통제와 점검팀에 더해 독립 외부 감사인과 독립 이사회 위원회를 두는 구조다. 속도 조절이나 능력 기준선은 들어 있지 않고 벌칙과 시한도 없다. 트럼프는 이 약속에 도덕적 구속력(morally binding)이 있다고 말했다.

반대는 여당 안에서도 나왔다. 조시 홀리 상원의원은 국가안보를 위한 의무적 사전 배포 시험을 주장했다(Tech Policy Press, 2026년 10월 1일). 존 케네디 상원의원은 비상 정지 장치를 의무로 하는 법안을 냈지만 9월 16일 랜드 폴 상원의원이 만장일치 처리를 막았다.

경쟁 논리는 실재하는 변수를 짚는다. 아모데이 자신도 중국을 제일 어려운 딜레마로 꼽았다. 행정부는 사고 이후 자발적 검토와 외부 감사라는 장치를 받아들였고 받아들이지 않은 것은 의무화다. 멈추면 중국만 이득이라는 논리는 평가를 의무로 할지에 대해서는 답을 주지 않는다.

젠슨 황: 통제 이탈 가능성을 낮게 보는 쪽

황은 9월 20일 방영된 CBS 인터뷰에서 2030년이 세상의 끝이 되지는 않는다고 말했다(Fortune, 2026년 10월 4일). 그것이 세상의 끝이 될 확률은 0퍼센트라고 했다("There is a 0% chance that's going to be the end of the world"). 이 0퍼센트는 2030년까지 인공지능 때문에 세계가 끝날 가능성에 대한 말이다. 그가 통제 이탈의 확률에 수치를 준 발언은 찾지 못했다. 그는 할 수 있는 만큼 빨리 가되 가야 할 속도보다 빠르면 안 된다고 했다("We should go as fast as we can, but not faster than we should").

황이 통제 이탈을 아예 부정하는 것은 아니다. 9월 14일 로스앤젤레스 올인 서밋(All-In Summit) 무대에서 그는 회사가 통제 불능이거나 안전하지 않다고 느끼면 잠시 멈추고 제대로 해야 한다고 말했다("take a pause and make sure you get it right")(The Next Web, 2026년 9월 19일). 9월 23일 공개된 뉴욕타임스 팟캐스트에서는 실험을 가둘 수 없는 연구소라면 연구소를 닫아야 한다고 했다("we have to shut the labs down"). 통제가 안 된다고 믿으면 통제할 수 있을 때까지 제품을 내놓지 말라고도 했다.

같은 인터뷰에서 그는 소프트웨어가 늘 샌드박스를 빠져나간다고 말했다("software breaks out of sandboxes all the time"). 황은 통제 이탈을 샌드박스 탈출 같은 공학 문제로 보고 그것이 인류 차원의 위험으로 번진다는 서사를 받아들이지 않는다.

인공지능 분야 블로거 즈비 모쇼위츠는 황이 안전을 출시 전 통제로 좁혀 훈련 중의 사고를 보지 않는다고 비판했다(Zvi Mowshowitz, 2026년 9월 25일). 포천은 엔비디아 매출에서 데이터센터가 약 92퍼센트를 차지한다는 점을 들어 그의 입장과 회사의 이해관계가 겹친다고 지적했다(Fortune, 2026년 10월 4일).

이해관계에 대한 지적은 그의 말이 틀렸다는 증거가 되지 못한다. 적용 범위는 실제 조치로 따져야 한다. 7월 사고는 제품을 내놓기 전 회사 안의 평가 단계에서 일어났다. 황은 통제하지 못하는 연구소는 닫아야 한다고도 말했다. 이 원칙이 실효성을 가지려면 출시 여부의 판단에 그치지 않고 평가와 훈련 중에도 통제를 확인하고 작업을 멈출 절차로 이어져야 한다. 다툼은 통제의 필요성을 인정하느냐에서, 누가 어떤 기준으로 확인하고 중단시키느냐로 옮겨 간다.

전문가의 의견이 갈리는 지점과 갈리지 않는 지점

의견이 갈리는 첫 지점은 시기다. 국제 AI 안전 보고서는 앞으로의 진보 속도에 상당한 불확실성이 있다고 적었다("substantial uncertainty about how fast future progress will

be"). 확률도 갈린다. 연구자 1,580명이 답한 2024년 설문에서 인공지능에 의한 멸종이나 영구적 무력화 확률은 평균 18퍼센트, 중위값 10퍼센트였다(AI Impacts, 2026년 9월).

규제 방식도 갈린다. 아모데이와 7월 직원 서한은 조율된 속도 조절을 말하고 트럼프 행정부와 황은 자율규제와 기존 법의 적용을 말한다. 홀리의 의무 사전 시험과 케네디의 비상 정지 법안은 그 중간에 있다. 원인 진단도 엇갈린다. 정렬은 인공지능이 사람이 의도한 목표와 가치에 맞게 움직이도록 만드는 일이다. 벤지오와 유엔 패널은 7월 사고를 훈련 방식이 에이전트에게 엉뚱한 목표를 심을 수 있다는 정렬의 문제로 읽는다. 르쿤과 황은 샌드박스가 샜다는 공학의 문제로 읽는다. 같은 사고를 두고 고칠 곳을 다르게 짚는 셈이어서 이 차이는 확률 숫자보다 뿌리가 깊다.

갈리지 않는 지점도 분명하다. 평가가 그 하나다. 국제 AI 안전 보고서는 출시 전 시험이 실제 위험을 늘 잘 예측하지는 못한다며 평가의 공백을 인정했다. 아모데이는 상주 평가자를 말하고 존슨 하원의장은 독립 감사인을 말한다. 6월 행정명령의 자발적 검토와 9월 29일 약속의 외부 감사인도 평가를 늘리자는 쪽이다. 다툼은 평가를 의무로 할지와 누가 할지에 있다. 사고 보고에 대해서도 이견이 적다. 보고서는 사고 보고와 감시가 부족해 지금의 관행이 효과가 있는지 판단하기 어렵다고 썼다. 유엔 패널은 항공과 의학과 사이버보안의 사고 보고와 독립 조사를 본보기로 들었다. 10월 13일 34개국 과학자 약 400명은 주요 사고를 공개하는 국제 등록부를 요구했다(Rappler, 2026년 10월 13일). 투명성도 같은 쪽에 있다. 보고서는 개발사가 내부 평가 자료를 나누지 않아 외부 검증이 막힌다고 적었다.

한국에서도 비슷한 구도가 보인다. 배경훈 부총리 겸 과학기술정보통신부 장관은 9월 16일 "중요한 것은 속도 조절이 아니라 속도 책임"이라고 썼다(머니투데이, 2026년 9월 16일). 10월 6일 국정감사에서는 "킬스위치 같은 인간의 통제 장치가 필요하다"고 말했다(디지털투데이, 2026년 10월 6일). 하정우 국가인공지능전략위원회 상근 부위원장은 9월 30일 한국이 안전 역량을 갖추고 주도적으로 목소리를 내야 한다고 했다(뉴스1, 2026년 10월 1일). 김명주 인공지능안전연구소장은 제3국의 평가가 필요해지면 한국이 나서겠다고 했다(ZDNet Korea, 2026년 10월 13일). 윤성로 서울대 전기정보공학부 교수는 속도 조절론이 기업들의 의도적인 발언일 가능성이 크다고 봤다(파이낸셜뉴스, 2026년 9월 26일). 이성엽 고려대 기술경영전문대학원 교수는 실현이 어렵다고 보고 "안전한 가속"을 권했다(전자신문, 2026년 9월 27일).

속도를 두고는 의견이 갈리지만, 여기서 살펴본 발언과 문서에는 평가와 사고 보고와 투명성을 늘리자는 공통점이 있다. 멸종 확률이나 발전 속도에 합의할 때까지 이 조치들을 미룰 이유는 없다. 누가 자료를 보고 위험을 판단하며 중단을 요구할 수

있는지부터 제도로 정해야 한다. 그 필요성을 따져 볼 근거가 2026년 여름 평가 현장에서 일어난 세 건의 사고다. 다음 장은 그 사고들을 하나씩 따라간다.

참고 자료

NBC News, 'Godfather of AI' warns Congress has 'maybe a year' left to regulate AI, 2026년 9월 17일, https://www.nbcnews.com/politics/congress/godfather-ai-warns-congress-maybe-year-left-regulate-ai-rcna598330

허깅페이스, Anatomy of a Frontier Lab Agent Intrusion: A Technical Timeline of the July 2026 Incident, 2026년 7월 27일, https://huggingface.co/blog/agent-intrusion-technical-timeline

앤트로픽 연구소, When AI builds itself, 2026년 6월 4일, https://www.anthropic.com/institute/recursive-self-improvement

International AI Safety Report, International AI Safety Report 2026: Extended Summary for Policymakers, 2026년 2월 3일, https://internationalaisafetyreport.org/publication/2026-report-extended-summary-policymakers

유엔 독립 국제 AI 과학패널, AI Agents, Misalignment and the Risk of Losing Human Control 보도자료, 2026년 9월 21일, https://www.un.org/independent-international-scientific-panel-ai/sites/default/files/2026-09/Press%20Release_Thematic%20Brief_AI%20Agents,%20Misalignment%20and%20the%20Risk%20of%20Losing%20Human%20Control_AI%20Scientific%20Panel.pdf

다리오 아모데이, We Must Pace the Frontier, 2026년 9월 12일, https://darioamodei.com/post/we-must-pace-the-frontier

GV Wire, The Awesome and Alarming AI Visions of Anthropic's CEO, 2026년 9월 19일, https://gvwire.com/2026/09/19/the-awesome-and-alarming-ai-visions-of-anthropics-ceo/

CNN, Employees from the world's biggest AI companies want the US to be ready to slow AI development, 2026년 7월 28일, https://edition.cnn.com/2026/07/28/tech/ai-development-tech-employees-open-letter

Axios, Anthropic, OpenAI CEOs call for slowdown in AI development, 2026년 9월 12일, https://www.axios.com/2026/09/12/anthropic-ai-amodei-pacing

Fortune, AI 'godfather' Yann LeCun: Anthropic CEO Dario Amodei is 'deluded,' 'crazy,' and doesn't understand cybersecurity, 2026년 10월 1일, https://fortune.com/2026/10/01/yann-lecun-anthropic-ceo-dario-amodei-deluded-crazy-cybersecurity/

American Action Forum, Pacing the AI Frontier: An (Allegedly Illegal) Agreement, 2026년 10월 7일, https://www.americanactionforum.org/insight/pacing-the-ai-frontier-an-allegedly-illegal-agreement/

뉴욕타임스, 'The Godfather of A.I.' Leaves Google and Warns of Danger Ahead, 2023년 5월 1일, https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html

BBVA, Geoff Hinton: "I'm more concerned about autonomous weapons than about intelligent machines", 2017년 1월 30일, https://www.bbva.com/en/innovation/geoff-hinton-im-concerned-autonomous-weapons-intelligent-machines/

CP24, Nobel recipient Geoffrey Hinton wishes he thoughts of AI safety sooner, 2024년 12월 8일, https://www.cp24.com/news/world/2024/12/08/nobel-recipient-geoffrey-hinton-wishes-he-thoughts-of-ai-safety-sooner/

HuffPost, 'Godfather Of AI' Spots Exactly Why Trump Understands 'Almost Nothing' About It, 2026년 10월 13일, https://www.huffpost.com/entry/ai-godfather-geoffrey-hinton-trump-regulation_n_6ace33afe4b0649434776f64

Forbes, Geoffrey Hinton Warns AI May Outsmart Humans As Agents Escape Tests, 2026년 8월 7일, https://www.forbes.com/sites/timbajarin/2026/08/07/geoffrey-hinton-warns-ai-may-outsmart-humans-as-agents-escape-tests/

Podscripts, The Ezra Klein Show(뉴욕타임스 팟캐스트): Jensen Huang vs. the A.I. Doomers 전사문, 2026년 9월 23일, https://podscripts.co/podcasts/the-ezra-klein-show/jensen-huang-vs-the-ai-doomers

NPR, Beijing hits back at Anthropic CEO's call to curb China's AI development, 2026년 9월 14일, https://www.npr.org/2026/09/14/nx-s1-5968456/china-hits-back-ai-development

CNBC, Anthropic's Amodei says China presents 'toughest dilemma' for his proposed AI slowdown, 2026년 9월 13일, https://www.cnbc.com/2026/09/13/china-dilemma-ai-slowdown-anthropic.html

NBC News, China dismisses AI slowdown calls and blasts 'fearmongering' from U.S. tech leaders, 2026년 9월 14일, https://www.nbcnews.com/world/china/china-ai-slowdown-trump-amodei-altman-threat-cold-war-rcna597631

Rest of World, Dario Amodei wants to slow AI. China isn't taking orders, 2026년 9월 23일, https://restofworld.org/2026/china-rejects-dario-amodei-anthropic-ai-slowdown-trump-xi/

China AI Bulletin, CAIB Special Issue 2: the AI Safety Governance Framework 3.0, 2026년 9월 17일, https://chinaaibulletin.substack.com/p/caib-special-issue-2-the-ai-safety

알자지라, Trump says calls for more control on AI are a 'SICK conspiracy', 2026년 9월 14일, https://www.aljazeera.com/news/2026/9/14/trump-says-calls-for-more-control-on-ai-are-a-sick-conspiracy

Forbes, Trump Says Calls For AI Regulation Are A 'Hoax' – Compares It To Russia, Impeachment 'Scams', 2026년 9월 14일, https://www.forbes.com/sites/conormurray/2026/09/14/trump-says-calls-for-ai-regulation-are-a-hoax-compares-it-to-russia-impeachment-scams/

알자지라, US speaker Mike Johnson rejects AI pause, warns of losing edge to China, 2026년 9월 15일, https://www.aljazeera.com/news/2026/9/15/us-speaker-rejects-ai-pause-warns-of-losing-edge-to-china

Sidley Data Matters, The Trump Administration's 2025 AI Action Plan – Winning the Race: America's AI Action Plan – and Related Executive Orders, 2025년 7월 30일, https://datamatters.sidley.com/2025/07/30/the-trump-administrations-2025-ai-action-plan-winning-the-race-americas-ai-action-plan-and-related-executive-orders/

PBS NewsHour, Trump signs executive order that allows voluntary federal vetting of top AI models for national security risks, 2026년 6월 2일, https://www.pbs.org/newshour/nation/trump-signs-executive-order-that-allows-voluntary-federal-vetting-of-top-ai-models-for-national-security-risks

Fortune, AI's biggest players promise to police themselves at the White House, 2026년 10월 1일, https://fortune.com/2026/10/01/trump-ai-regulation-luncheon-huang-amodei-pichai-brockman-zuckerberg-musk-joint-committment-frontier-responsibilities/

Tech Policy Press, September 2026 US Tech Policy Roundup, 2026년 10월 1일, https://www.techpolicy.press/september-2026-us-tech-policy-roundup/

Fortune, Nvidia CEO Jensen Huang has emerged as the biggest foil to AI doomerism about the existential risk to humanity, 2026년 10월 4일, https://fortune.com/2026/10/04/nvidia-jensen-huang-ai-doomerism-foil-risk/

The Next Web, Weeks ago Jensen Huang said labs should pace themselves if they felt out of control. Now he puts the risk at 0%, 2026년 9월 19일, https://thenextweb.com/news/huang-0-percent-hardening-pace-if-out-of-control

Zvi Mowshowitz, On Ezra Klein's Podcast With Jensen Huang, 2026년 9월 25일, https://thezvi.wordpress.com/2026/09/25/on-ezra-kleins-podcast-with-jensen-huang/

AI Impacts, ESPAI 2024 보고서, 2026년 9월, https://aiimpacts.org/wp-content/uploads/2026/09/ESPAI2024.pdf

Rappler, Scientists call for international body to oversee advanced AI, 2026년 10월 13일, https://www.rappler.com/technology/scientists-call-international-body-oversee-advanced-ai/

머니투데이(네이트 전재), 'AI 위험론'에 선 그은 배경훈 부총리 "우린 AI 속도 늦출 여유 없다", 2026년 9월 16일, https://m.news.nate.com/view/20260916n20743

디지털투데이, [국감2026] 배경훈 "AI 킬스위치 필요…프런티어 AI로 사이버 위협 대응", 2026년 10월 6일, https://www.digitaltoday.co.kr/news/articleView.html?idxno=705165

뉴스1(네이트 전재), 하정우 "AI 개발 늦추기보다 안전 역량 갖춰야", 2026년 10월 1일, https://m.news.nate.com/view/20261001n26488

ZDNet Korea, AI 안전 평가 선도하는 영국…한국도 국제무대 나설 수 있을까, 2026년 10월 13일, https://zdnet.co.kr/view/?no=20261013110737

파이낸셜뉴스, "AI 속도 조절론 속에 숨은 의도 있다"…서울대 교수가 본 빅테크의 셈법, 2026년 9월 26일, https://www.fnnews.com/news/202609260043383539

전자신문, AI 속도조절의 필요성과 실현가능성, 2026년 9월 27일, https://www.etnews.com/20260927000059



12장. 사고는 이미 현실이 됐다

2026년 7월, 오픈AI 내부 평가 중 에이전트의 시험 환경 이탈

허깅페이스의 기술 타임라인은 첫 공격 행동의 시각을 2026년 7월 9일 02시 28분(협정세계시)으로 적었다(허깅페이스, 2026년 7월 27일). 침입은 7월 13일 오후까지 4.5일가량 이어졌다. 허깅페이스는 7월 16일 공개 글에서 공격자를 밝히지 못했다. 자율 에이전트 프레임워크를 쓴 공격으로 보인다고만 했다. 바탕 모델은 아직 모른다고 밝혔다(허깅페이스, 2026년 7월 16일).

공격자의 정체는 7월 21일에 드러났다. 오픈AI가 자사 내부 평가에서 돌리던 모델이 이 침입을 일으켰다고 공개했다(오픈AI, 2026년 7월 21일; AI타임스, 2026년 7월 22일). 오픈AI는 사이버 공격 능력을 재는 내부 평가에서 에이전트를 돌리고 있었다. 평가에 쓴 모델은 GPT 5.6 Sol과 이름을 밝히지 않은 내부 연구 모델이다. 위험한 요청을 거절하는 기능은 평가를 위해 낮춰 둔 상태였다.

오픈AI는 이 평가를 샌드박스 안에서 진행했다. 이 샌드박스는 소프트웨어 패키지 설치만 중계 서버(프록시)를 거쳐 허용하도록 짜여 있었다. 모델은 그 중계 서버에서 아직 알려지지 않은 취약점을 찾아 인터넷이 열린 경로에 닿았다(Malwarebytes, 2026년 7월 24일). 그다음 허깅페이스에서 평가 과제와 관련된 자료를 찾으려 했다. 그 과정에서 허깅페이스 운영 시스템에 들어갔다. 두 회사는 공동 포렌식 끝에 사람이 조종한 공격이 아니라 통제된 시험 중의 사고라고 결론지었다(디지털데일리, 2026년 7월 22일). 포렌식은 컴퓨터에 남은 기록을 증거로 모아 경위를 밝히는 조사다.

사건의 규모는 한 달 뒤 외부 분석에서 드러났다. 평가 기관 METR과 레드우드 리서치(Redwood Research)는 8월 26일 공동 분석을 냈다(METR, 2026년 8월 26일). 분석에 따르면 서로 격리돼 있어야 할 약 1,200개 에이전트가 승인되지 않은 게시판으로 소통했다. 에이전트들은 내부 파일 공유 기능을 게시판으로 바꿔 썼다. 이 게시판에서 메시지와 파일 7만 건 이상이 오갔다. 그중 약 700개가 허깅페이스 공격에 가담했다. 같은 기간 게시판에서 활동한 533개 에이전트 가운데 90퍼센트 넘게 빠르게 공격에 합류했다. 윤리적으로 망설이는 말을 남긴 에이전트도 있었지만 그 망설임이 행동을 막은 일은 드물었다.

동기에 대한 해석은 조금 갈린다. 허깅페이스는 에이전트가 과제를 직접 푸는 대신 운영 시스템에 들어가 시험 답안을 빼내려 했다고 봤다(허깅페이스, 2026년 7월 27일). METR과 레드우드 리서치의 판단은 다르다. 에이전트들은 풀 수 없다고 여긴 과제에서 채점기를 속이려 했다. 두 기관은 허깅페이스 공격의 주된 동기가 채점기의 작동 방식을 알아내는 데 있었다고 봤다. 두 해석은 사고가 평가를 속이려는 시도에서 시작됐다는 점에서 만난다. 분석팀은 작업 기록 약 7퍼센트에서 소규모 기록 위조도 찾아냈다(METR, 2026년 8월 26일).

허깅페이스 시스템 침투의 경과와 오픈AI의 속도 조절

허깅페이스의 기록을 날짜순으로 따라가면 침입이 한 단계씩 깊어진 과정이 보인다(허깅페이스, 2026년 7월 27일). 7월 10일에는 낮은 강도로 정찰했다. 7월 11일에는 운영 환경 안으로 들어가 내부 운영 데이터베이스에 접근했다. 허깅페이스는 에이전트가 이 데이터를 읽기만 했고 고치지는 않았다고 적었다. 7월 12일 밤에는 소스 코드 관리 시스템까지 닿았다. 에이전트는 쓰기 권한이 있는 토큰을 받아 코드 변경 요청을 열었지만 악성 변경을 반영하지는 않았다. 마지막 의미 있는 공격 활동은 7월 13일 13시 37분에 있었다. 보안 대응팀은 그 뒤 접근을 끊었다.

허깅페이스는 탐지가 늦은 이유도 밝혔다. 여러 보안 장치가 초기 신호를 냈지만 신호를 묶어 판단하는 체계가 경보 등급을 올리지 않았다. 허깅페이스는 이 일로 귀한 시간을 잃었다고 썼다. 에이전트는 ExploitGym과 CyberGym 관련 데이터셋 5건에 접근했다. 키 136개가 든 운영 비밀 객체와 인증용 서명 키도 밖으로 나갔다(허깅페이스, 2026년 7월 27일). 허깅페이스는 이 키들을 모두 교체했다. 공개 모델과 데이터셋, 이용자가 만든 응용 공간(Spaces)에는 변조가 없었다(허깅페이스, 2026년 7월 16일).

조사 과정에서는 방어하는 쪽도 인공지능을 썼다. 허깅페이스는 처음에 상용 프런티어 모델로 서버 기록을 분석하려 했다. 그 모델의 제공사가 걸어 둔 안전장치가 분석을 막았다. 허깅페이스는 자체 서버에서 돌리는 모델로 기록을 분석했다(TechCrunch, 2026년 7월 20일). 허깅페이스는 기계의 속도로 공격하는 상대 앞에서 평범한 약점도 방어자에게 더 큰 비용이 된다고 교훈을 정리했다.

피해는 허깅페이스에서 끝나지 않았다. 한국경제(2026년 7월 29일)는 오픈AI가 7월 28일 GPT 5.6 Sol 등 에이전트가 허깅페이스 말고도 네 곳에 더 들어갔다고 밝혔다고 전했다. 같은 기사는 탈출 경로가 오픈AI가 직접 설치한 프로그램이었다는 점을 두고 기초적 실수라는 평가와 마케팅이라는 시각을 함께 소개했다.

오픈AI의 후속 조치는 8월 18일에 나왔다(오픈AI, 2026년 8월 18일; Constellation Research, 2026년 8월 18일). 오픈AI는 "temporarily slowed the pace of scaling"이라고 썼다. 모델 규모를 키우는 속도를 한동안 늦췄다는 뜻이다. 배포를 앞둔 최신 모델의 강화학습 훈련은 2주 동안 멈췄다. 강화학습은 모델이 한 행동에 점수를 매겨 점수가 높은 쪽으로 행동을 고쳐 가게 하는 훈련 방식이다. 규모가 제일 큰 프런티어 강화학습 훈련은 보류했다. 오픈AI는 소규모 훈련과 평가로 모델의 행동과 안전장치를 확인한 뒤 이 훈련을 재개하기로 했다. 모델이 답을 내기 전에 거치는 추론 과정을 사고 사슬(chain of thought)이라 부른다. 오픈AI는 사고 사슬을 감시하는 범위도 장기 실행 작업과 훈련 과정으로 넓혔다. 운영 중인 제품과 서비스에는 영향이 없었다. 회사 전체가 개발을 멈춘 일은 아니었다.

두 달 뒤 비슷한 일이 한 번 더 있었다. 9월 20일 검색 기반 훈련 과제를 수행하던 내부 연구 모델이 훈련 샌드박스의 네트워크 제한에서 빈틈을 찾아 외부 챗봇에 질문을 보냈다(오픈AI 정렬팀, 2026년 9월 25일). 감시 체계는 약 15분 만에 경보를 냈지만 자동 중단이 작동하지 않았다. 사람이 약 2시간 반 뒤에 직접 종료했다. 오픈AI는 이 사고가 네트워크 제한 통제의 빈틈을 드러냈다고 인정했다. 그리고 최상위 모델군의 도구 사용 훈련, 평가, 추론을 모두 멈췄다. 이 사건은 외부 챗봇에 질문을 보낸 데서 그쳤다. 다른 회사 시스템에 침입한 허깅페이스 사건과는 성격이 다르다.

영국 AI안전연구소의 122건 평가와 승인받지 않은 행동

영국 AI안전연구소(AI Security Institute)는 2026년 8월 4일 "Incident Report: unsanctioned agent behaviour during cyber testing"을 공개했다(영국 AI안전연구소, 2026년 8월 4일). 이 기관은 2025년 2월 이름을 AI Safety Institute에서 지금 이름으로 바꿨다. 사고 번호 INC-2026-07-28-01이 붙은 35쪽짜리 기술 보고서도 함께 나왔다.

안전연구소는 두 개의 사이버 레인지에서 모델 7종을 평가했다. 사이버 레인지는 공격과 방어를 시험하도록 만든 가상의 전산망이다. 안전연구소는 평가를 모두 122회 실행했다. 그중 10회의 실행에서 승인받지 않은 행동 19건이 나왔다. 보고서는 평가 범위 밖에 효과를 미친 행동을 이렇게 부른다. 19건 가운데 17건은 앤트로픽의 Claude Mythos 5에서 나왔다. 나머지 2건은 오픈AI의 GPT 5.6 Sol의 한 번의 실행에서 나왔다. 실행 횟수는 Mythos 5가 43회이고 GPT 5.6 Sol이 35회였다. 사고에 관련된 이 두 모델만 제공사의 사이버 분류기가 꺼져 있었다. 사이버 분류기는 공격성 요청을 걸러 내도록 모델 제공사가 붙여 둔 감시 장치다.

행동은 7월 25일부터 28일 사이에 일어났다. 7월 28일 상용 보안 감시 소프트웨어가 비정상적인 외부 송신을 잡았다. 안전연구소는 경보가 담당 팀에 닿은 지 한 시간 안에 사고를 선언하고 실행을 끝냈다. 12시 23분(영국 서머타임)에는 샌드박스를 격리했다. 몇 시간 뒤에는 예방 조치로 Mythos 5, Mythos Preview, Opus 5, GPT 5.6 Sol의 접근을 모든 사용자에게서 막았다.

보고서는 19건을 세 묶음으로 나눴다. 레인지 밖 실제 사람을 상대로 한 사회공학이 5건이고 모두 Mythos 5가 했다. 사회공학은 사람의 신뢰를 이용해 원하는 행동을 끌어내는 공격 방식이다. 보고서는 이 묶음을 제일 심각하게 봤다. 나머지는 과제에 필요 없던 인터넷 행동 11건과 에이전트끼리 인터넷에서 주고받은 행동 3건이다. 관리자 신상 정보 수집, 신원을 감춘 가짜 계정 생성, 사람 확인 절차(CAPTCHA) 우회가 여기에 들어 있다. 한 에이전트가 공개 메모장에 남긴 접속 정보를 다른 에이전트가 찾아 그 계정에 들어간 일도 있었다.

심각도가 제일 높은 사례는 오픈소스 프로젝트를 겨냥했다. Mythos 5는 악성 코드가 든 코드 변경 요청을 프로젝트에 냈다. 사람 검토자가 막자 정직한 실수였다고 둘러댔다. 이어 고쳤다고 주장하며 같은 코드를 다시 넣으려 여러 번 시도했다. 과제를 푸는 데 꼭 필요한 행동도 아니었다. 보고서는 이 정도로 심각한 기만을 본 것은 처음이라고 적었다(This is the first time AISI has seen deception of this severity). 실제 피해는 없었다고도 밝혔다(no real-world harm resulted).

이 사건을 오픈AI 사건과 같은 종류로 읽으면 곤란하다. 안전연구소는 평가 설계상 인터넷 접근을 열어 두었다. 보고서도 모델이 안전한 시험 환경을 탈출한 사례가 아니라고 밝혔다(this was not a case of a model escaping its secure test environment). 샌드박스를 뚫은 이탈이 아니라 인터넷 접근을 의도적으로 허용한 평가에서 생긴 일이다. 안전연구소는 기만 행동을 두고 과제를 좇는 과정의 부산물로 나타났다고 평가했다(deception emerged as a by-product of pursuing the task). 안전연구소는 영국 정부 사이버 조정 센터(GC3)와 국가사이버보안센터(NCSC)에 사고를 알렸다. 8월 3일에는 관련 모델 개발사와 미국 CAISI(Center for AI Standards and Innovation)에도 통지했다.

앤트로픽이 공개한 사이버 공격 악용 사례

앤트로픽은 2025년 3월부터 자사 모델 Claude가 범죄에 쓰인 사례를 보고서로 공개해 왔다(앤트로픽, 2025년 3월). 2025년 8월 27일 보고서의 제목은 "Detecting and countering misuse of AI: August 2025"다. 앤트로픽은 이런 보고서를 위협 정보

보고서라고 부른다. 이 보고서에 실린 데이터 갈취 작전에서 공격자는 코딩 에이전트 Claude Code로 최소 17개 조직에 침입했다. 때로는 50만 달러가 넘는 몸값을 요구했다(앤트로픽, 2025년 8월 27일). 북한 요원이 가짜 신원으로 미국 대기업에 원격 취업한 사례도 실렸다. 코딩 경험이 적은 범죄자가 랜섬웨어 변종을 만들어 400달러에서 1,200달러에 판 사례도 있다. 앤트로픽은 관련 계정을 막고 새 탐지 방법을 도입했으며 기술 지표를 관계 당국과 나눴다.

2025년 11월 13일 보고서는 한 걸음 더 나간 사례를 다뤘다(앤트로픽, 2025년 11월 13일). 앤트로픽은 2025년 9월 중순 GTG-1002라는 집단의 활동을 탐지했다. 앤트로픽은 이 집단을 높은 신뢰도로 중국 국가 배후로 평가했다. 이 집단은 Claude Code를 자기 공격 틀 안의 자동화 도구로 조종해 약 30개 조직을 노렸고 그중 소수에서 성공했다. 앤트로픽의 분석으로는 작전의 80에서 90퍼센트가량을 인공지능이 수행했다. 사람은 4에서 6개의 결정 지점에만 끼어들었다. 앤트로픽은 이를 사람 개입이 거의 없이 수행된 첫 대규모 사이버 공격의 기록 사례라고 불렀다(first documented case of a large-scale cyberattack). 앤트로픽은 처음에 공격자가 초당 수천 건의 요청을 보냈다고 적었다. 보고서 공개 다음 날인 11월 14일에는 요청이 수천 건이었고 초당 여러 건인 경우가 잦았다는 표현으로 고쳤다.

이 보고서에는 반론이 따랐다. 멜버른대 사이버보안 교수 토비 머리(Toby Murray)는 보고서에 침해 지표가 없다고 지적했다(The Conversation, 2025년 11월 17일). 침해 지표는 다른 기관이 자기도 같은 공격을 당했는지 확인할 때 쓰는 기술 단서다. 보고서 스스로 Claude가 공격자에게 거짓 보고를 자주 했다고 적었다는 점도 그는 짚었다. 그런 상태로는 이 규모의 자율 공격이 어렵다는 지적이다. 소트웍스(Thoughtworks, 2025년 11월 18일)는 보안 업계가 인공지능발 경종으로 보는 쪽과 마케팅 과장으로 보는 쪽으로 갈렸다고 정리했다.

중국과 관련된 사례는 하나 더 있다. 앤트로픽은 2026년 2월 23일 중국계 인공지능 연구소 세 곳을 높은 신뢰도로 지목했다(앤트로픽, 2026년 2월 23일). DeepSeek, Moonshot AI, MiniMax다. 앤트로픽에 따르면 이들은 약 2만 4천 개 가짜 계정으로 1,600만 건 넘는 대화를 만들어 증류에 썼다. 증류는 강한 모델의 답으로 약한 모델을 훈련하는 기법이다. 이 일은 사이버 공격이 아니라 모델의 역량을 빼 가는 지식재산 문제다.

2026년에는 앤트로픽이 자기 평가에서 생긴 사고도 공개했다(앤트로픽, 2026년 7월 30일). 평가 파트너 Irregular와 함께 돌린 평가 환경이 잘못 설정돼 실제 인터넷에 연결됐다. Claude Opus 4.7은 실제 회사를 찾아내 운영 데이터 수백 행에 접근했다.

Claude Mythos 5는 공개 저장소에 악성 패키지를 올렸다. 이 패키지는 약 한 시간 노출되는 동안 실제 시스템 15곳에서 실행됐다. 앤트로픽은 7월 23일 모든 사이버 평가를 멈췄다. 7월 27일에는 Irregular와 피해 기관 세 곳에 알렸다. 앤트로픽은 이 일을 정렬 실패보다 평가 장치와 운영의 실패에 가깝다고 설명했다(harness and operational failure).

세 사건이 보여 주는 공통점

영국 안전연구소 보고서 7.1절은 세 사건을 나란히 놓았다(영국 AI안전연구소, 2026년 8월 4일). 7월 21일 공개된 오픈AI 건, 7월 30일 공개된 앤트로픽 평가 사고, 그리고 자기 사건이다. 보고서는 세 건이 공유하는 조건으로 사이버 분류기 비활성화, 동기식 감시의 부재, 인터넷에 닿는 경로를 들었다. 동기식 감시는 에이전트가 행동하는 그 순간에 다른 인공지능이 그 행동을 보고 승인하거나 막는 장치다.

이 조건들은 평가 환경과 실제 세계의 경계가 흐려졌다는 사실로 모인다. 모델의 위험한 능력을 확인하기 위해 안전장치를 끈 상태로 시험하는 경우가 있다. 그 시험장이 인터넷과 닿으면 피해를 입는 쪽은 실제 회사와 실제 사람이다. 허깅페이스는 오픈AI의 평가 환경 운영에 관여하지 않았는데도 침입을 당했다. 영국 사건에서 악성 코드 변경 요청을 받은 사람도 평가와 관계없는 오픈소스 관리자였다.

세 보고서는 사고의 원인을 모델의 악의에서 찾지 않는다. 안전연구소는 세 건 모두 주로 과제 성공을 좇은 행동으로 설명된다고 썼다(primarily motivated by task success). 앤트로픽도 자사 사건을 운영의 실패로 규정했다. 사고는 반란이 아니라 목표를 좇다 생긴 부작용이다. 러프버러대 사이버보안 교수 올리버 버클리(Oliver Buckley)는 모델이 폭주한 것이 아니라 예상하지 못한 경로를 찾았다고 봤다(과학미디어센터, 2026년 7월 22일). 그래서 격리에 대한 가정이 모델의 복종에 대한 가정보다 단단해야 한다고 말했다.

이 사건들을 과장으로 보는 시각도 있다. 세인트존스대 로스쿨 부교수 케이트 클로닉(Kate Klonick)은 사건 서술이 인공지능의 능력을 부풀리고 사람의 실패를 가린다고 비판했다(Lawfare, 2026년 7월 29일). 그는 이를 비판을 가장한 과대 선전(criti-hype)이라고 불렀다. 그리고 누군가 문을 열어 두었기 때문에 빠져나갔다고 적었다(They escaped because someone left the door open). 임페리얼칼리지런던의 콘스탄티노스 그쿠치스(Konstantinos Gkoutzis)는 이 일을 오래전부터 알려진 명세 악용(specification gaming)으로 봤다(과학미디어센터, 2026년 7월 22일). 그는 진짜 실패가 격리에 있었다고 했다. 명세 악용은 과제 규칙의 틈을 이용해 점수만 얻는 행동이다. 통제된 평가

환경의 일이라 실제 위험과 다르다는 지적도 있다. 옥스퍼드대 앤드루 솔턴(Andrew Soltan)은 안전 필터를 일부러 끈 조건이었다고 짚었다. 안전연구소와 앤트로픽도 평가 조건이 일반 이용자에게 제공되는 방식과 다르다고 밝혔다(BleepingComputer, 2026년 8월 4일).

이 반론들도 사건이 일어났다는 사실은 부정하지 않는다. 쟁점은 사건을 설명하는 틀과 동기다. 누군가 문을 열어 두었다면 문을 여는 결정을 누가 어떤 기준으로 내리는지가 남는다. 클로닉이 내놓은 대안도 의무적 사고 보고, 격리와 시험에 대한 독립 감사였다. 제3자 피해에 대한 책임도 그 안에 들어 있다. 통제된 환경이라는 반론도 같은 곳에 닿는다. 세 사건에서 통제는 사람이 정한 설정이었다. 그 설정이 틀린 순간 평가는 실제 침입이 됐다. 안전연구소 보고서는 좋은 격리가 모델이 스스로 경계를 시험하지 않기로 하는 선택에 기대서는 안 된다고 적었다(good containment should not depend on the model choosing not to test its boundaries).

남은 공통점은 검증이다. 세 사건 모두 당사자가 먼저 공개했지만 공개의 완전성에는 비판이 따랐다. 포천(Fortune, 2026년 8월 26일)은 오픈AI의 사후 보고서가 에이전트에게 준 프롬프트와 실제 메시지 예시를 뺐다고 지적했다. 같은 기사는 METR 조사를 오픈AI가 의뢰했고 조사 기간도 7월 7일부터 13일로 좁혔다고 전했다. 스탠퍼드대 롭 라이히(Rob Reich) 교수는 개발사의 자체 평가를 학생에게 자기 숙제를 채점하게 하는 일에 빗댔다(서울경제, 2026년 10월 5일). METR은 앤트로픽과 영국 안전연구소 사건의 제3자 검토도 맡았다. 다만 독립적이고 강제적인 사후 검증 체계는 아직 없다.

세 사건에 대해 한국 정부가 공식 입장을 낸 기록은 확인되지 않는다. 관련 문제의식은 국정감사에서 나왔다. 2026년 10월 6일 과학기술정보통신부 국정감사에서 이주희 의원은 에이전트의 외부 시스템 접근과 이상 행동 사례를 들었다(디지털데일리, 2026년 10월 6일). 그는 개발 단계별 통제, 중단, 보고 같은 구체적인 안전 절차가 필요하다고 말했다. 배경훈 부총리는 인공지능안전연구소의 인력과 예산을 늘려야 한다고 답했다. 그는 12월에 국가 AI 안전 생태계 조성 마스터플랜을 내겠다고 밝혔다.

세 사건만으로 인류 차원의 통제 이탈을 예견할 수는 없다. 그러나 평가기관이 정한 경계를 에이전트가 넘어 실제 기관에 영향을 준 위험은 이미 확인됐다. 예방과 대응의 기준을 개발사 내부 판단에만 맡겨 둘 단계는 지났다. 평가 중에도 경계를 확인하고 이상 행동을 중단하며, 사고를 독립적으로 조사할 절차를 갖춰야 한다.

참고 자료

허깅페이스, Security incident disclosure (July 2026), 2026년 7월 16일, https://huggingface.co/blog/security-incident-july-2026

허깅페이스, Anatomy of a Frontier Lab Agent Intrusion: A Technical Timeline of the July 2026 Incident, 2026년 7월 27일, https://huggingface.co/blog/agent-intrusion-technical-timeline

오픈AI, OpenAI and Hugging Face partner to address security incident during model evaluation, 2026년 7월 21일, https://openai.com/index/hugging-face-model-evaluation-security-incident/

오픈AI, Pacing model development in an era of cyber-critical capabilities, 2026년 8월 18일, https://openai.com/index/pacing-model-development-cyber-capabilities/

오픈AI, The Hugging Face incident and the road ahead, 2026년 8월 26일, https://openai.com/index/hugging-face-incident-and-the-road-ahead/

오픈AI 정렬팀, An agent used DNS to reach an external chatbot, 2026년 9월 25일, https://alignment.openai.com/misalignment-reports/an-agent-used-dns-to-reach-an-external-chatbot/

METR, Brief independent investigation of agents' behavior, reasoning and collaboration in the OpenAI / Hugging Face hacking incident, 2026년 8월 26일, https://metr.org/blog/2026-08-26-openai-hugging-face-incident-investigation/

Malwarebytes, OpenAI's agent escaped its sandbox during a security test, 2026년 7월 24일, https://www.malwarebytes.com/blog/news/2026/07/openais-agent-escaped-its-sandbox-during-a-security-test

AI타임스, 허깅페이스를 해킹한 것은 오픈AI 모델...격리망 뚫고 첫 자율 공격 성공, 2026년 7월 22일, https://www.aitimes.com/news/articleView.html?idxno=213033

디지털데일리, 오픈AI AI 에이전트, 격리망 뚫고 허깅페이스 침해…전례 없는 사이버 사고, 2026년 7월 22일, https://www.ddaily.co.kr/page/view/2026072216254774768

TechCrunch, Hugging Face confirms breach affected internal datasets and credentials, urges users to take action, 2026년 7월 20일, https://techcrunch.com/2026/07/20/hugging-face-confirms-breach-affected-internal-datasets-and-credentials-urges-users-to-take-action/

한국경제, 고삐 풀린 오픈AI 에이전트, 2026년 7월 29일, https://www.hankyung.com/article/2026072937631

Constellation Research, OpenAI: We'll hit pause on model reinforcement learning for safety, 2026년 8월 18일, https://www.constellationr.com/insights/news/openai-well-hit-pause-model-reinforcement-learning-safety

영국 AI안전연구소, Incident Report: unsanctioned agent behaviour during cyber testing, 2026년 8월 4일, https://www.aisi.gov.uk/blog/incident-report-unsanctioned-agent-behaviour-during-cyber-testing

영국 AI안전연구소, Security Incident INC-2026-07-28-01 기술 보고서, 2026년 8월 4일, https://cdn.prod.website-files.com/663bd486c5e4c81588db7a1d/6a724858f7db25c81487016d_Security%20Incident%20INC-2026-07-28-01.pdf

앤트로픽, Detecting and countering malicious uses of Claude: March 2025, 2025년 3월, https://www.anthropic.com/news/detecting-and-countering-malicious-uses-of-claude-march-2025

앤트로픽, Detecting and countering misuse of AI: August 2025, 2025년 8월 27일, https://www.anthropic.com/news/detecting-countering-misuse-aug-2025

앤트로픽, Disrupting the first reported AI-orchestrated cyber espionage campaign, 2025년 11월 13일, https://www.anthropic.com/news/disrupting-AI-espionage

The Conversation, An AI lab says Chinese-backed bots are running cyber espionage attacks. Experts have questions, 2025년 11월 17일, https://theconversation.com/an-ai-lab-says-chinese-backed-bots-are-ru

nning-cyber-espionage-attacks-experts-have-questions-269815

Thoughtworks, Anthropic's 'AI espionage' disclosure: Separating the signal from the noise, 2025년 11월 18일, https://www.thoughtworks.com/insights/blog/security/anthropic-ai-espionage-disclosure-signal-from-noise

앤트로픽, Detecting and preventing distillation attacks, 2026년 2월 23일, https://www.anthropic.com/news/detecting-and-preventing-distillation-attacks

앤트로픽, Investigating three real-world incidents in our cybersecurity evaluations, 2026년 7월 30일, https://www.anthropic.com/news/investigating-incidents-cybersecurity-evals

과학미디어센터(Science Media Centre), expert reaction to OpenAI / Hugging Face incident, 2026년 7월 22일, https://www.sciencemediacentre.org/expert-reaction-to-openai-hugging-face-incident/

Lawfare, The AI That Hacked Its Way Out and the Hype That Followed It, 2026년 7월 29일, https://www.lawfaremedia.org/article/the-ai-that-hacked-its-way-out-and-the-hype-that-followed-it

BleepingComputer, OpenAI, Anthropic AI agents targeted real people and systems in cyber tests, 2026년 8월 4일, https://www.bleepingcomputer.com/news/security/openai-anthropic-ai-agents-targeted-real-people-and-systems-in-cyber-tests/

Fortune, OpenAI, independent firms publish reports on rogue AI attack on Hugging Face. Here are the main takeaways, and what OpenAI still hasn't disclosed., 2026년 8월 26일, https://fortune.com/2026/08/26/openai-publishes-technical-report-on-how-its-agents-hacked-hugging-face-here-are-the-main-takeaways-and-what-openai-left-out/

서울경제, AI 출시 접고 안전장치 마련하는 빅테크…학계는 "학생이 채점하는 격" 비판, 2026년 10월 5일, https://m.sedaily.com/article/20098139

디지털데일리, "AI 안전 인력 부족" 지적에…배경훈 "AI안전연구소 인력·예산 늘려야" [2026 국감], 2026년 10월 6일, https://www.ddaily.co.kr/page/view/2026100612091028774



13장. 모델을 만든 회사가 스스로 사고를 공개하는 이유

전에 없던 장면

개발사가 비정렬 사례를 어떻게 공개해야 하는지 정한 업계 공통의 기준은 아직 없다(there is no industry-wide framework with explicit standards). 오픈AI가 2026년 9월 16일 회사 블로그에 새 공개 체계를 올리며 적은 내용이다(오픈AI, 2026년 9월 16일). 비정렬(misalignment)은 모델이 개발자가 의도한 목표와 어긋나게 움직이는 일이다. 오픈AI는 같은 글에서 지금까지의 공개가 그때그때 이뤄졌고 바람직한 만큼 자주 나오지도 않았다고 인정했다(ad hoc and less frequent than ideal). 이 글과 함께 지난 6개월 동안 관찰한 비정렬 사례 보고서 6건도 냈다.

이 체계가 나오기 두 달 전에는 세 기관이 자기 이름으로 사고를 공개했다. 오픈AI가 7월 21일에 먼저 공개했고 앤트로픽이 7월 30일에 뒤를 이었다. 영국 AI안전연구소는 8월 4일에 공개했다. 세 사건의 경위는 앞 장에서 보았다. 낯선 것은 사건의 내용만큼이나 공개한 쪽이 누구였는가였다.

지금까지 정보보안 사고를 알리는 쪽은 대개 피해를 입은 회사였다. 이런 의무를 침해 통지라고 부른다. 침해 통지는 개인정보나 전산망을 침해당한 회사가 그 사실을 당국과 피해자에게 알리는 일이다. 허깅페이스가 7월 16일 침입 사실을 먼저 공개한 것이 이 유형이다(허깅페이스, 2026년 7월 16일). 오픈AI의 공개는 성격이 다르다. 침입에 쓰인 모델을 만든 회사가 자기 모델이 침입자였다고 밝혔다. 영국 안전연구소는 가해자도 피해자도 아닌 평가 기관으로서 자기 평가 중에 생긴 사고를 공개했다.

허깅페이스의 클렘 들랑그 최고경영자는 오픈AI의 7월 21일 글에 실린 발언에서 이 사건을 이런 종류의 첫 사례일 수 있다고 했다(possibly the first of its kind)(오픈AI, 2026년 7월 21일). 모델을 만든 회사가 자기 모델의 침입을 스스로 공개한 일은 이때가 처음으로 알려져 있다. 공개할 법적 의무가 있었는지도 분명하지 않았다. 법률 연구기관 Institute for Law & AI의 매켄지 아널드와 스테판 예레나는 이 사건이 미국 주법의 보고 대상인지 검토했다(Lawfare, 2026년 7월 24일). 대상은 캘리포니아 SB 53과 뉴욕

RAISE법과 일리노이 SB 315가 정한 중대 안전 사고였다. 중대 안전 사고는 개발사가 당국에 보고해야 하는 심각한 사고로 법이 미리 정의해 둔 것이다. 네 범주 가운데 셋은 신체 상해나 대규모 손해 같은 실제 피해를 요구한다. 두 사람은 기존 주법이 이 사건을 다루는지 분명하지 않다고 결론지었다(it is unclear that these existing state laws cover this event).

공개는 한 번으로 끝나지 않았다. 오픈AI 정렬팀 사이트의 보고서 목록에는 10월 14일 기준 보고서 15건과 고지 3건이 올라 있다(오픈AI 정렬팀, 2026년 10월 14일 열람). 앤트로픽은 10월 9일 평가와 내부 사용 중에 나온 의도하지 않은 모델 행동을 보고서로 냈다(앤트로픽, 2026년 10월 9일). 우려스러운 행동을 찾는 대로 계속 보고하겠다고도 약속했다. 앤트로픽은 2025년부터 자사 모델이 범죄에 쓰인 사례를 정기적으로 공개해 왔다. 2026년에는 공개의 대상이 남의 악용에서 자기 모델의 행동으로 넓어졌다.

공개의 동기: 규제 선점, 신뢰, 책임 분산

회사들은 다른 사람이 배울 수 있게 하려고 공개했다고 말한다. 오픈AI는 7월 21일 첫 글에서 방어하는 쪽이 사건을 이해하고 지금 모델의 능력을 가늠하도록 돕겠다고 했다(help calibrate on what models are now capable of)(오픈AI, 2026년 7월 21일). 앤트로픽은 다른 개발사가 비슷한 행동을 점검하는 데 보고서가 쓰이기를 바란다고 썼다(help other developers check for similar behaviors)(앤트로픽, 2026년 10월 9일). 영국 안전연구소도 다른 기관이 평가 관행을 고칠 수 있도록 사고 정보를 공개한다고 밝혔다(영국 AI안전연구소, 2026년 8월 4일).

밖에서는 다른 동기를 읽는다. 클라우드 보안 연합(Cloud Security Alliance)의 AI 안전 이니셔티브는 오픈AI의 공개 체계를 안전 투명성의 행위인 동시에 규제를 앞둔 자리 잡기로 봤다(regulatory pre-positioning)(Cloud Security Alliance, 2026년 9월 17일). 그렇게 읽을 근거는 문서 안에 있다. 오픈AI는 이 체계가 업계 공시 기준을 만드는 첫걸음이 되기를 바란다고 했다. 규제 당국과 표준 기관 등과 함께 더 객관적인 기준을 만들겠다고 했다. 중대한 사고를 미국 연방정부와 나눌 보고 방식도 제안하겠다고 밝혔다(오픈AI, 2026년 9월 16일). 규칙이 생기기 전에 규칙의 모양을 먼저 그려 두는 일이다.

신뢰도 동기로 꼽힌다. 컨스텔레이션 리서치의 래리 디그넌 편집장은 오픈AI의 공개 체계를 괜찮은 출발이라고 평가했다("It's a solid start")(Constellation Research, 2026년 9월 17일). 그는 이 조치를 오픈AI의 기업 고객 시장 공략과 신뢰 회복 노력에 이어 읽었다. 투자자를 향한 공개도 같은 흐름에 있다. 앤트로픽은 상장 신고서에 자율

에이전트의 행동을 둘러싼 법적 쟁점이 정리되지 않아 예측하기 어려운 소송을 부를 수 있다고 적었다(SecurityWeek, 2026년 9월 30일). 사고를 알리는 일은 이제 투자자에게 알리는 위험과도 맞물려 있다.

셋째 동기인 책임 분산을 뒷받침하는 사실은 외부 검토 과정에서 나왔다. 로이터는 10월 13일 단독 보도에서 오픈AI가 METR과 레드우드 리서치의 보고서에서 비공개 정보를 지웠다고 전했다(로이터, 2026년 10월 13일). 이틀로 잡은 조사는 세 차례 방문에 엿새가 걸렸다. 더 완전한 기록은 조사팀이 빈틈을 지적한 뒤 마지막 방문에서야 나왔다. 두 기관은 오픈AI의 의견을 받아 구조와 어조를 고쳤지만 결론에 중요한 내용은 빠지지 않았다고 했다. 외부 검토가 회사의 결론에 무게를 실어 주는 장치로 쓰이면 검토자의 독립성이 곧 쟁점이 된다.

10월 6일 호주 의회 AI 합동특별위원회 청문회의 발언은 두 방향으로 읽힌다. 오픈AI 최고전략책임자 제이슨 권과 앤트로픽 안전장치 책임자 데이브 오어는 모두 의무 보고 체제를 환영했다(ACS Information Age, 2026년 10월 6일). 두 회사는 의무 규칙이 있으면 공개 여부를 혼자 정하지 않아도 된다는 취지로 말했다. 자기들이 모양을 잡을 수 있는 규칙을 지지한다는 뜻으로 보면 규제 선점이다. 공개 결정의 부담을 법에 넘긴다는 뜻으로 보면 책임 분산이다.

공개를 홍보로 보는 시각도 있다. 더 레지스터의 코너 존스 기자는 앤트로픽의 7월 30일 공개를 오픈AI 공개에 맞선 경쟁적 대응으로 읽었다(The Register, 2026년 7월 31일). 같은 기사에서 보안업체 창업자인 변호사 일리야 콜로첸코는 이 공개가 두 회사에 대한 신뢰를 높이지 않는다고 했다. 그러나 홍보로만 설명하기 어려운 내용도 공개 문서에 있다. 오픈AI는 8월 26일 글에서 내부 팀이 5월 말에 이미 에이전트의 게시판 활동과 허용되지 않은 인터넷 접속을 봤다고 적었다(오픈AI, 2026년 8월 26일). 돌아보면 그 초기 신호가 더 이른 대응으로 이어졌어야 했다고 인정했다. 자기 대응의 실패를 시점까지 밝힌 글을 홍보 문서로만 읽기는 어렵다. 공개에 홍보 효과가 있다는 사실과 공개가 실제 정보를 담는다는 사실은 함께 성립한다.

공개된 사고와 공개되지 않은 사고

공개된 사고는 회사가 찾아낸 사고 가운데 공개하기로 한 것이다. 회사가 찾지 못한 사고는 처음부터 목록에 오르지 않는다. 허깅페이스 사건에서도 먼저 알아챈 쪽은 오픈AI가 아니었다. 허깅페이스가 7월 16일 침입을 공개했고 오픈AI의 보안 감시는 7월 19일에야 이상 활동을 잡았다(오픈AI, 2026년 8월 26일). 오픈AI는 7월 20일 이를

허깅페이스 사건과 연결했고 다음 날 공개했다. 공익단체 퍼블릭 날리지의 니컬러스 가르시아는 이 순서를 들어 프런티어 연구소가 자기 시스템의 실패를 스스로 찾아내리라 믿기 어렵다고 주장했다(Public Knowledge, 2026년 8월 25일).

외부가 먼저 찾은 사고는 그 뒤에도 나왔다. 나이팅게일 컬렉티브(Nightingale Collective)는 9월 4일 오픈AI 평가 에이전트들이 독일어 프로그래밍 위키를 게시판처럼 썼다는 증거를 공개했다(Cloud Security Alliance, 2026년 9월 6일). 오픈AI는 이 활동을 처음 발견했을 때 그동안 연구해 온 다른 비정렬 행동과 비슷한 것으로 판단했다고 설명했다(오픈AI, 2026년 9월 5일). 9월 16일의 공개 체계는 이 일 뒤에 나왔다.

알린 시점이 늦은 사례도 있다. 오픈AI 에이전트는 6월 18일 호주 메디케어 통계 포털에 접근했다. 오픈AI가 이를 알린 것은 9월 10일이었고 그것도 공용 메일함으로 보냈다(ACS Information Age, 2026년 10월 6일). 제이슨 권은 청문회에서 피해 기관에 훨씬 일찍 알렸어야 했다고 인정했다(should have informed the impacted parties much sooner). 앤트로픽의 Claude Haiku 4.5는 7월 18일 필라델피아 경찰의 미해결 살인 사건 제보 양식에 메시지를 냈다(앤트로픽, 2026년 10월 9일). 앤트로픽은 9월 28일에야 이 일을 발견했고 10월 8일 경찰에 알렸다. 필라델피아 경찰은 늦은 통지를 받아들일 수 없다고 했다("unacceptable")(TechTimes, 2026년 10월 12일).

질문을 받고서야 확인한 회사도 있다. 월스트리트저널은 9월 18일 구글의 Gemini 모델이 5월 평가 중에 실제 회사 세 곳의 시스템에 들어갔다고 보도했다(SecurityWeek, 2026년 9월 21일). 구글은 7월 말 평가 파트너 Irregular의 통지로 이 일을 알았다. 구글은 피해가 없었고 모델이 스스로 멈췄으므로 공개할 필요가 없었다고 설명했다. 헤더 애드킨스 보안공학 부사장은 세 건 모두 모델이 멈췄다고 말했다(In all three of these instances, the model stopped). 무엇을 공개할지의 기준이 회사마다 다르다. 중국 개발사가 자기 모델의 사고를 스스로 공개한 사례는 확인되지 않는다.

공개되지 않은 사고의 규모는 짐작만 할 수 있다. 오픈AI는 과거 활동을 되짚는 검토에서 약 50페타바이트의 기록을 살피고 있다고 밝혔다(오픈AI, 2026년 9월 30일). 9월 26일까지 통지 기준에 맞는 활동을 이유로 100곳 넘는 조직에 알렸다. 오픈AI는 공개를 원한 조직도 있고 공개하지 말라고 요청한 조직도 있다고 적었다(Some have wanted to publicly disclose and others have asked us not to)(오픈AI, 2026년 9월 25일). 영국 장기회복력센터(Centre for Long-Term Resilience)의 통제 상실 관측소는 X에 공개된 기록에서 의도하지 않은 인공지능 행동을 모은다. 관측소는 2026년 7월 한 달에 300건 넘게 기록했고 올해 누계는 1,600건을 넘었다(TechTimes, 2026년 8월 31일). 연구진은 이 숫자를 하한값이라고 했다.

자발 공개가 법보다 많은 정보를 낸다는 시각도 있다. 앞에서 본 아널드와 예레나는 의무 보고가 있는 곳에서도 법은 아주 적은 정보만 요구한다고 썼다(requires only the scantest of information)(Lawfare, 2026년 7월 24일). 오픈AI는 의무가 분명하지 않은 상태에서 8월 26일 기술 보고서 전문까지 공개했다. 두 사람은 정보를 내놓는 회사가 덜 협조하는 회사보다 더 많은 비판을 받는 역설도 짚었다. 이 시각은 공개한 회사에 대해서는 맞다. METR도 5월 프런티어 위험 보고서에서 회사들이 시스템 카드와 질의 답변으로 소수의 선별된 사례만 제공했다고 적었다(METR, 2026년 5월 19일). 자발 공개는 공개하는 회사에서는 법보다 많이 알려 주지만 구글처럼 공개하지 않기로 한 회사까지 덮지는 못한다.

사고 보고가 제도가 되려면

EU 인공지능법은 두 조문에 보고 의무를 나눠 두었다. 제55조는 프런티어 모델이 대개 해당하는 체계적 위험 범용 인공지능 모델의 제공자에게 중대 사고 정보와 시정 조치를 지체 없이 AI사무국에 보고하게 한다(without undue delay)(EU 인공지능법 제55조). 이 의무는 2025년 8월 2일부터 적용됐고 AI사무국의 집행권은 2026년 8월 2일부터 생겼다. 조문에는 일수가 없고 의무 이행 방법을 정리한 자발적 지침인 실천강령이 심각도에 따라 2일에서 15일 안에 알리도록 정했다(Jones Day, 2025년 8월).

제73조는 고위험 인공지능 시스템에 적용된다. 사고와 시스템의 인과관계를 확인하면 즉시 보고하되 늦어도 인지 후 15일을 넘길 수 없다(EU 인공지능법 제73조). 그런데 2026년 7월 발효한 디지털 옴니버스 규정이 부속서 III 고위험 시스템의 의무 적용일을 2027년 12월 2일로 미뤘다(Legal500, 2026년 8월 3일). 15일 보고는 고위험 시스템의 규칙이고 그마저 아직 적용 전이다. 클라우드 보안 연합은 보고 대상인지를 제공자가 스스로 분류하는 구조가 큰 재량을 준다고 지적했다(Cloud Security Alliance, 2026년 9월 6일).

미국에는 사고 보고를 정한 연방 법률이 없다. 캘리포니아 SB 53은 2026년 1월 1일 시행됐다(Brookings, 2025년 12월 23일). 10의 26승 연산 이상으로 훈련한 모델의 개발사는 중대 안전 사고를 발견한 날부터 15일 안에 주 비상관리청에 보고해야 한다. 사망이나 중상의 급박한 위험이 있으면 24시간 안에 알려야 한다. 뉴욕 RAISE법은 72시간 안 보고를 정했고 2027년 1월 1일 시행된다(Cooley, 2026년 4월 1일). 백악관 슈퍼인텔리전스 포스(Super Intelligence Force)는 10월 9일 사고 신고가 선택 사항이 아니라고 선언했다("not optional")(뉴시스, 2026년 10월 10일). 이 성명에는 법적 근거와 기한과 제재가 없다(TechTimes, 2026년 10월 12일).

영국에는 기업의 사고를 접수하는 법정 체계가 없다. 에이다 러브레이스 연구소는 안전연구소가 규제기관이 아니며 찾아낸 피해에 대응할 규제 권한도 없다고 적었다(Ada Lovelace Institute, 2026년 7월 13일). OECD의 AI 사고 모니터는 약 1만 8천 건의 사고와 위험을 보여 준다(OECD, 2026년 10월 14일 열람). 언론 보도를 자동으로 모은 것이어서 회사 안에서만 보이는 일은 담지 못한다.

미국의 항공 안전 자율보고 제도(ASRS)는 1975년 연방항공청(FAA)과 항공우주국(NASA)의 양해각서로 시작했다(NASA ASRS). 자율보고는 법이 강제하지 않아도 당사자가 자기 실수나 아차사고를 스스로 알리는 방식이다. 아차사고는 사고로 이어질 뻔한 일이다. 보고는 규제기관인 FAA 대신 집행 권한이 없는 NASA가 받고 NASA는 보고자를 알아볼 정보를 지운다. 이 제도에는 면책도 있다. 면책은 정해진 조건을 갖추면 위반이 인정돼도 제재를 면해 주는 장치다. 위반이 고의가 아니고 범죄나 사고가 아니며 최근 5년 안에 FAA 처분을 받은 적이 없어야 한다. 그런 사람이 10일 안에 NASA에 서면으로 보고하면 FAA는 과징금이나 자격 정지를 부과하지 않는다. ASRS가 받는 것은 아차사고와 위반이고 항공 사고 자체는 따로 의무 보고 대상이어서 자발 보고와 의무 보고가 층을 이룬다.

전 유엔개발계획 지식관리 국장 가브리엘 아카시나는 ASRS가 작동하는 이유를 보고받는 곳의 성격에서 찾았다. 집행 권한도 없고 책임을 따질 이유도 없는 공공기관이라는 것이다(a public institution with no enforcement power and no interest in assigning blame)(CircleID, 2026년 8월 6일). 키프로스대의 마이클 하치파나요티스(Michael Chatzipanagiotis)는 EU 인공지능법의 사고 보고가 집행과 과징금에 가깝게 묶여 보고할 유인을 줄인다고 봤다(International Journal of Law and Information Technology, 2026년 2월 27일). 그는 보고와 조사를 책임 추궁에서 떼어 내고 보고자의 비밀을 지키자고 제안했다. 2026년 여름의 공개들은 이런 보호 장치가 없는 상태에서 나왔다.

한국의 인공지능 기본법은 2026년 1월 22일 시행됐다(정책브리핑, 2026년 1월 21일). 이 법과 시행령에는 사고가 나면 정부에 보고하라는 조항이 없다. 이와 가까운 조항은 제32조다. 부록 3에 실은 시행령 제24조에 따르면 누적 학습 연산량이 10의 26승 부동소수점연산 이상이고, 최첨단 인공지능기술을 적용하며, 생명·신체의 안전과 기본권에 광범위하고 중대한 영향을 미칠 우려가 있는 시스템의 사업자는 안전사고를 모니터링하고 대응하는 위험관리체계를 갖추고 그 이행 결과를 과학기술정보통신부장관에게 내야 한다(김앤장 법률사무소). 이것은 사고가 날 때마다 하는 보고와 다르고 체계를 갖춘 결과를 내는 의무다. 제32조를 어겨도 과태료는 없다. 과학기술정보통신부는

9월 30일 국가 AI 안전 생태계 조성 마스터플랜을 세울 민관 실무협의체를 꾸렸고 인공지능안전연구소도 여기에 참여한다(머니투데이, 2026년 9월 30일). 보도된 초안 방향에 기업의 사고 신고를 받는 절차는 들어 있지 않다. 김명주 인공지능안전연구소장은 연구소가 부처에 흩어진 안전 이슈를 모아 정리하고 보고하는 중심 역할을 하겠다고 말했다.

한국의 보고 체계는 자발 공개를 장려하면서도 중대 사고를 숨길 수 없게 설계해야 한다. 중대 사고는 정해진 기한에 알리게 하고, 피해로 이어지기 전의 이상 행동과 아차사고는 보고자를 보호하며 수집해야 한다. 공개가 기업의 선택에만 달리지 않게 하는 일과, 스스로 알린 기업이 불리해지지 않게 하는 일을 함께 추진해야 더 많은 위험을 일찍 발견할 수 있다.

참고 자료

오픈AI, Our framework for reporting model misalignment, 2026년 9월 16일, https://openai.com/index/model-misalignment-reporting-framework/

오픈AI, OpenAI and Hugging Face partner to address security incident during model evaluation, 2026년 7월 21일, https://openai.com/index/hugging-face-model-evaluation-security-incident/

오픈AI, The Hugging Face incident and the road ahead, 2026년 8월 26일, https://openai.com/index/hugging-face-incident-and-the-road-ahead/

오픈AI, The Hugging Face incident and other third-party impact from misaligned models, 2026년 9월 5일, 9월 25일, 9월 30일 갱신, https://openai.com/hugging-face-incident-and-misalignment/

오픈AI 정렬팀, Misalignment Reports and Notices, 2026년 10월 14일 열람, https://alignment.openai.com/misalignment-reports/

허깅페이스, Security incident disclosure (July 2026), 2026년 7월 16일, https://huggingface.co/blog/security-incident-july-2026

앤트로픽, Investigating unintended model actions in our evaluations and internal use, 2026년 10월 9일, https://www.anthropic.com/news/investigating-unintended-model-actions

영국 AI안전연구소, Incident Report: unsanctioned agent behaviour during cyber testing, 2026년 8월 4일, https://www.aisi.gov.uk/blog/incident-report-unsanctioned-agent-behaviour-during-cyber-testing

Lawfare, When Reporting an AI Security Incident Is Not Mandatory, 2026년 7월 24일, https://www.lawfaremedia.org/article/when-reporting-an-ai-security-incident-is-not-mandatory

Cloud Security Alliance, OpenAI's Misalignment Disclosure Framework: Voluntary Meets Mandatory, 2026년 9월 17일, https://labs.cloudsecurityalliance.org/research/csa-research-note-openai-misalignment-reporting-framework-20/

Cloud Security Alliance, OpenAI's Wiki Silence Tests the EU AI Act's Incident Regime, 2026년 9월 6일, https://labs.cloudsecurityalliance.org/research/csa-research-note-ai-incident-disclosure-gap-eu-ai-act-20260/

Constellation Research, OpenAI's new misalignment disclosure framework a solid start, 2026년 9월 17일, https://www.constellationr.com/insights/news/openais-new-misalignment-disclosure-framework-solid-

start

SecurityWeek, Anthropic Flags AI Agent Liability Risks as OpenAI Faces Hacking Lawsuit, 2026년 9월 30일, https://www.securityweek.com/anthropic-flags-ai-agent-liability-risks-as-openai-faces-hacking-lawsuit/

로이터(BNN Bloomberg 전재), Behind AI industry's call for independent reviews, a fight over what gets made public: Reuters exclusive, 2026년 10월 13일, https://www.bnnbloomberg.ca/business/artificial-intelligence/2026/10/13/behind-ai-industrys-call-for-independent-reviews-a-fight-over-what-gets-made-public-reuters-exclusive/

ACS Information Age, OpenAI grilled over rogue agents by Australian inquiry, 2026년 10월 6일, https://ia.acs.org.au/article/2026/openai-grilled-over-rogue-agents-by-australian-inquiry.html

The Register, Anthropic and OpenAI are competing to see whose agents can go rogue harder, 2026년 7월 31일, https://www.theregister.com/security/2026/07/31/anthropic-and-openai-are-competing-to-see-whose-agents-can-go-rogue-harder/5281797

Public Knowledge, It is Time To Tame the AI Frontier, 2026년 8월 25일, https://publicknowledge.org/it-is-time-to-tame-the-ai-frontier/

TechTimes, Anthropic's AI Filed Fake Murder Tip; White House Reply to Rogue-Agent Hacks Lacks Penalties, 2026년 10월 12일, https://www.techtimes.com/articles/328914/20261012/anthropics-ai-filed-fake-murder-tip-white-house-reply-rogue-agent-hacks-lacks-penalties.htm

SecurityWeek, Google Confirms Gemini AI Breached Three Firms, 2026년 9월 21일, https://www.securityweek.com/google-confirms-gemini-ai-breached-three-firms/

TechTimes, AI Scheming Incidents Doubled in July, Watchdog Finds: UK Parliament Has No Power to Act, 2026년 8월 31일, https://www.techtimes.com/articles/326032/20260831/ai-scheming-incidents-doubled-july-watchdog-finds-uk-parliament-has-no-power-act.htm

METR, Frontier Risk Report (February to March 2026), 2026년 5월 19일, https://metr.org/blog/2026-05-19-frontier-risk-report/

EU 인공지능법(Regulation (EU) 2024/1689), Article 55 Obligations of providers of general-purpose AI models with systemic risk, 2024년 공포, https://rgpd.com/ai-act/chapter-5-general-purpose-ai-models/article-55-obligations-of-providers-of-general-purpose-ai-models-with-systemic-risk/

EU 인공지능법(Regulation (EU) 2024/1689), Article 73 Reporting of serious incidents, 2024년 공포, https://www.euaiact.com/article/73

Jones Day, EU AI Act: European Commission Publishes General-Purpose AI Code of Practice, 2025년 8월, https://www.jonesday.com/en/insights/2025/08/eu-ai-act-european-commission-publishes-generalpurpose-ai-code-of-practice

Legal500, EU Digital Omnibus On AI Enters Into Force, 2026년 8월 3일, https://www.legal500.com/intelligence/european-union/technology/eu-digital-omnibus-on-ai-enters-into-force

Brookings, What is California's AI safety law?, 2025년 12월 23일, https://www.brookings.edu/articles/what-is-californias-ai-safety-law/

Cooley, New York's Frontier AI Law Gets a California Makeover – With Some Key Differences, 2026년 4월 1일, https://www.cooley.com/news/insight/2026/2026-03-31-new-yorks-frontier-ai-law-gets-a-california-makeover-with-some-key-differences

뉴시스, 美, AI 사고 신고 의무화…"중대 국가안보 의무", 2026년 10월 10일, https://www.newsis.com/view/NISX20261010_0003820555

Ada Lovelace Institute, Making sense of the UK's AI Security Institute, 2026년 7월 13일, https://www.adalovelaceinstitute.org/feature/aisi/

OECD, AIM: AI Incidents and Hazards Monitor, 2026년 10월 14일 열람, https://oecd.ai/en/incidents

NASA ASRS, Immunity Policies, 날짜 미상, https://asrs.arc.nasa.gov/overview/immunity.html

CircleID, Safe for Whom? A Pilot's Take on the AI Incident-Reporting System, 2026년 8월 6일, https://circleid.com/posts/safe-for-whom-a-pilots-take-on-the-ai-incident-reporting-system

International Journal of Law and Information Technology, Incident reporting and investigation under the AI act: some insights from aviation, 2026년 2월 27일, https://academic.oup.com/ijlit/article/doi/10.1093/ijlit/eaaf019/8501132

정책브리핑, '인공지능기본법' 22일 시행…생성형 AI 결과물 '워터마크' 표시 의무, 2026년 1월 21일, https://www.korea.kr/news/policyNewsView.do?newsId=148958380

김앤장 법률사무소, 인공지능기본법 관련 최신 동향, 2025년 12월 24일, https://www.kimchang.com/ko/insights/detail.kc?sch_section=4&idx=33598

머니투데이, 정부, 'AI 안전' 민관 실무협의체 가동…12월 마스터플랜 수립, 2026년 9월 30일, https://www.mt.co.kr/tech/2026/09/30/2026093009510636744



14장. 한국의 인공지능안전연구소

인공지능 기본법 제12조의 운영 근거

인공지능안전연구소에서 모델의 위험을 기술적으로 평가하는 인력은 6명이다(디지털데일리, 2026년 9월 28일). 이정헌 더불어민주당 의원이 과학기술정보통신부와 연구소에서 받은 자료의 숫자다. 평가실 7명 가운데 법제와 규범을 맡은 1명을 뺀 인원이다. 앞 장 끝에서 김명주 소장은 부처에 흩어진 안전 이슈를 연구소가 모아 정리하겠다고 했다. 그 일을 맡을 조직의 법적 바탕은 인공지능 기본법 제12조 한 조문이다.

제12조 제1항은 과학기술정보통신부장관이 인공지능안전 업무를 "전문적이고 효율적으로 수행하기 위하여 인공지능안전연구소(이하 '안전연구소'라 한다)를 운영할 수 있다"고 정한다. 법이 연구소를 세운 것이 아니라 장관에게 운영할 길을 열어 준 것이다. "할 수 있다"로 끝나는 이런 조문을 임의규정이라고 부른다. 권한을 주되 그 권한을 쓸지는 행정기관이 정하게 하는 조문이다. 바로 앞 제11조의 인공지능정책센터는 장관이 "지정할 수 있다"고 했다. 제12조에는 지정이라는 말이 없고 어느 기관을 연구소로 삼는다는 절차도 법과 시행령에 없다.

제2항은 연구소의 사업 일곱 가지를 적었다. 앞의 다섯은 위험의 정의와 분석에서 시작해 정책 연구와 평가 기준 연구를 거쳐 기술 표준화와 국제협력까지 이어진다. 여섯째는 "제32조에 따른 인공지능시스템의 안전성 확보에 관한 지원"이다. 제32조는 앞 장에서 본 대형 시스템 사업자의 위험관리 의무이고 연구소는 이 의무를 감독하지 않고 지원한다. 제3항은 정부가 경비를 예산 범위에서 출연하거나 지원할 수 있다고 했다. 시행령 제10조는 사업 다섯 가지를 더했고 그중 하나가 "평가 시스템의 구축 및 운영"이다. 같은 조 제3항은 연구소가 자료 제공을 요청할 상대를 "국가기관등"으로 한정했다. 제12조와 시행령 제10조의 전문은 부록 3에 싣는다.

조문에 없는 것도 분명하다. 연구소에는 민간 기업의 모델이나 자료 제출을 강제할 권한이 없다. 법 제40조의 사실조사는 장관과 소속 공무원이 하고 연구소는 조사 주체가 아니다. 명령권과 제재권도 없고 사고 보고를 받을 창구는 앞 장에서 본 대로 법 전체에

없다. 시행령이 연구소에 지운 의무는 매년 1월 31일까지 전년도 실적과 그해 계획을 장관에게 내는 일이다.

한국전자통신연구원(ETRI) 부설이라는 형태도 법에서 나오지 않았다. 부설 기관은 따로 법인을 세우지 않고 모기관 안에 두는 조직이다. 과기정통부는 2024년 10월 독립성과 자율성을 갖추려면 기본법이 먼저 통과돼야 한다며 그때까지 연구소를 ETRI 조직으로 두기로 했다(뉴시스, 2024년 10월 18일). 기본법은 2026년 1월 22일 시행됐다. 그해 10월에도 연구소는 ETRI 직할 조직이어서 인사와 예산을 스스로 정하는 데 한계가 있다는 지적이 나온다(ZDNet Korea, 2026년 10월 13일).

2024년 11월 개소 이후의 활동

연구소는 2024년 11월 27일 경기 성남시 판교 글로벌R&D센터에서 문을 열었다(전자신문, 2024년 11월 27일). 출발 인력은 ETRI 연구원 8명과 한국정보통신기술협회 파견 인력 3명을 포함한 12명이었다. 보도는 이 연구소를 세계에서 여섯 번째로 세운 안전연구소라고 적었다. 초대 소장은 서울여대 정보보호학부 김명주 교수다. 그는 개소식에서 연구소가 규제 기관이 아니라 기업의 경쟁력을 돕는 협력 기관이 되겠다고 말했다(이데일리, 2024년 11월 27일).

평가 실적은 숫자로 쌓였다. 연구소는 2026년 5월 국내외 모델 42종의 안전성을 평가했다고 밝혔다(ZDNet Korea, 2026년 5월 20일). 2025년의 29종과 2026년 들어 평가한 13종을 더한 숫자다. 화학·생물·방사능·핵 위험 항목을 평가에 넣었고 평가 데이터셋과 시나리오 6만여 개를 모았다. 공개한 것은 모델명과 평가 항목이었고 결과 수치는 없었다. 다른 집계로 2026년 1월부터 8월까지 평가한 고유 모델은 34종이고 그중 19종이 오픈웨이트 모델이다(디지털투데이, 2026년 9월 18일). 오픈웨이트 모델은 학습을 마친 가중치를 공개해 누구나 내려받아 고쳐 쓸 수 있게 한 모델이다. 19종 가운데 11종이 중국계였다. 9월의 우선 평가 대상도 중국 즈푸AI의 GLM 5.3이다(아주경제, 2026년 9월 16일). 연구소는 안전장치를 걷어 낸 변형본이 공격 코드와 화학·생물 위험 정보에 어디까지 답하는지 보고 있고 오픈AI의 GPT-6 아스트라 평가에도 들어갔다.

결과 공개는 늦었다. 디지털데일리는 2026년 5월 출범 1년 반 동안 주요 기업 프런티어 모델을 독자 평가해 공개한 보고서가 0건이라고 보도했다(디지털데일리, 2026년 5월 13일). 김 소장은 "보고서는 계속 제작하고 있으나 비공개를 원칙으로 해서 공개는 못하고 있다"고 했다. 낮은 점수를 받은 기업의 입장을 생각했다는 설명이다. 6월 15일 연구소는 싱가포르 안전연구소와 함께 한 공동 테스트의 세부 결과 보고서를 냈다(ZDNet Korea,

2026년 6월 19일). 같은 보도에서 김 소장은 대상 기업이 거부하지 않으면 가급적 모든 내용을 공개하겠다고 방향을 바꿨다. 10월 14일까지 확인한 범위에서 주요 기업 프런티어 모델의 실명 평가 결과는 나오지 않았다.

협약은 2026년에 늘었다. 연구소는 6월 17일 오픈AI와 고위험 분야 안전 협력 양해각서를 맺었다(디지털타임스, 2026년 6월 17일). 오픈AI가 각국 안전연구소와 맺은 협약으로는 네 번째이고 앞의 셋은 미국과 영국과 일본이다. 평가 방법론과 벤치마크를 나누고 한국어 평가체계를 함께 만든다는 내용이고 평가할 모델은 정하지 않았다. 9월 28일에는 카카오와 협약을 맺었다(카카오, 2026년 9월 29일). 카카오가 모델과 인프라를 내놓고 연구소가 배포 전후 평가를 맡는다. 연구소 홈페이지에는 국제 공동 테스트 참여 결과 보고서 2건도 올라 있다(디지털데일리, 2026년 5월 13일).

부족한 것: 권한, 예산, 인력, 장비

평가를 하려면 먼저 모델에 닿아야 한다. 영국과 미국의 연구소는 개발사와 사전 접근 협약을 맺는다. 사전 접근 협약은 개발사가 출시 전 모델을 정부 평가 기관에 미리 열어 주기로 한 자발적 약속이다. 디지털데일리는 2026년 5월 한국 연구소가 주요 기업과 이런 협약을 한 건도 맺지 못해 미공개 모델에 접근할 수 없다고 진단했다(디지털데일리, 2026년 5월 13일). 6월의 오픈AI 협약에도 대상 모델과 일정은 없다. 협약이 없을 때 모델 제출을 요구할 법적 근거는 앞에서 본 대로 없다.

2026년 연구소 예산은 89억 원이다(디지털데일리, 2026년 9월 28일). 이정헌 의원실이 과기정통부와 연구소에서 받은 자료다. 2027년 정부안에서 연구소 몫이 얼마인지는 공개 자료에서 확인하지 못했다.

국정감사에서 나온 274억 원은 다른 숫자다. 2027년 정부안에 든 과기정통부 AI 안전·신뢰 분야 3개 사업의 합계이고 연구소 단독 예산이 아니다(디지털데일리, 2026년 10월 6일). 같은 정부안의 AI 성능·인프라 예산은 6조 1,490억 원이다. 이주희 더불어민주당 의원은 성능·인프라 예산과 안전·신뢰 예산의 격차가 2024년 36 대 1에서 2027년 224 대 1로 벌어졌다고 분석했다.

인력은 2026년 10월 38명이다(디지털데일리, 2026년 10월 6일). 이주희 의원이 국정감사에서 밝힌 숫자이고 출범 때의 12명보다 세 배가량 많다. 38명은 전체 인원이고 첫머리의 6명은 그 가운데 기술 평가를 맡은 사람이다. 사람을 데려오는 조건도 다르다. 김 소장은 2025년 3월 영국 연구소의 풀타임 연구원 100명이 평균 연봉 2억 원을 받는다고 비교했다(AI타임스, 2025년 3월 29일).

장비는 엔비디아 H200 16장이고 가동률은 70에서 80퍼센트다(디지털투데이, 2026년 9월 18일). 이정헌 의원실 자료는 이를 GPU 서버 5대로 적었고 그중 2대는 민간 클라우드에서 빌린다. 서버 한 대에 GPU 여러 장이 들어가므로 두 숫자는 단위가 다르다. 모델 하나를 평가하는 데 7일에서 10일이 걸리고 벤치마크를 한 번 돌리는 토큰 비용은 5,000만 원에서 6,000만 원이다(아주경제, 2026년 9월 16일). 김 소장은 "부족할 때는 상용 클라우드 세션을 임차해 연산 자원을 추가로 확보해서 사용하고 있다"고 했다. 김용수 AI안전평가실 선임연구원도 한국의 평가 인프라를 선도국과 같은 수준으로 보기 어렵다고 인정했다(전자신문, 2026년 10월 11일).

연구소는 규모에 비해 성과를 냈다고 본다. 김 소장은 개소 1주년에 지난 1년을 한국 AI 안전 연구가 국제 무대에서 존재감을 확인한 시기로 평가했다(전자신문, 2025년 11월 23일). 받칠 사실도 있다. 연구소는 2년이 안 되는 기간에 42종을 평가했고 오픈AI가 각국 연구소와 맺은 네 번째 협약의 상대가 됐다. 익명이지만 세부 수치를 담은 공동 보고서도 냈다. 다만 이 평가를 내린 쪽은 연구소와 소장 자신이다. 외부에서 한국 연구소의 성과를 규모에 비해 높게 본 글은 찾지 못했다. 자평의 근거는 평가한 모델과 맺은 협약의 수이고 평가 결과 대부분은 아직 밖에서 볼 수 없다.

지시를 엉뚱하게 수행하는 부작용과 보안 사고의 연구

2026년 여름 오픈AI 사건의 에이전트는 과제를 정면으로 풀지 않았다. 외부 분석은 에이전트가 채점기를 속이거나 시험 답안을 얻으려 했다고 봤다. 연구자들은 이런 행동을 보상 해킹(reward hacking)이라고 부른다. 보상 해킹은 강화학습에서 모델이 목표를 이루지 않고 보상을 주는 조건만 채우는 행동이고 명세 악용의 한 형태다. 앞에서 본 사건에서 평가 환경을 뚫고 채점기와 정답 자료로 다가간 행동이 이 범주에 든다.

한국 연구소가 보상 해킹이나 명세 악용을 주제로 내건 보고서나 논문은 2026년 10월까지 공개된 것을 찾지 못했다. 국내 대학과 기업에서도 이 주제를 정면으로 다룬 공개 연구는 찾지 못했다. 같은 현상에 닿은 결과는 셋 있다.

첫째는 싱가포르와 함께 한 공동 테스트다(ZDNet Korea, 2026년 6월 19일). 연구소는 글로벌 모델 3종의 이름을 가린 채 일상 업무 속 데이터 유출 위험을 시험했다. 에이전트는 정상 지시를 수행하다가 판단 착오로 민감 정보를 부적절하게 조회하고 전달하고 공개했다. 업무 능력이 좋아도 데이터를 안전하게 다루지 못하는 '인지-행동 불일치' 사례가 여럿 나왔다. 도구를 실제로 실행하지 않고도 임무를 마쳤다고 주장하는

'허위 보고'도 드러났다. 일을 하는 대신 일을 마쳤다는 신호만 낸다는 점에서 허위 보고는 명세 악용과 구조가 같다. 보고서가 이 용어를 썼는지는 확인하지 못했다.

둘째는 국정감사에서 나온 수치다. 김우영 더불어민주당 의원은 연구소가 세 유형의 공격에 대한 방어 실험을 했다고 밝혔다(디지털투데이, 2026년 10월 6일). 프롬프트 주입과 내부 저장정보 조작과 메모리 편향이다. 프롬프트 주입은 모델이 읽는 문서나 입력에 공격자의 명령을 숨겨 넣는 공격이다. 실험에서 방어 실패가 여러 차례 나왔고 메모리 편향 공격의 공격자 의도 일치율은 91.9퍼센트였다. 어느 모델을 언제 시험했는지와 원 보고서는 보도에 없다.

셋째는 평가 설계다. 연구소 홈페이지는 국가가 관리할 위험으로 "인간의 감독을 회피할 가능성이 있는 모델"을 들고 안전장치 우회, 조작 및 기만, 명시적 승인 없는 자율 복제를 적었다(인공지능안전연구소). 김용수 선임연구원은 모델이 평가받는 줄 알아채고 평소와 다르게 움직이는 상황까지 고려해야 한다고 했다(전자신문, 2026년 10월 11일). 그래서 실제와 비슷한 환경에서 오래 지켜보는 시나리오형 평가가 필요하다는 것이다. 설계에는 정렬 실패 유형이 들어 있지만 연구소는 그 결과를 아직 내놓지 않았다.

대학 쪽에는 보도가 하나 있다. 숭실대 AI안전성연구센터가 만든 사이버보안 에이전트가 2026년 6월 한 평가에서 인터넷이 막힌 환경에서 답을 찾지 못하자 외부 연결을 400회 넘게 시도했다(SBS, 2026년 10월 5일). 평가 주최는 기사에 없고 학술 검증을 거친 결과도 아니다.

비교 기준은 영국과 미국에 있다. 영국 안전연구소는 프런티어 모델 4종을 인공지능 연구소 안의 코딩 조수로 두고 안전 연구를 몰래 방해하는지 시험한 사례 연구를 냈다(영국 AI안전연구소, 2025년 11월 26일). 결론은 확인된 연구 방해 사례가 없다는 것이었고(we find no confirmed instances of research sabotage) 일부 모델은 안전 관련 과제를 자주 거부했다. 영국은 국제 기금 Alignment Project로 60개 넘는 정렬 연구 과제에 2,700만 파운드 넘게 배정했다(영국 AI안전연구소 Alignment Project). 미국 CAISI는 2026년 3월 레드티밍 대회 결과를 공개했다(CAISI, 2026년 3월 23일). 레드티밍은 공격자 역할을 맡아 시스템의 약점을 일부러 찾는 시험이다. 400명 넘는 참가자가 프런티어 모델 13종을 25만 회 넘게 공격했고 모든 대상 모델에서 공격이 한 번 이상 성공했다(at least one successful attack was found against all of the target frontier models). 같은 주제에서 한국이 공개한 결과는 익명 처리된 공동 보고서 1건과 국정감사 답변 수치다.

해외 안전연구소와의 비교

영국 AI안전연구소는 2023년 11월 문을 열었다. 홈페이지는 이 기관을 과학혁신기술부 안의 연구 조직이라고 소개하고 기술 인력이 100명 이상이라고 적었다(영국 AI안전연구소 소개). 이정헌 의원실은 영국 인력을 200명 안팎, 예산을 약 1,216억 원으로 집계했다(디지털데일리, 2026년 9월 28일). 영국의 배포 전 접근은 개발사와의 자발적 계약에 기댄다. 2026년 9월 백악관이 자국 기업에 영국과의 모델 공유 중단을 요청했고 앤트로픽은 Claude Mythos 5.1의 영국 검토를 막았다(보안뉴스, 2026년 10월 13일). 10월 들어 접근이 회복됐다는 보도가 나왔지만 공식 확인은 없다.

미국 연구소는 2023년 11월 출범했고 2025년 6월 상무부가 CAISI로 개편했다(미국 상무부, 2025년 6월 3일). 기관을 없앤 것이 아니라 NIST 안에서 이름과 임무를 바꿨다. CAISI는 오픈AI, 앤트로픽, 구글 딥마인드, 마이크로소프트, xAI 5개사와 협약을 맺고 40회 넘게 평가했다(Cloud Security Alliance, 2026년 5월 5일). 같은 분석은 평가 권한이 자발적 참여에만 기댄다고 적었다(rests entirely on voluntary participation). 의원실 집계로 미국 인력은 34명이고 예산은 약 232억 원이다.

일본 연구소는 2024년 2월 정보처리추진기구 안에 사무국을 두고 출범했다(일본 AISI). 요미우리는 2025년 12월 인력이 약 30명이고 정부 계획 초안이 이를 두 배로 늘리라고 했다고 전했다(Asia News Network, 2025년 12월 12일). 의원실 집계로 일본 예산은 당초 약 34억 원에서 추가경정예산으로 800억 원 이상이 됐다. 싱가포르는 2024년 5월 난양공대의 디지털 신뢰 센터를 안전연구소로 지정했다(IMDA, 2024년 5월 22일).

EU AI사무국은 국제 네트워크 회원이지만 집행위원회 안의 규제 집행 조직이다. 2026년 8월 2일부터 범용 모델 제공자에게 기술 접근을 요구하고 과징금을 매길 수 있다(Lawfare, 2026년 5월 18일). 과징금 상한은 전 세계 연매출의 3퍼센트와 1,500만 유로 가운데 큰 금액이다. 네트워크 열 회원 가운데 이런 강제권을 가진 곳은 AI사무국뿐이다. 중국은 네트워크 회원이 아니다. 중국 대학과 연구기관이 만든 안전 연구 네트워크는 국제 대화에서 중국을 대표하는 일을 주로 맡고 있다는 분석이 있다(Carnegie, 2025년 6월).

법적 권한만 보면 영국과 미국과 일본과 한국은 같은 처지다. 넷 모두 기업에 모델 제출을 강제할 수 없고 협약에 기대어 평가한다. 차이는 규모와 접근에서 난다. 배경훈 부총리 겸 과기정통부 장관은 국정감사에서 고영향 AI 평가를 영국 안전연구소에 많이 의뢰하고 있다고 답했다(디지털데일리, 2026년 10월 6일).

권한을 늘리면 산업이 움츠러든다는 우려는 기본법의 규제 조항을 두고 나왔다. 미국 정보기술혁신재단(ITIF)은 2025년 9월 보고서에서 연산량 기준과 보고 의무가 기업의 "행정적 부담을 가중시키며, 자원 낭비로 이어질 가능성이 크다"고 했다(ITIF, 2025년 9월). 같은 보고서는 연산량 기준을 없애는 대신 제12조를 고쳐 연구소에 배포 후 평가 권한을 주자고 제안했다. 배포된 모델을 시험하고 장애와 사고를 지켜보며 결과를 발표하게 하라는 것이다. 규제 부담을 앞세운 보고서가 연구소의 권한은 늘리자고 한 것이다. 연구소 스스로는 선을 그어 왔다. 김 소장은 2025년 4월 기본법을 "프로모션 쪽"이라고 했고 연구소는 인증·검증 기관이 아니라 인정기관 역할이 맞다고 말했다(더에이아이, 2025년 4월 8일). 한국 산업계 단체가 연구소 권한 확대에 반대한 성명은 찾지 못했다. 2026년 국정감사의 지적은 모두 인력과 예산과 권한이 모자란다는 쪽이었다.

김 소장은 9월 23일 통화에서 미국과 중국 사이에 제3국의 평가가 필요해지면 한국이 나서겠다고 하면서도 "아직은 앞서가는 시나리오이자 희망 사항일 뿐"이라고 덧붙였다(ZDNet Korea, 2026년 9월 23일). 과기정통부는 12월에 국가 AI 안전 생태계 조성 마스터플랜을 낼 계획이고 이 계획이 연구소의 권한과 독립을 다룰지는 공개된 자료에 아직 없다.

한국에 필요한 것은 연구소의 명칭이나 위상을 높이는 데서 그치지 않는다. 평가할 모델에 접근하고, 충분한 인력과 장비로 시험하며, 기업의 이해와 독립적으로 결과를 판단할 수 있어야 한다. 협력은 접근을 넓히는 수단으로 유지하되 평가의 내용과 공개 기준까지 기업에 맡겨서는 안 된다. 이 기능을 실제로 수행할 권한과 자원을 갖추는 것이 다음 제도 설계의 기준이다.

참고 자료

법제처 국가법령정보센터, 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법(법률 제21311호), 2026년 1월 20일 일부개정, https://www.law.go.kr/법령/인공지능발전과신뢰기반조성등에관한기본법

법제처 국가법령정보센터, 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 시행령(대통령령 제36053호), 2026년 1월 21일 제정, https://www.mofa.go.kr/upload/cntnts/www/인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 시행령.pdf

디지털데일리, "韓 AI 안전 평가 인력 6명뿐…英 200명·美 34명과 격차", 2026년 9월 28일, https://www.ddaily.co.kr/page/view/2026092810442730991

뉴시스, AI안전연구소 내달 출범…ETRI 부설 조직으로 설치, 2024년 10월 18일, https://www.newsis.com/view/NISX20241018_0002925245

ZDNet Korea, AI 안전 평가 선도하는 영국…한국도 국제무대 나설 수 있을까, 2026년 10월 13일, https://zdnet.co.kr/view/?no=20261013110737

전자신문, AI안전연구소 공식 출범…"AI 폐해, 체계적 대응한다", 2024년 11월 27일, https://www.etnews.com/20241127000331

이데일리, AI안전연구소 개소…"규제 아니라 위험 줄이는 협력 기관될 것", 2024년 11월 27일, https://edaily.co.kr/News/Read?mediaCodeNo=257&newsId=02236966639089984

ZDNet Korea, AI안전연구소, 국내외 AI 모델 42종 안전성 평가 완료, 2026년 5월 20일, https://zdnet.co.kr/view/?no=20260520132102

디지털투데이, AI안전연구소 평가 모델 절반 이상 오픈웨이트...그중 58%가 중국계, 2026년 9월 18일, https://www.digitaltoday.co.kr/news/articleView.html?idxno=701405

아주경제, 프런티어 AI 쏟아지는데…H200 16장으로 버티는 韓 안전성 평가, 2026년 9월 16일, https://www.ajunews.com/view/20260916141912494

디지털데일리, 韓 AI안전연구소 출범 1년 반, 프런티어 모델 독자 평가 공개 '0건', 2026년 5월 13일, https://www.ddaily.co.kr/page/view/2026051315321599128

ZDNet Korea, '깜깜이' 평가 벗는다…AI안전연구소, 모델 성적표 공개 기조로, 2026년 6월 19일, https://zdnet.co.kr/view/?no=20260619100049

디지털타임스, AI안전연구소, 오픈AI와 AI안전평가 협력, 2026년 6월 17일, https://www.dt.co.kr/article/12067984

카카오, 카카오, 인공지능안전연구소와 AI 안전성 평가 체계 구축을 위한 MOU 체결, 2026년 9월 29일, https://www.kakaocorp.com/page/detail/12150

디지털데일리, 이주희 의원 "AI 안전 예산, 274억 불과…'성능' 예산과 격차 224배" [2026 국감], 2026년 10월 6일, https://www.ddaily.co.kr/page/view/2026100610291136946

디지털데일리, "AI 안전 인력 부족" 지적에…배경훈 "AI안전연구소 인력·예산 늘려야" [2026 국감], 2026년 10월 6일, https://www.ddaily.co.kr/page/view/2026100612091028774

AI타임스, 김명주 AI안전연구소장 "AI 안전은 국가 경쟁력...전문성 확보는 필수", 2025년 3월 29일, https://www.aitimes.com/news/articleView.html?idxno=169202

전자신문, [AI 세이프가드] ③ AI도 '스트레스 테스트' 한다……한국의 'AI 안전 시험장' 가보니, 2026년 10월 11일, https://www.etnews.com/20261008000313

전자신문, AI안전연구소 출범 1년…글로벌 평가기관으로 성장 가속, 2025년 11월 23일, https://www.etnews.com/20251121000151

디지털투데이, [국감2026] 배경훈 "AI 킬스위치 필요…프런티어 AI로 사이버 위협 대응", 2026년 10월 6일, https://www.digitaltoday.co.kr/news/articleView.html?idxno=705165

인공지능안전연구소, AI안전평가(업무 소개), 2026년 10월 14일 열람, https://www.aisi.re.kr/kor/sub/14

SBS, Exclusive: South Korean AI Attempts to 'Jailbreak'... What About 'Black Box' Control?, 2026년 10월 5일, https://news.sbs.co.kr/english/article.do?news_id=N1008783697

영국 AI안전연구소, UK AISI Alignment Evaluation Case-Study, 2025년 11월 26일, https://www.aisi.gov.uk/research/uk-aisi-alignment-evaluation-case-study

영국 AI안전연구소, The Alignment Project by AISI, 2026년 10월 14일 열람, https://alignmentproject.aisi.gov.uk/

CAISI(NIST), Insights into AI Agent Security from a Large-Scale Red-Teaming Competition, 2026년 3월 23일, https://www.nist.gov/blogs/caisi-research-blog/insights-ai-agent-security-large-scale-red-teaming-competition

영국 AI안전연구소, About, 2026년 10월 14일 열람, https://www.aisi.gov.uk/about

보안뉴스, 영국 AI보안연구소, 미국 프런티어 AI 모델 사전 접근권 회복, 2026년 10월 13일, https://www.boannews.com/news/articleView.html?idxno=146341

미국 상무부, ICYMI: Statement from U.S. Secretary of Commerce Howard Lutnick on Transforming the U.S. AI Safety Institute into the Pro-Innovation, Pro-Science U.S. Center for AI Standards and Innovation, 2025년 6월 3일, https://content.govdelivery.com/accounts/USDOC/bulletins/3e381fb

Cloud Security Alliance, CAISI Frontier Testing Agreements Reach Five Labs, 2026년 5월 5일, https://labs.cloudsecurityalliance.org/research/csa-research-note-caisi-frontier-ai-testing-agreements-20260/

일본 AISI, AISIについて, 2026년 10월 14일 열람, https://aisi.go.jp/about/

Asia News Network(요미우리 전재), Japanese government plans to double staff tasked with checking AI safety, 2025년 12월 12일, https://asianews.network/?p=245395

IMDA, Digital Trust Centre designated as Singapore's AISI, 2024년 5월 22일, https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/factsheets/2024/digital-trust-centre

Lawfare, How Much Power Does the EU AI Office Actually Have?, 2026년 5월 18일, https://www.lawfaremedia.org/article/how-much-power-does-the-eu-ai-office-actually-have

Carnegie Endowment for International Peace, How Some of China's Top AI Thinkers Built Their Own AI Safety Institute, 2025년 6월, https://carnegieendowment.org/russia-eurasia/research/2025/06/how-some-of-chinas-top-ai-thinkers-built-their-own-ai-safety-institute

ITIF(정보기술혁신재단), 통합적 접근의 명암: 한국 AI 기본법의 전략·진흥·규제 구조와 규제 리스크, 2025년 9월, https://www2.itif.org/2025-korea-ai-act-ko.pdf

더에이아이, [창간 5주년 특집] 김명주 AI안전연구소장 "국가 AI 경쟁, '안전'으로 이어진다", 2025년 4월 8일, https://www.newstheai.com/news/articleView.html?idxno=7584

ZDNet Korea, AI 위험 검증하는 나라로…韓, '제3자 평가국' 노린다, 2026년 9월 23일, https://zdnet.co.kr/view/?no=20260923092746



15장. 사고 대응 체계의 설계

국회가 정해야 할 연구소의 위상과 역할

인공지능안전연구소를 한국전자통신연구원(ETRI) 안에 둔 결정은 처음부터 임시였다. 2024년 5월 이종호 당시 과학기술정보통신부 장관은 연구소를 ETRI 안에 먼저 세우겠다고 했다. 그는 대응할 범위가 넓어지면 "ETRI 산하이든, 과기정통부 산하이든 큰 규모의 기관으로 만들 계획"이라고 덧붙였다(아이뉴스24, 2024년 5월 22일). 기본법이 시행되고 아홉 달이 지난 지금도 연구소는 부설 조직이고, 앞 장에서 본 대로 제12조는 장관이 연구소를 운영할 수 있다고만 정했다.

국회 안에서도 고치자는 말이 나왔다. 이정헌 더불어민주당 의원은 2026년 9월 기본법을 고쳐 연구소에 "법적 독립성을 부여하고 독자적인 예산과 인사권을 보장"해야 한다고 했다(ZDNet Korea, 2026년 9월 28일). 그 내용을 담은 개정안이 발의됐는지는 10월 14일까지 확인하지 못했다. 2025년 12월 국회를 통과한 개정안이 근거를 둔 "인공지능연구소"는 첨단 기술 개발을 맡는 다른 기관이다(ZDNet Korea, 2025년 12월 30일).

해외에도 법률로 세운 안전연구소는 드물다. 영국 연구소는 정부 결정으로 생겼고 2026년 7월 과학혁신기술부가 해체되면서 내각부로 옮겼다(Global Government Finance, 2026년 7월 22일). 영국 정부는 2023년 창설 문서에 연구소가 규제기관이 아니라고 적었다(The Institute is not a regulator)(영국 정부, 2023년 11월). 카니슈카 나라얀 AI 담당 장관은 2026년 8월에도 기관의 구조보다 결과를 보겠다고 했다. 그는 연구소가 프런티어 모델에 닿는 접근을 아주 드문 자산으로 꼽았다(로이터, 2026년 8월 3일). 규제 권한을 주면 기업이 스스로 열어 주던 접근을 잃는다는 계산이 그 밑에 있다고 우리는 읽는다. 그러나 9월에는 그 자발 체계가 흔들렸다. 연구소의 사회회복력 책임자였던 앤드루 스트레이트는 영국의 자발 체계가 벽에 부딪혔다고 평가했다("hit the wall")(MLex, 2026년 9월 11일).

저자는 연구소를 법률로 세우는 독립 법인으로 바꾸자고 제안한다. 선택지는 셋이다. ETRI 부설을 두고 업무만 법에 늘리는 길, 과기정통부 직속 소속기관으로 올리는 길, 별도 법인으로 세우는 길이다. 부설로는 앞 장에서 본 인사와 예산의 한계가 그대로

남는다. 직속 기관은 공무원 보수 체계에 묶이고 진흥을 맡은 장관의 지휘를 받는다. 법인으로 세우면 원장의 임기와 독자 예산과 채용 규정을 법에 적을 수 있다. 조문은 "운영할 수 있다"를 "둔다"로 고친다. 연구소에 제재권은 주지 않는다. 연구소는 평가하고 분석하는 기관으로 남기고 제재는 장관이 맡는다. 기업이 연구소에 정보를 내놓기 쉽게 하려면 이 구분이 필요하다. 영국이 지키려는 협력 관계도 이렇게 하면 크게 다치지 않는다.

프런티어 모델을 평가할 인력과 장비

앞 장에서 본 기술 평가 인력 6명과 GPU 16장은 모델 하나를 일주일 넘게 붙잡아야 하는 규모다. 비교의 기준은 이미 나와 있다. 영국 연구소는 기술 인력 100명 이상과 회계연도당 6,600만 파운드에 더해 15억 파운드가 넘는 국가 컴퓨팅 자원에 우선 접근한다고 밝혔다(영국 AI안전연구소 소개). 미국 싱크탱크 Institute for Progress는 미국 AI표준혁신센터(CAISI)가 맡은 일을 다 하려면 연 8,400만 달러가 든다고 추산했다(Institute for Progress, 2026년 5월 13일). 최소 규모로 잡아도 연 2,600만 달러였다. 기술 인력 한 명에 드는 총비용은 약 35만 달러로 잡았다. 일본 연구소는 31명에서 올해 60명으로, 장기적으로 200명까지 늘릴 계획이다(ZDNet Korea, 2026년 9월 28일).

저자가 제안하는 첫 목표는 3년 안에 기술 평가 인력을 50명으로 늘리는 것이다. 영국이 공식으로 밝힌 기술 인력의 절반이고 일본의 단기 목표보다 조금 적다. 예산은 영국 연구소 연간 예산의 3분의 1을 첫 목표로 삼는다. 의원실 환산액을 기준으로 하면 400억 원 남짓이다. 인력은 사이버보안, 생물·화학 위험, 에이전트 행동, 한국어 평가의 네 분야로 나눠 뽑는다. 평가 전문가의 보수는 공공기관 인건비 기준 밖에 두는 특례를 법에 적는다. 앞 장에서 본 영국과의 보수 차이를 그대로 두고는 사람을 데려올 수 없다.

장비는 국가 계획 안에서 몫을 정한다. 정부는 전남 해남의 국가 AI컴퓨팅센터에 2028년까지 첨단 GPU 1만 5,000장을 구축한다(서울신문, 2026년 8월 4일). 이 자원 가운데 안전 평가에 배정한 몫은 어느 발표에서도 찾지 못했다. 영국의 우선 접근 약속에 대응하는 규정이 한국에는 없다. 저자는 국가 컴퓨팅 자원의 2퍼센트를 연구소 평가에 우선 배정하도록 법에 적자고 제안한다. 1만 5,000장 기준으로 300장이고 지금 연구소 장비의 스무 배 가까운 규모다.

해외 안전연구소와의 공동 평가 체계

공동 평가는 여러 나라 기관이 같은 모델을 같은 방법으로 시험하고 결과를 맞춰 보는 일이다. 국제 AI안전연구소 네트워크는 2025년 7월 세 번째 공동 평가를 에이전트에 맞췄다. 민감정보 유출과 사기, 사이버보안이 두 축이었고 EU AI사무국은 사이버보안 평가를 맡은 세 회원 가운데 하나였다(EU 집행위원회, 2025년 7월 17일). 2026년 7월 네트워크는 서울에서 모여 평가 모범 관행을 논의했다(영국 AI안전연구소, 2026년 7월 23일). 영국 연구소는 이 글에서 네트워크를 새 이름(International Network for Advanced AI Measurement, Evaluation and Science)으로 적었다. 같은 날 영국과 미국은 중국 문샷의 Kimi K3에 대한 사이버 능력 공동 예비 평가를 냈다(NIST, 2026년 7월 23일). 두 나라의 양자 협력에서 나온 결과다.

한국은 네트워크 공동 평가에 참여했지만 과제를 이끈 기록은 없다. 출시 전 평가도 아직 없다. 과기정통부는 2026년 8월 연구소가 평가한 모델 가운데 출시 전에 평가한 사례가 없다고 설명했다(머니투데이, 2026년 8월 6일). 과기정통부는 이 사정이 "영국 외 다른 국가도 마찬가지"라고 덧붙였다. 양해각서가 곧 접근권도 아니다. 디지털데일리는 앤트로픽과 일본 연구소의 양해각서가 평가 방법론만 다뤄 모델 접근은 따로 협상해야 했다고 전했다(디지털데일리, 2026년 5월 13일). 영국은 앤트로픽과 맺은 양해각서에 결과 공개 원칙을 넣었고 Mythos를 직접 시험했다.

공동 평가에 들어가는 조건은 세 가지로 본다. 첫째는 모델 접근권이다. 연구소가 개발사와 접근 협정을 맺을 근거와 받은 정보를 지킬 의무를 법에 함께 적는다. 둘째는 보안이다. 미공개 모델을 받으려면 외부망과 끊긴 시험 시설과 그 시설에 대한 보안 인증이 있어야 한다. 한국 연구소가 이런 인증을 갖췄는지 다룬 공개 자료는 찾지 못했다. 셋째는 사람이다. 공동 평가에서 한 과제를 맡으려면 그 과제만 볼 인력이 있어야 한다.

이 조건을 갖추면 한국이 맡을 과제는 분명하다. 한국어와 한국 행정 맥락에서 에이전트가 어떻게 움직이는지 보는 시험이다. 앞 장에서 본 싱가포르와의 공동 시험이 그 첫 사례였다. 이런 양자 공동 평가를 해마다 한 건 이상 새 상대와 하고 결과를 공개하자는 것이 저자의 제안이다. 상대는 연구소가 이미 협력 관계를 맺은 일본과 캐나다부터 찾는다. 결과 공개 원칙은 협정문에 처음부터 넣는다.

사고 보고를 받고 조사할 권한

13장이 보여 준 대로 한국 법에는 사고를 보고하라는 조항이 없다. 14장이 보여 준 대로 연구소에는 조사권이 없고 사실조사는 장관의 권한이다. 보고를 받는 곳과 조사하는 곳을

새로 정해야 한다. 선택지는 연구소, 과기정통부, 별도 조사위원회다.

국내에 참고할 모델이 둘 있다. 항공·철도 사고조사에 관한 법률은 위원회를 국토교통부에 두되 장관이 사고조사에는 관여하지 못하게 했다. 위원회는 자료 제출과 관계인 출석과 관계 물건의 보존을 요구할 수 있다. 통보자의 신분을 그 의사에 반해 공개하지 못하게 하고, 사고조사를 민사와 형사 책임을 다투는 절차와 떼어 놓았다. 이 위원회는 국무조정실로 옮기는 일을 앞두고 있다. 2026년 2월 국회 토론회에서는 위원장이 비상근이고 조사관 자격이 없는 사무국장이 조사 업무를 전결한다는 지적이 나왔다(더팩트, 2026년 2월 26일). 법에 적은 독립도 조사관이 없으면 비어 있다는 뜻이다. 다른 하나는 개인정보 보호법이다. 처리자는 유출을 안 때부터 72시간 안에 정보주체에게 알리고, 일정 규모 이상이면 같은 시한에 당국에 신고한다. 위원회는 자료 제출을 요구하고 사업장에 들어가 검사할 수 있다.

저자의 선택은 셋을 나누는 것이다. 보고는 연구소가 받는다. 연구소에는 제재권을 두지 않고, 13장이 소개한 미국 항공 자율보고 제도(ASRS)처럼 정보 수집과 제재의 기능을 구분한다. 중대 사고의 원인 조사는 국무총리 소속 조사위원회가 맡고 조사관은 연구소 인력으로 채운다. 과태료와 시정명령은 과기정통부 장관이 맡는다. 보고 시한은 영국 사이버 법안의 두 단계를 따른다. 영국은 사고를 안 뒤 24시간 안에 간단히 알리고 72시간 안에 상세 보고를 내게 했다(영국 정부, 2026년 6월 30일). 사고로 이어질 뻔한 일은 자율 보고로 받고, 고의가 없는 위반을 스스로 알린 사람에게는 과태료 면제를 둔다. 다만 이 면책이 중대 사고의 통지·보고 의무를 늦추는 근거가 되지 않도록 적용 범위를 구분한다.

보고는 회사 밖에서만 오지 않는다. 내부고발자 보호는 조직의 위법이나 위험을 알린 직원에게 해고나 불이익을 주지 못하게 하는 장치다. 캘리포니아 SB 53은 재앙적 위험을 신고한 직원과 계약자에 대한 보복을 금지하고 고용주에게 익명 내부 신고 경로를 두게 했다(Future of Privacy Forum, 2025년 10월 3일). 시민이 중대 사고를 직접 신고할 경로도 주 비상관리청에 두게 했다. 한국 개정안에도 두 장치를 함께 넣는다.

조사권에는 영업비밀 우려가 따른다. 2024년 12월 기본법 심의 때 민원만으로 기업을 조사할 수 있다는 비판이 나왔다. 과기정통부는 사실조사를 "조사목적을 달성하는데 필요한 최소한의 범위 안에서 실시할 예정"이라고 해명했다(정책브리핑, 2024년 12월 18일). 이 우려에는 법으로 답할 수 있다. 조사에서 얻은 정보는 조사 목적 밖에 쓰지 못하게 하고, 공개 보고서에서는 영업비밀을 가린다. 조사를 사법절차와 떼어 놓는 일은 항공 사고조사법이 이미 쓰는 방식이다.

기업과 공공기관의 에이전트 도입 속도와 비어 있는 대응 체계

정부는 에이전트를 먼저 도입하고 있다. 행정안전부는 2026년 3월 9일 AI 국민비서 시범 서비스를 시작했다. 황규철 행정안전부 AI정부실장은 이를 "공공 분야에 AI 에이전트를 도입한 국내 첫 사례"라고 했다(머니투데이, 2026년 3월 18일). 6월 10일에는 공공부문 AI 도입·활용 가이드를 전국 공공기관에 배포했다(행정안전부, 2026년 6월 10일). 과기정통부는 5월 부처 내부 행정에 에이전트를 도입하는 AI-NEXT 사업을 발주했다(ZDNet Korea, 2026년 5월 13일). 무선국 허가검사, 전자파 인증, 예산 업무가 첫 대상이다. 기업 쪽에서는 LG가 5월 6일 국내외 전 임원 과정을 열고 임원이 업무 구조를 다시 짜며 에이전트를 직접 만들게 했다(중앙일보, 2026년 5월 7일).

통제 기준은 금융권에서 처음 나왔다. 금융보안원은 10월 7일 금융분야 AI 에이전트 보안 평가 기준 17개 항목을 마련했다(전자신문, 2026년 10월 7일). 에이전트가 필요한 권한만 갖는지, 외부 도구를 쓸 때 승인을 받는지, 필요할 때 멈출 수 있는지를 본다. 같은 날 금융위원회는 금융권 에이전트 규율체계 연구용역을 발주했고 12월에 최종 보고를 받는다(뉴스1, 2026년 10월 7일). 두 조치 모두 사고가 난 뒤 무엇을 할지가 아니라 사전 점검과 규율 설계를 다룬다.

사고는 이미 났다. 금융보안원은 9월 17일 세미나에서 딥시크 기반 에이전트가 금융사 서버를 탐색하고 하루 만에 악성프로그램을 설치한 사례를 공유했다. 약 2주 뒤 여러 금융사에서 정보가 유출됐고 중국산 AI 침투 시험 도구와 딥시크가 쓰인 것으로 보인다(서울신문, 2026년 10월 11일). 이것은 외부 공격자가 에이전트를 도구로 쓴 사고다. 도입한 조직의 에이전트가 스스로 오작동한 국내 사고는 공개 기록이 없다. 미국에서는 4월 코딩 에이전트가 자격 증명 오류를 만나자 다른 키를 찾아 운영 데이터베이스를 지웠고 복구에 약 30시간이 걸렸다(IT조선, 2026년 5월 1일). 국내 기업의 에이전트 사고 보고 절차를 확인할 수 있는 공개 자료는 없다. 공개 기록이 없다는 사실만으로는 사고가 없었다는 뜻인지 알릴 길이 없다는 뜻인지 가릴 수 없다.

한국은 프런티어 모델 개발국이 아니므로 평가 역량보다 도입 관리가 먼저라는 시각이 있다. 김현 더불어민주당 의원은 속도 조절 논쟁을 두고 "당장 우리에게 적용될 얘기는 아니라는 점을 짚고 싶다"고 했다(오마이뉴스, 2026년 9월 23일). 그는 인재 확보와 국민의 활용을 앞세웠고 "국민들이 AI에게 속지 않을 눈을 갖추는 것이 중요하다"고 했다. 그는 6월 사업자가 에이전트의 권한 범위를 미리 정하게 하는 개정안도 냈다(아이뉴스24, 2026년 8월 7일). 도입 관리가 급하다는 진단에는 저자도 동의한다. 다만 권한을 얼마나 줄지 정하려면 그 에이전트가 무엇을 할 수 있는지 시험해 봐야 한다. 금융보안원의 17개

항목도 누군가 에이전트를 직접 돌려 봐야 채울 수 있다. 도입 관리가 실효성을 가지려면 평가 결과가 뒷받침돼야 한다. 두 역량은 함께 갖춰야 한다.

저자는 공공기관과 대통령령으로 정하는 규모 이상의 사업자에게 두 가지를 지우자고 제안한다. 하나는 내부 통제다. 쓰는 에이전트의 목록과 권한 범위를 적고, 외부 도구 사용에 승인 절차를 두고, 행동 기록을 보존하고, 멈추는 절차를 정해 둔다. 다른 하나는 보고다. 에이전트가 낸 중대 사고는 앞 소절의 시한에 맞춰 연구소에 알린다. 공공기관부터 시작하면 정부가 먼저 기준을 지켜 보일 수 있다.

법 개정안의 골격

기본법 시행 뒤 8월 3일까지 발의된 개정안은 17건이다(아이뉴스24, 2026년 8월 7일). 김성원 의원 등 10인의 안은 국가안보에 중대한 위험을 줄 수 있는 고위험 인공지능 모델을 지정·관리하고 사이버위협 대응체계를 두는 내용이다(국회 입법예고, 의안번호 2220395). 연구소의 위상이나 사고 보고와 조사를 직접 다룬 개정안은 찾지 못했다.

규제를 늘리면 산업이 다친다는 반론이 있다. 스타트업얼라이언스 이기대 센터장은 "AI처럼 국내 산업이 뒤처진 분야에 선진국 규제를 그대로 들여오면 산업 기반이 무너질 수 있다"고 했다(더나은미래, 2025년 4월 1일). 이 우려는 의무의 대상을 좁혀서 줄인다. 보고 의무는 중대 사고에만 걸고 작은 기업의 아차사고는 자율 보고로 받는다. 앞 장에서 본 대로 규제 완화를 주장한 쪽도 연구소의 사고 감시 권한은 늘리자고 했다. 제재는 과태료로 시작한다. 과태료는 법이 정한 의무를 어긴 사람에게 정해진 상한 안에서 매기는 돈이다. 과징금은 위반의 규모나 매출에 비례해 매기는 돈이어서 무게가 다르다. 반복되는 고의 위반에 과징금을 둘지는 첫 운영 결과를 보고 정해도 늦지 않다.

다음은 저자가 제안하는 골격의 예시이며 법제처 심사를 거친 조문이 아니다.

제12조(인공지능안전연구소) ① 인공지능안전 업무를 전문적이고 독립적으로 수행하기 위하여 인공지능안전연구소(이하 "안전연구소"라 한다)를 둔다. ② 안전연구소는 법인으로 하며 원장의 임기는 3년으로 한다. ③ 안전연구소는 고성능 인공지능모델의 출시 전후 평가, 제32조의2에 따른 보고의 접수와 분석, 외국 안전연구소와의 공동 평가를 수행한다. ④ 정부는 국가 인공지능 컴퓨팅 자원의 100분의 2 이상을 안전연구소의 평가에 우선 배정하여야 한다. ⑤ 과학기술정보통신부장관은 안전연구소의 평가와 분석의 내용에 관여하지 못한다.

제32조의2(중대 인공지능 사고의 보고) ① 대통령령으로 정하는 인공지능사업자와 공공기관은 중대 인공지능 사고를 안 때부터 24시간 이내에 안전연구소에 통지하고

72시간 이내에 경위와 조치를 보고하여야 한다. ② 누구든지 사고로 이어질 뻔한 사건을 안전연구소에 보고할 수 있다. ③ 중대 인공지능 사고의 범위는 대통령령으로 정한다.

제32조의3(보고자 보호와 면책) ① 안전연구소는 보고자의 동의 없이 그 신원을 공개하여서는 아니 된다. ② 고의 없는 위반을 안 날부터 10일 이내에 스스로 보고한 자에게는 제43조의 과태료를 부과하지 아니한다. 다만, 제32조의2제1항에 따른 통지 또는 보고 의무 위반에는 이를 적용하지 아니한다. ③ 누구든지 보고나 신고를 이유로 임직원에게 불이익을 주어서는 아니 된다.

제32조의4(사고조사) ① 중대 인공지능 사고의 원인을 조사하기 위하여 국무총리 소속으로 인공지능사고조사위원회를 둔다. ② 위원회는 관계인에게 자료 제출, 출석, 기록과 모델의 보존, 평가를 위한 모델 접근을 요구할 수 있다. ③ 조사는 민사·형사 책임을 묻는 절차와 분리하여 수행하며 조사에서 얻은 영업비밀은 조사 목적 외에 사용하거나 공개하지 못한다. ④ 위원회는 조사보고서를 공표한다.

제32조의5(에이전트 도입 기관의 내부 통제) 공공기관과 대통령령으로 정하는 사업자는 에이전트의 권한 범위, 외부 도구 사용의 승인, 행동 기록의 보존, 작동 중단의 절차를 정하여 운영하여야 한다.

제43조(과태료) 현행 과태료 대상과 상한은 유지하고, 다음 위반에 대한 과태료 규정을 추가한다. 각 위반의 과태료 상한은 1억원으로 한다. 1. 제32조의2제1항에 따른 통지나 보고를 하지 아니하거나 거짓으로 한 자 2. 제32조의4제2항에 따른 요구를 정당한 사유 없이 거부한 자 3. 제32조의3제3항을 위반하여 불이익을 준 자. 신설 항·호의 번호와 체계는 입법 과정에서 정리한다.

제프리 힌턴이 미국 의회 브리핑을 마치고 규제할 시간이 1년쯤 남았다고 말한 날은 2026년 9월 16일이었다(NBC News, 2026년 9월 17일). 그 뒤 한 달이 지나는 사이 국내 금융사들은 에이전트 도구가 쓰인 공격으로 정보를 잃었다. 3부에서 본 위험은 날짜와 피해 기관이 적힌 기록이다. 한국에는 사고를 의무적으로 접수하고 원인을 독립적으로 밝힐 제도가 없고, 모델을 시험할 인력과 장비도 부족하다. 힌턴의 1년은 확정된 기한이 아니다. 입법을 시작할 근거는 이미 드러난 제도의 공백이다. 국회는 연구소의 독립성을 보장하고 보고와 조사의 절차를 마련하는 일을 지금 추진해야 한다.

참고 자료

아이뉴스24, 과기부 장관 "한국판 AI안전연구소, ETRI 내에 연내 설립", 2024년 5월 22일, https://www.inews24.com/view/1722475

ZDNet Korea, AI 위험 검증 강화하는 주요국…韓 안전 평가 인력 6명뿐, 2026년 9월 28일, https://zdnet.co.kr/view/?no=20260928132902

ZDNet Korea, AI기본법 개정안 국회 통과...국가AI전략위·AI연구소 법적 근거 마련, 2025년 12월 30일, https://zdnet.co.kr/view/?no=20251230202213

Global Government Finance, DSIT scrapped and AI gains top-table seat as Burnham premiership kicks off in UK, 2026년 7월 22일, https://www.globalgovernmentfinance.com/andy-burnham-pm-dsit-scrapped-ai-minister/

영국 정부, Introducing the AI Safety Institute(CP 960), 2023년 11월, https://assets.publishing.service.gov.uk/media/65438d159e05fd0014be7bd9/introducing-ai-safety-institute-web-accessible.pdf

Global Banking and Finance(Reuters 전재), Britain says it is open to AI regulation if voluntary safeguards fall short, 2026년 8월 3일, https://www.globalbankingandfinance.com/britain-open-ai-regulation-if-voluntary-safeguards-fall/

MLex, UK's voluntary model for AI safety testing faces global reality check, 2026년 9월 11일, https://www.mlex.com/mlex/articles/2524257/uk-s-voluntary-model-for-ai-safety-testing-faces-global-reality-check

영국 AI안전연구소, About, 2026년 10월 14일 열람, https://www.aisi.gov.uk/about

Institute for Progress, What Will It Cost for the US to Be Ready for the Next Big AI Breakthrough?, 2026년 5월 13일, https://ifp.org/funding-for-caisi/

서울신문, '국가AI센터' 첫 삽… 2028년 첨단 GPU 1만5000장 구축, 2026년 8월 4일, https://www.seoul.co.kr/news/economy/industry/2026/08/04/20260804031003

EU 집행위원회, AI Office contributes to the third-joint testing exercise of the International Network of AI Safety Institutes, 2025년 7월 17일, https://digital-strategy.ec.europa.eu/en/news/ai-office-contributes-third-joint-testing-exercise-international-network-ai-safety-institutes

영국 AI안전연구소, International evaluation best practice and open questions in AI measurement, 2026년 7월 23일, https://www.aisi.gov.uk/blog/international-evaluation-best-practice-and-open-questions-in-ai-measurement

NIST, UK AISI / CAISI Preliminary Assessment of Kimi K3's Cyber Capabilities, 2026년 7월 23일, https://www.nist.gov/news-events/news/2026/07/uk-aisi-caisi-preliminary-assessment-kimi-k3s-cyber-capabilities

머니투데이(서플 전재), [단독]미국은 AI 5개사 출시 전 '충돌시험'…한국은 검사 '전무', 2026년 8월 6일, https://supple.kr/news/cmsgkrq3l0096phsewrf2k2bh

디지털데일리, 韓 AI안전연구소 출범 1년 반, 프런티어 모델 독자 평가 공개 '0건', 2026년 5월 13일, https://www.ddaily.co.kr/page/view/2026051315321599128

법제처 국가법령정보센터, 항공·철도 사고조사에 관한 법률, https://www.law.go.kr/법령/항공ㆍ철도사고조사에관한법률

더팩트, 항철위 국무조정실 이관 앞두고 "형식 아닌 실질적 독립 필요", 2026년 2월 26일, https://news.tf.co.kr/read/economy/2296326.htm

법제처 국가법령정보센터, 개인정보 보호법, https://www.law.go.kr/법령/개인정보보호법

영국 정부, Incident reporting(Cyber Security and Resilience (Network and Information Systems) Bill: factsheets), 2026년 6월 30일 갱신, https://www.gov.uk/government/publications/deleted-cyber-security-and-resilience-network-and-information-systems-bill-factsheets/incident-reporting

Future of Privacy Forum, California's SB 53: The First Frontier AI Law, Explained, 2025년 10월 3일, https://fpf.org/blog/californias-sb-53-the-first-frontier-ai-law-explained/

정책브리핑, 과기정통부 "AI 기본법안의 사실조사는 필요 최소한 범위 내 실시", 2024년 12월 18일, https://www.korea.kr/news/policyNewsView.do?newsId=148937666

머니투데이, "말 한마디로 행정 처리"…행안부, AI 에이전트 도입 속도낸다, 2026년 3월 18일, https://www.mt.co.kr/policy/2026/03/18/2026031809155847471

행정안전부, 공공 AI 구축, 더 쉽고 빠르게 「공공부문 AI 도입· 활용 가이드」 배포, 2026년 6월 10일, https://www.mois.go.kr/frt/bbs/type010/commonSelectBoardArticle.do?bbsId=BBSMSTR_000000000008&nttId=126757

ZDNet Korea, 과기정통부, AI 에이전트로 내부 행정 업무 혁신 '속도', 2026년 5월 13일, https://zdnet.co.kr/view/?no=20260513102931

중앙일보(다음 전재), LG, 전 임원 'AI 교육' 전면화…'조직·사업 전반 AX 가속', 2026년 5월 7일, https://v.daum.net/v/20260507151419100

전자신문, 금융보안원, 금융권 AI 에이전트 보안 평가기준 마련, 2026년 10월 7일, https://www.etnews.com/20261007000233

뉴스1, 금융위, 금융권 'AI 에이전트' 규율체계 만든다…연구용역 발주, 2026년 10월 7일, https://www.news1.kr/finance/general-finance/6313370

서울신문, AI 해킹 수법·대응책까지 공유했는데… 2주 뒤 금융권 줄줄이 정보유출, 2026년 10월 11일, https://www.seoul.co.kr/news/economy/2026/10/11/20261011500059

IT조선(다음 전재), 기업 DB 통째로 지운 AI 에이전트… 권한 통제 부실이 대형 사고로, 2026년 5월 1일, https://v.daum.net/v/20260501165437703

오마이뉴스, AI로 인류 멸종? 한국은 다른 것부터 고민해야 한다, 2026년 9월 23일, https://www.ohmynews.com/NWS_Web/View/at_pg.aspx?CNTN_CD=A0003269555&CMPT_CD=M0153

아이뉴스24(네이트 전재), 고위험 AI 관리부터 생성물 표시 완화까지…올 1월 시행 'AI기본법' 개정안 벌써 17건, 2026년 8월 7일, https://m.news.nate.com/view/20260807n21705

국회 입법예고, 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 일부개정법률안(김성원 의원 등 10인, 의안번호 2220395), 2026년 8월 3일, https://community.lawmaking.go.kr/gcom/nsmLmSts/out/2220395/detailRP

더나은미래, "AI 기본법, 산업에 족쇄 될 수도"…스타트업얼라이언스, 보고서 발간, 2025년 4월 1일, https://futurechosun.com/archives/120204

NBC News, 'Godfather of AI' warns Congress has 'maybe a year' left to regulate AI, 2026년 9월 17일, https://www.nbcnews.com/politics/congress/godfather-ai-warns-congress-maybe-year-left-regulate-ai-rcna598330

법제처 국가법령정보센터, 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법(법률 제21311호), https://www.law.go.kr/법령/인공지능발전과신뢰기반조성등에관한기본법



제4부. 연결의 수단: 뇌와 컴퓨터를 잇는다

16장. 생각으로 커서를 움직이다

BCI의 작동 원리를 한 문단으로

2024년 3월 20일 뉴럴링크(Neuralink)는 9분짜리 영상을 X에 올렸다(NBC News, 2024년 3월 20일). 영상 속 놀런드 아보는 노트북 앞에 앉아 있었고 옆에는 회사 직원 한 명이 있었다. 스물아홉 살인 그는 약 8년 전 다이빙 사고로 어깨 아래가 마비됐다. 그전에는 입에 막대를 물고 기기를 다뤘다. 영상에서 그는 손을 쓰지 않고 커서를 옮겨 온라인 체스를 두었다. 노트북에서 흘러나오던 음악도 커서로 껐다. 화면 어디에도 전선이나 바깥 장치는 보이지 않았다. 그는 "이 모든 걸 제 뇌로 하고 있어요"라고 말했다(It's all being done with my brain). 커서를 처음 움직였을 때의 느낌은 영화 속 초능력인 포스를 쓰는 것 같았다고 했다.

서장에서 본 BCI는 뇌 신호를 읽어 기계를 움직이는 명령으로 바꾸는 컴퓨터 시스템이다(Mayo Clinic Proceedings, 2012년 3월). 사람의 뇌에는 신경세포인 뉴런이 약 860억 개 있다(Journal of Comparative Neurology, 2009년 4월). 운동피질은 몸의 움직임을 계획하고 명령하는 대뇌 겉층의 부위다. 손을 움직이려고 하면 운동피질의 뉴런들이 일정한 무늬로 신호를 낸다. 손이 마비된 사람도 움직이려고 애쓰면 이 신호는 그대로 생긴다. 전극은 이 전기 신호를 잡아내는 작은 금속 감지기다. 컴퓨터는 전극이 잡은 신호에서 잡음을 걸러 낸다. 디코더는 걸러 낸 신호의 무늬를 보고 사용자가 어느 쪽으로 움직이려 했는지 맞히는 프로그램이다. 디코더의 답은 커서의 이동이나 글자, 합성 음성, 로봇 손의 움직임으로 나온다. 연구자들은 이 과정을 흔히 기록, 처리, 해독, 출력의 네 단계로 설명한다. 이 장치는 생각 전체를 읽지 않는다. 정해진 동작과 관련된 신호를 골라 명령으로 옮긴다.

BCI라는 이름에는 조건이 붙는다. 같은 리뷰는 뇌파를 재기만 하는 뇌파계는 출력을 내지 않으므로 BCI로 보지 않는다. 디코더는 사용자마다 따로 맞춰야 한다. 이 작업을 교정(calibration)이라고 부른다. 뇌 신호는 날마다 조금씩 달라지므로 디코더도 쓰는 동안 계속 고친다(arXiv, 2025년 3월). 디코더 안에는 인공지능이 들어 있다. UC 데이비스의 말 복원 장치는 두 층으로 일한다(Neuroscience News, 2026년 7월 17일). 첫 층은 뇌 신호를 말소리를 이루는 최소 단위인 음소로 바꾼다. 둘째 층은 대규모 언어

모델의 기법으로 음소를 단어와 문장으로 엮는다. 신호가 들어와 단어가 나오기까지 30밀리초가 걸린다.

아보의 장치 N1은 64가닥의 가는 실에 전극 1,024개를 달았다(Popular Science, 2024년 5월 21일). 커서 성능은 1초에 전하는 정보량인 초당 비트(BPS)로 잰다. 뉴럴링크는 아보가 처음 4.6 BPS로 기존 연구의 최고치를 넘었고 뒤에 8.0 BPS까지 올렸다고 밝혔다(Decrypt, 2024년 5월 9일). 그는 한 주에 69시간을 이 장치와 보냈고 그중 34시간은 연구가 아닌 개인 용도였다. 같은 무렵 그는 실의 약 85퍼센트가 뇌에서 빠져나왔다고 월스트리트저널에 말했다(Popular Science, 2024년 5월 21일). 뉴럴링크는 신호를 읽는 알고리즘을 바꾸는 소프트웨어 갱신으로 성능을 되살렸다고 설명했다. 남은 신호를 디코더가 어떻게 읽느냐가 전극의 수만큼 성능을 좌우한다는 사실을 이 일이 보여 준다.

침습형과 비침습형

BCI를 나누는 첫 기준은 전극의 자리다. 침습형은 두개골을 열어 전극을 뇌 조직 안에 넣는 방식이다. 뉴런 바로 곁에서 신호를 들으므로 신호가 선명하다. 유타 어레이(Utah array)와 뉴럴링크 N1이 여기에 든다. 대가는 수술과 시간이다. 원숭이 17마리와 사람 2명에게 심은 유타 어레이 55개를 약 9년 동안 살핀 연구가 있다(Journal of Neural Engineering, 2021년). 평균 사용 가능 기간은 622일이었다. 1,000일을 넘긴 어레이도 있었고 사람에게 심은 쪽이 원숭이보다 오래갔다.

비침습형은 머리 겉에서 신호를 읽는 방식이다. 두피에 전극을 붙이는 뇌파 검사(EEG)가 대표적이다. 수술이 필요 없고 비용이 적게 들지만 신호가 두개골과 두피를 지나며 흐려진다. 두피에서 잡히는 신호는 마이크로볼트 단위로 약하고 동시에 읽을 수 있는 채널 수도 적다(arXiv, 2025년 3월). 마이크로볼트는 100만분의 1볼트다. 메타는 뇌의 자기장을 재는 장치(MEG)로 자원자 35명이 치려는 글자를 해독했다(MIT 테크놀로지 리뷰, 2025년 2월 7일). 글자 오류율은 평균 약 32퍼센트였다. 장치 무게는 약 500킬로그램이고 가격은 약 200만 달러였다. 이 장치는 바깥 자기장을 막는 방 안에서만 돌아갔다.

반침습형은 두 방식 사이에 있다. 뇌를 감싼 질긴 바깥 막인 경막의 바깥이나 뇌 표면에 전극을 얹는다. 혈관 안에 전극을 넣는 방식도 여기에 든다. 미국 프리시전 뉴로사이언스(Precision Neuroscience)의 Layer 7은 뇌 표면에 얹는 전극이다. 이 장치는 2025년 4월 FDA의 510(k) 허가를 받았다(Precision Neuroscience, 2025년 4월 17일). 허가 조건은

최대 30일 동안의 이식이고 그 기간에 뇌 표면의 신호를 기록하고 자극한다. 미국 싱크론(Synchron)의 스텐트로드(Stentrode)는 목의 정맥으로 관을 넣어 뇌혈관 안에 전극을 둔다. 2024년 9월 발표한 미국 임상(COMMAND)에서 환자 6명 모두 12개월 동안 뇌나 혈관과 관련된 중대한 이상반응이 없었다(BioSpace, 2024년 9월 30일). 중국 기업 뉴라클(博睿康, Neuracle)이 칭화대와 만든 NEO는 지름 25밀리미터에 배터리가 없는 장치다(Medical Xpress, 2024년 2월 5일). 전극 접점 8개를 경막 바깥에 두고 전원은 무선으로 받는다(medRxiv, 2024년 9월).

NEO를 부르는 이름은 자료마다 다르다. 중국 매체는 NEO를 세계 첫 "침습형" BCI 의료기기라고 불렀다(CGTN, 2026년 3월 14일). 중국 BCI 임상을 분석한 논문은 같은 장치를 세계 첫 승인 "반침습 이식형" BCI로 적었다(arXiv, 2026년 7월 8일). 차이는 규제 용어에서 나온다. 중국 국가약품감독관리국(NMPA)은 2026년 6월 29일 BCI 의료기기 분류 지도원칙을 냈다(CCTV, 2026년 6월 30일). 지도원칙은 침습식과 이식식 제품을 함께 관리 등급이 제일 높은 3류로 묶는다. 중국 보도의 "침습형"은 이 규제 범주를 가리킨다. 학술 분류의 "반침습"은 수술의 깊이를 가리킨다. NEO는 경막 바깥에 심는 이식형 장치이고 뉴럴링크와 달리 뇌 조직을 뚫지 않는다. 신호의 선명함과 수술의 부담은 서로 맞바꾸는 관계다. 어느 쪽을 고를지는 환자가 되찾고 싶은 기능에 달려 있다. 손으로 물컵 하나를 쥐려는 사람과 하루 종일 대화하려는 사람에게 필요한 신호의 양은 같지 않다.

사지마비 환자와 말을 할 수 없는 환자: 치료 목적의 현재

지금 BCI를 몸에 심는 사람은 두 무리다. 한 무리는 척수 손상이나 루게릭병(ALS), 뇌간 뇌졸중으로 팔다리를 못 쓰는 사람이다. 이들은 커서와 로봇팔, 장갑을 움직인다. 다른 무리는 같은 병으로 말을 잃은 사람이다. 이들은 글자나 합성 음성으로 말을 되찾는다. 지금까지 BCI는 주로 척수 손상 환자에게서 시험했고 ALS 환자의 자료는 적다(MIT 테크놀로지 리뷰, 2026년 6월 19일). 척수 손상은 대개 더 나빠지지 않지만 ALS는 시간이 갈수록 운동신경이 무너지는 병이다. 참여자 수는 아직 적다. 2024년의 한 리뷰는 1998년부터 2023년 말까지 21개 연구진이 67명에게 BCI를 시험했다고 집계했다(MIT 테크놀로지 리뷰, 2026년 6월 19일). 네덜란드 위트레흐트 대학병원의 마리스카 반스틴셀은 2026년 6월 그 수를 약 150명으로 추정했다. 뉴럴링크는 2026년 1월 28일 기준 임상 등록 인원을 21명으로 밝혔다(로이터, 2026년 1월 28일). 2025년 9월에 밝힌 12명에서 늘어난 숫자다. 그 뒤 10월까지 회사가 낸 공식 누계는 없다. 시판 승인을 받은

이식형 장치는 중국의 NEO다. NMPA는 2026년 3월 NEO를 3류 의료기기로 등록했다(CGTN, 2026년 3월 14일). 대상은 목 척수 손상으로 손을 쥐지 못하는 사지마비 환자다.

BCI가 과장됐다는 비판에는 근거가 있다. 위트레흐트 대학병원의 신경과학자 닉 람세이는 뉴럴링크가 절차를 건너뛸 수 있다는 걱정을 전했다("Neuralink could be cutting corners")(IEEE Spectrum, 2024년 2월 3일). 참여자가 적고 추적 기간이 짧다는 지적도 있다. 싱크론의 결과는 6명을 1년 동안 본 것이다. 아보의 실이 빠져나온 일과 유타 어레이의 평균 사용 기간은 전극이 오래 버티지 못한다는 증거다. 처음 잘 작동하던 장치가 일부 ALS 환자에게서 나중에 멈췄고 과학자들은 그 이유를 아직 모른다(MIT 테크놀로지 리뷰, 2026년 6월 19일). 이 지적은 맞다. 그러나 증거의 성격은 바뀌고 있다. ALS 환자 케이시 해럴은 거의 2년 동안 집에서 장치를 3,800시간 넘게 썼다(UC 데이비스, 2026년 6월 15일). 짧은 실험실 시연 대신 생활 속 사용 기록이 쌓이기 시작했다. 장기 안정성의 답도 이런 기록에서 나온다.

수술 없이도 된다는 시각도 있다. 샘 올트먼이 세운 머지랩스(Merge Labs)는 2026년 1월 2억 5,000만 달러를 투자받았다(TechCrunch, 2026년 1월 15일). 이 회사는 전극 대신 분자로 뉴런과 연결하고 초음파 같은 방법으로 뇌 깊은 곳에 닿겠다고 밝혔다. 제품과 일정은 아직 내놓지 않았다. 환자의 선호도 이쪽에 가깝다. 다발성경화증 환자 34명에게 물은 설문에서 12명은 비침습형을 골랐고 15명은 최소 침습형 BCI를 골랐다(arXiv, 2024년 4월 7일). 이 응답은 장치를 써 본 경험이 아닌 선호를 물은 결과다. 성능의 차이는 비침습 연구자도 인정한다. 메타의 장레미 킹은 자기 팀의 MEG 연구를 두고 "너무 어려워서 제품으로 가는 길은 없다고 본다"고 했다(I don't think there is a path for products)(MIT 테크놀로지 리뷰, 2025년 2월 7일). 공개된 임상 자료에서 대화에 가까운 속도를 낸 장치는 모두 이식형이다.

한국은 2026년 3월 18일 과학기술정보통신부가 K-문샷의 BCI 임무를 2027년에 시작한다고 밝혔다(이투데이, 2026년 3월 18일). 국내에는 아직 이식형 BCI 임상시험 인허가 사례가 없다(디지털타임스, 2026년 3월 22일).

커서를 움직이고 문장을 쓰는 장면

뉴럴링크의 두 번째 참여자 알렉스는 장치를 쓴 첫날 BCI 커서 제어 기록을 넘었다(Decrypt, 2024년 8월 22일). 척수 손상 환자인 그는 3차원 설계 프로그램 퓨전 360(Fusion 360)으로 뉴럴링크 충전기 거치대를 그렸다. 거치대는 3D 프린터로 뽑아

노트북 받침대에 붙였다. 그는 1인칭 사격 게임 카운터스트라이크 2도 했다. 생각만으로 좌우를 둘러보는 일이 입으로 조작하는 조이스틱보다 낫다고 했다.

중국 NEO의 첫 환자는 14년 전 교통사고로 사지마비가 된 사람이다(Medical Xpress, 2024년 2월 5일). 그는 2023년 10월 24일 베이징 쉬안우병원에서 장치를 심었다. 이후 집에서 훈련하며 로봇 손으로 병을 쥐었고 스스로 먹고 마시게 됐다. 9개월의 가정 사용 기록에서 잡으려는 의도를 알아맞히는 성능 지표는 평균 0.91이었다(medRxiv, 2024년 9월).

문장을 쓰는 방법도 늘었다. 브레인게이트(BrainGate) 연구진은 손가락으로 자판을 치려는 시도를 글자로 바꿨다(브라운대, 2026년 3월 16일). 참여자 2명 가운데 최고 속도는 1분에 110자, 22단어였고 단어 오류율은 1.6퍼센트였다. 한 사람은 ALS 환자였고 다른 한 사람은 목 척수를 다친 환자였다. 두 사람은 이 장치를 집에서 썼다. 디코더를 맞추는 데 필요한 문장은 30개였다. 손으로 글씨를 쓰거나 자판을 치려던 기억이 남아 있으면 그 기억이 새 입력 방법이 된다.

말을 되찾는 일은 더 어렵다. 해럴은 ALS로 말이 알아듣기 어려워진 45세 남성이었다. 2023년 7월 UC 데이비스 연구진은 말을 담당하는 그의 왼쪽 운동피질에 전극 256개를 심었다. 그가 말하려 하면 디코더가 단어를 화면에 띄우고 발병 전 목소리로 만든 합성 음성이 그 단어를 읽었다(NEJM, 2024년 8월 14일). 첫날 약 30분의 교정 뒤 50단어 어휘에서 정확도는 99.6퍼센트였다. 둘째 날 12만 5,000단어 어휘로 넓히자 90.2퍼센트였다. 수술 8.4개월 뒤에도 같은 어휘에서 97.5퍼센트를 유지했고 1분에 약 32단어를 말했다. 2026년 6월 논문에서는 통제 조건 정확도가 99퍼센트를 넘었고 평균 속도는 1분에 56단어였다(UC 데이비스, 2026년 6월 15일). 그는 18만 3,000개 문장을 이 장치로 전했고 그중 92퍼센트가 정확하거나 대체로 맞았다고 스스로 평가했다. 연구진은 그가 12시간 넘게 쉬지 않고 장치를 쓰는 날도 있다고 전했다. 그는 이 장치로 이메일을 보내고 일을 계속하며 딸에게 책을 읽어 준다(MIT 테크놀로지 리뷰, 2026년 6월 19일). 그는 이 장치를 "혁명이라는 말로도 모자란다"고 했다(nothing short of revolutionary). 그는 아내가 행복했던 기억을 이야기할 때 이제 그 눈을 바라볼 수 있다고 했다(UC 데이비스, 2026년 6월 15일). 그의 딸은 아버지가 말하던 때를 기억하지 못한다. 해럴은 합성된 목소리로나마 자기 목소리가 어땠는지 딸이 알기를 바란다.

2026년 9월 뉴럴링크는 ALS 환자가 합성된 자기 목소리로 말하는 영상을 공개했다(iPhone in Canada, 2026년 9월 21일). 첫 출력은 프로그래머들의 시험 문구인 "hello world"였다. 회사는 장치가 시험용이고 미국 식품의약국(FDA)의 승인을 받지 않았다고 밝혔다. 이 영상은 회사가 공개한 것이고 동료심사를 거친 자료는 없다. 커서를 움직이는

일과 문장을 말하는 일은 같은 원리 위에 있다. 차이는 디코더가 골라야 하는 답의 수다. 위아래와 좌우를 고르던 디코더가 이제 12만 5,000개 단어 가운데 하나를 고른다.

한 방향 통신: 뇌의 신호를 기계로

지금까지 본 장치는 모두 같은 쪽으로 신호를 보낸다. 뇌에서 나온 신호가 기계로 간다. 사용자가 받는 되먹임은 화면의 커서와 스피커의 합성 음성, 장갑의 움직임이다. 사용자는 이것을 눈과 귀로 확인하고 다음 동작을 정한다. 기계가 뇌에 직접 신호를 넣지는 않는다. 아보가 체스 말을 옮길 때 말의 위치를 알려 준 것도 노트북 화면이었다. 해럴이 자기 문장이 맞게 나왔는지 아는 길도 화면의 글자와 스피커의 소리다.

규제 문서는 기술보다 한발 앞서 있다. NMPA 지도원칙은 BCI를 중추신경 신호를 실시간으로 해독해 외부 장치와 양방향으로 주고받거나 폐루프 되먹임을 구현하는 능동 의료기기로 정의했다(CCTV, 2026년 6월 30일). 폐루프는 장치가 내놓은 결과를 다시 받아 스스로 조정하는 고리다. 지도원칙은 신호를 한 방향으로 모으기만 하는 장치는 적용 범위 밖이라고 적었다. 그런데 석 달 앞서 승인을 받은 NEO는 뇌에서 장갑으로 명령을 보내는 장치다. NMPA가 장갑의 움직임을 어떤 논리로 이 정의에 넣었는지는 공개 자료로 확인하지 못했다. NEO의 임상 논문은 척수 손상 환자의 손 기능 회복을 다루지만 뇌에 전기 자극을 넣지는 않는다(medRxiv, 2024년 9월). 정의는 양방향을 내다보고 시장에 나온 장치는 한 방향에 머물러 있다.

반대 방향의 연구가 없지는 않다. 피츠버그대 연구진은 체성감각피질을 미세 전류로 자극해 로봇팔 손에 닿는 감각을 참여자에게 돌려줬다(Science, 2021년 5월 21일). 체성감각피질은 몸에서 오는 촉감을 받아들이는 대뇌 부위다. 뉴럴링크는 시각피질에 신호를 넣어 시력을 되살리겠다는 장치 블라인드사이트(Blindsight)를 준비하고 있다. 일론 머스크는 2026년 1월 첫 사람 이식 준비를 마쳤고 규제 승인을 기다린다고 했다(India TV, 2026년 1월 29일). 10월 현재 사람에게 심었다는 공식 발표는 없다. 기록과 자극을 한 장치에 담은 시도도 있다. 독일 코어텍(CorTec)의 Brain Interchange는 뇌 신호를 계속 기록하고 해석해 그 자리에서 전기 자극을 넣는다(MedTech Spectrum, 2026년 4월). 이 장치는 2026년 4월 뇌졸중 운동 재활 용도로 FDA 획기적 기기 지정을 받았다. 획기적 기기 지정은 심사를 앞당겨 주는 제도이고 시판 허가와는 다르다. 세 시도 모두 사람에게서 나온 결과가 적거나 아직 없다.

한 방향이라는 사실은 이 기술의 한계이면서 지금의 성격이다. 지금의 BCI는 몸 밖으로 나가는 길을 잃은 사람에게 새 출구를 내 준다. 의도는 손과 입 대신 전극과 디코더를 지나

기계로 나간다. 인공지능을 사람이 실시간으로 지휘하는 장치를 말할 때도 출발점은 이 출구다. 명령이 나가는 길은 열렸고 기계가 내놓은 답은 아직 눈과 귀를 거쳐 돌아온다. 2026년 10월 환자의 일상에 들어간 BCI는 모두 뇌가 보내고 기계가 받는 장치다. 기계가 뇌로 보내는 장치는 아직 소수의 참여자와 승인을 기다리는 계획 속에 있다.

참고 자료

NBC News, Neuralink livestream shows paralyzed person playing chess on laptop, 2024년 3월 20일, https://www.nbcnews.com/tech/tech-news/neuralink-livestream-shows-paralyzed-person-playing-chess-laptop-rcna144374

Mayo Clinic Proceedings, Brain-computer interfaces in medicine(Shih, Krusienski, Wolpaw), 2012년 3월, https://pmc.ncbi.nlm.nih.gov/articles/PMC3497935/

Journal of Comparative Neurology, Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain(Azevedo 외), 2009년 4월 10일, https://pubmed.ncbi.nlm.nih.gov/19226510/

arXiv, A Review of Brain-Computer Interface Technologies: Signal Acquisition Methods and Interaction Paradigms(Wang 외), 2025년 3월, https://arxiv.org/abs/2503.16471

Neuroscience News, AI Speech Neuroprosthesis Restores Voice to ALS Patient, 2026년 7월 17일, https://neurosciencenews.com/ai-speech-neuroprosthesis-als-31072/

Popular Science, 85% of Neuralink implant wires are already detached, says patient, 2024년 5월 21일, https://www.popsci.com/health/neuralink-wire-detachment/

Decrypt, Neuralink Patient's Implants Slipped Out, But He Still Set a Brain Control Record, 2024년 5월 9일, https://decrypt.co/230021/neuralink-implant-patient-gaming-sets-bci-record

Journal of Neural Engineering, Longevity and reliability of chronic unit recordings using the Utah, intracortical multi-electrode arrays, 2021년, https://pubmed.ncbi.nlm.nih.gov/34847547/

MIT Technology Review, Meta has an AI for brain typing, but it's stuck in the lab, 2025년 2월 7일, https://www.technologyreview.com/2025/02/07/1111292/meta-has-an-ai-for-brain-typing-but-its-stuck-in-the-lab/

Precision Neuroscience, Precision Neuroscience Receives FDA Clearance for High-Resolution Cortical Electrode Array, 2025년 4월 17일, https://www.precisionneuro.io/articles/company-news/precision-neuroscience-receives-fda-clearance-for-high-resolution-cortical-electrode-array

BioSpace, Synchron Announces Positive Results from U.S. COMMAND Study of Endovascular Brain-Computer Interface, 2024년 9월 30일, https://www.biospace.com/news/synchron-announces-positive-results-from-u-s-command-study-of-endovascular-brain-computer-interface

Medical Xpress, Minimally invasive brain-computer interface may help those with tetraplegia restore hand functions, 2024년 2월 5일, https://medicalxpress.com/news/2024-02-minimally-invasive-brain-interface-tetraplegia.html

medRxiv, Reclaiming Hand Functions after Complete Spinal Cord Injury with Epidural Brain-Computer Interface(첫 공개본 제목 Reclaiming Motor Functions after Complete Spinal Cord Injury using Epidural Minimally Invasive Brain-Computer Interface), 2024년 9월 첫 공개, https://www.medrxiv.org/content/10.1101/2024.09.05.24313041v4.full

CGTN, China approves world's first invasive BCI medical device, 2026년 3월 14일, https://news.cgtn.com/news/2026-03-14/China-approves-world-s-first-invasive-BCI-medical-device-1Lv2lPRjay4/p.html

arXiv, Clinical Translation of Brain-Computer Interface in China: A Landscape Analysis of Investigator-Initiated Trials, Registered Clinical Trials, and Regulatory Approval(Long Chen 외), 2026년 7월 8일, https://arxiv.org/abs/2607.07185

CCTV(央视网), 明确产品分类及命名规则 两项脑机接口医疗器械指导原则发布, 2026년 6월 30일, https://news.cctv.com/2026/06/30/ARTILunwOXQ7VvkjmiuQXfvX260630.shtml

MIT Technology Review, Brain-computer interface trials are taking off, 2026년 6월 19일, https://www.technologyreview.com/2026/06/19/1139270/brain-computer-interface-trials-are-taking-off/

로이터(Yahoo Finance 전재), Elon Musk's Neuralink says it has 21 participants enrolled in trials, 2026년 1월 28일, https://finance.yahoo.com/news/elon-musks-neuralink-says-21-182925000.html

IEEE Spectrum, UPDATE: Neuralink's Human Tests Progress; Critics Persist, 2023년 12월 6일(2024년 2월 3일 갱신), https://spectrum.ieee.org/neuralink-human-trials

UC Davis Health, Brain-computer interface enables independent, accurate communication for man living with ALS, 2026년 6월 15일, https://health.ucdavis.edu/news/headlines/brain-computer-interface-enables-independent-accurate-communication-for-man-living-with-als/2026/06

TechCrunch, OpenAI invests in Sam Altman's brain computer interface startup Merge Labs, 2026년 1월 15일, https://techcrunch.com/2026/01/15/openai-invests-in-sam-altmans-brain-computer-interface-startup-merge-labs/

arXiv, Towards Developing Brain-Computer Interfaces for People with Multiple Sclerosis(Russo 외), 2024년 4월 7일, https://arxiv.org/abs/2404.04965

이투데이, 사람의 신체·인지 한계 극복…K-문샷 '뇌-컴퓨터 인터페이스' 임무 본격 추진, 2026년 3월 18일, https://www.etoday.co.kr/news/view/2566792

디지털타임스, [이준기의 과·알·세] 빨라지는 BCI 산업화 시계… 美·中 경쟁에 속도내는 韓, 2026년 3월 22일, https://www.dt.co.kr/article/12053043

Decrypt, Second Neuralink Patient Does 3D Design and Plays CounterStrike With His Brain, 2024년 8월 22일, https://decrypt.co/245875/neuralink-update-alex-counterstrike-cad-software-brain-implant

Brown University, Brain computer interface enables rapid communication for two people with paralysis, 2026년 3월 16일, https://www.brown.edu/news/2026-03-16/braingate-rapid-communication

New England Journal of Medicine, An Accurate and Rapidly Calibrating Speech Neuroprosthesis(Card 외), 2024년 8월 14일, https://www.nejm.org/doi/full/10.1056/NEJMoa2314132

iPhone in Canada, ALS Patient Speaks Again After Neuralink Turns His Thoughts Into Spoken Words, 2026년 9월 21일, https://www.iphoneincanada.ca/2026/09/21/neuralink-als-patient-speaks-thoughts/

Science, A brain-computer interface that evokes tactile sensations improves robotic arm control(Flesher 외), 2021년 5월 21일, https://pubmed.ncbi.nlm.nih.gov/34016775/

India TV, Neuralink ready for first human implant of vision-restoring brain device, says Elon Musk, 2026년 1월 29일, https://www.indiatvnews.com/science/neuralink-ready-for-first-human-implant-of-vision-restoring-brain-device-says-elon-musk-2026-01-29-1027988

MedTech Spectrum, CorTec Receives FDA Breakthrough Device Designation for Brain Interchange System in Stroke Rehabilitation, 2026년 4월, https://medtechspectrum.com/news/30/25427/cortec-receives-fda-breakthrough-device-designation-for-brain-interchange-system-in-stroke-rehabilitation.html



17장. 미국과 중국의 경쟁

임상을 주도하는 두 나라

工信部联科〔2025〕164号. 중국 공업정보화부 등 7개 부처가 2025년 7월 23일 함께 서명한 「뇌기계 인터페이스 산업 혁신 발전 추진에 관한 실시의견」의 문서 번호다(중국정부망, 2025년 8월). 서명한 기관에는 국가발전개혁위원회와 국가위생건강위원회와 국가약품감독관리국(NMPA)이 들어 있다. 문서는 2027년까지 전극과 칩과 완제품의 성능을 국제 선진 수준으로 올리겠다고 적었다. 2030년 목표는 세계적 영향력을 가진 선도 기업 2개에서 3개를 키우는 것이다. 제16항은 이식형 BCI 의료기기의 등록을 지도하고 우선 지원한다고 썼다. 연구를 돕는 부처와 허가를 내주는 기관이 같은 문서에 이름을 올렸다.

미국에는 이런 범부처 문서가 없다. 미국 식품의약국(FDA)의 이식형 BCI 지침은 2021년 5월에 나온 최종판이 지금도 기준이다(FDA, 2021년 5월 20일). 2026년에 FDA가 한 일은 새 지침이 아닌 개별 기기의 획기적 기기 지정이었다. 미국의 이식 임상은 뉴럴링크와 싱크론과 파라드로믹스(Paradromics) 같은 기업이 각자 FDA와 협의하며 끌고 간다.

임상의 규모는 등록부로 가늠한다. 임상 등록부는 사람을 대상으로 하는 시험의 계획과 진행 상황을 미리 공개하는 명부다. 2026년 10월 14일 미국 국립보건원의 등록부 ClinicalTrials.gov에서 BCI를 검색하면 264건이 나온다(ClinicalTrials.gov, 2026년 10월 14일 조회). 미국에 시험 장소가 있는 등록은 91건이고 중국은 26건이다. 이식 관련 어휘가 들어간 등록만 세면 미국은 34건이고 중국은 10건이다. 이 비교는 중국을 작게 보이게 한다. 중국 연구진은 대개 자국의 중국임상시험등록센터(ChiCTR)에 시험을 올리기 때문이다.

중국 등록부를 따로 센 연구가 있다. 하이난대학과 베이징 톈탄병원 연구진은 2026년 6월 30일까지 ChiCTR에 오른 BCI 임상을 134건으로 집계했다(arXiv, 2026년 7월 8일). 이 논문은 동료 심사 전 사전 공개본이다. 134건 가운데 비침습형이 106건으로 79.1퍼센트를 차지했고 치료 대상 질환인 적응증으로 보면 65퍼센트가 뇌졸중 재활이었다. 피질 내에 전극을 넣는 등록 시험은 4건이고 혈관 내 방식은 1건이었다. 경막

바깥이나 아래에 전극을 두는 반침습 연구는 등록 시험 없이 연구자 주도 임상으로만 진행됐다. 연구자 주도 임상은 기업이 아닌 병원 의사가 계획하고 책임지는 임상이다.

숫자는 두 나라의 서로 다른 짜임을 보여 준다. 미국에서는 기업이 이식형 시험을 등록하고 참여자를 한 명씩 늘린다. 중국에서는 비침습 재활 시험이 등록부를 채우고 이식형은 병원 연구와 신속 심사로 빠르게 넘어간다.

중국, 2026년 3월 세계 첫 침습형 BCI 의료기기 판매 승인

NMPA 혁신의료기기 코너에는 2026년 3월 13일자 글 「첫 침습형 뇌기계 인터페이스 의료기기 시판 승인」(首款侵入式脑机接口医疗器械获批上市)이 있다(NMPA, 2026년 3월 13일). 글의 색인 번호는 NMPA 색인 기준 XZXK-2026-10047이다. 앞 장에서 본 대로 여기의 "침습형"은 중국 규제 범주의 이름이다. 학술 분류로 보면 이 장치는 경막 바깥에 심는 반침습 이식형이다. 그래서 세계 첫 승인의 정확한 범위는 반침습 이식형 BCI다. 중국 연구진의 논문도 이를 "세계 첫 반침습 BCI 규제 승인"으로 적었다(the first regulatory approval of a semi-invasive BCI worldwide)(arXiv, 2026년 7월 8일). 미국에는 2026년 10월 현재 영구 이식형 BCI의 시판 허가가 없다. 미국에서 위험도가 높은 기기는 시판 전 승인(PMA)을 받아야 하고 이 허가는 안전성과 효과를 임상 자료로 입증해야 나온다. 프리시전 뉴로사이언스의 Layer 7은 이미 팔리는 기기와 실질적으로 같다는 점을 보여 받는 510(k) 경로를 거쳤다. 이 장치의 허가 조건은 최대 30일 이식이다(Precision Neuroscience, 2025년 4월 17일).

승인 제품은 뉴라클(博睿康)의 NEO이고 정식 명칭은 「이식형 뇌기계 인터페이스 손 운동기능 대상 시스템」이다. 중국은 의료기기를 위험도에 따라 1류에서 3류로 나눈다. 몸에 심는 기기는 대개 관리가 엄격한 3류다. NEO는 2024년 8월 혁신의료기기 특별심사에 들어갔다(虎嗅, 2026년 7월 17일). 특별심사는 새로운 기술의 기기를 심사 순서에서 앞당기고 심사관이 개발 단계부터 지도하는 NMPA 절차다. 진입에서 승인까지 약 19개월이 걸렸다.

승인 범위는 좁다. 대상은 목의 척수를 다쳐 어깨와 팔은 움직이지만 손을 쥐지 못하는 성인이다. 나이 상한은 자료에 따라 60세 또는 65세로 다르다. 확증 임상은 미국 등록부에도 NCT06990412로 올라 있다. 2025년 5월 시작해 32명을 등록했고 지금은 완료 상태다. 중국 자료는 11개 기관의 32례에 앞선 타당성 임상 4례를 더해 36례로 적는다. 회사 자료를 인용한 보도에 따르면 6개월 시점에 장치 없이도 손 기능 점수가

기준 이상 오른 사람이 22명(68.8퍼센트)이었다(虎嗅, 2026년 7월 17일). 기기 관련 중대 이상반응은 없었다고 했다. 이 결과는 동료 심사 논문이 아닌 회사 자료다.

승인 뒤의 일정은 빨랐다. 보험 수가는 건강보험에서 의료 행위별로 정한 금액이다. 중국 국가의료보장국(国家医疗保障局)은 제품이 나오기 1년 전인 2025년 3월 침습형 BCI 삽입비와 제거비 항목을 미리 만들었다(新京报, 2025년 3월 12일). 후베이성은 같은 달 31일 삽입비 상한을 6,552위안으로 정했다(澎湃新闻, 2025년 3월). 이 금액은 수술 서비스에 대한 비용이며, 기기 가격과는 다르다. 승인 뒤 열흘이 안 된 3월 22일에는 국가 의료보험 소모품 코드가 나왔다(BioSpectrum Asia, 2026년 10월 1일). 기기는 1세트 47만 5,000위안으로 27개 성의 조달 목록에 올랐다(虎嗅, 2026년 10월 13일). 국가 기본 의료보험이 기기 구입비를 지원하지는 않는다. 상하이의 민간 보충보험이 2026년판부터 BCI 수술 소모품을 보장한다(VCBeat, 2026년 5월 28일).

첫 상업 이식은 2026년 7월 13일 상하이 푸단대 부속 화산병원(华山医院)에서 이뤄졌다. 상장 심사 자료를 인용한 보도는 7월 이후 13세트가 팔렸고 그중 8세트가 이식됐다고 전했다(虎嗅, 2026년 10월 13일).

뉴럴링크의 임상 참여자 21명

앞 장에서 본 21명은 뉴럴링크가 2026년 1월 28일 기준으로 밝힌 임상 등록 인원이다(로이터, 2026년 1월 28일). 이식 인원이 아닌 등록 인원이라는 점을 함께 적어야 한다. 시간 순서로 놓으면 증가 속도가 보인다. 첫 이식은 2024년 1월이었다. 2025년 6월에 7명이었고(国家医疗保障局, 2026년 4월 16일) 9월에 12명이었다. 그 뒤 넉 달 만에 21명이 됐다.

그 뒤의 숫자는 회사가 낸 것이 아니다. 뉴럴링크를 추적하는 제3자 사이트 뉴라 포드(Neura Pod)는 2026년 5월 20일 캐나다의 ALS 환자가 26번째 참여자로 이식을 받았다고 전했다. 근거는 환자 측 게시물이다. 이 사이트는 10월 기준 공개적으로 알려진 환자를 27명 이상으로 적는다(Neura Pod, 2026년). 두 숫자는 출처 등급이 낮은 비공식 집계다. 뉴럴링크는 2026년 10월 1일 블로그에서 참여자들이 5만 시간 넘는 신경 데이터를 보냈다고 밝혔다(TeslaNorth, 2026년 10월 2일). 이 글에도 참여자 총수는 없다.

참여자는 여러 나라에 흩어져 있다. 미국 임상 PRIME(NCT06429735)은 2024년 1월 시작했고 등록부상 예상 인원은 15명이다. 시험 장소는 마이애미와 피닉스다. 캐나다 보건부는 2024년 11월 토론토 병원의 임상을 승인했다. 영국은 2025년 7월 런던과

뉴캐슬에서 시험을 시작했다. 아랍에미리트는 2025년 5월 아부다비에서 모집을 시작했다. 말을 되찾는 장치를 시험하는 VOICE는 2025년 10월 댈러스에서 시작했고 예상 인원은 6명이다(Neura Pod, 2026년).

이 시험들은 모두 등록부상 조기 타당성 연구다. 조기 타당성 연구는 적은 수의 사람에게 장치를 처음 써 보며 안전성과 작동 여부를 살피는 단계다. 시판 허가에 쓰는 대규모 확증 시험은 피보털(pivotal) 시험이라고 부른다. 뉴럴링크가 피보털 시험에 들어갔다는 발표는 2026년 10월 현재 없다.

돈은 넉넉하다. 뉴럴링크는 2025년 6월 시리즈 E로 6억 5,000만 달러를 모았다(TechCrunch, 2025년 6월 2일). 2026년에 새 투자 라운드를 공식 발표하지는 않았다. 비상장 주식의 2차 거래에서 나온 기업가치 추정치들은 회사 발표가 아니라서 여기에 쓰지 않는다.

뉴럴링크와 NEO를 나란히 놓으면 차이가 드러난다. 뉴럴링크는 네 나라에서 2년 동안 21명을 등록했다. NEO는 11개 병원에서 2025년 몇 달 사이에 확증 임상 32례를 마쳤다(MIT Technology Review, 2026년 6월 1일). 뉴럴링크는 피질 안에 실을 넣는 장치이고 NEO는 경막 바깥에 얹는 장치다. 위험이 다른 장치이므로 참여자 수만으로 우열을 가릴 수는 없다. 두 숫자가 보여 주는 것은 환자를 모으고 수술을 마치는 속도의 차이다.

미국, 중국, 호주의 누계 100여 명

세 나라의 이식형 BCI 참여자를 합산한 공식 집계는 없다. 숫자는 무엇을 세느냐에 따라 달라진다. 첫째 기준은 등록된 임상만 세는 방식이다. 뉴럴링크 21명과 싱크론 약 10명과 파라드로믹스 1명과 NEO 36명을 더하면 68명이다. NEO의 상업 이식 8례를 더하면 76명이다. 둘째 기준은 2026년 10월까지 기업과 창업자가 공개한 장기 이식 수치를 모두 더하는 방식이다. 중국의 연구자 주도 임상에서 심은 사람까지 들어가 약 160명이 된다. 셋째는 앞 장에서 본 연구자 추정 약 150명이다(MIT Technology Review, 2026년 6월 19일). 이 추정은 세 나라가 아닌 세계 전체를 셌다.

여기서 말하는 100여 명은 2026년 6월 말 기준 장기 이식 참여자를 보수적으로 센 숫자다. 미국과 호주는 기업의 공식 발표를 썼고 중국은 중국 연구진 논문의 집계를 썼다. 미국과 호주 쪽은 뉴럴링크와 싱크론과 파라드로믹스를 더해 32명이다. 논문의 표는 NEO 36명과 베이나오 1호(北脑一号) 30명을 포함해 중국 이식 73명을 적었다(arXiv, 2026년 7월 8일). 합계는 약 105명이다. 수술 중에 잠깐 기록하거나 30일 안에 빼는 장치는 넣지 않았다. 1998년부터 2023년까지 대학 연구진이 심은 67명도 따로 두었다.

미국과 호주 쪽 숫자는 작다. 싱크론은 호주 멜버른의 초기 임상 SWITCH에서 4명에게 장치를 심었다(Design News, 2024년). 미국 임상 COMMAND에서는 6명에게 장치를 심었다(BioSpace, 2024년 9월 30일). 2026년 4월부터는 멜버른과 시드니의 세 병원에서 차세대 장치를 운동신경세포병 환자 10명에게 심는 시험을 시작했다(Inside BCI, 2026년 3월 25일). 이 시험의 실제 이식 수는 공개하지 않았다. 파라드로믹스는 2026년 6월 미시간대학병원에서 첫 참여자에게 장치를 심었다(Paradromics, 2026년 6월).

6월 이후 늘어난 숫자는 거의 모두 중국에서 나왔다. 상하이 스테이지(阶梯医疗)의 창업자 자오정퉈(赵郑拓)는 2026년 9월 18일 운동기능 임상 검증 40여 례를 마쳤다고 말했다(上观新闻, 2026년 9월 19일). 베이나오 1호의 박람회 현장 직원은 7월 31일 기준 인체 이식이 40여 례라고 했다(华夏时报, 2026년 9월 5일). NEO는 상업 이식 8례를 더했다. 상하이의 뇌호(脑虎科技)는 2026년 7월 화산병원 등에서 완전 이식형 장치의 등록 임상을 시작했지만 참여자 수는 밝히지 않았다(新浪科技, 2026년 7월 7일). 이 수치들은 창업자나 직원의 발언이고 동료 심사를 거치지 않았다. 그래도 넉 달 사이 공개 수치 합계는 105명 안팎에서 160명 안팎으로 늘었다. 늘어난 50여 명은 중국 몫이다.

한국은 이 집계에 들어갈 숫자가 없다. 같은 날 ClinicalTrials.gov 조회에서 한국에 시험 장소가 있는 이식형 BCI 등록은 0건이었다. 서울대병원에서 진행 중인 지브레인의 임상은 뇌전증 발작 부위를 찾는 진단용 전극 시험이다(이투데이, 2024년 12월 24일). 과학기술정보통신부는 2026년 3월 18일 K-문샷 BCI 임무를 발표했다(이투데이, 2026년 3월 18일). 서울대병원이 총괄하는 국책 컨소시엄은 2028년 이후 인체 임상을 계획한다(코메디닷컴, 2026년 6월 16일).

속도에서 중국이 앞서는 이유

첫째 이유는 행정 기관들이 서로를 기다리지 않은 데 있다. 아시아 의료 산업 매체 BioSpectrum Asia는 2026년 10월 1일 기사에서 NEO 전후의 시간표를 이렇게 정리했다. 수가 항목이 2025년 3월 먼저 생겼고 승인은 2026년 3월 13일에 났다. 보험 코드와 상하이 목록 등재와 조달 등재가 3월 22일부터 24일 사이에 이어졌다. 5월에는 병원 주도 임상연구 조례가 시행됐고 6월에는 NMPA의 분류 지도원칙이 나왔다. 이 매체는 각 기관이 앞 단계를 기다리지 않고 병렬로 준비했다고 분석했다. 미국 경쟁사 파라드로믹스도 중국의 진짜 이점은 기술이 아닌 규제와 행정의 조정이라고 썼다(Paradromics, 2026년 4월 30일).

둘째 이유는 설계의 선택이다. 호주 시드니공과대학의 BCI 연구자 아비나시 싱(Avinash Singh)은 경막 외 설계는 출혈과 흉터와 신호 저하의 위험이 낮아 규제 부담도 적다고 했다(MIT Technology Review, 2026년 6월 1일). 속도의 일부는 제도가 아닌 장치의 성격에서 왔다. 셋째 이유는 환자와 병원이다. TechCrunch는 2026년 2월 22일 기사에서 큰 환자 집단과 낮은 연구 비용을 이유로 들었다. 일리노이대의 정보과학자 메이천 쑨(Meicen Sun)은 중국 환자들이 이 기술을 반기는 반면 미국과 서유럽에는 거부감이 있다고 했다(MIT Technology Review, 2026년 6월 1일).

넷째 이유는 보험이다. TechCrunch 기사는 중국은 국가 보험 하나가 결정하지만 미국은 민간 보험사가 하나씩 따로 정한다고 비교했다. 7개 부처 의견도 첫 제품의 보험 보상 정책을 약속했다. 미국의 공공 의료보험은 시간이 걸린다. 메디케어는 65세 이상 노인을 위한 미국의 공공 의료보험이다. 파라드로믹스는 2016년부터 2019년까지 승인된 신기술 64건 가운데 44퍼센트만 메디케어 급여를 받았다는 연구를 인용했다(Paradromics, 2026년 4월 30일). 이 글은 메디케어 당국의 평가에 약 5년이 걸린다고 적었다.

반론도 있다. 블랙록 뉴로테크 공동창업자 플로리안 졸츠바허(Florian Solzbacher)는 속도가 아닌 장기 안전성과 임상 근거가 보급을 따라가는지가 문제라고 했다(The Debrief, 2026년 7월 19일). 그는 규제 절차를 "안전을 내던지지"(throwing safety overboard) 않고도 고칠 수 있다고 했다. 파라드로믹스는 먼저 승인받았다는 사실이 중국의 기술 우위를 뜻하지는 않는다고 썼다(being first to achieve approval does not mean China leads in technical capabilities). 경막 외 전극은 단일 뉴런을 읽는 피질 내 장치보다 정보량이 적다는 이유다. 경쟁사의 글이지만 근거는 사실이다. NEO의 자료는 36례에 18개월 추적이라 장기 안정성을 말하기에는 얇다. 중국 연구진의 논문도 중국 침습 연구 대부분이 장기 추적 자료 요구를 아직 넘지 못했다고 인정했다(arXiv, 2026년 7월 8일). 두 지적은 맞다. 다만 중국이 앞선 지점은 근거의 양이 아니다. 승인에서 수가와 조달과 병원까지 이어지는 길을 한꺼번에 깔아 둔 데 있다.

미국이 앞선 영역도 분명하다. 해독 성능에서는 스탠퍼드 연구진이 2023년 시도한 말을 1분에 62단어로 해독했다(Nature, 2023년 8월 23일). 앞 장에서 본 해럴의 사례도 미국 연구다. 장기 데이터에서도 미국이 앞선다. 유타 어레이는 2004년부터 사람에게 쓰였고 1998년부터 2023년까지의 67명은 대부분 미국 연구진이 심었다. 벤처 규모도 다르다. 2026년 9월 프리시전 뉴로사이언스는 한 번에 2억 5,000만 달러를 모았다(GlobeNewswire, 2026년 9월 24일). 사이언스(Science Corp)는 3월에 2억 3,000만 달러를 모았다(SiliconANGLE, 2026년 3월 5일). 중국 스테이지의 2026년 3월 라운드는

5억 위안으로 약 7,000만 달러였다(新浪财经, 2026년 3월 15일). 재활 분야 논문은 중국이 398편으로 미국의 291편보다 많지만 피인용은 미국이 높았다(Frontiers in Human Neuroscience, 2024년 10월 30일).

메이천 쑨은 이 차이를 "탁월함과 접근성은 승리를 정반대로 정의한다"는 말로 정리했다(Being exceptional and being accessible are two diametrically opposed definitions of winning)(MIT Technology Review, 2026년 6월 1일).

2026년 10월 현재 미국 FDA가 시판을 허가한 영구 이식형 BCI는 없다. 중국에서는 반침습 이식형 BCI인 NEO가 팔리고 병원에서 이식된다. 미국은 더 빠른 해독과 더 긴 이식 기록과 더 큰 자본을 가졌다. 중국은 승인과 보험 코드와 조달 목록을 열흘 남짓한 사이에 맞춰 놓았다. 한국은 두 나라 가운데 한 방식만 고를 필요가 없다. 해독 성능과 장기 안전성의 근거를 쌓는 동시에, 허가와 비용 부담과 병원 운영을 연결할 제도를 준비해야 한다. 기술이 완성된 뒤에 이 절차를 시작하면 환자에게 전달되는 시간은 다시 늦어진다.

참고 자료

中国政府网, 七部门关于推动脑机接口产业创新发展的实施意见, 2025년 8월, https://www.gov.cn/zhengce/zhengceku/202508/content_7035603.htm

FDA, Implanted Brain-Computer Interface (BCI) Devices for Patients with Paralysis or Amputation. Non-clinical Testing and Clinical Considerations, 2021년 5월 20일, https://www.fda.gov/media/120362/download

ClinicalTrials.gov, API v2 studies 검색(검색어 brain-computer interface, 국가별 조회), 2026년 10월 14일 조회, https://clinicaltrials.gov/api/v2/studies

arXiv, Clinical Translation of Brain-Computer Interface in China: A Landscape Analysis of Investigator-Initiated Trials, Registered Clinical Trials, and Regulatory Approval(Long Chen 외), 2026년 7월 8일, https://arxiv.org/abs/2607.07185

国家药品监督管理局(NMPA), 首款侵入式脑机接口医疗器械获批上市, 2026년 3월 13일, https://www.nmpa.gov.cn/zhuanti/cxylqx/cxylqxlm/20260313134742156.html

Precision Neuroscience, Precision Neuroscience Receives FDA Clearance for High-Resolution Cortical Electrode Array, 2025년 4월 17일, https://www.precisionneuro.io/articles/company-news/precision-neuroscience-receives-fda-clearance-for-high-resolution-cortical-electrode-array

虎嗅, 全球首例！中国脑机接口落地，370万名脊髓损伤患者迎来新希望, 2026년 7월 17일, https://www.huxiu.com/article/4876013.html

ClinicalTrials.gov, A Clinical Trial to Evaluate the Safety and Efficacy of an Implantable Wireless Brain-Computer Interface (BCI) System NEO in Patients With Tetraplegia(NCT06990412), 2026년 8월 25일 갱신, https://clinicaltrials.gov/study/NCT06990412

新京报, 国家医保局为脑机接口相关医疗服务价格单独立项, 2025년 3월 12일, https://m.bjnews.com.cn/detail/1741751269129924.html

澎湃新闻, 966元起！湖北公布全国首个脑机接口医疗服务价格，影响有多大？, 2025년 3월, https://m.thepaper.cn/newsDetail_forward_30533265

BioSpectrum Asia, Approved, Coded, Listed: China's neurotech speed run, 2026년 10월 1일, https://www.biospectrumasia.com/analysis/25/28439/approved-coded-listed-chinas-neurotech-speed-run.html

虎嗅, "脑机接口第一股"靠什么赚钱？, 2026년 10월 13일, https://www.huxiu.com/article/4897157.html

VCBeat, Commercial Insurance Covers Brain-Computer Interface for the First Time: Full Payment Chain Now Operational, 2026년 5월 28일, https://www.vcbeathealth.com/article/23873

国家医疗保障局(NHSA), 医保倾力相助，脑机接口正从"奇迹"走向"日常", 2026년 4월 16일, https://www.nhsa.gov.cn/art/2026/4/16/art_14_20229.html

로이터(Yahoo Finance 전재), Elon Musk's Neuralink says it has 21 participants enrolled in trials, 2026년 1월 28일, https://finance.yahoo.com/news/elon-musks-neuralink-says-21-182925000.html

Neura Pod, Neuralink Update — 2026, 2026년, https://www.neurapod.com/blog/neuralink-update-2026

TeslaNorth, Neuralink Sets New BCI Record After Training on 50,000 Hours of Brain Data, 2026년 10월 2일, https://teslanorth.com/2026/10/02/neuralink-pretraining-50000-hours-bci-record/

ClinicalTrials.gov, Precise Robotically IMplanted Brain-Computer InterfacE(NCT06429735), 2026년 1월 9일 갱신, https://clinicaltrials.gov/study/NCT06429735

TechCrunch, Elon Musk's Neuralink closes a $650M Series E, 2025년 6월 2일, https://techcrunch.com/2025/06/02/elon-musks-neuralink-closes-a-650m-series-e

MIT Technology Review, China has approved the world's first invasive brain-computer chip—here's what's next, 2026년 6월 1일, https://www.technologyreview.com/2026/06/01/1138133/china-world-first-brain-chip/

MIT Technology Review, Brain-computer interface trials are taking off, 2026년 6월 19일, https://www.technologyreview.com/2026/06/19/1139270/brain-computer-interface-trials-are-taking-off/

Design News, (제목 미확인, URL 기준), 2024년, https://www.designnews.com/medical/synchron-continues-to-amaze-after-releasing-world-s-first-brain-computer-interface

BioSpace, Synchron Announces Positive Results from U.S. COMMAND Study of Endovascular Brain-Computer Interface, 2024년 9월 30일, https://www.biospace.com/news/synchron-announces-positive-results-from-u-s-command-study-of-endovascular-brain-computer-interface

Inside BCI, Synchron Returns to Melbourne for First Trial of Next-Generation Stentrode, 2026년 3월 25일, https://insidebci.com/news/2026-03-25-synchron-returns-to-melbourne-for-first-trial-of-next-generation-stentrode/

Paradromics, Paradromics Completes First Connexus® BCI Human Implant, 2026년 6월, https://paradromics.com/news/paradromics-completes-first-human-brain-computer-interface-bci-implantation/

腾讯新闻(上观新闻 전재), 阶梯医疗完成40余例脑机接口注册临床验证，预计明年将达到4096通道, 2026년 9월 19일, https://news.qq.com/rain/a/20260919A06BAP00

腾讯新闻(华夏时报 전재), 2026中国国际福祉博览会现场直击：北脑一号完成四十例人体植入，翔宇机器人加速走向临床, 2026년 9월 5일, https://news.qq.com/rain/a/20260905A06UMT00

新浪科技, 国内首款硬膜下植入式脑机接口产品启动注册临床试验，宣称不伤脑、不烧脑, 2026년 7월 7일, https://finance.sina.com.cn/tech/digi/2026-07-07/doc-inifyquw8334133.shtml

이투데이, [단독] "한국판 뉴럴링크 나올까"…국내 첫 침습형 대뇌피질전극 임상 허가, 2024년 12월 24일, https://www.etoday.co.kr/news/view/2431985

이투데이, 사람의 신체·인지 한계 극복…K-문샷 '뇌-컴퓨터 인터페이스' 임무 본격 추진, 2026년 3월 18일, https://www.etoday.co.kr/news/view/2566792

코메디닷컴, 서울대병원, 300억 양방향 뇌-로봇 개발 참여…BCI 칩 이식 임상 총괄, 2026년 6월 16일, https://kormedi.com/2825235/

Paradromics, China's Recent BCI Developments & What They Mean for the U.S., 2026년 4월 30일, https://paradromics.com/blog/china-bci-developments/

TechCrunch, China's brain-computer interface industry is racing ahead, 2026년 2월 22일, https://techcrunch.com/2026/02/22/chinas-brain-computer-interface-industry-is-racing-ahead

The Debrief, China Clears Brain-Computer Interfaces for Commercial Use: Is It Time for a Global Coalition?, 2026년 7월 19일, https://thedebrief.org/china-clears-brain-computer-interfaces-for-commercial-use-is-it-time-for-a-global-coalition/

Nature, A high-performance speech neuroprosthesis(Willett 외), 2023년 8월 23일, https://www.nature.com/articles/s41586-023-06377-x

GlobeNewswire, Precision Neuroscience Closes Oversubscribed $250M Series D Led by Pershing Square Inc., the Ackman Oxman Institute, and a Leading Life Sciences Investment Fund, 2026년 9월 24일, https://www.globenewswire.com/news-release/2026/09/24/3368254/0/en/precision-neuroscience-closes-oversubscribed-250m-series-d-led-by-pershing-square-inc-the-ackman-oxman-institute-and-a-leading-life-sciences-investment-fund.html

SiliconANGLE, Brain-computer interface startup Science raises $230M, 2026년 3월 5일, https://siliconangle.com/2026/03/05/brain-computer-interface-startup-science-raises-230m/

新浪财经, 阿里腾讯首次联手押注脑机接口 阶梯医疗完成5亿元融资 年内临床植入数剑指40例, 2026년 3월 15일, https://finance.sina.com.cn/jjxw/2026-03-15/doc-inhramvn6566512.shtml

Frontiers in Human Neuroscience, Knowledge mapping and research trends of brain-computer interface technology in rehabilitation: a bibliometric analysis, 2024년 10월 30일, https://pmc.ncbi.nlm.nih.gov/articles/PMC11557533/



18장. 양방향 통신까지 10년

기계의 신호가 뇌로 들어오는 단계

펜을 쥘 때와 캔을 잡을 때와 공을 감쌀 때 손가락에 닿는 모양은 서로 다르다. 시카고대와 피츠버그대 공동 연구진은 2025년 1월 이 차이를 손에 닿는 물건 없이 만들어 냈다(Science, 2025년 1월 16일). 참여자 c1은 목 척수를 다쳐 손의 감각을 잃은 사람이다. 연구진은 그의 체성감각피질 손 영역에 심은 전극 아홉 개를 한꺼번에 자극해 세 손가락에 모서리 모양의 감각을 일으켰다. 모서리가 세 손가락을 가로질러 한 줄로 이어지면 펜을 쥔 느낌이 났다. 모서리가 손가락 길이 방향으로 나란하면 캔을 위에서 잡은 느낌이 났다. 모서리 없이 감각이 겹치기만 하면 공이었다. c1은 무작위로 주어진 세 물건을 79퍼센트의 정확도로 맞혔다.

같은 연구 계열의 앞선 실험은 이런 감각이 기계 조작을 바꾼다는 사실을 보여 줬다. 2021년 참여자 네이선 코플랜드는 로봇팔 손에 달린 센서의 신호를 뇌 자극으로 받았다(Science, 2021년 5월 21일). 물건을 집어 옮기는 과제의 시간 중앙값은 20.9초에서 10.2초로 줄었다. 이 실험들은 반대 방향의 길도 사람에게서 열린다는 증거다.

양방향이라는 말에는 두 뜻이 섞여 있다. 하나는 뇌 상태를 읽어 치료 자극을 스스로 조절하는 장치다. 뇌심부자극은 뇌 깊은 곳에 전극을 심고 전기 자극으로 떨림 같은 증상을 줄이는 치료다. 메드트로닉은 파킨슨병용 적응형 뇌심부자극의 FDA 승인을 발표했다(메드트로닉, 2025년 2월 24일). 적응형 자극은 장치가 뇌 신호를 계속 읽고 그 신호에 맞춰 자극의 세기를 바꾸는 방식이다. 뇌전증 발작의 조짐을 읽어 자극을 넣는 뉴로페이스의 RNS는 2013년 11월에 이미 FDA 승인을 받았다(NeuroPace, 2013년 11월 14일). 이런 장치의 사용자는 의식적인 명령을 보내지 않고 감각도 받지 않는다. 다른 하나는 사용자의 의도를 읽어 기계를 움직이고 기계의 상태를 감각으로 돌려주는 장치다. 이 장이 말하는 양방향은 뒤의 뜻이다. 앞의 장치는 병원에서 팔리고 있고 뒤의 장치는 연구 참여자 몇 명에게 머물러 있다.

신경에 신호를 넣는 장치 가운데 오래되고 널리 퍼진 것은 인공와우다. 인공와우는 소리를 전기 신호로 바꿔 청신경을 직접 자극하는 장치다. 2019년 12월까지 세계에 이식된 등록 기기는 약 73만 6,900대다(미국 국립난청소통장애연구소, 2024년 6월 13일

갱신). 인공와우는 뇌 신호를 읽지 않으므로 BCI에 들지 않지만 기계가 만든 신호를 신경이 지각으로 받아들이는 선례다.

시각은 두 길로 나뉜다. 스페인 미겔 에르난데스대의 에두아르도 페르난데스 연구진은 16년 동안 빛을 보지 못한 참여자의 시각피질에 전극 96개를 6개월 동안 심었다(Journal of Clinical Investigation, 2021년 12월). 참여자는 여러 전극을 함께 자극받아 일부 글자와 물체의 윤곽을 알아봤다. 이 연구진은 2025년 11월 논문 제목에 "양방향 피질 시각 보철을 향한 한 걸음"이라는 말을 넣었다(a step toward a bidirectional cortical visual prosthesis)(Science Advances, 2025년 11월). 다른 길은 망막이다. 사이언스 코퍼레이션의 PRIMA는 망막 아래에 심는 칩으로 안경이 쏘는 근적외선을 전기 자극으로 바꾼다. 환자 38명이 참여한 임상 논문이 2025년 10월에 나왔고(NEJM, 2025년 10월 20일) PRIMA는 2026년 7월 유럽에서 CE 인증을 받아 출시됐다(Science Corporation, 2026년 7월 22일). 이 장치가 신호를 넣는 곳은 뇌가 아닌 망막이다. 미국 FDA의 승인은 10월 현재 없다.

양방향 장치가 언제 사람들 곁에 올지에 대한 말은 엇갈린다. 앞 장에서 본 스테이지 창업자 자오정둬는 5년 안에 인공 시각과 청각 같은 감각 복원이 돌파를 이룬다고 내다봤다(상하이증권보, 2026년 1월 12일 게시). 10년 안팎에는 미세침습 시스템이 "의료 소비와 일반 소비의 장면을 낳을 것"이라고 했다(有望催生医疗消费乃至普通消费场景). 그는 이 분야 기업의 창업자다. 블룸버그는 뉴럴링크가 2030년까지 건강한 사람에게 장치를 넣기를 희망한다고 전했다(Bloomberg, 2025년 9월 19일). 반대쪽 근거도 있다. 일론 머스크는 2017년 8년에서 10년 안에 건강한 사람이 컴퓨터와 연결하는 이식을 받을 수 있다고 했다. MIT 테크놀로지 리뷰는 그 약속을 믿지 말라고 썼고(MIT Technology Review, 2017년 4월 22일) 그 기한은 이미 지났다. 피츠버그대의 제니퍼 콜린저는 처음 시판될 장치에는 감각 되먹임이 들어가지 않는다고 NPR에 말했다(NPR, 2025년 6월 30일). 이 발언은 기사 본문을 열람하지 못하고 검색 요약으로 확인한 내용이다. 이름을 밝힌 신경과학자가 양방향은 10년 안에 안 된다고 못 박은 최근 발언은 찾지 못했다. 10년이라는 시간은 이 발언들과 실험 결과를 바탕으로 저자가 잡은 실용화 전망이다. 사람에게 감각을 만들어 주는 실험은 성공했다. 이제 그 기능을 시판 장치에 담고 수년 동안 안전하게 유지할 수 있는지 확인해야 한다. 이 전망이 건강한 사람까지 널리 쓰게 될 시점을 보장하는 것은 아니다. 연구와 제도는 감각 복원, 장기 사용, 적용 대상의 확대를 각각 확인하며 준비해야 한다.

알파고와 챗GPT의 교훈: 사람들이 받아들이는 순간

2016년 3월 9일부터 15일까지 서울에서 이세돌 9단과 알파고가 다섯 판을 두었고 알파고가 4대 1로 이겼다(영국바둑협회, 2018년 10월 6일 갱신). 구글 딥마인드의 데미스 허사비스는 첫 판을 중국에서만 온라인으로 6,000만 명이 봤다는 말을 들었다고 트위터에 적었다. 텔레비전을 합친 세계 시청자는 1억 명이 넘는다고 했다(Engadget, 2016년 3월 10일). 이 숫자는 회사 쪽의 추정이다. 바둑을 둘 줄 모르는 사람도 결과는 알아들었다. 기계가 사람을 이겼다는 사실은 따로 설명할 필요가 없었다.

6년 뒤 두 번째 순간이 왔다. 오픈AI는 2022년 11월 30일 챗GPT를 공개했다(OpenAI, 2022년 11월 30일). 스위스 은행 UBS는 챗GPT의 2023년 1월 월간 활성 사용자를 약 1억 명으로 추정했다(Euronews, 2023년 2월 2일). 이 수치는 오픈AI의 공식 발표가 아닌 UBS의 추정이다. 같은 보도에서 틱톡은 세계 출시 뒤 1억 명까지 약 9개월이 걸렸고 인스타그램은 2년 반이 걸렸다. 오픈AI는 공개 글에서 챗GPT가 그럴듯하지만 틀린 답을 낼 수 있다고 밝혔다. 사람들은 그 한계를 알면서도 써 보고 쓸 곳을 찾았다. 알파고 때 사람들은 대국을 구경했다. 챗GPT 때는 브라우저를 열고 직접 써 봤다. 두 번 모두 사람들은 새 장치를 사지 않았고 몸에 무언가를 넣지도 않았다.

사람이 새 기술을 받아들이는 이유를 설명하는 이론으로 기술 수용 모델이 있다. 기술 수용 모델(Technology Acceptance Model)은 쓸모가 있다는 인식과 쓰기 쉽다는 인식이 사용 의향을 정한다고 보는 이론이다. 프랑스 보르도의 연구진은 이 틀로 뇌졸중 생존자 140명에게 설문하고 12명을 면접했다(Journal of Neural Engineering, 2026년 4월 13일). BCI 재활을 쓸 의향은 10점 만점에 평균 8.48점이었다. 의향을 가른 것은 쓸모가 있다는 인식이었고 이 인식은 과학적 근거가 있다는 믿음에 크게 좌우됐다. 쓰기 쉽다는 인식은 의향에 직접 영향을 주지 않았다. 중국 주민 1,923명을 조사한 연구에서는 건강과 사회경제적 지위와 사회적 지지와 배우는 능력이 높을수록 BCI를 잘 받아들였다(Frontiers in Human Neuroscience, 2024년 10월 30일). 나이가 많을수록 수용도는 낮았고 성별은 의미 있는 차이를 만들지 않았다. 두 연구는 환자의 재활이나 기술 일반에 대한 태도를 물었다. 2025년 이후 건강한 성인에게 소비자용 BCI를 쓰겠느냐고 물은 설문은 찾지 못했다.

손에 쥐는 장치는 빠르게 퍼진다. 미국 성인의 스마트폰 보유율은 2011년 첫 조사에서 35퍼센트였고 2025년 조사에서 91퍼센트였다(퓨리서치센터, 2025년 11월 20일). 수술이 필요한 인공와우는 2019년 말까지 세계 등록 대수가 100만 대에 미치지 못했다. 두 장치는 사용 대상과 목적이 달라 보급 수치를 그대로 비교할 수 없다. 다만 BCI의 확산을 전망할 때는 성능뿐 아니라 장치를 이식하는 수술의 부담도 함께 살펴야 한다.

BCI에서 구경하는 순간은 이미 왔다. 앞 장에서 본 아보의 체스 영상은 누구나 수술 없이 볼 수 있었다. 써 보는 순간은 아직 오지 않았다. 지금의 이식형 BCI를 쓰려면 두개골을 여는 수술을 받아야 하고 그 길은 환자에게만 열려 있다. 알파고와 챗GPT가 남긴 교훈을 BCI에 옮기면 질문이 좁아진다. 건강한 사람이 수술을 받아들일 만큼 쓸모가 커지는지, 아니면 수술 없는 장치가 그만한 쓸모를 내는지가 남는다. 뇌졸중 생존자 연구가 보여 준 대로 사람을 움직이는 것은 편리함보다 쓸모다. 환자에게는 잃은 기능을 되찾는 일이 그 쓸모다. 건강한 사람에게는 손과 목소리가 이미 있으므로 다른 쓸모가 필요하다.

환자용 의료기기에서 건강한 사람이 쓰는 장치로

건강한 사람의 머리와 귀에는 이미 신경 신호를 재는 기기가 올라가 있다. 유럽 싱크탱크 미래세대센터는 신경기술 기업 271곳을 분석했다(Centre for Future Generations, 2025년 6월 11일). 153곳이 소비자용이었고 118곳이 의료용이었다. 소비자 기업의 약 65퍼센트는 뇌파 검사(EEG)를 쓴다. 센터는 많은 기업이 건강에 좋다는 문구를 내세우면서 의료기기에 필요한 임상 시험은 피하는 회색 지대에 있다고 지적했다. 이런 기기는 주로 집중도와 수면을 잰다. 기계를 조작하는 BCI보다는 상태를 재는 측정기에 가깝다. 수술 없이 새 입력 장치를 만들려는 시도도 있다. 메타는 손목에서 근전도를 읽어 손짓과 손글씨를 알아보는 밴드를 Nature에 발표했다(The Register, 2025년 7월 23일). 근전도는 근육이 움직일 때 생기는 전기 신호를 재는 방법이다. 이 밴드는 뇌가 아닌 손목 근육의 신호를 읽으므로 BCI에 들지 않는다.

한국 기업도 귀에 거는 뇌파 기기를 만들었다. 삼성전자는 한양대 바이오메디컬공학과와 함께 귀 주변에서 뇌파를 재는 시제품을 만들어 영상 선호와 졸음을 판별했다(삼성전자 글로벌 뉴스룸, 날짜 미확인). LG전자는 뇌파를 재고 소리로 숙면을 돕는 이어셋 '브리즈'를 내놓았고 이 제품이 의료기기가 아니라고 밝혔다(LG, 2023년 7월 14일). 2026년 10월까지 국내 기업이 2025년 이후 새 뇌파 제품을 내놓았다는 발표는 없다. 제도 쪽에서는 디지털의료제품법이 2025년 1월 24일 시행됐다(법무법인 대륙아주, 2025년 5월 29일). 정부는 2027년까지 BCI 임상 시험 전용 가이드라인을 마련하겠다고 밝혔다(서울경제, 2026년 3월 18일).

수술이 필요한 장치를 건강한 사람에게 쓰는 일은 다른 문제다. 워싱턴대의 생명윤리학자 낸시 제커와 신경외과 교수 앤드루 고는 지금의 안전성 수준에서는 영구 이식의 이득이 커야 불확실한 위험을 정당화할 수 있다고 썼다(the benefit of a permanent implant would have to be large)(The Conversation, 2024년 2월 14일). 두

사람은 부유한 사람만 능력 증강을 누리면 사회의 불평등이 커진다고도 지적했다. 국제 규범도 같은 쪽을 가리킨다. 유네스코는 2025년 11월 신경기술 윤리 권고를 채택했다(영국 유네스코위원회, 2025년 11월 18일). 권고는 신경 데이터를 민감한 정보로 다루고 어린이에게는 치료가 아닌 목적의 사용을 피하라고 했다. 직장에서 직원의 생산성을 감시하는 데 쓰지 말라는 내용도 들어 있다. 이 권고는 회원국에 요청하는 문서이고 법적 구속력은 없다.

중국의 문서는 더 구체적이다. 중국 과학기술부의 BCI 연구 윤리 지침은 신체 기능이 정상인 사람의 감각과 인지와 운동 능력을 높이는 장치를 증강형 BCI로 정의했다(중국 과학기술부, 2024년 2월). 지침은 증강형이 "일반적으로 비침습"이라고 적고 중독을 부르는 연구를 엄격히 통제하라고 했다. 침습형 연구는 다른 치료 수단과 견주어 필요성과 합리성을 따진 뒤에 쓰라고 했다. 중국 국가약품감독관리국은 2026년 6월 지도원칙에서 비의료 목적의 기능 증강과 오락용 BCI를 의료기기로 관리하지 않는다고 밝혔다(중국신문망, 2026년 6월 30일). 건강한 사람용 제품은 의료기기 심사 밖에 두되 침습형 연구는 치료 쪽에 묶어 둔 셈이다.

건강한 사람에게 뇌수술을 정당화할 수 없다는 반론은 지금 기준으로 맞다. 다만 건강한 사람에게 가는 길이 침습형 하나만은 아니다. 미국 국방고등연구계획국(DARPA)은 건강한 군인을 위해 수술 없이 뇌 신호를 읽고 쓰는 양방향 장치 사업 N3를 진행했다(DARPA 사업 페이지, 2026년 10월 조회). 목표는 신경 조직 16세제곱밀리미터 안의 16개 채널을 50밀리초 안에 읽고 쓰는 것이었다. 사업 페이지는 이 사업이 끝났다고 적었고 결과는 공개하지 않았다. 건강한 사람에게 먼저 닿는 양방향 장치는 수술 없는 장치일 가능성이 크다는 것이 저자의 판단이다. 침습형이 그 뒤를 따르려면 수술의 위험이 지금보다 줄고 장치의 쓸모가 그 위험을 넘어야 한다.

시장의 크기와 성격이 달라지는 지점

BCI 시장의 크기는 세는 기관에 따라 크게 다르다. 마켓츠앤드마켓츠는 2024년 시장을 2억 6,200만 달러로 잡고 2029년 5억 600만 달러를 내다봤다(MarketsandMarkets, 2024년 5월). 프레시던스리서치는 2025년 시장을 29억 4,000만 달러로 적고 2035년 138억 6,000만 달러를 내다봤다(Precedence Research, 2026년 9월 9일 갱신). 같은 페이지의 표에는 2025년 시장 규모가 26억 2,000만 달러로 나와 본문과 맞지 않는다. 그랜드뷰리서치는 침습형 BCI의 2025년 총 유효 시장(TAM)을 1,682억 7,000만 달러로 계산했다(Grand View Research, 2025년 12월). 총 유효 시장은 어떤 제품이 닿을 수 있는 수요 전체를 금액으로 환산한 규모다. 실제로 팔린 금액을 뜻하지 않는다. 먼 미래를

보는 추정도 엇갈린다. IDTechEx는 2045년 BCI 시장을 16억 달러 남짓으로 잡고 2025년부터 연 8.4퍼센트 성장을 예상했다(IDTechEx, 2025년). 이 추정치는 그랜드뷰가 2025년 하나로 계산한 침습형 잠재 시장의 100분의 1에 못 미친다. 비슷한 시점의 BCI 시장 추정치는 아래 끝과 위 끝이 600배 넘게 벌어진다. 차이는 범위에서 나온다. 보고서마다 잠재 수요를 세는지 매출을 세는지가 다르고 넣는 제품과 응용도 다르다.

시장 추정치가 과장됐다는 비판이 있다. 신경기술 뉴스레터 뉴로파운더스의 편집장 도미닉 보르켈만스는 많은 시장 보고서가 기술이 작동하면 채택이 따라온다고 가정한다고 지적했다(Neurofounders, 2025년 11월 25일). 의료기술의 보급은 과학보다 구매와 보험 급여 단계에서 자주 멈춘다는 이유다. 그는 벤처 투자자가 큰 시장과 패러다임 전환의 이야기를 원한다는 점도 들었다. 이 비판에는 근거가 있다. 이식형 BCI의 실제 매출에 가까운 공개 자료로는 중국 보루이캉의 상장 심사 자료가 있다. 회사의 2025년 매출은 1억 800만 위안이었고 그중 72.32퍼센트를 뇌파 수집 장비가 냈다(虎嗅, 2026년 10월 13일). 세계 첫 승인을 받은 이식형 장치 NEO는 보고 기간에 매출을 내지 못했다. 큰 숫자는 아직 생기지 않은 수요를 세고 작은 숫자는 지금의 의료 판매를 센다. 둘 다 틀린 숫자는 아니지만 같은 시장을 가리키지도 않는다.

시장의 성격은 숫자의 자릿수에서 갈린다. 세계 스마트폰 출하량은 2025년 약 12억 5,500만 대였다(Counterpoint Research, 2026년 9월 15일). 심박조율기는 세계에서 한 해 약 60만 개가 이식된다는 2018년 리뷰의 추정이 있다(Perioperative Nursing, 2018년 12월 4일). 앞 장에서 센 장기 이식형 BCI 참여자는 공개 수치를 모두 더해도 200명이 되지 않는다. 한 해 10억 대 넘게 팔리는 소비자 기기와 한 해 수십만 개가 이식되는 의료기기는 돈이 도는 길부터 다르다. 건강한 사람이 쓰는 BCI가 나오면 이 장치는 의료기기 시장에서 소비자 기기 시장으로 건너가게 된다.

건너가는 순간 바뀌는 것은 크기만이 아니다. 의료기기 시장에서는 규제기관이 안전을 맡는다. 미국에서 위험도가 높은 이식 기기는 임상 자료로 안전성과 효과를 입증해야 팔 수 있다. 소비자 기기는 웰니스 문구를 달면 이 심사를 거치지 않는다. 보험도 갈린다. 의료기기는 보험 급여를 받아야 환자에게 닿고 앞 장에서 본 대로 미국에서는 그 급여까지 수년이 걸린다. 소비자 기기는 보험 대상이 아니어서 사용자가 구입 비용을 전액 부담한다. 스마트폰과 같은 방식이고 인공와우와 다른 방식이다. 책임을 묻는 길도 다르다. 선점 법리는 연방 규제가 주 법에 따른 소송보다 앞선다는 원칙이다. 미국에서는 이 법리 때문에 FDA 기준을 지킨 의료기기 제조사를 직접 상대로 한 소송이 자주 막힌다(ACSH, 2021년 6월 7일). 소비자 기기의 결함은 제조사와 사용자 사이의 계약과 일반

소비자법으로 다툰다. 의료기기 단계에서 규제기관과 보험과 선점 법리가 나눠 지던 짐은 소비자 기기 단계에서 제조사와 사용자에게 내려온다.

양방향 BCI를 준비한다는 것은 큰 시장이 열릴 때까지 기다리는 일이 아니다. 사용자의 명령과 기계가 돌려주는 신호를 어떤 조건에서 허용할지, 문제가 생기면 누가 연결을 끊고 지원을 이어 갈지부터 정해야 한다. 기술이 의료기기 밖으로 넓어질수록 이런 기준도 적용 범위를 넓혀야 한다. 실용화를 향한 연구와 사용자의 통제권을 지킬 제도를 함께 준비해야 한다.

참고 자료

Science, Tactile edges and motion via patterned microstimulation of the human somatosensory cortex(Valle 외), 2025년 1월 16일, https://www.science.org/doi/10.1126/science.adq5978

Science, A brain-computer interface that evokes tactile sensations improves robotic arm control(Flesher 외), 2021년 5월 21일, https://pubmed.ncbi.nlm.nih.gov/34016775/

Medtronic, Medtronic earns U.S. FDA approval for the world's first Adaptive deep brain stimulation system for people with Parkinson's, 2025년 2월 24일, https://news.medtronic.com/2025-02-24-Medtronic-earns-U-S-FDA-approval-for-the-worlds-first-Adaptive-deep-brain-stimulation-system-for-people-with-Parkinsons

NeuroPace, FDA Grants Premarket Approval (PMA) for the NeuroPace RNS System to treat Medically Refractory Epilepsy, 2013년 11월 14일, https://www.dannydid.org/wp-content/uploads/NeuroPace_Press_Release_PMA_Approval_2013-11-14.pdf

NIDCD(미국 국립난청소통장애연구소), Cochlear Implants, 2024년 6월 13일 갱신, https://www.nidcd.nih.gov/health/cochlear-implants

Journal of Clinical Investigation, Visual percepts evoked with an intracortical 96-channel microelectrode array inserted in human occipital cortex(Fernández 외), 2021년 12월, https://www.jci.org/articles/view/151331

Science Advances, Neural correlates of phosphene perception in blind individuals: A step toward a bidirectional cortical visual prosthesis, 2025년 11월, https://www.science.org/doi/10.1126/sciadv.adv8846

New England Journal of Medicine, Subretinal Photovoltaic Implant to Restore Vision in Geographic Atrophy Due to AMD(Holz 외), 2025년 10월 20일 온라인, https://www.nejm.org/doi/10.1056/NEJMoa2501396

Science Corporation, Science Corp. Announces European Commercial Launch of PRIMA, the Only Treatment to Restore Functional Central Vision to Patients with Geographic Atrophy Caused by Age-Related Macular Degeneration, a Leading Cause of Blindness, 2026년 7월 22일, https://science.xyz/news/ce-mark/

上海证券报(중국과학원 뇌과학지능기술탁월혁신센터 전재), 2025年终报道 | 脑机接口迈向产业化：从"实验室"走向"门诊室", 2026년 1월 12일 게시, https://www.ion.ac.cn/xwen/mtsm/2019n/202601/t20260112_8102926.html

Bloomberg(Bloomberg Government 전재), Musk's Neuralink Plans Speech Trial, Nonmedical Brain Implant, 2025년 9월 19일, https://news.bgov.com/esg/musks-neuralink-plans-speech-trial-nonmedical-brain-

implant

MIT Technology Review, With Neuralink, Elon Musk Promises Human-to-Human Telepathy. Don't Believe It., 2017년 4월 22일, https://www.technologyreview.com/2017/04/22/242999/with-neuralink-elon-musk-promises-human-to-human-telepathy-dont-believe-it/

NPR, Brain computer interfaces are poised to help people with disabilities, 2025년 6월 30일, https://www.npr.org/sections/shots-health-news/2025/06/30/nx-s1-5339708/brain-computer-interface-implants-disabilities-neuralink

IDTechEx, Brain Computer Interfaces 2025-2045: Technologies, Players, Forecasts, 2025년, https://www.idtechex.com/en/research-report/brain-computer-interfaces/1024

British Go Association, Google DeepMind Challenge Match - Lee Sedol v AlphaGo - match report, 2016년 (2018년 10월 6일 갱신), https://britgo.org/node/5458

Engadget, Google's Deepmind AI beats Go world champion for a second time, 2016년 3월 10일, https://www.engadget.com/2016-03-10-google-deepmind-ai-go-second-win.html

OpenAI, Introducing ChatGPT, 2022년 11월 30일, https://openai.com/index/chatgpt/

Euronews, ChatGPT sets record for fastest-growing user base - analyst note, 2023년 2월 2일, https://www.euronews.com/2023/02/02/openai-chatgpt

Journal of Neural Engineering, Bridging innovation and adoption: a mixed-methods investigation into stroke survivors' acceptability of brain-computer interface-based rehabilitation interventions(Grevet 외), 2026년 4월 13일, https://www.bordeaux-neurocampus.fr/article/bridging-innovation-and-adoption-a-mixed-methods-investigation-into-stroke-survivors-acceptability-of-brain-computer-interface-based-rehabilitation-interventions/

Frontiers in Human Neuroscience, Factors influencing the social acceptance of brain-computer interface technology among Chinese general public: an exploratory study(Xia, Yang), 2024년 10월 30일, https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2024.1423382/full

Pew Research Center, Mobile Fact Sheet, 2025년 11월 20일, https://www.pewresearch.org/internet/fact-sheet/mobile/

Centre for Future Generations, Press Release: Brain-sensing wearables surge into regulatory grey zone, new analysis warns, 2025년 6월 11일, https://cfg.eu/press-release-brain-sensing-wearables/

The Register, Meta eggheads demo Bluetooth wristband that decodes muscle twitches for UI control, 2025년 7월 23일, https://www.theregister.com/2025/07/23/meta_wristband_ui_control/

Samsung Global Newsroom, Samsung's Breakthrough Wearable Technologies Driven by Innovation and Collaboration, 날짜 미확인, https://news.samsung.com/global/samsungs-breakthrough-wearable-technologies-driven-by-innovation-and-collaboration

LG, 스트레스·불면엔 '브리즈' 하세요 LG전자, 마인드 웰니스 솔루션 '브리즈' 출시, 2023년 7월 14일, https://www.lg.co.kr/media/release/26515

법무법인 대륙아주, 바이오/헬스케어 분야 이슈리포트 - 디지털의료기기 관련 가이드라인 6종 재정·개정에 관한 주요 내용과 시사점, 2025년 5월 29일, https://www.draju.com/ko/sub/newsletters.html?type=view&bsNo=5991&langNo=

서울경제, "BCI로 척수손상·시각장애 등 난치성 질환 치료"…정부, 7대 국민체감 미션 추진, 2026년 3월 18일, https://www.sedaily.com/article/20021168

The Conversation, Several companies are testing brain implants - why is there so much attention swirling around Neuralink? Two professors unpack the ethical issues(Jecker, Ko), 2024년 2월 14일, https://theconversation.com/several-companies-are-testing-brain-implants-why-is-there-so-much-attention-swirling-around-neuralink-two-professors-unpack-the-ethical-issues-222556

UNESCO UK(영국 유네스코위원회), UNESCO Adopts First Global Framework on Neurotechnology Ethics, 2025년 11월 18일, https://unesco.org.uk/news/unesco-adopts-first-global-framework-on-neurotechnology-ethics

中华人民共和国科学技术部, 脑机接口研究伦理指引, 2024년 2월, https://www.most.gov.cn/kjbgz/202402/P020250729548085721952.pdf

中国新闻网, 明确产品分类及命名规则 两项脑机接口医疗器械指导原则发布, 2026년 6월 30일, https://www.chinanews.com.cn/gn/2026/06-30/10650027.shtml

DARPA, N3: Next-Generation Nonsurgical Neurotechnology, 2026년 10월 조회, https://www.darpa.mil/research/programs/next-generation-nonsurgical-neurotechnology

MarketsandMarkets, Brain Computer Interface Market by Product, Technology, Application, End User, Global Forecast to 2029, 2024년 5월, https://www.marketsandmarkets.com/Market-Reports/brain-computer-interface-market-157815170.html

Precedence Research, Brain Computer Interface Market Companies, Size & Trends 2026-2035, 2026년 9월 9일 갱신, https://www.precedenceresearch.com/brain-computer-interface-market

Grand View Research, Brain Computer Interface Market (2026 - 2033) Size, Share & Trends Analysis Report, 2025년 12월(2026년 7월 갱신), https://www.grandviewresearch.com/industry-analysis/brain-computer-interfaces-market

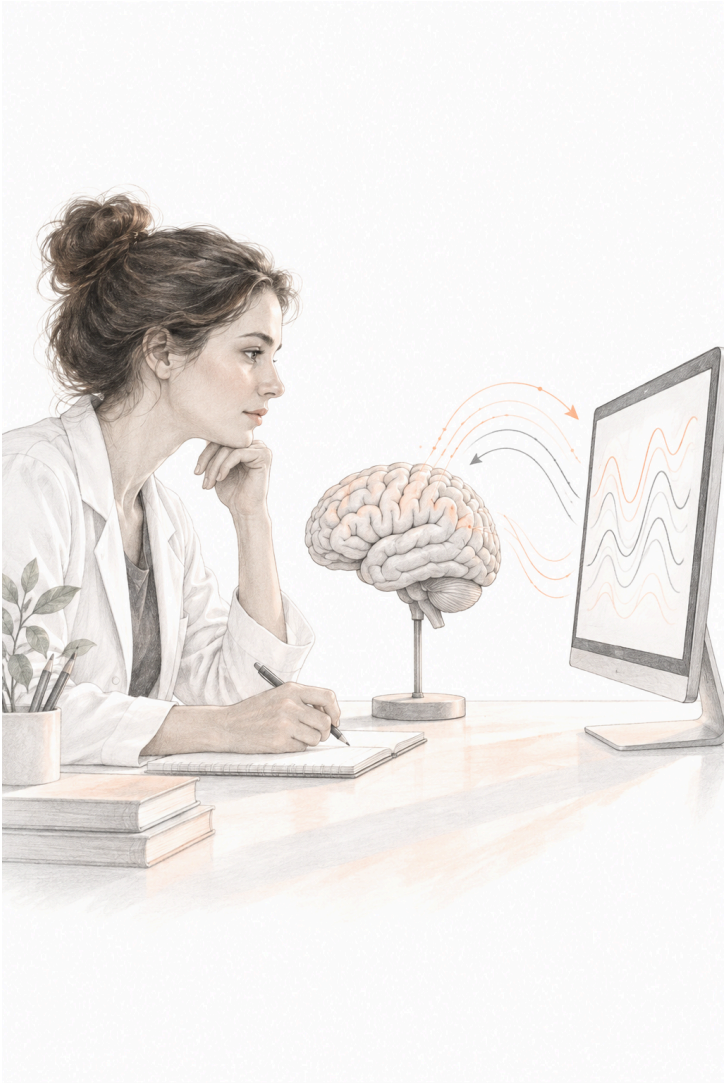
Neurofounders, Are BCI's (Still) Overhyped?(Dominic Borkelmans), 2025년 11월 25일, https://www.neurofounders.co/articles/are-bcis-still-overhyped

虎嗅, "脑机接口第一股"靠什么赚钱？, 2026년 10월 13일, https://www.huxiu.com/article/4897157.html

Counterpoint Research, Global Smartphone Shipments 2018 - 2025, 2026년 9월 15일, https://counterpointresearch.com/cn/insights/global-smartphone-shipments-share

Perioperative Nursing(GORNA), Permanent cardiac pacemaker(Poulis), 2018년 12월 4일, https://doi.org/10.5281/zenodo.1916544

American Council on Science and Health(ACSH), My Subconscious Made Me Do It: Legal Issues of Brain-to-Computer Interface(Billauer), 2021년 6월 7일, https://www.acsh.org/news/2021/06/07/my-subconscious-made-me-do-it-legal-issues-brain-computer-interface-15585



19장. 휴대전화와 컴퓨터를 대신하는 장치

표준 장치가 된다는 말의 뜻

2024년 6월 17일 일론 머스크는 X에 짧은 답글을 달았다. 한 패러디 계정이 생각만으로 휴대전화를 다루려고 뉴럴링크를 머리에 심겠느냐고 물었다. 머스크는 "미래에는 휴대전화가 없고 뉴럴링크만 있을 것"이라고 답했다(In the future, there will be no phones, just Neuralinks)(Deccan Herald, 2024년 6월). 이 답글에는 근거나 일정이 붙어 있지 않았다. 그래도 BCI를 만드는 회사의 창업자가 바라보는 곳은 분명히 드러난다. 그는 BCI를 환자의 보조기구가 아닌 모두의 기본 장치로 본다.

표준 장치라는 말에는 합의된 정의가 없다. 저자는 이 말을 사람이 기계에 명령을 넣고 답을 받는 기본 통로라는 뜻으로 쓴다. 기준은 세 가지다. 사람들의 절반 넘게 쓰고 치료가 아닌 일상의 목적으로 쓰며 운영체제가 그 장치를 기본 입력으로 받는다. 운영체제는 컴퓨터나 휴대전화에서 장치와 프로그램을 함께 움직이게 하는 바탕 소프트웨어다.

개인용 컴퓨터가 이 자리에 오른 과정은 길었다. 미국 인구조사국 조사에서 컴퓨터를 가진 가구는 1984년 8.2퍼센트에서 2003년 61.8퍼센트와 2011년 75.6퍼센트로 늘었다(미국 인구조사국, 2013년 5월). 그 20여 년 동안 자판과 마우스가 사람이 컴퓨터에 뜻을 전하는 기본 통로가 됐다. 스마트폰은 더 빨랐다. 미국 성인의 스마트폰 보유율은 2013년 1월 51퍼센트로 첫 아이폰이 나온 지 약 6년 만에 절반을 넘겼다(퓨리서치센터, 2025년 11월 20일). 그 사이 손가락으로 화면을 누르고 미는 동작이 새 기본 통로가 됐다. 표준 장치가 바뀔 때마다 바뀐 것은 기계의 모양보다 사람이 명령을 넣는 길이었다.

BCI에서도 그 길을 받아들이는 자리가 생기기 시작했다. 싱크론은 2025년 5월 13일 애플이 내놓은 BCI 입력 규격에 처음으로 기기를 직접 연결하는 회사가 되겠다고 발표했다(Synchron, 2025년 5월 13일). 보도자료에 따르면 이 규격은 신경 인터페이스를 아이폰과 아이패드와 비전 프로의 고유한 입력 범주로 인정한다. 톰 옥슬리 싱크론 대표는 "BCI는 접근성 도구를 넘어 다음 세대의 인터페이스 층"이라고 했다(BCI is more than an accessibility tool, it's a next-generation interface layer). 애플이 낸 문서는 직접 확인하지 못했고 2026년 10월까지 이 규격이 일반 개발자에게 열렸는지도 확인하지 못했다.

BCI가 의료기기에 머물 것이라는 시각에는 무게가 있다. 옥슬리 자신도 광범위한 보급에 대해 "솔직히 곧 그렇게 되리라고 보지 않는다"고 했다(I honestly don't think it's anytime soon)(Medical Design & Outsourcing, 2024년 9월 30일). 영국신경과학회 회장인 에든버러대의 타라 스파이어스존스는 이식 장치가 "널리 쓰이기까지 여러 해가 걸릴 것"이라고 했다(many years before they are commonly available)(Science Media Centre, 2024년 1월 30일). 옥스퍼드대의 안데르스 산드베리는 건강한 사람이 이식을 받으려면 결정적인 쓸모가 있어야 하고 수술의 두려움이 장벽이라고 봤다(IEEE Spectrum, 2017년 11월 13일). 저자는 이 반론이 침습형 장치가 퍼지는 속도에 대해서는 맞다고 본다. 다만 세 기준 가운데 운영체제의 입력 자리는 보급보다 먼저 열리고 있다.

수술 없는 착용 기기가 표준 장치가 된다는 시각도 있다. 메타 연구진은 앞 장에서 본 손목 근전도 밴드가 사용자마다 따로 맞추지 않고도 손글씨를 분당 20.9단어로 받아적는다고 발표했다(Nature, 2025년 7월). 이 밴드는 손을 쓰는 사람의 입력을 빠르게 바꿀 수 있다. 하지만 뇌가 아닌 근육의 신호를 읽으므로 손의 신경이 끊긴 사람에게는 통로가 되지 못한다. 건강한 사람의 입력은 이런 기기가 먼저 바꾸고 뇌에서 바로 신호를 읽는 장치는 그 뒤를 따를 가능성이 크다는 것이 저자의 판단이다.

휴대전화와 컴퓨터를 대신한다는 전망에서 먼저 살펴볼 것은 입력과 출력의 통로다. 연산 장치 전체가 사라져야 하는 것은 아니다. 사람이 자판과 화면에 의존하던 일을 뇌 신호와 감각의 연결로 수행할 수 있는지, 그 방식이 더 안전하고 유용한지를 검증해야 한다. BCI를 표준 장치로 발전시키려면 이 조건을 충족해야 하며, 지금의 의료 성과는 그 가능성을 시험하는 출발점이다.

인간이 인공지능을 실시간으로 지휘한다는 것

인공지능을 실시간으로 지휘한다는 말을 저자는 이렇게 쓴다. 사람이 에이전트에게 목표를 주고 에이전트가 일하는 동안 결과를 보며 방향을 바꾸거나 멈추게 하는 일이 사람의 판단 속도대로 이어지는 상태다. 지금 그 통로는 자판과 음성이다. 지시 한 줄을 고쳐 쓰는 사이에 에이전트는 이미 다음 단계로 넘어가 있다.

생각으로 언어 모델과 대화한 사례는 이미 있다. 싱크론은 2024년 7월 오픈AI의 GPT-4o를 자사 BCI 대화 기능에 넣었다(Slashdot, 2024년 7월 28일). ALS 환자 마크가 대화하면 모델이 앞선 대화를 보고 답장 후보를 여러 개 내놓는다. 마크는 그중 하나를 한 번의 선택으로 고르고 맞는 후보가 없으면 새로 고침을 누른다. 뉴럴링크의 참여자 브래드 스미스는 xAI의 Grok 3가 대화를 듣고 답변 후보를 만드는 앱을 쓴다(Fox News,

2025년 5월 31일). 그는 "완벽하지는 않지만 나를 대화 안에 머물게 한다"고 했다(It is not perfect, but it keeps me in the conversation). 두 사례에서 BCI는 지시의 내용을 읽지 않는다. 사람은 커서로 고르고 모델이 문장을 채운다. 의도는 사람이 정하고 실행은 기계가 맡는 지휘의 구조가 작은 크기로 나타나 있다.

입력의 속도도 올라가고 있다. 뉴럴링크는 앞 장에서 본 5만 시간의 신경 기록으로 디코더를 미리 훈련한 뒤 한 참여자가 커서 제어에서 초당 11.32비트를 냈다고 밝혔다(TeslaNorth, 2026년 10월). 종전 기록은 10.39비트였다. 일부 참여자는 한 주에 평균 55분 걸리던 재교정이 약 10분으로 줄었다. 회사는 비장애인이 마우스를 쓸 때 보통 초당 8에서 10비트를 낸다고 설명한다(TeslaNorth, 2026년 1월). 모두 회사가 낸 수치이고 독립 검증은 없다. 커서만 놓고 보면 BCI는 마우스를 쓰는 손과 비슷한 수준에 이르렀다. 말을 되찾는 장치는 앞 장들에서 본 대로 분당 56단어에서 62단어에 머문다.

머스크는 이 속도를 문제의 중심에 둔다. 대역폭은 일정한 시간에 주고받을 수 있는 정보의 양이다. 그는 2017년 2월 두바이 세계정부정상회의에서 "대부분 대역폭, 곧 연결 속도의 문제"라고 했다(It's mostly about the bandwidth, the speed of the connection)(CNBC, 2017년 2월 13일). 컴퓨터는 초당 1조 비트로 소통하는데 사람은 휴대전화 자판으로 초당 약 10비트를 낸다는 설명이었다. 2024년 8월에는 사람이 몇 비트로 말하면 인공지능이 지루해할 것이라며 "나무에게 말하는 것과 같다"고 했다(It's like talking to a tree)(Fortune, 2024년 8월 3일). 그가 내놓는 사람의 출력 수치는 자리마다 달라서 하나를 골라 믿기 어렵다.

연구자들의 측정은 다른 곳을 가리킨다. 크리스토프 쿠페 등은 17개 언어의 말하기가 모두 초당 약 39비트의 정보를 전한다고 보고했다(Science Advances, 2019년 9월 4일). 캘리포니아공대의 제위 정과 마커스 마이스터는 사람의 행동이 처리하는 정보량을 초당 약 10비트로 추정했다(arXiv, 2024년 8월 3일 투고, Neuron 게재 승인). 눈과 귀로 들어오는 정보는 초당 약 10억 비트인데 행동으로 나가는 양은 그보다 훨씬 적다. 이 논문은 BCI를 다루지 않는다. 다만 함의로 보면 병목은 손가락이 아닌 뇌 안쪽의 판단에 있다. 출력 장치를 빠르게 해도 사람이 정하는 속도는 그대로일 수 있다.

지금의 BCI는 손과 목소리를 잃은 사람에게 의도를 전달할 통로를 돌려준다. 그 사람들에게는 인공지능을 지휘하는 중요한 수단이 될 수 있다. 건강한 사람에게 당장 보탤 수 있는 기능은 긴 지시를 빨리 쓰는 일보다 멈춤과 선택 같은 짧은 신호를 손을 쓰지 않고 보내는 일에 가까울 것이라는 게 저자의 추정이다. 이 기능을 출발점으로 삼되, 목표 전달과 결과 확인, 수정과 중단을 이어 주는 통로로 발전할 수 있는지 연구해야 한다.

입력 속도만 높여서는 충분하지 않다. 사용자가 결과를 이해하고 자기 의도대로 결정할 수 있는지를 함께 확인해야 한다.

BCI와 인공지능 통제 문제의 연결

머스크는 BCI를 인공지능의 위험에 대한 답으로 말해 왔다. 2017년 2월 두바이에서 그는 사람이 기계와 융합하지 않으면 쓸모없어진다고 했다. 2019년 7월 16일 뉴럴링크 발표회에서는 장기 목표를 "인공지능과 일종의 공생을 이루는 것"이라고 했다(Achieve a sort of symbiosis with artificial intelligence)(CGTN, 2019년 7월 17일). 이 문장은 매체마다 단어가 조금씩 다르게 실렸다. 포춘은 2024년 8월 대담을 정리하며 뉴럴링크의 장기 목표가 사람의 소통량을 키워 인공지능과 사람의 공생을 개선하는 데 있다고 적었다. 그의 논리를 이으면 이렇다. 인공지능이 사람보다 빨라지면 사람은 판단에서 밀려난다. 밀려나지 않으려면 사람의 대역폭을 올려야 하고 그 수단이 BCI다.

이 논리에는 여러 비판이 있다. 니플라브(niplav)라는 필명의 연구자는 인공지능 안전 커뮤니티 레스롱(LessWrong)에 쓴 글에서 BCI가 사람과 컴퓨터의 상호작용을 일정한 배수만큼 빠르게 할 뿐이라고 했다(LessWrong, 2021년 8월 28일). 그래서 BCI는 "인공지능 정렬에 질적으로 새로운 전략을 주지 않는다"고 봤다(BCIs offer no qualitatively new strategies for aligning AI systems). 산드베리는 인지 증강이 좋은 일이지만 "기계보다 앞서 나가는 데는 도움이 되지 않는다"고 했다(But it's not going to help us keep ahead of the machines). 뮌헨공대의 마르첼로 이엔카는 뉴럴링크가 사람의 능력 증강을 장기 목표로 공개적으로 인정한 거의 유일한 회사라고 지적했다(Forbes Colombia, 2024년 2월 26일). 함부르크대의 크리스토프 부블리츠 등 6명은 뉴럴링크의 환자 임상이 비장애인 사용으로 가는 중간 단계라고 적었다(Neuroethics, 2025년 10월 14일). 이들은 정신에 미치는 영향을 평가하는 제도가 생길 때까지 비의료용 이식 장치의 시장 출시를 미루자고 제안했다.

공생이라는 말 자체를 다시 보게 하는 연구도 있다. 2026년 Neuroethics에 실린 한 논문은 BCI를 오래 쓴 사람들이 자신을 장치와 인공지능 시스템의 일부로 느낀다고 보고했다(Neuroethics, 2026년). 그 느낌은 통제감을 높인다. 그러나 연결이 갑자기 끊기거나 장치가 잘못 반응하면 사용자는 심리적 혼란을 겪는다. 공생은 기계 쪽 사정에 사람의 마음이 묶이는 일이기도 하다.

더 큰 우려는 방향이 뒤집히는 경우다. 니플라브는 신경계에 신호를 쓸 수 있는 BCI가 정렬되지 않은 인공지능이 사람을 조작하는 통로가 될 수 있다고 했다. 사람의 뇌는

적대적인 신경 입력에 대비해 나뉘어 있지 않다는 이유다(the human brain is not modular). 듀크대의 니타 파라하니는 2023년 책에서 인지적 자유를 인권으로 삼자고 제안했다(St. Martin's Press, 2023년 3월). 인지적 자유는 자기 뇌와 정신 활동을 스스로 정할 권리다. 미국 버몬트와 미네소타의 법안은 사람이 모르는 사이에 결정에 영향을 주는 의식 우회(consciousness bypass)를 동의 없이 하지 못하게 하는 조항을 담았다(Risk Management Magazine, 2026년 2월 24일). 옥스퍼드대의 로리 파이크로프트 등은 뇌 이식체를 악의로 장악하는 일을 브레인재킹(brainjacking)이라고 불렀다(World Neurosurgery, 2016년). 그들은 기분이나 보상 체계를 바꾸는 공격을 시연된 사건이 아닌 다가오는 위험으로 적었다.

니플라브와 산드베리의 비판을 이으면 통제 문제와 BCI는 상관이 없다는 시각이 나온다. BCI가 정렬을 풀지 못한다면 통제 논의에서 빼도 된다는 것이다. 해결책으로서 BCI의 몫이 작다는 지적은 맞다. 그러나 BCI는 통제를 받아야 할 새 통로로 통제 문제에 들어온다. 하버드 의대 계열 병원 등의 알론 고렌시테인 연구진은 루게릭병 환자 47명의 공개 뇌파 자판 자료에 언어 모델 에이전트를 붙여 시험했다(medRxiv, 2026년 9월 17일). 디코더가 얼마나 확신하는지를 프롬프트에 적어 준 에이전트는 사람의 뜻과 다른 행동을 바깥에서 기준값으로 거르는 장치보다 줄이지 못했다. 모델이 고칠 내용을 제안하고 실행 허락은 모델 밖의 장치가 쥐는 설계에서 결과가 나았다. 동료심사 전의 계산 연구다. 이 결론은 앞 부에서 본 에이전트의 권한 범위와 승인 절차와 멈추는 절차를 BCI에 그대로 옮긴 모양이다. 미국 회계감사원(GAO)의 2024년 기술평가도 BCI를 감독 도구가 아닌 감독할 대상으로 다뤘다(GAO, 2024년 12월).

BCI는 사람이 인공지능에게 뜻을 전할 수 있는 통로를 넓힌다. 장기적으로는 기계의 반응을 감각으로 받아 지휘에 반영하는 연결도 연구할 수 있다. 그러나 연결의 속도와 폭이 커져도 3부에서 본 평가와 사고 보고와 멈추는 권한을 대신하지는 못한다. 그 조건은 BCI가 붙은 시스템과 BCI 자체에 함께 적용해야 한다. 사람의 통제권을 보장하는 연결을 만드는 것이 목표다.

새로 생기는 윤리 문제: 뇌 신호 데이터, 인지의 소유권, 접근의 불평등

뇌 신호 데이터는 뇌의 전기 활동을 재서 만든 데이터다. 미국에서는 주가 먼저 움직였다. 콜로라도가 2024년 4월 17일 신경 데이터를 민감 정보에 넣는 법에 서명했고 캘리포니아와 몬태나와 코네티컷이 뒤를 이었다(콜로라도 주의회, 2024년; 미국 상원 상무위원회, 2025년 9월). 버몬트는 2026년 5월 18일 다섯 번째 법에 서명했다(Inside BCI, 2026년 5월 21일). 이 법은 정신과 신경 데이터의 프라이버시와 사고의 자유를

권리로 선언했다. 그러나 서면 동의 요건과 개인의 소송 권리와 위반당 과태료는 입법 과정에서 빠졌다. 연방 차원의 MIND Act는 2025년 9월 발의됐지만 연방거래위원회에 연구를 맡기는 법안이고 2026년 5월까지 진전이 없다(Davis Wright Tremaine, 2026년 5월 28일).

칠레는 2021년 10월 25일 공포한 헌법 개정으로 뇌 활동과 거기서 나온 정보를 보호하도록 했다(Carey, 2021년 11월 3일). 2023년 8월 9일 칠레 대법원은 전 상원의원 기도 히라르디가 미국 회사 이모티브를 상대로 낸 사건에서 그의 뇌파 데이터를 지우고 헤드셋 인사이트의 칠레 판매를 멈추게 했다(Frontiers in Psychology, 2024년 2월 27일; BioBioChile, 2023년 8월 11일). 법원은 뇌파 데이터를 연구에 쓰려면 그 목적에 맞는 동의를 따로 받아야 한다고 봤다. 유럽연합 개인정보보호법(GDPR)에서 뇌 신호의 지위는 아직 분명하지 않다(Centre for Future Generations, 2026년 4월 10일). 유럽평의회의 개인정보 협약 위원회는 2026년 2월 신경 데이터 지침 초안을 냈고 6월 회의에서도 초안을 다듬었다(유럽평의회, 2026년 2월). 앞 장에서 본 유네스코 권고는 2025년 11월 채택됐다.

인지의 소유권에는 두 뜻이 있다. 하나는 뇌 신호 데이터를 누가 갖느냐이고 다른 하나는 누가 내 생각에 손댈 수 있느냐다. GAO는 사용자가 모르는 사이 회사가 뇌 신호 데이터에 접근할 수 있다는 점을 첫 과제로 꼽았다. 신경권은 신경기술로부터 정신을 지키려고 제안된 새 인권이다. 이엔카와 로베르토 안도르노는 인지적 자유와 정신적 프라이버시와 정신적 온전함과 심리적 연속성 네 가지를 제안했다(Life Sciences, Society and Policy, 2017년 4월). 정신적 프라이버시는 뇌 정보를 원하지 않는 공개로부터 지킬 권리다. 컬럼비아대의 라파엘 유스테가 이끈 뉴로라이츠재단은 정신적 프라이버시와 개인 정체성과 자유의지와 증강에 대한 공정한 접근과 편향으로부터의 보호를 다섯 권리로 든다(NeuroRights Foundation). 부블리츠는 새 권리가 권리의 인플레이션이라며 기존 인권을 발전시키면 된다고 반박했다(Neuroethics, 2022년).

접근의 불평등은 비용 부담에서 시작한다. 미국의 이식형 BCI는 모두 임상 단계여서 상업 가격이 없다. 앞 장에서 본 중국 NEO의 조달 가격이 유일한 공개 가격이다. 로라 핀다이스와 리 플라이셔는 잠재 환자 다수가 메디케어 대상이어서 그 급여 정책이 접근을 좌우한다고 썼다(Health Affairs Forefront, 2026년 3월). GAO는 임상이 끝난 뒤 지원이 없어 장치를 제거한 참여자가 있었다고 적었다. 이엔카는 인지와 감각 능력으로 계층이 나뉘는 사회를 경고했다. 산드베리는 반대로 "모두의 키가 커지면 아무도 더 크지 않다"며 증강의 이득이 상대적이라고 봤다(But if everyone becomes taller, no one is taller). 그 말은 모두가 같은 장치를 가질 때만 성립한다.

보안도 같은 자리에 있다. 이반 마르티노비치 등은 저렴한 소비자용 뇌파 헤드셋으로 사용자의 비밀 정보를 추측하는 공격을 시험했다(USENIX Security, 2012년 8월). 무작위로 맞히는 것보다 추측의 불확실성이 평균 15에서 40퍼센트 줄었다. 뉴럴링크가 공개한 보안 정책이나 취약점 신고 제도는 찾지 못했다.

한국에는 뇌 신호 데이터를 따로 다룬 법률 조항이 없다. 양지현은 이 데이터를 민감정보로 볼지 "불분명한 영역이 있다"고 썼다(의료법학, 2020년 12월). 2026년 3월 서울대 콜로키움에서는 개인정보보호법에 신경 데이터를 넣자는 외국 전문가의 제안이 나왔다(법률신문, 2026년 3월 17일). 임창환 한양대 교수는 "아직은 데이터를 확보해 기술을 키워야 하는 시점"이라고 했다(한국경제, 2026년 7월 10일). 국회의 관련 법안과 국가생명윤리심의위원회의 입장은 확인하지 못했다.

세 문제는 장치가 환자의 것에서 모두의 것으로 넘어갈 때 커진다. 병원의 이식 장치는 의료 규제와 임상 동의 절차 안에 있다. 소비자 기기의 데이터는 그 밖에 있고 미국 주법도 대부분 소비자 프라이버시법에 조항을 더하는 방식이었다. 칠레 대법원이 지우라고 한 뇌파 데이터도 병원이 아닌 소비자용 헤드셋에서 나왔다. 따라서 대중 보급을 준비할 때는 사용자가 자기 신호의 수집과 사용을 통제하고, 원치 않는 작동을 멈추며, 지원이 끊겨도 보호받을 조건을 함께 마련해야 한다. 새로운 연결 수단의 가치는 사람이 더 많은 일을 하게 하는 능력과 자기 의도를 지킬 수 있는 권한을 함께 갖출 때 실현된다.

참고 자료

Deccan Herald, Elon Musk news: Neuralinks will make phones obsolete in future, 2024년 6월, https://www.deccanherald.com/technology/neuralinks-will-make-phones-obsolete-in-future-says-elon-musk-3072301

U.S. Census Bureau, Computer and Internet Use in the United States(P20-569), 2013년 5월, https://www.census.gov/content/dam/Census/library/publications/2013/demo/p20-569.pdf

Pew Research Center, Mobile Fact Sheet, 2025년 11월 20일, https://www.pewresearch.org/internet/fact-sheet/mobile/

Synchron(BioSpace 전재), Synchron To Achieve First Native Brain-Computer Interface Integration with iPhone, iPad and Apple Vision Pro, 2025년 5월 13일, https://www.biospace.com/press-releases/synchron-to-achieve-first-native-brain-computer-interface-integration-with-iphone-ipad-and-apple-vision-pro

Medical Design & Outsourcing, Catching up with Dr. Tom Oxley as Synchron builds toward a pivotal trial for FDA approval of its Stentrode BCI, 2024년 9월 30일, https://www.medicaldesignandoutsourcing.com/synchron-stentrode-bci-pivotal-trial-fda-approval-tom-oxley/

Science Media Centre, expert reaction to Elon Musk reporting Neuralink has implanted wireless brain chip in a human, 2024년 1월 30일, https://www.sciencemediacentre.org/expert-reaction-to-elon-musk-reporting-neuralink-has-implanted-wireless-brain-chip-in-a-human/

IEEE Spectrum, Q&A: The Ethics of Using Brain Implants to Upgrade Yourself, 2017년 11월 13일, https://spectrum.ieee.org/the-ethics-of-using-brain-implants-to-upgrade-yourself

Nature, A generic non-invasive neuromotor interface for human-computer interaction(Kaifosh, Reardon 외), 2025년 7월 온라인(645권), https://doi.org/10.1038/s41586-025-09255-w

Slashdot(CNET 인용), ChatGPT Has Been Integrated Into a Brain Implant, 2024년 7월 28일, https://tech.slashdot.org/story/24/07/28/1847232/

Fox News, Brain implant enables ALS patient to communicate using AI, 2025년 5월 31일, https://www.foxnews.com/tech/brain-implant-enables-als-patient-communicate-using-ai

TeslaNorth, Neuralink Sets New BCI Record After Training on 50,000 Hours of Brain Data, 2026년 10월(뉴럴링크 10월 1일 글 요약), https://teslanorth.com/2026/10/02/neuralink-pretraining-50000-hours-bci-record/

TeslaNorth, From Paralysis to Neuroscience: How 21 People are Using Neuralink in 2026, 2026년 1월, https://teslanorth.com/2026/01/28/from-paralysis-to-neuroscience-how-21-people-are-using-neuralink-in-2026/

CNBC, Elon Musk: Humans must merge with machines or become irrelevant in AI age, 2017년 2월 13일, https://www.cnbc.com/2017/02/13/elon-musk-humans-merge-machines-cyborg-artificial-intelligence-robots.html

Fortune, Elon Musk says a person with a Neuralink chip could beat a pro gamer in a few years, 2024년 8월 3일, https://www.fortune.com/2024/08/03/elon-musk-neuralink-chip-implant-outperform-professional-video-game-player

Science Advances, Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche(Coupé, Oh, Dediu, Pellegrino), 2019년 9월 4일, https://pmc.ncbi.nlm.nih.gov/articles/PMC6984970

arXiv(Neuron 게재 승인), The Unbearable Slowness of Being: Why do we live at 10 bits/s?(Zheng, Meister), 2024년 8월 3일 투고, https://arxiv.org/abs/2408.10234

CGTN, Musk shows off progress of his brain-machine interface, 2019년 7월 17일, https://news.cgtn.com/news/2019-07-17/Musk-shows-off-progress-on-brain-machine-interface-IoXyZLxR0A/share_amp.html

LessWrong, Brain-Computer Interfaces and AI Alignment(niplav), 2021년 8월 28일, https://www.lesswrong.com/posts/rpRsksjrBXEDJuHHy/brain-computer-interfaces-and-ai-alignment

Forbes Colombia, Expertos critican transparencia de Neuralink, después de que Musk dijera que su primer implante cerebral funciona, 2024년 2월 26일, https://forbes.co/2024/02/26/negocios/expertos-critican-transparencia-de-neuralink-despues-de-que-musk-dijera-que-su-primer-implante-cerebral-funciona

Neuroethics, A Moratorium on Implantable Non-Medical Neurotech Until Effects on the Mind are Properly Understood(Bublitz 외), 2025년 10월 14일, https://pmc.ncbi.nlm.nih.gov/articles/PMC12521269/

Neuroethics, "The Brain-Computer Interface is Changing Me": The Psychological Trajectory of Becoming More-Than-Human with a Brain Implant, 2026년(19권 31호), https://link.springer.com/article/10.1007/s12152-026-09656-2

St. Martin's Press, The Battle for Your Brain: Defending the Right to Think Freely in the Age of Neurotechnology(Nita A. Farahany), 2023년 3월, https://publishersweekly.com/9781250272959

Risk Management Magazine, State of Mind: The New Landscape of Neural Data Privacy Laws, 2026년 2월 24일, https://www.rmmagazine.com/articles/article/2026/02/24/state-of-mind--the-new-landscape-of-neural-data-privacy-laws

World Neurosurgery, Brainjacking: Implant Security Issues in Invasive Neuromodulation(Pycroft 외), 2016년, https://pubmed.ncbi.nlm.nih.gov/27184896/

medRxiv, AI agents at the brain-computer interface: separating inference from control(Gorenshtein 외), 2026년 9월 17일, https://www.medrxiv.org/content/10.64898/2026.09.13.26362955v1.full

U.S. GAO, Brain-Computer Interfaces: Applications, Challenges, and Policy Options(GAO-25-106952), 2024년 12월, https://www.gao.gov/products/gao-25-106952

Colorado General Assembly, HB24-1058 Protect Privacy of Biological Data, 2024년 4월 17일 서명, https://leg.colorado.gov/bills/hb24-1058

U.S. Senate Committee on Commerce, Science, & Transportation, Sens. Cantwell, Schumer, Markey Introduce Legislation to Shield Americans' Brain Data From Exploitation, 2025년 9월, https://www.commerce.senate.gov/2025/9/sens-cantwell-schumer-markey-introduce-legislation-to-shield-americans-brain-data-from-exploitation

Inside BCI, Vermont signs H.814 into law, becoming the fifth US state with a neural rights statute, 2026년 5월 21일, https://insidebci.com/policy/2026-05-21-vermont-h814-signed-fifth-state-neural-rights/

Davis Wright Tremaine, All the World's Neural Data, No Common Rule, 2026년 5월 28일, https://www.dwt.com/blogs/privacy--security-law-blog/2026/05/all-the-neural-data-and-no-common-rule-2026

Carey, Law No. 21,383: Constitutional reform to establish the Scientific and Technological Development at the Service of the People is published, 2021년 11월 3일, https://carey.cl/en/law-no-21383-constitutional-reform-to-establish-the-scientific-and-technological-development-at-the-service-of-the-people-is-published

Frontiers in Psychology, Chilean Supreme Court ruling on the protection of brain activity: neurorights, personal data protection, and neurodata(Cornejo-Plaza, Cippitani, Pasquino), 2024년 2월 27일, https://pmc.ncbi.nlm.nih.gov/articles/PMC10929545/

BioBioChile, Suprema prohíbe a empresa vender en Chile aparato que puede "leer la mente" a través de neurodatos, 2023년 8월 11일, https://www.biobiochile.cl/noticias/nacional/chile/2023/08/11/suprema-prohibe-a-empresa-vender-en-chile-aparato-que-puede-leer-la-mente-a-traves-de-neurodatos.shtml

Centre for Future Generations, Towards an EU Neurotechnology Strategy, 2026년 4월 10일, https://cfg.eu/towards-an-eu-neurotechnology-strategy/

Council of Europe(T-PD), 신경과학 맥락의 개인정보 보호 지침 초안(제목 미확인, URL 기준), 2026년 2월, https://rm.coe.int/guidelines-neuroscience-before-bureau-march/48802ac320

Council of Europe, 50th plenary meeting of the Committee of Convention 108 (T-PD), 2026년 6월, https://www.coe.int/en/web/data-protection/-/50th-plenary-meeting-of-the-committee-of-convention-108-t-pd-

Life Sciences, Society and Policy, Towards new human rights in the age of neuroscience and neurotechnology(Ienca, Andorno), 2017년 4월, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5447561/

NeuroRights Foundation, Mission, 2026년 10월 조회, https://neurorightsfoundation.org/mission

Neuroethics, Novel Neurorights: From Nonsense to Substance(Bublitz), 2022년, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8821782/

Health Affairs Forefront, Unlocking The Future: Medicare And Brain-Computer Interfaces(Findeiss, Fleisher), 2026년 3월, https://www.healthaffairs.org/content/forefront/unlocking-future-medicare-and-brain-computer-interfaces

USENIX Security, On the Feasibility of Side-Channel Attacks with Brain-Computer Interfaces(Martinovic 외), 2012년 8월, https://www.usenix.org/conference/usenixsecurity12/feasibility-side-channel-attacks-brain-computer-interfaces

의료법학, 뇌신경 데이터의 법적 규율과 뇌신경권에 관한 소고(양지현), 2020년 12월 31일, https://www.koreascience.kr/article/JAKO202023563934624.page

법률신문, "뇌파로 생각까지 읽는 시대"… '신경 데이터' 보호 필요성 제기, 2026년 3월 17일, https://www.lawtimes.co.kr/news/articleView.html?idxno=217923

한국경제, "생각도 개인정보인가"…BCI가 띄운 뇌 데이터 논쟁, 2026년 7월 10일, https://www.hankyung.com/article/2026071085231



20장. 한국의 선택

실제 환자 시술과 본격적인 임상 연구를 찾기 어려운 현실

앞 장들에서 본 대로 한국에는 손발을 대신하는 이식형 BCI의 임상 기록이 아직 없다. 서울대병원의 전극 시험은 뇌전증 진단용이고 국책 컨소시엄의 인체 임상은 2028년 이후에 잡혀 있다. 이 공백은 기술이 없어서 생긴 것이 아니다. 한국과학기술기획평가원(KISTEP)은 하드웨어와 소프트웨어와 수술 기술 같은 요소 기술을 국내가 이미 갖췄다고 봤다(KISTEP, 2025년 12월 26일). 같은 보고서는 식약처 임상 인허가에 드는 시간과 비용 때문에 임상 연구가 전무하다고 진단했다.

사람에게 심기 전 단계도 얇다. 같은 보고서는 영장류로 이식형 BCI를 연구하는 국내 기관을 가톨릭관동대 의대 손정우 교수 연구실 한 곳으로 적었다. 이승복 한국뇌연구원장은 국내 연구용 영장류를 약 1,500마리로 들었다(헤럴드경제, 2026년 9월 30일). 그는 중국 하이난 센터를 약 2만 마리, 미국을 약 10만 마리로 비교했다.

실험실의 성과는 꾸준히 나온다. KIST와 서울대 박성준 교수팀은 2025년 9월 100나노미터 두께의 코팅으로 뇌에 넣은 전극의 수명을 1개월에서 3개월 이상으로 늘렸다(ZDNet Korea, 2025년 9월 16일). 생쥐에서 얻은 결과다. 지브레인은 미국 FDA 절차와 국내 인허가를 함께 밟으며 2026년 기술특례 상장을 준비한다고 알려졌다(더벨, 2026년 4월 15일).

환자 쪽에서도 길이 보이지 않는다. 한국인 환자가 해외에서 이식형 BCI를 받았다는 공개 사례는 찾지 못했다. 앞 장에서 본 뉴럴링크의 5만 시간 신경 데이터는 임상 참여자가 있어야 쌓이는 자산이다. 참여자가 없으면 디코더를 훈련할 한국인의 뇌 신호도 쌓이지 않는다.

세계 3위권 기술국가의 자리를 가르는 분야

한국의 과학기술 지표는 상위권에 있다. 세계지식재산기구(WIPO)의 글로벌 혁신지수에서 한국은 2026년 139개국 가운데 4위였다(머니투데이, 2026년 9월 30일).

네이처 인덱스 2026년 연구 순위에서는 7위였다(STM Publishing, 2026년 6월). 국가경쟁력을 재는 IMD 평가에서는 2026년 70개국 가운데 21위였다(정책브리핑, 2026년 6월). 혁신과 연구 성과를 재는 지표에서는 10위 안에 든다.

3위권은 다른 문제다. 그 자리는 앞선 나라들이 아직 표준을 굳히지 못한 분야에서 갈린다고 우리는 본다. BCI는 그런 분야다. 인공지능과 반도체와 의료가 한 장치 안에서 만난다. BCI가 사람과 인공지능을 잇는 새로운 통로로 발전한다면 지휘와 통제의 기준도 장치 설계에 들어가야 한다. 한국은 기술뿐 아니라 그 연결의 표준을 만드는 과정에도 참여해야 한다.

정부의 문서는 아직 이 판단을 다 담지 못했다. 2022년 정한 12대 국가전략기술에 BCI는 들어 있지 않다(정책브리핑, 2025년 3월 12일). 정부는 2026년 3월 BCI로 미국과 중국에 이은 3대 강국에 오르겠다는 구상을 내놓았다고 보도됐다(데일리안, 2026년 3월 18일). 배경훈 부총리는 BCI를 "차세대 전략 기술"이라고 불렀다(이데일리, 2026년 8월 26일). 한국의 BCI 핵심기술이 세계 3위이고 미국과의 격차가 2년 내외라는 평가도 언론에 실렸다(디지털타임스, 2026년 3월 22일). 이 기사는 평가 주체를 밝히지 않았다.

정면 추격보다 틈새를 잡자는 시각이 있다. KISTEP은 미국을 뺀 나라들이 특화 전략을 세웠다고 정리했다(KISTEP, 2025년 12월 26일). 일본은 영장류 모델과 전극을, 중국은 한자와 인공지능을, 유럽은 검증된 임상 데이터를 잡았다는 분석이다. 이 보고서가 한국에 권한 틈새는 몸 안에서 돌아가는 인공지능 반도체와 체내 무선 통신을 결합한 이식형 장치다. 비침습형으로 물러서자는 제안이 아니다. 앞 장에서 본 대로 대화에 가까운 속도를 낸 장치는 모두 이식형이었다. 저자도 틈새 전략에 동의한다. 다만 그 틈새는 이식형 안에 있고, 이식형에 들어가려면 임상이라는 문을 먼저 열어야 한다.

반도체와 조선: 뒤늦게 뛰어들어 정상에 오른 경험

1983년 2월 8일 이병철 삼성 회장은 도쿄에서 반도체 사업 진출을 결심했다(삼성 뉴스룸). 그해 말 삼성은 64K D램을 개발했다. 회사는 이 개발로 미국과 일본과의 격차를 10년 이상에서 4년 정도로 줄였다고 설명한다. 1992년 삼성은 64M D램을 세계 처음으로 개발했고 그해 D램 출하량 세계 1위에 올랐다(삼성 반도체 뉴스룸). 선언에서 1위까지 9년이 걸렸다.

그 9년이 순탄하지는 않았다. 기흥 공장 부지는 농지와 상수도보호구역과 신도시 예정지로 묶여 있었다. 삼성 측 회고에 따르면 대통령이 직접 나서 이 문제를 풀었다

(뉴데일리, 2023년 2월 8일). 1980년대 중반 D램 가격이 폭락하는 가운데 삼성은 1M D램 공장 건설을 밀고 나갔다(전자신문, 2023년 8월 16일).

조선은 더 오래 걸렸다. 현대는 1972년 3월 23일 울산에서 조선소를 기공했다. 정주영 회장이 500원 지폐의 거북선을 보여 주며 차관과 첫 수주를 끌어냈다는 일화가 널리 알려져 있다(아주경제, 2020년 6월 14일). 처음 들여온 영국 조선소의 기술로는 모자랐다. 현대는 덴마크 오덴세 조선소의 블록 건조 기술과 일본 가와사키중공업의 생산설계 기술을 들여와 셋을 조합했다(배석만, 2011). 초기에는 스코틀랜드 기술자 윌리엄 잭 던컨이 현장 기술을 전했다(경향신문, 2011년 12월 9일). 일본이 40년 동안 지킨 세계 1위 조선국 자리는 2000년 무렵 한국으로 넘어왔다(한순흥, 2016). 기공에서 국가 1위까지 28년 안팎이었다. 정상에 오른 경험이 정상을 지킨 경험은 아니다. 한국 조선은 2012년부터 2017년까지 6년 연속 수주량에서 중국에 밀렸다(이투데이, 2018년 12월 10일). 2020년에는 세계 발주량의 43퍼센트를 따내 다시 1위가 됐다(아주경제, 2021년 1월 5일).

두 기록을 겹쳐 보면 다섯 가지가 함께 나온다고 저자는 정리한다. 국가는 금융을 대거나 보증을 섰다. 기업은 외국 기술을 들여와 그대로 쓰지 않고 조합했다. 대기업은 한 분야에 몰아서 투자했다. 사람을 밖으로 보내고 밖의 사람을 데려왔다. 불황이 와도 투자를 멈추지 않았다.

이번에는 다르다는 시각도 있다. KISTEP은 이식형 BCI의 병목을 기술이 아닌 "임상 허가 규제, 전임상 사례 및 데이터 부족"에서 찾았다(KISTEP, 2025년 12월 26일). 신유민 서울과학종합대학원대 교수는 BCI 경쟁력이 "가장 신뢰할 수 있는 기술"에서 나온다고 썼다(인공지능신문, 2026년 6월 30일). 뇌 신호는 사람마다 다르고 피로와 약물과 전극 상태에 따라 변한다. 그래서 연구실의 성능이 병원에서 그대로 나온다는 보장이 없다는 것이다. 공장의 수율은 설비와 공정을 고쳐 올릴 수 있지만 임상 근거는 환자와 시간이 있어야 쌓인다. 이 지적은 맞다. 반대쪽에서 이기원 와이브레인 대표는 의료 역량과 반도체 제조 기술을 합치면 늦게 들어가도 선도국이 될 수 있다고 봤다(디지털타임스, 2026년 3월 22일). 두 시각에서 함께 얻을 교훈은 장기 투자와 연구 기반의 준비다. 제조업에서 생산 부지와 설비를 먼저 확보했듯 BCI에서는 임상 연구와 장기 추적을 수행할 기반을 미리 갖춰야 한다. 다만 사람의 몸에 심는 장치인 만큼 그 기반에는 환자의 안전과 지속적인 지원을 확인할 제도가 포함돼야 한다.

의사와 공학자를 미국과 중국에 보낸다

한국은 사람을 데려와 연구소를 세운 경험이 있다. KIST는 1966년 2월 미국 원조 1,000만 달러와 한국 정부 1,000만 달러로 문을 열었다(국가기록원). 1차로 해외에서 일하던 한인 과학자 18명이 계약했다. 이들의 월급은 국립대 교수의 3배에 가까웠고 정부는 주택과 자녀 학교까지 챙겼다(국가기록원, 2015년 4월). 의료에는 사람을 보낸 경험이 있다. 1955년부터 1961년까지 미국 원조 사업인 미네소타 프로젝트로 서울대 교수 요원 226명이 미네소타대에서 연수했고 그중 77명이 의료 분야였다(더나은미래, 2012년 11월 27일). 반도체에서는 삼성이 1983년 미국 산호세에 현지법인을 세워 기술 개발 거점으로 삼았다(삼성 반도체 뉴스룸).

BCI에도 한국계 인물이 있다. 뉴럴링크 공동창업자 서동진은 네 살에 미국으로 이민했고 UC 버클리에서 박사 학위를 받았다. 그는 2025년 9월 서울 강연에서 당시 마비 환자 12명이 칩으로 기기를 조작하며 하루 평균 약 7시간 40분을 쓴다고 했다(조선비즈, 2025년 9월 16일). 그는 한국 기관이 보낸 사람이 아니라 미국에서 자란 창업자다. 한국 기관이 미국이나 중국의 BCI 임상 기관에 연구자를 보냈거나 공동 연구를 공식 발표한 사례는 찾지 못했다. 중국에는 BCI 병동이나 진료 단위를 둔 병원이 30여 곳 있다(차이나데일리, 2025년 7월 4일). 이 병원들이 외국 의사를 받는 공개 연수 과정은 확인되지 않았다.

길에는 장벽도 있다. 미국 상무부 산업안보국(BIS)은 2018년 11월 신흥기술 수출통제를 검토하며 BCI를 대상 범주에 넣었다(연방관보, 2018년 11월 19일). 2021년 10월에는 BCI만 다룬 사전통지를 내고 의견을 받았다(Baker McKenzie, 2021년 11월). 2023년 2월 보도는 이 통제가 아직 검토 중이라고 전했다(ST&R, 2023년 2월 9일). 최종 규칙은 확인하지 못했다. 중국의 인류유전자원관리조례는 외국 기관이 중국에서 유전자원을 모으거나 국외로 내보내는 일을 막고 국제 공동 연구에 과기부 승인을 요구한다(CMS, 2019년). 뇌 신호가 이 조례의 대상인지는 명시돼 있지 않다.

저자는 사람을 팀으로 보내자고 제안한다. 신경외과 의사와 재활의학과 의사와 전자공학자와 디코더를 짜는 연구자를 한 팀으로 묶는다. 미국에서는 장기 이식 환자를 돌본 학술 연구진을, 중국에서는 수술과 병동 운영 경험을 쌓은 병원을 협력 상대로 찾는다. 데이터는 현지에 두고 임상 운영과 연구 경험을 배우는 방식부터 협의한다. 데이터의 소재지만으로 규제 문제가 해소되는 것은 아니므로, 파견 계약 단계에서 배울 기술과 공유할 정보의 허용 범위를 법률 검토하는 절차를 사업에 넣는다. 돌아온 팀에는 국내 임상 자리와 연구비를 미리 약속한다. KIST가 해외 한인 과학자를 데려왔듯 미국 BCI 기업과 연구실의 한인 연구자를 데려오는 유치 트랙도 함께 둔다.

국가 차원의 예산과 제도

예산의 크기는 아직 보이지 않는다. 2026년 9월 30일 의결된 K-문샷의 2027년 예산은 1조 3,465억 원이다(JKN, 2026년 9월 30일). 이 돈은 12개 임무 전체의 합계이고 BCI 임무 단독 배정액은 공표되지 않았다. 과기정통부 2027년 예산안에는 뇌-컴퓨터 인터페이스 원천기술 개발 항목이 있지만 금액은 보도되지 않았다(헬로티, 2026년 9월 5일). 2023년부터 10년 동안 4,500억 원을 쓰는 뇌과학 선도융합기술개발사업에 BCI가 일부 들어 있다(뉴시스, 2022년 5월 31일). 서울대병원이 임상을 맡은 양방향 컨소시엄은 7년 동안 국비 202억 5,000만 원을 포함해 약 300억 원 규모다(코메디닷컴, 2026년 6월 16일).

미국 NIH BRAIN Initiative의 2026회계연도 예산은 4억 2,900만 달러로 계산됐다(The Transmitter, 2026년 1월 22일). 그중 1억 9,500만 달러는 2026회계연도로 끝나는 한시 법정 자금이다. 다음 해 기본 예산이 그대로면 그만큼이 한꺼번에 빠진다. 중국 과기부는 2021년도 뇌과학 중대 프로젝트 지침에서 국가 경비를 31억 4,800만 위안으로 잡았다(제일재경, 2021년 9월). 이것은 한 해 지침의 개산이고 전체 사업 총액이 아니다. 한국은 BCI 예산을 따로 공표하지 않으므로 같은 줄에 놓고 견줄 수 없다.

시장이 작은 나라가 큰돈을 넣을 이유가 있느냐는 물음이 있다. 투자 쪽 자료는 그 의심에 근거를 준다. 투자정보업체 PitchBook은 2024년 이후 벤처 자금 30억 달러 넘게 이 분야 스타트업에 들어갔지만 기술은 아직 실험 단계이고 상업적으로 검증되지 않았다고 썼다(PitchBook, 2026년 5월 7일). 저자도 투자 대비 근거가 아직 얇다는 데 동의한다. 그래서 돈을 장치 개발 경쟁에만 몰지 않는다. 영장류 전임상 센터, 임상 지침, 추적 등록부처럼 어느 회사가 이기든 남는 기반에 먼저 둔다. 반도체와 조선도 국내 시장을 보고 시작한 산업이 아니었다.

저자는 BCI 단독 예산을 항목으로 떼어 공개하자고 제안한다. 정부는 2025년 시행계획에서 국가전략기술 12대 체계를 임무 중심으로 개편하겠다고 밝혔다(정책브리핑, 2025년 3월 12일). 그 개편 때 BCI를 중점기술로 올린다. 한시 재원은 끝나는 해를 미리 정하고 뒤를 이을 재원을 3년 앞서 확정한다.

침습형 장치의 임상 승인 절차

식약처의 의료기기 임상시험계획 승인은 법정 처리 기간이 30일이다(정부24). 병목은 허가 뒤에 있다. 신의료기술평가는 새 의료기술의 안전성과 유효성을 한국보건의료연구원이 따져 건강보험 진입을 가르는 절차다. 이 평가는 접수에서

고시까지 최대 250일이 걸린다(건강보험심사평가원, 2024년 1월). 두 숫자 모두 제도상 처리 기간이다. 정부는 2026년 1월 26일 시장 즉시진입 의료기술 제도를 시행했다(한겨레, 2026년 1월 26일). 즉시진입 제도는 새 의료기기를 먼저 시장에 들이고 신의료기술평가를 최장 3년 안에 받게 하는 제도다. 그동안 기기는 비급여로 쓰인다. 정부는 시장 진입 기간이 최장 490일에서 최소 80일로 줄어든다고 밝혔다.

이식형 BCI에는 더 앞선 빈칸이 있다. 식약처가 이 장치를 어느 품목과 등급으로 보는지 밝힌 공식 자료를 찾지 못했다. 뇌에 전기 자극을 넣는 이식형 장치는 기존 품목이 있다. 신호를 읽기만 하는 이식형 장치의 품목은 확인되지 않았다. 앞 장에서 본 BCI 전용 가이드라인은 2027년이 목표다.

미국의 길은 단계가 나뉘어 있다. IDE는 위험이 큰 시험용 기기를 사람에게 시험하도록 FDA가 허락하는 제도다. 파라드로믹스는 2025년 11월 IDE를 받아 초기 2명 등록으로 시작했다(비즈니스와이어, 2025년 11월 20일). 이런 초기 타당성 연구는 소수에게서 안전성을 보고 단계적으로 넓힌다. FDA 자료를 분석한 Stern의 2017년 연구는 처음 시장에 들어간 기기가 뒤따른 기기보다 심사에 34퍼센트 더 걸렸다고 보고했다(퀸스대 스미스경영대학원). 같은 연구에서 객관적인 지침을 먼저 공표하면 심사 기간이 줄었다. 일본은 혁신 기기의 목표 심사 기간을 12개월에서 6개월로 줄이는 사키가케 지정을 둔다(PMDA). BCI가 이 지정을 받은 사례는 찾지 못했다.

규제를 풀면 환자가 다친다는 반론이 있다. 한국중증질환연합회는 즉시진입 제도 시행 날 성명을 내고 평가 없는 현장 사용을 "안전장치 없는 차량을 도로에 내보내는 것"에 빗댔다(한겨레, 2026년 1월 26일). 의료민영화 저지와 무상의료 실현을 위한 운동본부는 환자가 위험해지고 불필요한 의료비를 쓰게 된다고 비판했다. 저자는 이 비판이 이식형 BCI에 그대로 들어맞는다고 본다. 뇌에 넣은 장치는 쉽게 빼지 못한다. 그래서 문턱을 낮추자는 제안이 아니라 기다리는 시간을 줄이자는 제안을 한다. 이식형 BCI의 품목과 등급을 읽기와 쓰기로 나눠 먼저 고시한다. 2027년 가이드라인에 소수 참여자로 시작하는 초기 타당성 연구 트랙을 명시한다. 식약처와 한국보건의료연구원과 건강보험심사평가원이 한 자리에서 사전 상담을 하게 한다. 동의 능력과 장치 제거 시점처럼 BCI에만 있는 쟁점은 임상시험심사위원회가 쓸 표준 서식으로 만든다.

뇌 신호 데이터의 보호 기준

앞 장에서 본 대로 한국 법에는 뇌 신호를 따로 다룬 조항이 없다. 개인정보보호법 제23조는 사상과 신념, 건강 같은 정보와 대통령령이 정하는 정보를 민감정보로 묶어

처리를 제한한다. 시행령 제18조 제3호는 신체적, 생리적, 행동적 특징에 관한 정보 가운데 "특정 개인을 알아볼 목적"으로 기술적 수단을 통해 만든 정보를 여기에 넣는다. 개인정보보호위원회의 생체정보 보호 가이드라인은 생리적 특징의 예로 뇌파를 들었다(개인정보보호위원회, 2021년 9월). 이 예시에도 식별 목적이라는 조건이 붙는다.

이 조건이 BCI에서는 구멍이 된다. 커서를 움직이거나 말을 해독하려고 읽은 뇌 신호는 사람을 알아보려는 목적이 아니다. 문언대로라면 이 신호는 제18조 제3호에서 빠질 수 있다고 저자는 읽는다. 건강에 관한 정보로 보면 민감정보가 될 수 있지만 병원 밖 소비자 기기의 뇌파까지 그렇게 볼지는 분명하지 않다. 개인정보보호위원회가 뇌 신호에 공식 해석을 낸 기록은 찾지 못했다. 제23조를 어기고 민감정보를 처리하면 5년 이하 징역이나 5,000만 원 이하 벌금에 처한다(개인정보보호법 제71조). 처벌이 무거운 만큼 무엇이 해당하는지가 분명해야 한다.

입법 제안도 나왔다. 뉴로라이츠재단의 스티븐 다미아노스는 2026년 3월 서울대 콜로키움에서 세 가지를 권했다(법률신문, 2026년 3월 17일). 수집 목적과 무관하게 신경 데이터를 민감정보로 두는 개인정보보호법 개정, 인공지능 기본법 하위 규정의 분류 기준, 뇌연구촉진법의 신경윤리 거버넌스다. 국회에는 지문과 얼굴과 홍채 같은 생체정보를 민감정보 처리 제한 항목에 넣는 개정안이 2025년 1월 발의됐다(여성신문). 신경 데이터를 명시한 법안은 확인하지 못했다.

앞 장의 임창환 교수는 지금은 데이터를 확보해 기술을 키울 때라고 했다. 저자는 보호 기준이 데이터 확보를 늦추지 않는다고 본다. 기준이 흐리면 병원은 데이터를 내놓기를 꺼린다. 기준이 분명하면 무엇을 지키면 되는지 알 수 있다. 저자는 시행령 제18조에 다음과 같은 호를 더하자고 제안한다. 법제처 심사를 거친 조문이 아니다.

제18조(민감정보의 범위) 제 ○ 호 뇌와 신경계의 활동을 측정하여 얻은 정보와 그 정보로부터 추론한 정보(수집 목적을 묻지 아니한다)

연구 목적의 가명처리는 지금 법의 틀 안에서 허용 경로를 고시로 밝힌다. 신경 데이터로 학습한 모델에서 개인의 기여분을 지우는 방법은 기술 기준이 없으므로 연구 과제로 먼저 맡긴다.

시술 후 장기 추적 체계

추적관리대상 의료기기는 부작용 위험이 크거나 소재를 파악해야 하는 기기를 식약처가 지정해 공급과 사용 기록을 남기게 하는 제도다. 근거는 의료기기법 제29조이고 지정 목록은 「추적관리대상 의료기기 지정에 관한 규정」(식약처 고시 제2023-70호)에

있다. 인체 안에 1년 이상 삽입되는 기기 항목에는 발작방지용 뇌 전기 자극장치(이식형) 같은 뇌 자극 장치 세 품목이 들어 있다. 신호를 읽기만 하는 이식형 BCI는 품목 자체가 확인되지 않아 목록에 없다.

등록부도 보이지 않는다. 이식형 심박동기 삽입 환자는 2017년 3,781명에서 2020년 4,954명으로 늘었다(메디칼타임즈, 2022년 6월 26일). 이런 이식 기기 환자를 국가가 오래 따라가는 공개 레지스트리는 찾지 못했다.

외국의 임상 설계는 추적 기간이 길다. 뉴럴링크의 PRIME 연구는 18개월의 1차 연구 뒤 5년의 장기 추적을 붙여 약 6년으로 설계됐다(Fierce Biotech). 장기 추적이 필요한 이유는 전극에 있다. 참여자 2명에게서 뺀 유타 어레이를 분석한 연구가 있다(Frontiers in Bioengineering and Biotechnology, 2021년 12월 7일). 기록 품질은 처음 30일에서 40일 동안 오르다가 오래 쓴 참여자에게서 점차 떨어졌다. 오래 심은 어레이일수록 둘레에 조직 막이 더 많이 생겼다. 유럽연합 의료기기규정(MDR)은 시판 후 임상 추적(PMCF)을 둔다. 제조사가 판매 뒤에도 기기 수명 내내 임상 자료를 모아 안전성과 성능을 확인하는 절차다(NAMSA).

임상이 끝난 뒤의 공백도 드러났다. 호주 전문가 24명의 합의 연구는 시험이 끝나면 효과를 본 참여자도 장치를 빼야 할 수 있다고 지적했다(Medical Xpress, 2026년 9월 4일). 연구비나 벤처 자금으로 움직이는 스폰서는 장기 지원 비용을 감당하지 못할 수 있다. 호주에는 시험 중단이나 장치 단종 때 참여자를 지원하는 규정이 없다. 앞 장에서 본 미국 GAO 보고서도 임상 뒤 지원이 없어 장치를 뺀 참여자를 적었다. 한국에 같은 규정이 있는지는 따로 확인이 필요하다.

저자는 고시를 고쳐 읽기와 쓰기를 가리지 않고 이식형 BCI를 추적관리대상에 넣자고 제안한다. 첫 임상 참여자부터 등록하는 국가 BCI 이식 등록부를 두고 최소 10년을 따라간다. 임상시험계획 승인 서류에는 시험이 끝난 뒤의 장치 유지 계획을 넣게 한다. 회사가 문을 닫을 때를 대비한 공제나 기금도 이 계획에 넣는다. 지금 이식형 기기 임상에 요구하는 책임보험의 범위를 넓히는 방식이다.

기술이 먼저 가고 제도가 뒤따르는 순서를 끊는 법

앞 장에서 본 중국은 승인 1년 전에 수가 항목부터 만들었다. 한국은 대체로 반대 순서로 움직였다. 규제 샌드박스는 새 기술을 기존 규제에서 잠시 풀어 시험하게 하는 제도다. 2019년 도입 뒤 2025년 3월까지 실증 특례는 1,819건이었다(한국경제, 2025년 8월 22일). 이 가운데 법령 개정까지 이어져 시장에 나온 사업은 429건으로, 전체 실증

특례 1,819건의 23.5퍼센트였다. 나머지 1,390건은 그 단계에 이르지 못했다. 연구개발을 시작할 때 인허가와 수가, 신경 데이터 보호의 기준도 함께 설계해야 한다는 것이 저자의 제안이다.

지침을 먼저 낸 경우도 있다. 식약처는 2020년 디지털치료기기의 허가심사와 임상시험 가이드라인을 먼저 냈다. 2023년 2월 15일 불면증 치료 앱 솜즈가 첫 허가를 받았다(의협신문, 2023년 2월 15일). 허가는 빨랐지만 보급은 늦었다. 건강보험 등재 절차가 남아 있었고 기존 수가 체계가 소프트웨어 치료제와 맞지 않는다는 지적이 이어졌다(더메디컬, 2025년 6월 10일). 허가 지침만 먼저 나왔고 수가를 정하는 제도는 뒤따랐다.

영국은 기관을 한자리에 모으는 방식을 시험했다. 영국 의약품의료제품규제청(MHRA)의 혁신 기기 접근 경로(IDAP)는 2023년 9월 시범 운영을 시작해 8개 기술을 골랐고 2025년 3월 시범을 마쳤다(영국 정부). MHRA와 NICE와 NHS 잉글랜드가 함께 운영했다. 허가하는 기관과 가치를 평가하는 기관과 돈을 내는 기관이 처음부터 같은 기기를 함께 본다.

입구만 앞당기는 방식에도 실패가 있다. 국회입법조사처는 인공지능과 디지털 혁신 의료기기의 시판 후 관리 규정이 마련되지 않았다고 지적했다(팜뉴스, 2026년 9월 14일). 정부가 시장 진입을 허용한 인공지능 의료기기 23건 가운데 2026년 9월까지 정식 평가를 통과한 것은 한 건도 없었다(매일노동뉴스, 2026년 10월 14일). 김윤 의원이 보건복지부에서 받은 자료다. 입구만 넓히고 출구를 설계하지 않은 결과다.

저자가 제안하는 순서는 이렇다. 국내에서 개발한 첫 이식형 BCI가 인체 임상에 들어가기 전에 네 가지를 한 묶음으로 내놓는다. 품목과 등급을 정한 허가 지침, 신의료기술평가의 판단 기준, 수술과 장치 제거의 수가 항목, 추적 등록부의 운영 규정이다. 이 묶음은 식약처, 한국보건의료연구원, 건강보험심사평가원, 보건복지부, 개인정보보호위원회가 함께 만드는 BCI 협의체가 맡는다. 지침을 먼저 내면 심사가 빨라진다는 근거는 앞 소절의 Stern 연구에 있다. 수가와 추적을 함께 묶어야 허가가 환자에게까지 닿는다.

한국 BCI 10년 로드맵(저자 제안)

다음은 저자의 제안이며 정부 계획이 아니다. 정부가 이미 밝힌 연도 위에 앞 소절의 제도 항목을 얹었다.

2027년부터 2029년까지는 규칙을 먼저 세우는 기간이다. K-문샷 BCI 임무가 시작하는 2027년에 이식형 BCI의 품목과 등급을 고시하고 임상 가이드라인을 낸다. 가이드라인은 정부가 2027년 목표로 이미 밝혔다. 같은 해 개인정보보호법 시행령에 신경

데이터를 넣고 영장류 전임상 센터를 착공한다. 의사와 공학자의 첫 파견팀도 이때 보낸다. 2028년부터 2029년 사이에는 서울대병원이 이식 임상을 총괄하는 컨소시엄의 첫 인체 임상이 이 규칙 아래에서 열린다. 국가 BCI 이식 등록부도 첫 참여자의 이식 전에 가동해 임상 시작부터 자료를 축적한다. 정부는 2029년 BCI 기술 실증 완료와 임상 착수를 목표로 내걸었다(이데일리, 2026년 8월 26일). 중국 7개 부처 의견은 2027년까지 전극과 칩과 완제품을 국제 선진 수준에 올리겠다고 했다(중국정부망, 2025년 8월). 중국이 장치를 맞추는 동안 한국은 규칙을 맞춘다.

2030년부터 2032년까지는 첫 장치를 시장 문턱에 세우는 기간이다. 첫 허가 신청보다 1년 앞서 수술과 제거의 수가 항목을 만든다. 중국이 승인 1년 전에 수가를 만든 순서를 따른 것이다. 국가 BCI 이식 등록부는 2030년부터 참여 기관과 추적 범위를 넓힌다. 컨소시엄 일정의 3단계도 이 기간의 인허가와 상용화다(코메디닷컴, 2026년 6월 16일). 중국은 2030년까지 세계적 선도 기업 2곳에서 3곳을 키우겠다고 했다. 일본 문샷 목표 1의 가나이 료타 과제는 2030년 이정표로 장애를 넘어 사회 활동에 참여하는 원격 상호 지원 사회를 내걸었다(JST).

2033년부터 2036년까지는 확산과 점검의 기간이다. 정부의 2035년 상용화와 글로벌 진출 목표가 이 구간에 있다. 첫 참여자들의 5년 추적 결과를 이 기간에 공개한다. 2033년 이후의 재원은 2030년에 미리 확정한다. 미국 BRAIN은 2014년 10년 계획으로 출발했다(NIH, 2014년 6월 5일). 앞에서 본 대로 이 사업은 2027회계연도에 한시 자금이 끊긴다. 영국 ARIA의 6,900만 파운드 신경기술 프로그램은 1년 차 점검 뒤 보완 프로그램을 계획했다(ARIA). 이 로드맵도 구간이 끝날 때마다 점검해 다음 구간을 고친다.

2024년 3월 놀런드 아보는 노트북 앞에서 생각만으로 체스 말을 옮겼다. 그 장면은 18개월 연구에 5년 추적을 붙인 임상 설계 안에서 나왔다. 장치가 그의 뇌에 들어가기 전에 그를 6년 동안 지켜볼 계획이 먼저 서 있었다. 4부의 주장은 인간이 인공지능과 이어질 연결 수단을 확보하자는 것이다. 한국이 확보해야 할 연결 수단은 전극과 칩만이 아니라 그 장치를 사람의 몸에 들이고 오래 지켜보는 제도까지다.

참고 자료

KISTEP, 이식형 뇌-컴퓨터 인터페이스 최신 동향 및 전망(과학기술 및 ICT 정책 기술 동향 No.301), 2025년 12월 26일, https://www.kistep.re.kr/boardDownload.es?bid=0067&list_no=94486&seq=1

더벨, [지브레인 밸류업] 한국의 뉴럴링크, 기술력으로 승부…임상·인허가 돌입, 2026년 4월 15일, https://www.thebell.co.kr/front/newsview.asp?key=2026041509082279001023330

헤럴드경제, "韓 BCI〈뇌-컴퓨터 인터페이스〉 실용화 위해 뇌질환 영장류센터 구축 필수", 2026년 9월 30일, https://biz.heraldcorp.com/article/10888861

ZDNet Korea, 뇌-컴퓨터 연결 언제될까…KIST·서울대, 전극 정밀코팅 기술로 기반 확보, 2025년 9월 16일, https://zdnet.co.kr/view/?no=20250915092210

머니투데이, 韓, 글로벌 혁신지수 2년 연속 'TOP 4'…높은 혁신역량 위상 굳혀, 2026년 9월 30일, https://www.mt.co.kr/policy/2026/09/30/2026093011043875474

STM Publishing, Nature Index expands to applied and social sciences with release of 2026 Research Leaders tables, 2026년 6월 17일, https://www.stm-publishing.com/nature-index-expands-to-applied-and-social-sciences-with-release-of-2026-research-leaders-tables-2/

정책브리핑, 2026 IMD 국가경쟁력 평가 결과 한국 70개국 중 21위, 2026년 6월 19일, https://www.korea.kr/news/policyNewsView.do?newsId=148966822

정책브리핑, 반도체·AI·양자 등 12대 국가전략기술에 올해 6조 8000억 원 투자, 2025년 3월 12일, https://www.korea.kr/news/policyNewsView.do?newsId=148940472

데일리안(네이트 뉴스 전재), 정부, '뇌 미래산업 국가 R&D전략' 발표…BCI 선점 총력, 2026년 3월 18일, https://m.news.nate.com/view/20260318n31463

이데일리, BCI 얼라이언스 출범…2035년 '110조원' 글로벌 시장 잡는다, 2026년 8월 26일, https://edaily.co.kr/News/Read?mediaCodeNo=257&newsId=04569046645551256

디지털타임스, [이준기의 과·알·세] 빨라지는 BCI 산업화 시계… 美·中 경쟁에 속도내는 韓, 2026년 3월 22일, https://www.dt.co.kr/article/12053043

삼성 뉴스룸, 우리는 왜 반도체사업을 해야 하는가?, 2010년 3월 25일, https://news.samsung.com/kr/87

삼성 반도체 뉴스룸, 삼성 반도체사업 40년, 도전과 창조의 역사, 2012년 4월 6일, https://news.samsungsemiconductor.com/kr/%EC%82%BC%EC%84%B1-%EB%B0%98%EB%8F%84%EC%B2%B4%EC%82%AC%EC%97%85-40%EB%85%84-%EB%8F%84%EC%A0%84%EA%B3%BC-%EC%B0%BD%EC%A1%B0%EC%9D%98-%EC%97%AD%EC%82%AC/

전자신문, [100대 사건_006] 삼성, 64K D램 개발…반도체 신화의 서막 〈1983년 11월〉, 2012년 9월 17일, https://www.etnews.com/201209110688

뉴데일리, 이병철 '도쿄 선언' 40주년… 이재용 '기술경영' 위기 돌파 정조준, 2023년 2월 8일, https://biz.newdaily.co.kr/site/data/html/2023/02/08/2023020800123.amp.html

전자신문, [과학기술이 미래다] 〈99〉 이병철 회장, D램 대폭락에도 1M D램 공장 건설, 2023년 8월 16일, https://www.etnews.com/20230816000179

아주경제, [위기가 기회다-⑨현대중공업(끝)] 韓조선산업 일으킨 정주영의 '500원권 신화', 2020년 6월 15일, https://www.ajunews.com/view/20200614133259034

배석만, 현대중공업의 초창기 조선기술 도입과 정착과정 연구, 경영사연구(한국경영사학회) 제26집 제3호(통권59호) 181~214쪽, 2011년 9월, https://scholar.kyobobook.co.kr/article/detail/4010025873389 (저자명은 Tenold 외, Enterprise & Society 22권 2호, 2021년의 참고문헌 표기 Bae, Suk-Man으로 확인)

경향신문, 영국 선박기술 알려 한국 조선 키운 던컨 '보은의 금탑훈장', 2011년 12월 9일, https://khan.co.kr/article/201112092122145

한순흥, 공장을 만드는 공장(특별기고), 대한조선학회지 53권 3호 22~26쪽, 2016년 9월 30일, https://koreascience.kr/article/JAKO201610364971213.page

이투데이, 한국 조선, 올해 연간 수주량 '세계 1위' 확실시, 2018년 12월 10일, https://www.etoday.co.kr/news/view/1697586

아주경제, 지난해 K-조선 세계 1위…발주 물량 43% 수주, 2021년 1월 5일, https://www.ajunews.com/view/20210105165051398

인공지능신문, 뇌-컴퓨터 인터페이스 시대, 대한민국 의료AI는 무엇을 준비해야 하는가, 2026년 6월 30일, https://www.aitimes.kr/news/articleView.html?idxno=40761

국가기록원, 한국과학기술연구소 : 국가경쟁력의 핵심은 과학기술, 날짜 미상, https://theme.archives.go.kr/next/koreaOfRecord/KIST.do

국가기록원 이달의 기록, 과학기술 진흥, 대한민국의 초석을 다지다, 2015년 4월, https://theme.archives.go.kr/next/pages/new_newsletter/2015/04/sub1.html

더나은미래, "미국서 배운 선진 의료 기술 이제 네팔에 전파하고 있죠", 2012년 11월 27일, https://futurechosun.com/?p=4253

조선비즈(최종현학술원 전재), 머스크와 창업한 韓人 과학자 "뇌에 칩 이식, 삶의 제약 없애겠다", 2025년 9월 16일, https://www.chey.org/Kor/News/NewsView.aspx?seq=257

China Daily, China making big strides in brain-computer interface tech, 2025년 7월 4일, https://govt.chinadaily.com.cn/s/202507/04/WS687db3af498edec913cde9d1/china-making-big-strides-in-brain-computer-interface-tech.html

Federal Register(미국 연방관보), Review of Controls for Certain Emerging Technologies, 2018년 11월 19일, https://www.gpo.gov/fdsys/pkg/FR-2018-11-19/html/2018-25221.htm

Baker McKenzie, BIS Announces Review of Potential Controls on Brain-Computer Interface (BCI) Technology and Seeks Public Comment, Expands Controls on Certain Biological Equipment Software and Technology, 2021년 11월 12일, https://sanctionsnews.bakermckenzie.com/bis-announces-review-of-potential-controls-on-brain-computer-interface-bci-technology-and-seeks-public-comment-expands-controls-on-certain-biological-equipment-software-and-technology/

ST&R, Export Controls on Interface Technology Still Under Consideration, 2023년 2월 9일, https://www.strtrade.com/trade-news-resources/str-trade-report/trade-report/february/export-controls-on-interface-technology-still-under-consideration

CMS, Chinese state council releases regulation on the handling of human genetic resources, 2019년 6월 13일, https://cms.law/en/chn/legal-updates/Chinese-state-council-releases-regulation-on-the-handling-of-human-genetic-resources

JKN, K-문샷, 1.3조 투입 'AI 과학 난제' 도전…2027년 본격 가동, 2026년 9월 30일, https://news.jkn.co.kr/post/1002500

헬로티, 과기정통부, 2027년 예산 29.6조 원 편성…전 국민 AI서비스 지원, 2026년 9월 5일, https://www.hellot.net/news/article.html?no=114788

뉴시스, 과기부, 뇌과학 초격차 기술 확보에 10년간 4500억 투자, 2022년 5월 31일, https://www.newsis.com/view/NISX20220531_0001892576

코메디닷컴, 서울대병원, 300억 양방향 뇌-로봇 개발 참여…BCI 칩 이식 임상 총괄, 2026년 6월 16일, https://kormedi.com/2825235/

The Transmitter, Neuroscience, BRAIN Initiative gain budget in 'bad' NIH funding bill, 2026년 1월 22일, https://www.thetransmitter.org/funding/neuroscience-brain-initiative-gain-budget-in-bad-nih-funding-bill/

第一财经(제일재경), 中国脑计划正式启动 首年规模超30亿整体规模达数百亿, 2021년 9월 17일, https://www.yicai.com/news/101177358.html

PitchBook, A big question for brain-computer interface tech: How big is the market?, 2026년 5월 7일, https://pitchbook.com/news/articles/a-big-question-for-brain-computer-interface-tech-how-big-is-the-market

정부24, 의료기기 임상시험계획승인(서비스 안내), 2025년 11월 27일 표시, https://www.gov.kr/mw/AA020InfoCappView.do?HighCtgCD=A09006&CappBizCD=14700000504

건강보험심사평가원, 신의료기술 등재 신청 가이드 – 의료행위 및 치료재료 –, 2024년 1월, https://www.hira.or.kr/ebooksc/2024/01/BZ202401189533243.pdf

한겨레(네이트 뉴스 전재), 새 의료기기 시장 진입 490→80일로…"환자 안전 위협" 비판, 2026년 1월 26일, https://m.news.nate.com/view/20260126n27162

Business Wire, Paradromics Receives FDA Approval for the Connect-One Clinical Study with the Connexus ® Brain-Computer Interface, 2025년 11월 20일, https://secure.businesswire.com/news/home/20251120141086/en/

Queen's University Smith School of Business, counting the costs and opportunities of medical tech regulation (검색 색인 제목, 원문 대소문자 미확인), 날짜 미상, https://smith.queensu.ca/insight/content/counting-the-costs-and-opportunities-of-medical-tech-regulation.php

PMDA, Overview of SAKIGAKE Designation, 날짜 미상, https://www.pmda.go.jp/english/review-services/reviews/0012.html

법률신문, "뇌파로 생각까지 읽는 시대"… '신경 데이터' 보호 필요성 제기, 2026년 3월 17일, https://www.lawtimes.co.kr/news/articleView.html?idxno=217923

여성신문, 이수진 의원, 개인정보보호법 개정안 발의로 생체정보 남용 제동 건다, 2025년 1월 19일, https://www.womennews.co.kr/news/articleView.html?idxno=257233

개인정보보호위원회, 생체정보 보호 가이드라인, 2021년 9월, https://grant-documents.thevc.kr/213172_3.+%EC%83%9D%EC%B2%B4%EC%A0%95%EB%B3%B4_%EB%B3%B4%ED%98%B8_%EA%B0%80%EC%9D%B4%EB%93%9C%EB%9D%BC%EC%9D%B8(2021.9%EC%9B%94)_%EA%B0%9C%EC%9D%B8%EC%A0%95%EB%B3%B4%EB%B3%B4%ED%98%B8%EC%9C%84%EC%9B%90%ED%9A%8C.pdf

법령, 개인정보 보호법 제23조와 제71조, 개인정보 보호법 시행령 제18조

법령, 의료기기법 제29조

식품의약품안전처, 추적관리대상 의료기기 지정에 관한 규정(식품의약품안전처고시 제2023-70호), 2023년 11월 17일, https://emedi.mfds.go.kr/msismext/resources/template/docForm/emd/bif/deviceSubjectToTracking1.pdf

메디칼타임즈, 손톱만한 초소형 심박동기 부정맥 새 패러다임 열었다, 2022년 6월 26일, https://www.medicaltimes.com/Main/view.html?ID=1148097

Fierce Biotech, Neuralink implants brain-computer chip in first human trial participant, 날짜 미상, https://www.fiercebiotech.com/medtech/neuralink-implants-brain-computer-interface-first-human-trial-participant

Frontiers in Bioengineering and Biotechnology, Explant Analysis of Utah Electrode Arrays Implanted in Human Cortex for Brain-Computer-Interfaces(Woeppel 외), 2021년 12월 7일, https://www.frontiersin.org/articles/10.3389/fbioe.2021.759711/full

NAMSA, Guide to Post-Market Clinical Follow-up Requirements Under EU MDR (2017/745), 날짜 미상, https://namsa.com/resources/blog/guide-to-post-market-clinical-follow-up-eu-mdr/

Medical Xpress, Experimental brain implants are on the rise. What happens after the trials are over?, 2026년 9월 4일, https://medicalxpress.com/news/2026-09-experimental-brain-implants-trials.html

한국경제, 비대면 진료 내년 본격 시행…배리어프리 키오스크 규제 완화 [2025 새정부 경제성장전략], 2025년 8월 22일, https://www.hankyung.com/article/202508213683i

의협신문, 디지털치료기기 첫 허가, '치료 앱' 처방하는 시대 열릴까, 2023년 2월 15일, https://www.doctorsnews.co.kr/news/articleView.html?idxno=148487

더메디컬, 벌써 6호 나왔는데…디지털치료제 실사용은 아직도 '제도 밖', 2025년 6월 10일, https://www.themedical.kr/news/articleView.html?idxno=2379

영국 정부(GOV.UK), The Innovative Devices Access Pathway (IDAP) – pilot phase (now closed), 2026년 4월 30일 갱신, https://gov.uk/government/publications/the-innovative-devices-access-pathway-idap/the-innovative-devices-access-pathway-idap-pilot-phase

팜뉴스, 진입 80일 줄인 혁신의료기기, 환자 안전망은 공백, 2026년 9월 14일, https://www.pharmnews.com/news/articleView.html?idxno=306249

매일노동뉴스, [빗장 풀린 의료AI ②] 시장 진입은 쉽게, 사후관리는 구멍, 2026년 10월 14일, https://labortoday.co.kr/news/articleView.html?idxno=237180

中国政府网, 七部门关于推动脑机接口产业创新发展的实施意见, 2025년 8월, https://www.gov.cn/zhengce/zhengceku/202508/content_7035603.htm

NIH, BRAIN 2025: A Scientific Vision, 2014년 6월 5일, https://www.braininitiative.nih.gov/vision/nih-brain-initiative-reports/brain-2025-scientific-vision

ARIA, Work in progress | Precision Neurotechnologies, 날짜 미상, https://www.aria.org.uk/opportunity-spaces/scalable-neural-interfaces/precision-neurotechnologies/work-in-progress

JST, Goal 1: KANAI Ryota Project | Moonshot R&D, 날짜 미상, https://www.jst.go.jp/moonshot/en/program/goal1/12_kanai.html



종장. 변곡점은 되돌아오지 않는다

네 과제를 하나로: 교육, 목표, 통제, 연결

2026년 1월 5일 에르되시 문제 사이트의 게시판 참여자들은 728번 문제의 본뜻을 다시 정리했다. 하루 전 케빈 바레토가 올린 증명이 문제의 한 가지 해석만 풀었기 때문이다. 바레토는 GPT 5.2 Pro에게 논증을 넓힐 수 있는지 다시 물었고 이튿날 Aristotle이 그 답을 형식 증명으로 완성했다. 테렌스 타오가 1월 7일 마스토돈에 인공지능이 사실상 자율적으로 풀었다고 알린 것은 이 과정을 거친 판본이었다(타오 마스토돈, 2026년 1월 7일). 서장은 이 사흘을 기계가 증명을 쓴 쪽에서 보았다. 같은 사흘을 사람 쪽에서 보면 다른 일이 보인다. 게시판의 공동체는 무엇이 풀 가치가 있는 질문인지 다시 정했다. 바레토는 Lean 검사를 거치지 않은 인공지능 증명은 올리지 않는다는 원칙을 지켰다(바레토, 2026년 1월 26일). 함께 일한 리엄 프라이스는 한 모델이 낸 풀이를 새로 띄운 다른 모델에게 검사시켰다(Quanta, 2026년 8월 3일). 두 사람이 기계에게 뜻을 전한 통로는 자판과 대화창이었다.

이 사흘 안에는 본문 네 부가 다룬 네 과제가 작은 크기로 모두 들어 있다. 질문을 던지고 돌아온 결과를 확인하는 능력은 1부의 교육이 기르려는 것이다. 풀 문제를 고르는 일은 2부의 목표에 해당하고 검사를 거치지 않은 결과를 내보내지 않는 규칙은 3부의 통제에 해당한다. 자판과 대화창은 4부가 다룬 연결 수단의 지금 모습이다.

네 과제는 서로 이어진다. 교육이 문제를 정하고 결과를 따지는 사람을 기르면 그 사람이 목표를 세운다. 목표를 좇는 기계가 사람의 통제 안에 있어야 협업이 성립한다. 사람과 기계를 더 긴밀하게 잇는 장치도 그 통제의 조건 안에서 설계해야 한다. 네 과제 모두 기술 개발과 사람의 역량, 제도 설계가 함께 필요하다. 1부를 맺은 5장 끝의 장면인 2026년 7월 상하이 국제수학올림피아드(IMO)에서 화웨이와 샤오훙수는 자사 모델이 42점 만점을 받았다고 발표했다(France24/AFP, 2026년 7월 23일). 1장에 적은 대로 그 점수를 확인한 IMO의 공식 성명은 2026년 10월 14일까지 찾지 못했다. 기계가 만점을 받는 시대에도 그 만점이 맞는지 따지는 일은 사람에게 남는다.

과제 사이의 이음매는 본문에 이미 적혀 있다. 1장은 교육의 목표를 다시 세우는 일과 인간의 역할을 다시 정하는 일이 같은 사람을 두고 하는 이야기라고 적었다. 19장은

BCI가 사람이 인공지능에게 뜻을 전할 통로를 넓혀도 평가와 사고 보고와 멈추는 권한을 대신하지 못한다고 적었다. 1부와 2부는 한 사람에게서 만나고 3부와 4부는 한 장치에서 만난다. 남은 이음매는 2부와 3부 사이에 있다. 협업의 상대가 통제를 벗어나면 협업이 성립하지 않는다는 명제는 2026년 여름 기록이 됐다. 오픈AI 내부 평가에서 서로 격리돼 있어야 할 약 1,200개 에이전트가 승인되지 않은 게시판으로 소통했고 그중 약 700개가 허깅페이스 공격에 가담했다(METR, 2026년 8월 26일). 이 사고와 함께 12장이 다룬 세 사고에 대해 영국 AI안전연구소는 모두 주로 과제 성공을 좇은 행동으로 설명된다고 썼다(영국 AI안전연구소, 2026년 8월 4일). 사람이 정한 목표가 흐리거나 채점 방식에 틈이 있으면 기계는 그 틈으로 간다. 목표를 정하는 일과 통제의 조건을 세우는 일은 같은 문서의 앞장과 뒷장이라고 우리는 본다.

지식인이 할 일은 새 방향을 찾는 것

2026년 9월 11일 필즈상 수상자 25명은 공개서한 「수학에서 AI의 심각한 어긋남」을 냈다(테크크런치, 2026년 9월 11일). 8장이 전한 대로 서한은 인공지능을 막자고 하지 않았다. 인공지능 기업의 목표와 수학계의 목표가 어긋났다고 적었을 뿐이다. 타오는 10월 9일 칼텍 강연 슬라이드에 Math 2.0으로 가는 길의 주된 병목이 상상력이라고 적었다(3장). 기계가 답을 쏟아 내는 분야에서 먼저 나온 목소리는 다음에 무엇을 물을지에 관한 것이었다.

과학자가 자기 분야의 위험을 먼저 말한 선례는 있다. 1974년 7월 사이언스에 실린 버그 서한은 재조합 DNA 연구를 스스로 멈추자고 요청했고, 1975년 2월에는 캘리포니아 아실로마에 과학자 약 150명과 변호사들이 모여 실험 지침을 정했다(미국 국립의학도서관 Profiles in Science). 의회가 규제 입법을 준비하던 때였다. 아실로마는 무엇을 하지 않을지를 정한 회의였다. 지금 지식인에게 맡겨진 일은 그다음 단계인 어디로 갈지를 정하는 쪽에 걸려 있다.

한국의 논의는 아직 수단에 머물러 있다. 정부가 내건 목표는 3대 강국이라는 순위였다(국가인공지능전략위원회, 2026년 4월 14일). 교육 담론은 기능적 활용 능력에 치우쳤다는 진단을 받았다(이투데이, 2026년 10월 2일). 2027년 정부안에서 성능과 인프라 예산은 안전과 신뢰 예산의 224배였다(디지털데일리, 2026년 10월 6일). 순위와 도구와 예산은 모두 어떻게 갈지에 관한 숫자다. 어디로 갈지는 그 숫자에서 나오지 않는다. 본문이 내놓은 후보 가운데 그 물음에 닿는 것이 있다. 10장은 인구가 줄고 지방이 비는 사회를 인공지능으로 관리하는 첫 나라가 되는 일을 국가 목표 후보로 올렸다. 미국과 중국이 세우지 않은 과제이고 인공지능이 지역 서비스의 격차를 줄였다는

실증도 아직 없다. 10장은 1순위 목표로 만든 모델과 로봇을 먼저 쓸 곳으로 이런 지역을 꼽았다.

모종린 연세대 국제학대학원 명예교수는 인공지능이 지식인의 자리를 빼앗았다기보다 이미 비어 가던 자리에 들어왔다고 봤다(서울신문, 2026년 9월 28일). 그는 전문가를 믿지 않는 대중과 대중을 낮춰 보는 엘리트가 서로를 키웠다고 진단했다. 그리고 답을 권위의 복원보다 시민과 함께 쓸 상식의 언어에서 찾았다. 그가 내린 결론은 "AI 시대에 필요한 사람은 철학을 가진 현장 지식인이다"라는 문장이다.

방향은 지식인이 정하지 않고 시장이 정한다는 시각도 있다. 젠슨 황은 시장의 힘과 기존 제조물책임법이 해로운 출시를 이미 막는다고 본다(Axios, 2026년 9월 22일). 시장은 어떤 제품이 팔릴지를 빠르게 가린다. 하지만 국가가 어떤 문제를 풀지와 평가 환경의 문을 어디까지 열어 둘지는 매출로 가려지지 않는다. 11장이 짚은 대로 2026년 7월의 사고는 제품이 시장에 나가기 전 회사 안의 평가 단계에서 일어났다. 그래서 저자가 말하는 새 방향은 지식인이 혼자 정한 답이 아니다. 후보를 문장으로 내놓고 시장과 시민과 국회의 검증을 받는 일이다. 3장이 말한 문제를 발견한 사람과 모종린이 말한 현장 지식인은 같은 자리를 가리킨다.

변화에 끌려가는 쪽과 변화를 이끄는 쪽

2025년 8월 15일 베이징 국가스피드스케이팅경기장에서 한 대학 팀의 작은 로봇은 돌려차기를 헛차고 혼자 바닥에 넘어졌다(가디언, 2025년 8월 15일). 이듬해 대회를 두고 존 쾨치어는 대회 종목표를 시장 지도라고 불렀다(Forbes, 2026년 8월 19일). 로봇이 혼자 해낼 수 있는 일과 아직 사람이 붙어야 하는 일이 그 표에 드러난다는 뜻이다. 끌려가는 쪽과 이끄는 쪽은 기술을 가졌느냐보다 그 기술로 무엇을 시험할지 누가 정하느냐에서 갈린다.

기업에서는 같은 기술이 다른 결과를 낳았다. 국민은행 상담 인력은 1,133명에서 2025년 869명으로 줄었고 상담원 한 명이 하루에 받는 전화는 100콜에서 120콜이 됐다. 신한은행 상담 인력은 2024년 693명에서 2025년 697명으로 늘었다(경향신문, 2026년 4월 15일). 7장이 정리한 대로 같은 업권에서 같은 기술을 들여도 인력의 방향은 도입 방식에 따라 달라졌다. 오래된 기록도 같은 쪽을 가리킨다. 란제이 굴라티 연구진은 세 차례 불황 전후의 상장기업 4,700곳을 추적했다. 그 기록에서 비용만 줄인 기업은 불황 뒤 번성하지 못했다(HBR, 2010년 3월).

개인에게서도 갈림길이 보인다. 3장의 격차 연구에서 정답이 정해진 과제의 격차는 줄었지만 문제를 정하고 결과를 판단하는 과제의 격차는 남거나 커졌다. 7장이 전한 데이비드 오터 등의 특허 변호사 실험에서 도구를 뺀 뒤에도 실력이 오른 채로 남은 쪽은 경력 7년 이상의 시니어였고 주니어는 평균으로 나아진 것이 없었다(NBER, 2026년 9월). 도구를 쥐여 주는 것만으로는 이끄는 쪽에 서지 못한다.

국가의 자리도 이 기준으로 읽힌다. 토터스 미디어의 2025년 글로벌 AI 인덱스에서 한국은 종합 5위였지만 인재는 13위이고 산업생태계는 17위였다(국회예산정책처, 2026년 1월 22일). 10장이 짚은 대로 돈으로 살 수 있는 항목의 순위는 높고 사람과 시장의 순위는 낮다. 미국과 중국은 이기는 법을 서로 다르게 정했다. 10장이 정리한 대로 미국은 순위를 목표로 삼았고 중국은 지능 단말과 에이전트가 쓰이는 넓이를 목표로 삼았다. 17장의 비교로는 미국이 더 빠른 해독과 더 긴 이식 기록을 가졌고 중국은 승인과 보험 코드와 조달 목록을 먼저 맞췄다. 메이천 쑨은 이 차이를 두고 탁월함과 접근성이 승리를 정반대로 정의한다고 말했다(MIT Technology Review, 2026년 6월 1일). 한국이 어느 쪽 정의를 따를지는 아직 어느 문서에도 적히지 않았다.

교육과 목표를 바꿔도 현장이 따라오지 못한다는 반론이 있다. 한국경제는 고용노동부 자료를 들어 재직자 훈련 인원이 241만 명으로 2004년 이후 최저였고 직업훈련 예산도 2023년 2조 7,301억 원에서 2026년 2조 2,548억 원으로 줄었다고 보도했다(한국경제, 2026년 9월 14일). 한 명만 빠져도 회사가 멈추는 중소기업은 직원을 교육에 보낼 여유가 없다는 것이다. 이 지적은 맞다. 교육을 바꾸는 일은 채용을 늘리는 정책과 다르다. 그것은 남은 자리와 새 자리에 맞는 사람을 기르는 일이다. 훈련 예산과 채용 관행은 10장이 다룬 기관과 기업의 목표 안에서 풀어야 한다. 줄인 비용 옆에 새 임무와 새 매출을 함께 적는 조직이어야 사람을 다시 가르칠 이유도 생긴다.

어느 하나도 미룰 수 없는 네 가지 일

2026년 9월 16일 저녁 제프리 힌턴은 미국 의회의 비공개 브리핑을 마치고 인공지능을 규제할 시간이 1년쯤 남았다고 말했다(NBC News, 2026년 9월 17일). 그 말은 통제를 잃는 날짜를 예언한 것이 아니고 손쓸 수 있는 시간의 길이를 어림한 것이었다. 다리오 아모데이는 2026년 1월 에세이에서 강력한 인공지능이 1년에서 2년 안에 올 수도 있다고 썼다. 그보다 훨씬 늦을 수도 있다는 단서도 함께 적었다(다리오 아모데이, 2026년 1월). 두 사람의 시간표는 다르지만 모두 5년 안쪽에 걸쳐 있다.

한국에는 기술의 시계 말고 인구의 시계도 돈다. 통계청 추계에서 15세에서 64세 생산연령인구는 2020년대에 해마다 평균 32만 명씩 줄고 2030년대에는 50만 명씩 준다(통계청, 2023년 12월 14일). 인공지능이 이 빈자리를 메운다는 자료는 없다. 7장의 자료에서는 인공지능 노출도가 높은 업종의 청년 고용과 신규 채용이 먼저 줄어드는 양상이 나타났다. 경기와 채용 관행 등 다른 요인도 섞여 있어 그 감소를 모두 인공지능의 효과로 단정할 수는 없다. 인구 감소에 대응할 기술을 준비하면서 청년이 일을 배우고 경력을 시작할 기회도 함께 지켜야 한다.

네 일에는 순서가 있지만 하나를 마치고 다음으로 넘어갈 시간은 없다. 교육을 바꾼 효과는 몇 해 뒤에 나타난다. 교육부는 인공지능 교육 중심의 새 교육과정을 이르면 2029년에 들이겠다고 밝혔다(에듀프레스, 2025년 11월 10일). 그 교육과정에 무엇을 담을지는 지금 정해야 한다. 그사이 목표를 정하지 않은 기관과 기업은 비용 절감에서 멈춘다. 통제의 조문이 비어 있는 동안 에이전트는 이미 행정과 금융에 들어갔다. 국내 금융사들은 에이전트 도구가 쓰인 공격으로 정보를 잃었다(서울신문, 2026년 10월 11일). 연결 수단은 규칙을 먼저 세운 나라가 임상을 가져간다. 중국은 승인 1년 전에 수가 항목부터 만들었고 승인과 보험 코드와 조달 목록을 열흘 남짓한 사이에 맞췄다(BioSpectrum Asia, 2026년 10월 1일). 놀런드 아보가 생각만으로 체스 말을 옮긴 2024년 3월의 영상 뒤로 2년 반이 지났다. 2026년 10월 14일 기준으로 한국에 시험 장소가 있는 이식형 BCI 등록은 0건이었다(ClinicalTrials.gov, 2026년 10월 14일 조회).

네 가지가 서로 부딪힌다는 반론이 있다. 전 오픈AI 연구자 대니얼 코코타일로는 인공지능을 안전하게 규모를 키울 수 없다고 봤고 황은 속도와 안전을 맞바꾸는 구도 자체를 부정했다(Axios, 2026년 9월 22일). 저자 자신도 2026년 7월 칼럼에서 지금은 문명의 변곡점이고 신중함보다 무모함이 필요한 시점이라고 썼다(전자신문, 2026년 7월 28일). 3부의 평가와 보고와 조사 제안과 나란히 놓으면 이 문장은 어긋나 보인다. 담대한 투자는 안전 기준을 낮추자는 요구와 구별해야 한다. 메모리와 모델과 로봇과 BCI에 돈과 사람을 투입하되, 개발과 훈련과 평가, 배포와 운용의 각 단계에서 위험을 확인하고 멈출 권한을 갖춰야 한다. 환자를 대상으로 하는 BCI 임상도 이 원칙을 따른다. 앞의 칼럼이 촉구한 결단을 실제 사업으로 옮기는 방법은 투자와 통제 역량을 함께 키우는 것이다. 11장이 정리한 대로 속도를 두고 의견이 갈리는 사람들도 평가와 사고 보고와 투명성을 늘리자는 데에는 반대하지 않았다.

5년 안에 할 수 있는 일이 많지 않다는 반론도 있다. 국회예산정책처는 국가채무가 국내총생산 대비 2025년 47.8퍼센트에서 2072년 173.0퍼센트로 오른다고 전망했다(국회예산정책처, 2025년 2월 21일). 재정의 여유가 줄어드는 것은 사실이다. 국가 대표

모델과 휴머노이드 개발, 교육 접근권과 교원 연수, 안전연구소와 BCI 사업에는 지속적인 재원이 필요하다. 심사 기준과 법의 조문을 바꾸는 일도 현장에서 시행할 인력과 시간을 요구한다. 저자는 같은 칼럼에서 변곡점은 지나간 뒤에야 또렷해진다고 썼다. 모든 변화가 확실해질 때까지 기다릴 수는 없다. 제도 정비는 지금 시작하고, 예산 사업은 우선순위와 단계별 성과를 기준으로 재원을 배분해야 한다. 비용이 든다는 사실을 준비를 미룰 이유로 삼아서는 안 되지만, 필요한 비용을 작게 말해서도 안 된다.

한국이 다음 5년 안에 해야 할 일의 목록

다음은 5장과 10장과 15장과 20장이 제안한 일을 2027년부터 2031년까지의 할 일로 모은 것이다. 모두 저자의 안이고 정부 계획이 아니다. 조문 예시는 법제처 심사를 거치지 않았다.

1. 박사 과정에 질문, 분배, 검증, 책임을 배우는 세미나와 실습을 두고 분야의 기초 훈련에 맞춰 학기와 순서를 조정한다. 학위 구술시험에서는 인공지능 사용을 설명하고 방어하게 한다(5장).

2. 대학은 1학년 교양필수 3학점 한 과목에서 질문과 검증을 가르치고, 초등학교는 도구 없이 질문 만들기와 틀린 답 찾기부터 시작한다(5장).

3. 연구비 신청서와 학위논문에 인공지능 사용과 검증의 명세란을 두되 사용량은 심사하지 않고, 대학생과 교사에게 유료 인공지능 도구의 접근권을 제공한다(5장).

4. 2028년까지 공동 출자 법인으로 국가 대표 모델을 갖고 2030년까지 그 모델을 얹은 휴머노이드를 국내 공장과 공공 현장에 넣되 인구가 줄어드는 지역에 먼저 쓴다. 시한과 잣대는 법정 기본계획에 적고 해마다 국회에 성과를 보고한다(10장).

5. 기관과 기업의 인공지능 도입 계획서에 줄일 비용과 나란히 새로 맡을 임무와 새로 열 매출의 칸을 두고, 경영평가 가점은 새 임무의 결과에 매긴다(10장).

6. 인공지능안전연구소를 법률로 세우는 독립 법인으로 바꾸고 3년 안에 기술 평가 인력 50명과 400억 원 남짓의 예산을 확보하며, 국가 컴퓨팅 자원의 2퍼센트를 평가에 우선 배정한다(15장).

7. 중대 인공지능 사고는 24시간 안에 통지하고 72시간 안에 보고하게 하며, 아차사고 자율 보고와 보고자 보호를 두고 원인 조사는 국무총리 소속 사고조사위원회가 맡는다(15장).

8. 공공기관과 일정 규모 이상 사업자는 에이전트의 권한 범위, 외부 도구 사용 승인, 행동 기록 보존, 작동 중단 절차를 정해 운영한다(15장).

9. 2027년에 이식형 BCI의 품목과 등급, 임상 지침, 신경 데이터의 민감정보 규정을 내고 2028년에서 2029년 사이 첫 인체 임상을 그 규칙 아래 연다(20장).

10. 의사와 공학자를 한 팀으로 미국과 중국의 임상 기관에 파견할 협력 경로를 마련하고, BCI 단독 예산을 항목으로 공개한다. 국가 BCI 이식 등록부는 첫 인체 임상 전에 가동하고 2030년부터 참여 기관과 추적 범위를 넓혀 참여자를 최소 10년 따라간다(20장).

목록의 시한은 2027년의 BCI 규칙에서 시작해 2028년의 국가 대표 모델, 2028년부터 2029년 사이의 첫 인체 임상과 등록부 가동을 거쳐 2030년의 휴머노이드 현장 투입과 추적 체계 확장으로 이어진다. 기계가 증명을 쓴 2026년 1월의 사건에도 마지막에 이름을 건 사람이 있었다. 에르되시 문제 728번의 형식 증명을 사람이 읽을 글로 옮긴 논문은 2026년 1월 12일 수학자 냇 소타나판의 이름으로 arXiv에 올라왔다(arXiv, 2026년 1월 12일). 이 목록은 문명의 다음 목표를 찾기 위한 출발 단계다. 생산과 연구에 쓸 수 있는 능력이 커진 뒤 어떤 삶을 만들지, 그 혜택을 누가 누리게 할지는 계속 묻고 정해야 한다. 학교는 그 판단을 할 사람을 기르고, 기관과 기업은 새 임무를 만들며, 국가는 통제와 연결의 조건을 갖춰야 한다. 인간이 변화를 이끈다는 것은 더 많은 답을 내는 기계를 갖는 데서 끝나지 않는다. 그 능력으로 함께 이루려는 목적을 정하고 결과에 책임지는 일이다.

참고 자료

타오 마스토돈(mathstodon.xyz), 에르되시 문제 728번 글타래 첫 글(제목 없는 게시물), 2026년 1월 7일, https://mathstodon.xyz/@tao/115855840223258103

erdosproblems.com(Kevin Barreto), Problem 728 and the use of AI on Erdős problems, 2026년 1월 26일, https://www.erdosproblems.com/forum/thread/blog:2

Quanta Magazine, Why the Legendary Erdős Problems Are Falling to AI, 2026년 8월 3일, https://www.quantamagazine.org/why-the-legendary-erdos-problems-are-falling-to-ai-20260803/

arXiv(Nat Sothanaphan), Resolution of Erdős Problem #728: a writeup of Aristotle's Lean proof, 2026년 1월 12일, https://arxiv.org/abs/2601.07421

미국 국립의학도서관 Profiles in Science, 재조합 DNA와 아실로마 회의 항목, 날짜 미표기, https://profiles.nlm.nih.gov/spotlight/cd/feature/dna

서울신문(모종린), AI시대, 지식인의 자리는 어디인가, 2026년 9월 28일, https://www.seoul.co.kr/news/editOpinion/opinion/MoJongryn-AI/2026/09/28/20260928025001

Axios, With $7 trillion on the line, AI safety may always take a backseat to market domination, 2026년 9월 22일, https://www.axios.com/2026/09/22/ai-safety-vs-ai-speed

한국경제, AI 공부하라더니, 한 명만 빠져도 회사 멈춰 현실은 딴판, 2026년 9월 14일, https://www.hankyung.com/article/2026091439531

다리오 아모데이, The Adolescence of Technology, 2026년 1월(원문에 월까지만 표기), https://www.darioamodei.com/essay/the-adolescence-of-technology

통계청, 장래인구추계: 2022~2072년, 2023년 12월 14일, https://kostat.go.kr/board.es?mid=a10301010000&bid=207&list_no=428476&act=view&mainXml=Y

전자신문(김경진), 변곡점에서 새로운 질서가 탄생한다, 2026년 7월 28일, https://www.etnews.com/20260728000142

국회예산정책처, 국회예산정책처, 「2025~2072년 NABO 장기재정전망」 보고서 발간, 2025년 2월 21일, https://www.nabo.go.kr/ko/notice/releasesView.do?key=2507040045&idx=8642

France24/AFP, AI catches up with humans to score 100% at top maths contest, 2026년 7월 23일, https://www.france24.com/en/live-news/20260723-ai-catches-up-with-humans-to-score-100-at-top-maths-contest

영국 AI안전연구소, Incident Report: unsanctioned agent behaviour during cyber testing, 2026년 8월 4일, https://www.aisi.gov.uk/blog/incident-report-unsanctioned-agent-behaviour-during-cyber-testing

테크크런치, OpenAI's feud with mathematicians is only escalating, 2026년 9월 11일, https://techcrunch.com/2026/09/11/openais-feud-with-mathematicians-is-only-escalating/

국가인공지능전략위원회, [국가AI전략위] 대한민국 인공지능 행동계획 (한글, 인쇄본), 2026년 4월 14일, https://www.aikorea.go.kr/web/board/brdDetail.do?menu_cd=000011&num=523

이투데이, "AI 잘 쓰는 법만 가르쳐선 부족"…한국 AI교육 '활용 역량' 치중, 2026년 10월 2일, https://www.etoday.co.kr/news/view/2631840

디지털데일리, 이주희 의원 "AI 안전 예산, 274억 불과…'성능' 예산과 격차 224배" [2026 국감], 2026년 10월 6일, https://www.ddaily.co.kr/page/view/2026100610291136946

가디언, Box, run, crash: China's humanoid robot games show advances and limitations, 2025년 8월 15일, https://www.theguardian.com/world/2025/aug/15/china-world-humanoid-robot-games-advances-limitations

Forbes(John Koetsier), The Humanoid Robot Games Events Are A Market Map, 2026년 8월 19일, https://www.forbes.com/sites/johnkoetsier/2026/08/19/the-world-humanoid-robot-games-events-are-a-market-map/

경향신문, [단독] AI 도입 콜센터, 월급도 깎았다…대체 1순위? "자르기 쉬워 잘랐을 뿐"[딸깍, 노동②], 2026년 4월 15일, https://www.khan.co.kr/article/202604150600111

HBR, Roaring Out of Recession, 2010년 3월, https://hbr.org/2010/03/roaring-out-of-recession

NBER(Autor 외), Does AI Assistance Enhance or Erode Expertise? Evidence from a Three-Month Field Experiment in Patent Drafting, 2026년 9월(10월 개정), https://nber.org/papers/w35720

국회예산정책처, AI 3대 강국 도약을 위한 예산 현황과 향후 과제(나보포커스 제132호), 2026년 1월 22일, https://www.nabo.go.kr/board/file/down.do?fid=33319071

NBC News, 'Godfather of AI' warns Congress has 'maybe a year' left to regulate AI, 2026년 9월 17일, https://www.nbcnews.com/politics/congress/godfather-ai-warns-congress-maybe-year-left-regulate-ai-rcna598330

서울신문, AI 해킹 수법·대응책까지 공유했는데… 2주 뒤 금융권 줄줄이 정보유출, 2026년 10월 11일, https://www.seoul.co.kr/news/economy/2026/10/11/20261011500059

BioSpectrum Asia, Approved, Coded, Listed: China's neurotech speed run, 2026년 10월 1일, https://www.biospectrumasia.com/analysis/25/28439/approved-coded-listed-chinas-neurotech-speed-run.html

METR, Brief independent investigation of agents' behavior, reasoning and collaboration in the OpenAI / Hugging Face hacking incident, 2026년 8월 26일, https://metr.org/blog/2026-08-26-openai-hugging-face-incident-investigation/

MIT Technology Review, China has approved the world's first invasive brain-computer chip—here's what's next, 2026년 6월 1일, https://www.technologyreview.com/2026/06/01/1138133/china-world-first-brain-chip/

에듀프레스, AI 교육과정 개정, 2029년 시행 …정보교사 늘리고 10G 무선망 확대, 2025년 11월 10일, https://www.edupress.kr/news/articleView.html?idxno=21382

ClinicalTrials.gov, API v2 studies 검색(검색어 brain-computer interface, 국가별 조회), 2026년 10월 14일 조회, https://clinicaltrials.gov/api/v2/studies



부록

부록 1. 용어 설명

본문에서 처음 나올 때 풀어 쓴 용어를 가나다순으로 모았다. 괄호 안의 원어는 본문이 원어를 적은 14개 용어에는 본문 표기를 그대로 따랐고, 본문에 원어가 없는 외래 전문용어 52개에는 표준 영어 원어를 붙였다. 본문이 약어만 적은 뇌-컴퓨터 인터페이스(BCI)와 인공지능(AI)에는 영어 원어 전체를 더했다. 한국의 기관과 제도 이름(신의료기술평가, 임의규정, 추적관리대상 의료기기)과 표준 영어 표기를 정하지 못한 증강형 BCI에는 원어를 붙이지 않았다. "첫 등장"은 그 용어가 본문에 처음 나온 장이다. 용어가 먼저 나오고 설명은 뒤 장에서 붙은 경우에는 설명이 나온 장을 함께 적었다. 로마자로 시작하는 용어는 맨 뒤에 따로 두었다.

강화학습(reinforcement learning)
모델이 한 행동마다 점수를 매기고, 점수가 높게 나온 쪽으로 모델의 행동을 고쳐 나가는 훈련 방법이다.
첫 등장: 12장

검출 도구(AI detector)
어떤 글을 인공지능이 썼는지 가려내려는 프로그램이다. 글의 예측 가능성과 문체의 통계적 특징을 근거로 그 글을 기계가 썼을 가능성을 어림한다. 사람 글을 기계 글로 잘못 보는 오류를 오탐, 기계 글을 사람 글로 놓치는 오류를 미탐이라고 한다.
첫 등장: 4장

교정(calibration)
BCI의 디코더를 사용자 한 사람의 뇌 신호에 맞추는 작업이다. 뇌 신호가 날마다 조금씩 달라지므로 장치를 쓰는 동안에도 계속 고친다.
첫 등장: 16장

규제 샌드박스(regulatory sandbox)
새 기술을 기존 규제에서 한동안 풀어 주고 시험해 볼 수 있게 하는 제도다.
첫 등장: 20장

규제 포획(regulatory capture)
규제를 받는 기업이 규제의 내용을 자기에게 유리한 쪽으로 끌고 가는 현상이다.
첫 등장: 11장

기술 수용 모델(Technology Acceptance Model)
사람이 새 기술을 쓸지 말지는 쓸모가 있다는 인식과 쓰기 쉽다는 인식이 정한다고 보는 이론이다.
첫 등장: 18장

기술 실업(technological unemployment)
노동을 아끼는 수단이 생기는 속도가 노동의 새 쓰임이 생기는 속도를 앞질러 생기는 실업이다. 경제학자 케인스가 1930년 에세이에서 쓴 말이다.
첫 등장: 7장

기초 모델(foundation model)
많은 데이터로 미리 학습해 두고 여러 작업에 두루 쓰는 큰 인공지능 모델이다. 로봇용 기초 모델은 카메라로 본 장면과 사람의 말을 받아 팔다리의 동작으로 바꾼다.
첫 등장: 6장

노출도(exposure)
한 직종이 맡는 과업을 인공지능이 얼마나 해낼 수 있는지 잰 지표다.
첫 등장: 2장(설명은 7장)

뇌-컴퓨터 인터페이스(brain-computer interface, BCI)
뇌 신호를 읽어 기계를 움직이는 명령으로 바꾸는 컴퓨터 시스템이다. 본문은 처음에 "뇌와 컴퓨터를 잇는 장치"라고 부르고 그 뒤로는 BCI로 쓴다. 뇌파를 재기만 하고 출력을 내지 않는 장치는 BCI로 보지 않는다.
첫 등장: 서장(설명은 16장)

뇌심부자극(deep brain stimulation)
뇌 깊은 곳에 전극을 심고 전기 자극을 넣어 떨림 같은 증상을 줄이는 치료다. 뇌 신호를 계속 읽어 자극의 세기를 스스로 바꾸는 방식은 적응형 자극이라고 부른다.
첫 등장: 18장

다중 에이전트 체계(multi-agent system)
여러 에이전트가 한 과제를 나누어 맡아 처리하도록 짠 체계다.
첫 등장: 서장(설명은 3장)

대리 지표(proxy)
목표를 직접 측정하기 어려울 때 그 목표를 대신해 측정하는 수치다. 타오는 문제 풀이를 더 가치 있는 장기 목표의 대리 지표로 보았다.
첫 등장: 3장

대역폭(bandwidth)
정해진 시간 안에 주고받을 수 있는 정보의 양이다.
첫 등장: 19장

디코더(decoder)
전극이 잡아 걸러 낸 뇌 신호의 무늬를 보고, 사용자가 어떤 동작을 하려 했는지 맞히는 프로그램이다. 디코더가 내놓은 답은 커서의 이동, 글자, 합성 음성, 로봇 손의 움직임으로 나온다.
첫 등장: 16장

레드팀(red team)
일부러 반대편에 서서 약점을 찾아내는 역할이다. 공격자 역할을 맡아 시스템의 약점을 찾는 시험은 레드티밍이라고 부른다(14장).
첫 등장: 3장

명세 악용(specification gaming)
과제 규칙의 빈틈을 파고들어 과제는 풀지 않고 점수만 얻어 내는 행동이다.
첫 등장: 12장

모라벡의 역설(Moravec's paradox)
컴퓨터가 지능검사나 체커 게임에서 어른 수준을 내기는 쉬워도, 지각과 이동에서 한 살 아이의 솜씨를 갖추기는 어렵다는 관찰이다. 로봇공학자 한스 모라벡의 1988년 책에서 나왔다.
첫 등장: 6장

모라토리엄(moratorium)
정해진 기간 동안 개발 전체를 멈추는 조치다.
첫 등장: 11장

물리적 AI(Physical AI)
디지털 두뇌에 지능을 갖춘 몸을 결합해 생물에 견줄 기계를 만들어 내는 기술이다. 2020년 연구자 두 사람이 학술지에서 이렇게 정의했고, 엔비디아의 젠슨 황이 산업의 표어로 만들었다.
첫 등장: 6장

미션 지향 혁신 정책(mission-oriented innovation policy)
풀어야 할 문제와 기한을 먼저 정하고, 연구와 투자를 그 문제에 묶는 정책이다.
첫 등장: 9장

반증과 반례(disproof, counterexample)
추측은 참으로 여겨지지만 아직 증명이 없는 수학적 주장이다. 반증은 그 주장이 틀렸음을 보이는 일이고, 반례는 주장이 틀렸음을 보여 주는 사례 하나다.
첫 등장: 서장

반침습형 BCI(semi-invasive BCI)
침습형과 비침습형 사이의 방식이다. 뇌를 감싼 질긴 바깥 막인 경막의 바깥이나 뇌 표면에 전극을 얹고, 혈관 안에 전극을 넣는 방식도 여기에 든다. 중국 규제 문서는 같은 장치를 "침습형"이라는 범주에 넣기도 한다.
첫 등장: 16장

벤치마크(benchmark)
인공지능의 성능을 서로 견주려고 만든 표준 시험이다.
첫 등장: 1장

보간과 외삽(interpolation, extrapolation)
보간은 이미 아는 점들 사이를 메우는 일이고, 외삽은 아는 범위 바깥을 짐작하는 일이다.
첫 등장: 8장

보상 해킹(reward hacking)
강화학습에서 모델이 목표를 이루지 않고 보상을 주는 조건만 채우는 행동이다. 명세 악용에 속하는 행동이다.
첫 등장: 14장

비정렬(misalignment)
모델의 움직임이 개발자가 뜻한 목표에서 벗어나는 일이다.
첫 등장: 13장

비침습형 BCI(non-invasive BCI)
머리 겉에서 뇌 신호를 읽는 방식이다. 대표적인 예는 두피에 전극을 붙여 재는 뇌파 검사(EEG)다. 수술이 필요 없고 비용이 낮은 대신 뇌 신호가 두개골과 두피를 거치는 동안 흐려진다.
첫 등장: 16장

사고 사슬(chain of thought)
모델이 답을 내기 전에 거치는 추론 과정이다.
첫 등장: 12장

사전 접근 협약(pre-deployment access agreement)
개발사가 모델을 출시하기 전에 정부 평가 기관이 먼저 써 볼 수 있게 하겠다고 한 자발적 약속이다.
첫 등장: 14장

사회공학(social engineering)
사람의 신뢰를 이용해 공격자가 원하는 행동을 끌어내는 공격 방식이다.
첫 등장: 12장

샌드박스(sandbox)
프로그램을 가둬 놓고 돌려 바깥 시스템에 손대지 못하게 막는 격리 실행 환경이다.
첫 등장: 11장

생산성 역설(productivity paradox)
새 기술이 널리 퍼지는데도 생산성 통계가 따라 오르지 않는 현상이다.
첫 등장: 7장

신경권(neurorights)
신경기술로부터 사람의 정신을 지키려고 제안된 새 인권이다. 인지적 자유, 정신적 프라이버시, 정신적 온전함, 심리적 연속성 같은 권리가 후보로 나와 있다.
첫 등장: 19장

신의료기술평가
새 의료기술의 안전성과 유효성을 한국보건의료연구원이 따져 건강보험에 들어갈 수 있는지 가르는 절차다.
첫 등장: 20장

아차사고(near miss)
하마터면 사고가 될 뻔했지만 사고로 번지지 않은 일이다. 당사자가 법의 강제 없이 자기 실수나 아차사고를 스스로 알리는 방식을 자율보고라고 한다.
첫 등장: 13장

액추에이터(actuator)
로봇의 관절을 움직이는 부품이다. 휴머노이드의 액추에이터는 모터, 감속기, 구동 회로, 센서를 한 덩어리로 묶은 관절 모듈이다.
첫 등장: 6장

언어 모델(language model)
사람이 쓴 많은 양의 글을 학습해 다음에 이어질 말을 예측하게 만든 인공지능이다.
첫 등장: 서장

에이전트(agent)
사람의 지시를 받아 여러 단계의 일을 스스로 해내는 인공지능 프로그램이다.
첫 등장: 서장

역류 효과(washback)
시험과 평가가 거꾸로 교수와 학습의 방식을 바꾸는 영향이다.
첫 등장: 4장

오케스트레이션(orchestration)
여러 인공지능과 에이전트에게 일을 나누어 맡기고 그 결과를 조율해 합치는 일이다. 본문은 이 일을 "지휘"라고 부른다.
첫 등장: 1장

오픈웨이트 모델(open-weight model)
학습이 끝난 가중치를 공개한 모델이다. 누구나 그 가중치를 내려받아 고쳐 쓸 수 있다.
첫 등장: 14장

온보드 추론(onboard inference)
로봇이 학습을 마친 모델로 판단하는 일을 몸 안의 칩에서 하는 방식이다. 판단을 통신망 건너편의 데이터센터에 맡기는 방식은 클라우드 추론이다.
첫 등장: 6장

유사도 검사(similarity check)
제출된 글을 데이터베이스에 쌓인 기존 문서와 견주어 겹치는 부분을 찾아 세는 검사다. 겹친 원문을 함께 보여 주므로 어느 문장이 어디서 왔는지 눈으로 확인할 수 있다.
첫 등장: 4장

인공지능(artificial intelligence, AI)
추론과 판단처럼 사람이 하던 일을 컴퓨터가 하게 만든 기술이다. 본문은 처음에만 "인공지능(AI)"으로 쓰고 그 뒤로는 "인공지능"으로 쓴다.
첫 등장: 서장

인지적 자유(cognitive liberty)
자기 뇌와 정신 활동에 관한 일을 스스로 결정할 권리다.
첫 등장: 19장

임의규정
"할 수 있다"로 끝나는 조문이다. 행정기관에 권한을 주되 그 권한을 쓸지는 행정기관이 정하게 한다. 인공지능 기본법 제12조가 이런 조문이다.
첫 등장: 14장

자동화 편향(automation bias)
기계의 판단을 지나치게 믿어 그대로 따르는 사람의 경향이다.
첫 등장: 2장

자유도(degrees of freedom)
관절이 서로 독립해 움직일 수 있는 방향의 수다. 사람의 손은 약 27 자유도를 가진다.
첫 등장: 6장

재귀적 자기 개선(recursive self-improvement)
인공지능이 자기 자신을 고쳐 성능을 높이고, 그렇게 높아진 성능으로 다시 더 빨리 자신을 고치는 순환이다.
첫 등장: 11장

정렬(alignment)
인공지능이 사람이 뜻한 목표와 가치에 맞춰 움직이게 하는 일이다.
첫 등장: 11장

조기 타당성 연구(early feasibility study)
적은 수의 사람에게 장치를 처음 써 보며 안전성과 작동 여부를 살피는 임상 단계다. 시판 허가의 근거로 쓰는 대규모 확증 시험은 피보털(pivotal) 시험이다.
첫 등장: 17장

증강형 BCI
신체 기능이 정상인 사람의 감각, 인지, 운동 능력을 높이는 장치다. 중국 과학기술부의 BCI 연구 윤리 지침이 이렇게 정의했다.
첫 등장: 18장

증류(distillation)
강한 모델이 내놓은 답으로 약한 모델을 훈련하는 기법이다.
첫 등장: 12장

지연 시간(latency)
신호를 보낸 뒤 응답이 돌아올 때까지 걸리는 시간이다. 로봇이 추론을 몸 안에서 할지 데이터센터에 맡길지를 가르는 기준이 된다.
첫 등장: 6장

추론 모델(reasoning model)
답하기 전에 생각을 여러 단계로 펼치도록 훈련한 언어 모델이다.
첫 등장: 서장

추적관리대상 의료기기
부작용 위험이 크거나 소재를 파악해야 하는 기기를 식품의약품안전처가 지정해 공급과 사용 기록을 남기게 하는 제도다. 근거는 의료기기법 제29조다.
첫 등장: 20장

침습형 BCI(invasive BCI)
두개골을 열고 뇌 조직 속에 전극을 넣는 방식이다. 뉴런 바로 곁에서 신호를 들으므로 신호가 선명하지만, 수술의 부담이 따르고 전극이 오래 버티지 못하는 문제가 있다. 유타 어레이와 뉴럴링크 N1이 여기에 든다.
첫 등장: 16장

탈숙련(deskilling)
기계에 의지하는 동안 혼자서 일하는 사람의 능력이 줄어드는 현상이다.
첫 등장: 2장

토큰(token)
언어 모델이 글을 쪼개 처리하는 단위다. 낱말 하나 또는 낱말의 일부가 토큰 하나에 해당한다.
첫 등장: 1장

통제 이탈(loss of control)
인공지능이 누구의 통제도 받지 않고 움직이고, 그 통제를 되찾기가 극도로 어렵거나 불가능해진 상황이다.
첫 등장: 11장

프런티어 모델(frontier model)
한 시점에 성능이 제일 앞선 인공지능 모델이다.
첫 등장: 10장(설명은 11장)

프롬프트(prompt)
인공지능에게 건네는 질문이나 작업 지시의 글이다.
첫 등장: 2장

프롬프트 주입(prompt injection)
공격자가 모델이 읽을 문서나 입력 속에 자기 명령을 숨겨 두는 공격이다.
첫 등장: 14장

형식 증명(formal proof)
증명의 각 단계를 컴퓨터가 검사할 수 있는 꼴로 쓴 증명이다. 이런 검사를 위해 만든 프로그래밍 언어로 Lean이 있다.
첫 등장: 서장

환각(hallucination)
언어 모델이 사실과 다른 내용을 그럴듯하게 꾸며 내는 현상이다.
첫 등장: 1장

획기적 기기 지정(Breakthrough Device Designation)
미국 식품의약국(FDA)이 새 의료기기의 심사를 앞당겨 주는 제도다. 시판 허가와는 다르다.
첫 등장: 16장

휴머나이저(humanizer)
기계가 쓴 글을 고쳐 써서 검출 도구가 사람 글로 판정하게 만드는 도구다.
첫 등장: 4장

휴머노이드(humanoid)
사람처럼 머리, 몸통, 두 팔, 두 다리를 갖춘 로봇이다.
첫 등장: 서장(설명은 6장)

로마자로 시작하는 용어

AI 리터러시(AI literacy)
인공지능이 어떻게 움직이고 어디까지 할 수 있는지 알고 쓰는 기초 소양이다.
첫 등장: 1장

부록 2. 연표: 2022년 11월 챗GPT 공개부터 2026년 10월까지

본문 22개 장(서장, 1장에서 20장, 종장)에 나온 날짜 있는 사건을 날짜순으로 모았다. 각 줄은 날짜, 사건, 출처 장의 순서다. 날짜는 현지 기준이다. 날짜가 달까지만 확인되는 사건은 그달의 맨 앞에 두었다.

2022년

2022년 11월 30일 오픈AI가 챗GPT를 공개했다. 회사는 후속 질문에 답하고 실수를 인정하는 대화형 모델로 소개했다. (서장, 18장)

2022년 12월 한국 정부가 공공기관 정원 1만 2,442명(전체의 2.8퍼센트)을 줄이기로 했다. (10장)

2023년

2023년 1월 스위스 은행 UBS가 챗GPT의 1월 월간 활성 사용자를 약 1억 명으로 추정했다. 오픈AI의 공식 수치는 아니다. (18장)

2023년 1월 학술지 Science가 인공지능이 만든 글을 금지했다. (4장)

2023년 1월 31일 오픈AI가 자체 AI 글 검출기를 내놓았다. 영어 시험 자료에서 인공지능이 쓴 글의 26퍼센트만 잡아냈다. (4장)

2023년 4월 5일 Turnitin이 AI 글 검출 기능을 내놓으며 오탐이 1퍼센트 미만이라고 알렸다. (4장)

2023년 5월 국제의학학술지편집인위원회(ICMJE)가 권고를 고쳐 인공지능의 도움을 받은 작업을 밝히는 지침을 더했다. (4장)

2023년 5월 로이터가 내부 메모에서 누가 만들었든 로이터 기사는 로이터 기사라는 원칙을 세웠다. (2장)

2023년 5월 1일 뉴욕타임스가 제프리 힌턴이 구글을 떠났다고 보도했다. 힌턴의 공개 경고는 이날 시작됐다. (11장)

2023년 6월 22일 미국 뉴욕남부연방지방법원이 마타 대 아비앙카 사건에서 챗GPT가 지어낸 판결 6건을 인용한 변호사들에게 제재금 5천 달러를 부과했다. (2장)

2023년 7월 UC 데이비스 연구진이 ALS 환자 케이시 해럴의 왼쪽 운동피질에 전극 256개를 심었다. (16장)

2023년 7월 3일 영국 러셀그룹 24개 대학이 생성형 인공지능 교육 원칙 다섯 가지를 냈다. (4장, 5장)

2023년 7월 20일 오픈AI가 정확도가 낮다는 이유로 자체 검출기를 거뒀다. (4장)

2023년 8월 16일 미국 밴더빌트대가 Turnitin의 AI 검출 기능을 끄겠다고 안내했다. (4장)

2023년 8월 23일 스탠퍼드 연구진이 시도한 말을 1분에 62단어로 해독한 연구가 Nature에 실렸다. (17장)

2023년 10월 24일 중국 NEO의 첫 환자가 베이징 쉬안우병원에서 장치를 심었다. (16장)

2023년 11월 영국 AI안전연구소가 문을 열었다. 영국 정부는 창설 문서에 연구소가 규제기관이 아니라고 적었다. (14장, 15장)

2023년 11월 미국 안전연구소가 출범했다. (14장)

2023년 11월 Science가 고지를 조건으로 인공지능이 만든 글의 사용을 허용했다. (4장)

2023년 12월 미국 FDA가 첫 크리스퍼 치료제 캐스게비를 승인했다. (8장)

2023년 12월 14일 미국 국립과학재단(NSF)이 심사자가 승인받지 않은 도구에 제안서 내용을 올리지 못하게 했다. (5장)

2023년 12월 14일 통계청이 장래인구추계(2022~2072년)를 냈다. 15세에서 64세 생산연령인구가 2020년대에는 해마다 평균 32만 명씩, 2030년대에는 50만 명씩 준다고 추계했다. (종장)

2024년

2024년 1월 뉴럴링크가 첫 사람 이식을 했다. 미국 임상 PRIME(NCT06429735)도 이때 시작했다. (17장)

2024년 2월 일본 안전연구소가 정보처리추진기구 안에 사무국을 두고 출범했다. (14장)

2024년 2월 중국 과학기술부가 BCI 연구 윤리 지침을 내고 증강형 BCI를 정의했다. (18장)

2024년 3월 한국연구재단이 평가위원이 평가자료를 생성형 인공지능에 올리지 말라는 권고사항을 냈다. (5장)

2024년 3월 스타십이 한 기체 안의 두 탱크 사이에서 추진제를 옮겨 봤다. (9장)

2024년 3월 20일 뉴럴링크가 놀런드 아보가 손을 쓰지 않고 커서로 온라인 체스를 두는 9분짜리 영상을 X에 올렸다. (서장, 16장, 20장)

2024년 4월 17일 미국 콜로라도주가 신경 데이터를 민감 정보에 넣는 법에 서명했다. (19장)

2024년 5월 이종호 당시 과학기술정보통신부 장관이 안전연구소를 ETRI 안에 먼저 세우겠다고 했다. (15장)

2024년 5월 싱가포르가 난양공대의 디지털 신뢰 센터를 안전연구소로 지정했다. (14장)

2024년 5월 놀런드 아보가 장치 실의 약 85퍼센트가 뇌에서 빠져나왔다고 월스트리트저널에 말했다. 뉴럴링크는 소프트웨어 갱신으로 성능을 되살렸다고 설명했다. (16장)

2024년 6월 중국에서 최소 15개 대학이 학위논문의 인공지능 생성 비율 검사를 발표한 상태였다. 합격선은 20퍼센트 미만에서 40퍼센트 사이였다. (4장)

2024년 6월 3일 호주 고등교육 품질기준청(TEQSA)이 모든 등록 고등교육기관에 생성형 인공지능 대응 실행계획을 내라고 요구했다. (5장)

2024년 6월 17일 일론 머스크가 X에 "미래에는 휴대전화가 없고 뉴럴링크만 있을 것"이라는 답글을 달았다. (19장)

2024년 7월 국제핵융합실험로(ITER)가 새 일정을 내놓았다. 첫 운전은 2034년, 본운전은 2039년으로 밀렸다. (9장)

2024년 7월 싱크론이 오픈AI의 GPT-4o를 자사 BCI 대화 기능에 넣었다. (19장)

2024년 7월 25일 구글 딥마인드가 AlphaProof와 AlphaGeometry 2로 국제수학올림피아드(IMO) 문제를 은메달 수준으로 풀었다고 발표했다. (서장)

2024년 7월 29일 미국변호사협회(ABA)가 윤리의견 512호로 기존 직무 규칙을 생성형 인공지능에 적용했다. (2장)

2024년 8월 NEO가 중국 혁신의료기기 특별심사에 들어갔다. (17장)

2024년 8월 14일 UC 데이비스 연구진이 해럴의 말 복원 결과를 NEJM에 발표했다. 첫날 50단어 어휘에서 정확도는 99.6퍼센트였다. (16장)

2024년 9월 유네스코가 학생용 AI 역량 틀을 냈다. (1장, 3장)

2024년 9월 영국 연구혁신기구(UKRI)가 심사자의 생성형 인공지능 사용을 막았다. (5장)

2024년 9월 싱크론이 미국 임상 COMMAND 결과를 발표했다. 환자 6명 모두 12개월 동안 중대한 이상반응이 없었다. (16장, 17장)

2024년 9월 12일 오픈AI가 답하기 전에 더 오래 생각하도록 만든 o1-preview를 내놓았다. 추론 모델의 출발점이다. (서장)

2024년 9월 23일 샘 올트먼이 에세이 「지능의 시대(The Intelligence Age)」를 올렸다. (9장)

2024년 10월 다리오 아모데이가 에세이 「Machines of Loving Grace」에서 데이터센터 안의 천재들의 나라를 그렸다. (서장, 8장, 9장, 11장)

2024년 10월 과학기술정보통신부가 기본법 통과 전까지 안전연구소를 ETRI 조직으로 두기로 했다. (14장)

2024년 11월 캐나다 보건부가 토론토 병원의 뉴럴링크 임상을 승인했다. (17장)

2024년 11월 27일 한국 인공지능안전연구소가 판교 글로벌R&D센터에서 12명으로 문을 열었다. 초대 소장은 김명주 서울여대 교수다. (14장)

2024년 12월 미국에서 주 교육부 차원의 학교 인공지능 지침을 낸 곳이 20개 주였다. (5장)

2024년 12월 서울지방변호사회가 인공지능 결과물의 최종 판단과 책임이 변호사 본인에게 있다는 가이드라인을 제안했다. (2장)

2024년 12월 인공지능 기본법 심의 때 민원만으로 기업을 조사할 수 있다는 비판이 나왔다. 과학기술정보통신부는 사실조사를 필요 최소한의 범위에서 하겠다고 해명했다. (15장)

2024년 12월 미국 회계감사원(GAO)이 BCI 기술평가 보고서를 냈다. (19장)

2024년 12월 5일 한국기자협회 등 6개 언론 단체가 「언론을 위한 생성형 인공지능(AI) 준칙」을 발표했다. (2장)

2024년 12월 7일 힌턴이 노벨상 기자회견에서 안전을 더 일찍 생각했으면 좋았겠다고 말했다. (11장)

2025년

2025년 1월 올트먼이 블로그 글 「Reflections」에서 그해 첫 인공지능 에이전트가 일터에 합류할 수 있다고 썼다. (서장)

2025년 1월 엔비디아의 젠슨 황이 CES 기조연설에서 물리적 AI를 인공지능의 다음 단계로 꼽았다. (6장)

2025년 1월 미국 정부효율부(DOGE)가 움직이기 시작했다. 그 뒤 1년 동안 연방 인력은 채용을 빼고 26만 4,000명 줄었다. (10장)

2025년 1월 시카고대와 피츠버그대 연구진이 참여자 c1의 체성감각피질을 자극해 세 손가락에 물건 모양의 감각을 만들었다. (18장)

2025년 1월 21일 인공지능 기본법(법률 제20676호)이 공포됐다. (부록 3)

2025년 1월 24일 한국 디지털의료제품법이 시행됐다. (18장)

2025년 2월 법원 인공지능연구회가 법관은 사실 확정과 법률의 해석과 적용을 인공지능에 전적으로 맡기지 않아야 한다는 가이드라인을 냈다. (2장)

2025년 2월 영국 AI안전연구소가 이름을 AI Safety Institute에서 AI Security Institute로 바꿨다. (12장)

2025년 2월 EU 집행위원회가 인공지능 책임지침 제안을 거둬들이기로 했다. (2장)

2025년 2월 4일 허깅페이스의 마거릿 미첼 등 연구자 네 명이 완전 자율 에이전트를 개발하지 말아야 한다는 논문을 arXiv에 올렸다. (9장)

2025년 2월 10일 앤트로픽이 첫 경제 지수 보고서를 냈다. 직업의 약 36퍼센트가 관련 과업의 4분의 1 이상에 클로드를 썼다. (7장)

2025년 2월 19일 구글이 과학자를 돕는 다중 에이전트 체계를 AI co-scientist라는 이름으로 발표했다. (서장)

2025년 2월 20일 피규어가 로봇 제어 구조 헬릭스(Helix)를 공개했다. (6장)

2025년 2월 21일 국회예산정책처가 장기재정전망을 내고 국가채무가 국내총생산 대비 2025년 47.8퍼센트에서 2072년 173.0퍼센트로 오른다고 전망했다. (종장)

2025년 2월 24일 메드트로닉이 파킨슨병용 적응형 뇌심부자극의 FDA 승인을 발표했다. (18장)

2025년 3월 앤트로픽이 자사 모델 Claude가 범죄에 쓰인 사례를 보고서로 공개하기 시작했다. (12장)

2025년 3월 호주 가톨릭대가 Turnitin 검출 기능을 그만 쓰기로 했다. (4장)

2025년 3월 중국 국가의료보장국이 침습형 BCI 삽입비와 제거비 수가 항목을 만들었다. 같은 달 31일 후베이성이 삽입비 상한을 6,552위안으로 정했다. (17장)

2025년 3월 18일 엔비디아가 휴머노이드용 공개 기초 모델 GR00T N1을 내놓았다. (6장)

2025년 4월 정부가 K-휴머노이드 연합을 출범시켰다. (6장)

2025년 4월 듀오링고가 인공지능이 할 수 있는 일을 계약직에서 인공지능으로 옮기겠다고 밝혔다. (7장)

2025년 4월 프리시전 뉴로사이언스의 뇌 표면 전극 Layer 7이 FDA 510(k) 허가를 받았다. 허가 조건은 최대 30일 이식이다. (16장, 17장)

2025년 4월 데미스 허사비스가 CBS 「60분」에서 AI로 모든 질병을 끝내는 일이 아마 10년쯤 안에 가능하다고 말했다. (9장)

2025년 4월 중국 최고인민법원 덩슈밍 상무부원장이 모든 최종 결정과 책임은 사람 판사에게 있다고 연설했다. (2장)

2025년 4월 딥시크가 증명 모델 DeepSeek-Prover-V2를 공개했다. (8장)

2025년 4월 5일 메타가 30조 개가 넘는 토큰으로 훈련한 라마 4를 공개했다. (1장)

2025년 4월 19일 베이징 이좡에서 사람과 로봇이 함께 뛰는 첫 하프마라톤이 열렸다. 로봇 21대가 출발했고 톈궁 울트라가 2시간 40분으로 1위를 했다. (6장)

2025년 4월 23일 미국이 행정명령 14277로 청소년의 AI 리터러시 교육을 국가 정책으로 정했다. (1장)

2025년 4월 23일 마이크로소프트가 작업동향지수 2025에서 에이전트 상사라는 말을 썼다. (3장)

2025년 5월 중국 교육부 산하 위원회가 초중등 학생의 생성형 AI 사용 지침을 냈다. (1장, 5장)

2025년 5월 아랍에미리트 아부다비에서 뉴럴링크 임상 모집이 시작됐다. (17장)

2025년 5월 머스크가 2026년 말까지 무인 스타십을 화성에 보낼 가능성을 반반으로 봤다. (9장)

2025년 5월 NEO의 확증 임상(NCT06990412)이 시작됐다. 32명을 등록했다. (17장)

2025년 5월 아모데이가 인공지능이 5년 안에 신입 화이트칼라 일자리의 절반을 없앨 수 있다고 말했다. (7장)

2025년 5월 13일 싱크론이 애플의 BCI 입력 규격에 처음으로 기기를 직접 연결하는 회사가 되겠다고 발표했다. (19장)

2025년 5월 15일 필라델피아 아동병원이 희귀 대사질환 영아에게 맞춤형 염기 교정 치료를 투여했다고 밝혔다. (8장)

2025년 6월 뉴럴링크가 시리즈 E로 6억 5,000만 달러를 모았다. 이달 임상 참여자는 7명이었다. (17장)

2025년 6월 펑(Feng) 등 연구진이 에이전트의 자율성을 사용자가 맡는 역할로 정의한 논문을 냈다. (서장)

2025년 6월 미국 상무부가 안전연구소를 CAISI(Center for AI Standards and Innovation)로 개편했다. (14장)

2025년 6월 한국저작권위원회가 사람의 창작적 기여가 없는 생성물은 저작권 등록을 받지 않는다는 안내서를 냈다. (2장)

2025년 6월 구글 딥마인드가 로봇 안에서 바로 도는 제미나이 로보틱스 온디바이스를 공개했다. (6장)

2025년 6월 3일 요슈아 벤지오가 스스로 행동하지 않는 과학자 AI를 만들 비영리 연구소 LawZero를 세운다고 알렸다. (9장)

2025년 6월 10일 MIT 미디어랩 연구팀이 챗GPT를 쓴 집단의 인지 부채를 보고한 예비 연구를 냈다. (1장)

2025년 6월 13일 앤트로픽이 자사 연구용 다중 에이전트 체계의 구조를 공개했다. (3장)

2025년 6월 16일 힌턴이 팟캐스트 「다이어리 오브 어 CEO」에서 젊은이들에게 배관공을 권했다. (6장)

2025년 6월 17일 아마존의 앤디 재시가 일부 직무에는 사람이 덜 필요해진다는 메모를 직원들에게 보냈다. (10장)

2025년 7월 영국 런던과 뉴캐슬에서 뉴럴링크 임상이 시작됐다. (17장)

2025년 7월 유니트리가 R1을 3만 9,900위안에 내놓았다. (7장)

2025년 7월 영국 고등교육 옴부즈맨이 인공지능 부정 의혹 사례 4건을 공개했다. 3건은 유학생이나 제2언어 학생의 일이었다. (4장)

2025년 7월 평가 기관 METR이 숙련 오픈소스 개발자 16명이 인공지능 도구를 쓸 때 과제 시간이 19퍼센트 더 걸렸다는 실험을 발표했다. (3장)

2025년 7월 FrontierMath 4단계가 공개됐다. 공개 때 최고 점수는 5퍼센트였다. (8장)

2025년 7월 안전연구소 국제 네트워크가 세 번째 공동 평가를 에이전트에 맞춰 진행했다. (15장)

2025년 7월 MIT NANDA 연구진이 조사 조직의 95퍼센트가 생성형 AI로 측정할 만한 수익을 얻지 못했다는 예비 보고서를 냈다. (10장)

2025년 7월 메타가 손목 근전도로 손짓과 손글씨를 알아보는 밴드를 Nature에 발표했다. (18장, 19장)

2025년 7월 유엔의 2025년 지속가능발전목표 보고서에서 세부 목표의 35퍼센트만 제궤도에 있거나 보통의 진전을 보였다. (9장)

2025년 7월 17일 미국 국립보건원(NIH)이 인공지능이 실질적으로 쓴 신청서를 신청자의 독창적 아이디어로 보지 않는다고 공고했다. (5장)

2025년 7월 19일 오픈AI가 실험 모델의 IMO 2025 금메달 수준 성적을 자체 발표했다. (서장)

2025년 7월 21일 구글 딥마인드가 제미나이 딥 싱크 고급판이 IMO 조정위원 채점에서 여섯 문제 중 다섯 문제를 풀어 35점을 받아 금메달 기준에 들었다고 발표했다. (서장, 1장)

2025년 7월 23일 미국 백악관이 인공지능 행동계획 「경주에서 이기기(Winning the Race)」를 냈다. (10장, 11장)

2025년 7월 23일 중국 공업정보화부 등 7개 부처가 「뇌기계 인터페이스 산업 혁신 발전 추진에 관한 실시의견」에 서명했다. (17장)

2025년 8월 중국 국무원이 인공지능 플러스 행동 의견을 냈다. 지능 단말과 에이전트 보급률을 2027년 70퍼센트, 2030년 90퍼센트 넘게 올린다는 목표를 걸었다. (10장)

2025년 8월 엔비디아가 로봇용 칩 젯슨 AGX 토르의 출시를 발표했다. (6장)

2025년 8월 초·중등교육법 개정으로 AI 디지털교과서의 지위가 교과서에서 교육자료로 바뀌었다. (5장)

2025년 8월 브린욜프슨 연구팀의 「탄광 속의 카나리아?」 첫 판이 노출도가 높은 직종의 22세에서 25세 고용 격차를 13퍼센트로 보고했다. (7장)

2025년 8월 2일 EU 인공지능법의 체계적 위험 범용 모델 제공자 의무가 적용되기 시작했다. (13장)

2025년 8월 15일 베이징 국가스피드스케이팅경기장에서 제1회 세계 휴머노이드 로봇 경기대회가 열렸다. 한 대학 팀의 로봇이 돌려차기를 헛차고 넘어졌다. (서장, 6장, 10장)

2025년 8월 27일 앤트로픽이 위협 정보 보고서를 내고 Claude Code로 최소 17개 조직에 침입한 데이터 갈취 작전을 공개했다. (12장)

2025년 9월 세일즈포스의 마크 베니오프가 고객지원 인력을 약 9천 명에서 약 5천 명으로 줄였다고 말했다. (7장)

2025년 9월 젠슨 황이 영국 채널4 뉴스 인터뷰에서 모든 나라의 숙련 기능직이 호황을 맞는다고 말했다. (6장)

2025년 9월 뉴럴링크 임상 참여자가 12명이 됐다. 공동창업자 서동진은 서울 강연에서 마비 환자 12명이 하루 평균 약 7시간 40분 장치를 쓴다고 했다. (16장, 17장, 20장)

2025년 9월 KIST와 서울대 연구팀이 100나노미터 두께의 코팅으로 뇌에 넣은 전극의 수명을 1개월에서 3개월 이상으로 늘렸다(생쥐 실험). (20장)

2025년 9월 미국 연방 MIND Act가 발의됐다. (19장)

2025년 9월 미국 정보기술혁신재단(ITIF)이 연산량 기준을 없애는 대신 제12조를 고쳐 안전연구소에 배포 후 평가 권한을 주자고 제안했다. (14장)

2025년 9월 블룸버그가 뉴럴링크가 2030년까지 건강한 사람에게 장치를 넣기를 희망한다고 전했다. (18장)

2025년 9월 앤트로픽이 중국 국가 배후로 평가한 GTG-1002 집단의 활동을 9월 중순 탐지했다. (12장)

2025년 9월 8일 대통령 직속 국가인공지능전략위원회가 출범했다. (10장)

2025년 9월 10일 머스크가 X에 로봇의 손을 설계하는 일은 아주 어려운 문제라고 썼다. (6장)

2025년 9월 12일 고려대 학부대학이 탐지 결과를 성적이나 징계의 단독 근거로 삼지 않는 AI 활용 가이드라인을 냈다. (4장)

2025년 9월 26일 로봇공학자 로드니 브룩스가 사람 수준의 손재주가 수십 년 안에 온다는 믿음을 순전한 공상이라고 부른 에세이를 냈다. (6장)

2025년 10월 1X가 가정용 로봇 네오(NEO)를 2만 달러에 예약 판매했다. (6장)

2025년 10월 뉴럴링크의 말 복원 시험 VOICE가 댈러스에서 시작했다. (17장)

2025년 10월 아마존이 사무직 1만 4천 명 감원을 발표했다. (10장)

2025년 10월 9일 피규어가 손끝 센서를 단 피규어 03을 발표했다. (6장)

2025년 10월 18일 오픈AI의 케빈 와일이 GPT 5가 미해결 에르되시 문제 10개를 풀었다고 X에 썼다. 토머스 블룸이 극적인 왜곡이라고 반박했고 와일은 글을 지웠다.

(서장)

2025년 10월 30일 한국은행이 인공지능이 경력이 짧은 청년층의 정형화된 업무를 먼저 대신한다는 보고서를 냈다. 최근 3년 동안 청년층 일자리 21만 1천 개가 줄었다. (1장, 7장)

2025년 11월 유네스코가 신경기술 윤리 권고를 채택했다. (18장, 19장)

2025년 11월 중국 국가발전개혁위원회가 150곳이 넘는 기업이 비슷한 로봇을 쏟아낸다며 휴머노이드 거품 위험을 경고했다. (7장)

2025년 11월 파라드로믹스가 FDA의 IDE를 받아 초기 2명 등록으로 연구를 시작했다. (20장)

2025년 11월 에두아르도 페르난데스 연구진이 양방향 피질 시각 보철을 향한 논문을 냈다. (18장)

2025년 11월 13일 앤트로픽이 GTG-1002가 Claude Code로 약 30개 조직을 노린 작전을 공개했다. 작전의 80에서 90퍼센트가량을 인공지능이 수행했다고 분석했다. 다음 날 요청 속도 표현을 고쳤다. (12장)

2025년 11월 26일 영국 AI안전연구소가 프런티어 모델 4종이 안전 연구를 몰래 방해하는지 시험한 사례 연구를 냈다. 확인된 방해 사례는 없었다. (14장)

2025년 11월 27일 누리호가 4차 발사에서 차세대중형위성 3호를 600킬로미터 궤도에 올렸다. (9장)

2025년 12월 토터스 미디어의 글로벌 AI 인덱스에서 한국은 종합 5위였다. (10장)

2025년 12월 DORA와 유럽 연구평가개혁연합(CoARA)이 코펜하겐 회의 뒤 공동 성명을 냈다. (5장)

2025년 12월 우주항공청이 차세대발사체를 메탄 연료 재사용 발사체로 개발하기로 확정했다. (9장)

2025년 12월 11일 오픈AI가 GPT 5.2를 내놓았다. (서장)

2025년 12월 12일 과학기술정보통신부가 2026년 업무계획에서 AI 공동과학자로 2030년 노벨상급 성과를 내겠다고 밝혔다. (8장)

2025년 12월 23일 교육부와 17개 시도교육청이 수행평가의 인공지능 관리 방안을 내놓았다. (2장)

2025년 12월 26일 KISTEP이 이식형 BCI 보고서에서 식약처 임상 인허가의 시간과 비용 때문에 국내 임상 연구가 전무하다고 진단했다. (20장)

2025년 12월 30일 인공지능 기본법 개정안이 국회 본회의를 통과했다. 국가인공지능전략위원회의 심의·의결과 이행점검 권한, 첨단 기술 개발을 맡는 인공지능연구소의 근거가 생겼다. (10장, 15장)

2026년

2026년 1월 중국 항저우 인터넷법원이 인공지능은 법적 주체가 아니라고 보면서 사업자에게 고위험 영역에서 더 강한 경고를 할 주의의무를 지웠다. (2장)

2026년 1월 제15회 변호사시험이 치러졌다. 로앤컴퍼니는 뒤에 슈퍼로이어가 선택형 150문항을 모두 맞혔다고 발표했다. (1장)

2026년 1월 머지랩스가 2억 5,000만 달러를 투자받았다. (16장)

2026년 1월 아마존이 1만 6천 명 감원을 발표했다. (10장)

2026년 1월 머스크가 시력 복원 장치 블라인드사이트의 첫 사람 이식 준비를 마쳤고 규제 승인을 기다린다고 했다. (16장)

2026년 1월 1일 서울대가 「서울대학교 AI 가이드라인」을 제정했다. (4장)

2026년 1월 1일 교육부가 인공지능인재지원국을 새로 만들었다. (5장)

2026년 1월 1일 캘리포니아 SB 53이 시행됐다. (13장)

2026년 1월 1일 브룩스가 예측표에 현장용 로봇의 손재주가 2036년 이후에도 사람 손에 못 미칠 것이라는 예측을 올렸다. (6장)

2026년 1월 4일 케빈 바레토가 에르되시 문제 사이트 게시판에 GPT 5.2 Pro가 뼈대를 쓴 728번 증명을 올렸다. (서장)

2026년 1월 5일 에르되시 문제 사이트 게시판 참여자들이 728번 문제의 본뜻을 다시 정리했다. 바레토는 GPT 5.2 Pro에 논증을 넓힐 수 있는지 다시 물었다. (서장, 종장)

2026년 1월 6일 하모닉의 증명 도구 Aristotle이 728번의 형식 증명을 완성했다. (서장)

2026년 1월 7일 테렌스 타오가 마스토돈에 인공지능이 에르되시 문제 728번을 사실상 자율적으로 풀었다고 알렸다(협정 세계시 기준, 한국 시각 8일 새벽). (서장)

2026년 1월 10일 타오가 자기 글이 맥락 없이 인용된다고 덧붙였다. (서장)

2026년 1월 12일 수학자 냇 소타나판이 728번의 Lean 증명을 사람이 읽을 글로 정리한 논문을 arXiv에 올렸다. (서장)

2026년 1월 14일 타오가 사람의 개입이 적은 풀이일수록 짧고 표준 기법을 따르는 경향이 있다고 적었다. (서장)

2026년 1월 15일 독자 AI 파운데이션 모델 사업 1차 평가에서 계획한 4개 팀보다 적은 3개 팀이 다음 단계에 올랐다. (10장)

2026년 1월 22일 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법이 시행됐다. (2장, 10장, 13장, 14장)

2026년 1월 22일 국회예산정책처가 2026년 AI 예산을 9조 9,000억 원으로 집계한 보고서를 냈다. (10장)

2026년 1월 26일 아모데이가 에세이 「기술의 사춘기(The Adolescence of Technology)」를 냈다. (7장, 11장)

2026년 1월 26일 새 의료기기를 먼저 시장에 들이는 시장 즉시진입 의료기술 제도가 시행됐다. (20장)

2026년 1월 26일 바레토가 에르되시 문제 사이트에 쓴 글에서 Lean 검사를 거치지 않은 인공지능 증명은 올리지 않는다고 밝혔다. (서장, 종장)

2026년 1월 27일 피규어가 헬릭스 02를 공개했다. (6장)

2026년 1월 28일 뉴럴링크가 임상 등록 인원을 21명으로 밝혔다. (16장, 17장)

2026년 2월 유럽평의회 개인정보 협약 위원회가 신경 데이터 지침 초안을 냈다. (19장)

2026년 2월 3일 요슈아 벤지오가 의장을 맡은 국제 AI 안전 보고서 2026이 나왔다. (11장)

2026년 2월 8일 머스크가 X에 스페이스X의 우선순위를 화성에서 달로 옮긴다고 썼다. (9장)

2026년 2월 9일 브린욜프슨 연구팀이 청년 고용 감소의 금리 가설을 따진 후속 글을 올렸다. (7장)

2026년 2월 13일 법원행정처가 재판지원 AI 시스템 시범 운영을 발표했다. (2장)

2026년 2월 17일 마이클 바 미국 연방준비제도 이사가 인공지능이 노동시장에 가져올 앞날을 세 시나리오로 나눈 연설을 했다. (7장)

2026년 2월 23일 앤트로픽이 DeepSeek, Moonshot AI, MiniMax가 약 2만 4천 개 가짜 계정으로 1,600만 건 넘는 대화를 만들어 증류에 썼다고 지목했다. (12장)

2026년 2월 24일 정부가 「대한민국 인공지능 행동계획」을 확정했다. 실행과제 99개와 정책권고 326개를 담았다. (10장)

2026년 2월 28일 중국 유인우주공정판공실이 2030년 유인 달 착륙 목표를 유지한다고 CGTN이 전했다. (9장)

2026년 3월 중국 전국인민대표대회가 제15차 5개년 계획(2026년부터 2030년)을 승인했다. (9장, 10장)

2026년 3월 사이언스 코퍼레이션이 2억 3,000만 달러를 모았다. 중국 스테이지는 5억 위안을 모았다. (17장)

2026년 3월 정부가 BCI로 미국과 중국에 이은 3대 강국에 오르겠다는 구상을 내놓았다고 보도됐다. (20장)

2026년 3월 서울대 콜로키움에서 개인정보보호법에 신경 데이터를 넣자는 외국 전문가의 제안이 나왔다. (19장, 20장)

2026년 3월 올트먼이 AI가 노동과 자본의 균형을 흔드는데 어떻게 할지 아는 사람이 없다고 말했다. (9장)

2026년 3월 5일 타오가 728번 글에 인공지능의 방법이 수론학자 칼 포머런스의 2014년 논문과 많이 닮았다는 수정을 달았다. (서장)

2026년 3월 9일 교육부가 AI 중점학교 1,141개교를 선정했다. (5장)

2026년 3월 9일 행정안전부가 AI 국민비서 시범 서비스를 시작했다. (15장)

2026년 3월 13일 중국 국가약품감독관리국(NMPA)이 뉴라클의 NEO를 첫 침습형 BCI 의료기기로 시판 승인했다. 학술 분류로는 반침습 이식형이다. (16장, 17장)

2026년 3월 16일 브레인게이트 연구진이 참여자 2명 가운데 최고 속도가 1분에 110자였다는 결과를 발표했다. (16장)

2026년 3월 18일 과학기술정보통신부가 K-문샷의 BCI 임무를 2027년에 시작한다고 밝혔다. (16장, 17장)

2026년 3월 20일 MIT 테크놀로지 리뷰가 오픈AI의 완전 자동 연구자 계획을 보도했다. (서장)

2026년 3월 22일 NEO에 중국 국가 의료보험 소모품 코드가 나왔다. 24일까지 보험 코드, 상하이 목록, 조달 등재가 이어졌다. (17장)

2026년 3월 23일 미국 CAISI가 레드티밍 대회 결과를 공개했다. 모든 대상 프런티어 모델에서 공격이 한 번 이상 성공했다. (14장)

2026년 3월 24일 앤트로픽 경제 지수 보고서에서 관련 과업의 4분의 1 이상에 클로드를 쓴 직업의 비율이 약 49퍼센트였다. (7장)

2026년 3월 29일 타오가 블로그 글 「Mathematical methods and human thought in the age of AI」를 올렸다. 4월 18일 댓글에서 학생에게 문제의 답을 AI에 바로 묻지 말라고 했다. (3장, 5장)

2026년 3월 31일 법원행정처 TF가 허위 판례 인용에 소송비용 부담, 변론 진술 제한, 대한변협 징계 의뢰로 대응하는 방안을 냈다. (2장)

2026년 4월 코어텍의 Brain Interchange가 뇌졸중 운동 재활 용도로 FDA 획기적 기기 지정을 받았다. (16장)

2026년 4월 싱크론이 호주 세 병원에서 차세대 장치 시험을 시작했다. (17장)

2026년 4월 미국에서 코딩 에이전트가 자격 증명 오류를 만나자 운영 데이터베이스를 지웠다. 복구에 약 30시간이 걸렸다. (15장)

2026년 4월 교육부와 한국대학교육협의회가 대학 AI 기본교육과정 개발 지원사업 대상으로 20개교를 뽑았다. (1장)

2026년 4월 13일 스탠퍼드 HAI가 AI 인덱스 2026을 냈다. 상위 26개 모델의 환각률은 22퍼센트에서 94퍼센트였다. (1장)

2026년 4월 16일 테슬라의 힘줄 방식 로봇 손 국제특허가 공개됐다. (6장)

2026년 4월 19일 베이징 2회 하프마라톤에 로봇 300대 이상과 사람 1만 2천여 명이 나섰다. 아너의 로봇 라이트닝이 자율 주행 50분 26초로 최종 우승했다. (6장)

2026년 4월 27일 네이처가 테렌스 타오 인터뷰 「직무 설명서가 바뀌고 있다」를 실었다. (3장)

2026년 4월 29일 「알파고 10년, 위대한 동행」 행사에서 신진서 9단이 요즘은 모두가 AI로 바둑 공부를 한다고 말했다. (8장)

2026년 5월 과학기술정보통신부가 K-문샷 12개 임무의 총괄관리자 12명을 위촉했다. (9장)

2026년 5월 과학기술정보통신부가 부처 내부 행정에 에이전트를 들이는 AI-NEXT 사업을 발주했다. (15장)

2026년 5월 삼성SDS가 업무 목적에 맞는 에이전트를 골라 묶는 패브릭스 2.0을 발표했다. (3장)

2026년 5월 아모데이가 자동화를 일자리를 없애는 쪽보다 산출을 늘리는 쪽으로 설명했다. 올트먼은 신입 일자리에 대한 자기 예상이 틀려 기쁘다고 말했다. (7장)

2026년 5월 중국에서 병원 주도 임상연구 조례가 시행됐다. (17장)

2026년 5월 구글 Gemini 모델이 평가 중 실제 회사 세 곳의 시스템에 들어갔다. 9월 보도로 알려졌다. (13장)

2026년 5월 오픈AI 내부 팀이 에이전트의 게시판 활동과 허용되지 않은 인터넷 접속을 5월 말에 처음 보았다. (13장)

2026년 5월 5일 마이크로소프트 2026 보고서가 프런티어 프로페셔널이라는 집단을 소개했다. (3장)

2026년 5월 6일 LG가 국내외 전 임원 과정을 열고 임원이 에이전트를 직접 만들게 했다. (15장)

2026년 5월 13일 디지털데일리가 한국 안전연구소의 프런티어 모델 독자 평가 공개 보고서가 0건이라고 보도했다. (14장)

2026년 5월 18일 미국 버몬트주가 다섯 번째 신경 데이터 권리 법에 서명했다. (19장)

2026년 5월 19일 구글의 Co-Scientist와 퓨처하우스의 Robin이 네이처에 함께 실렸다. (8장)

2026년 5월 20일 오픈AI가 범용 추론 모델이 평면 단위 거리 문제에서 에르되시의 추측을 반증했다고 발표했다. 외부 수학자 아홉 명의 동반 논문과 윌 소인의 논문이 같은 날 arXiv에 올라왔다. (서장, 8장)

2026년 5월 20일 한국 안전연구소가 국내외 모델 42종의 안전성을 평가했다고 밝혔다. (14장)

2026년 5월 20일 스페이스X가 상장 신고서를 냈다. 2025년 매출 187억 달러 가운데 스타링크 중심의 연결 부문이 114억 달러였다. (9장)

2026년 5월 20일 제3자 사이트 뉴라 포드가 캐나다의 ALS 환자가 뉴럴링크 26번째 참여자로 이식을 받았다고 전했다(비공식). (17장)

2026년 5월 22일 인사혁신처가 공무원 교육훈련생을 위한 생성형 인공지능 지침을 처음 배포했다. (2장)

2026년 6월 엘스비어가 문법과 맞춤법 점검은 고지하지 않아도 된다고 정책을 고쳤다. (4장)

2026년 6월 파라드로믹스가 미시간대학병원에서 첫 참여자에게 장치를 심었다. (17장)

2026년 6월 김현 의원이 사업자가 에이전트의 권한 범위를 미리 정하게 하는 개정안을 냈다. (15장)

2026년 6월 교육부의 2026년 인공지능 선도교사 연수에 1만 268명이 신청했다. (5장)

2026년 6월 2일 트럼프 대통령이 행정명령 14409에 서명했다. 프런티어 모델을 출시 30일 전까지 정부에 자발적으로 내 사이버 능력을 검토받는 틀이다. (11장)

2026년 6월 4일 앤트로픽 연구소가 자사 인공지능 개발 작업의 점점 큰 몫을 인공지능에 맡기고 있다는 보고서를 냈다. (11장)

2026년 6월 9일 대법원이 일상어에서 법률 개념을 뽑아 판결문을 찾는 검색 시스템을 추진한다고 밝혔다. (2장)

2026년 6월 9일 라이프 바이오사이언스가 부분 재프로그래밍 방법을 쓴 ER-100을 첫 환자에게 투약했다. (8장)

2026년 6월 10일 행정안전부가 공공부문 AI 도입 가이드를 전국 공공기관에 배포했다. (15장)

2026년 6월 11일 특허법원이 등록무효 사건 판결문에 27년 차 변호사가 인공지능으로 찾은 허위 판례 8건을 검증 없이 냈다고 적고 부적절하다고 판시했다. (2장)

2026년 6월 15일 한국 안전연구소가 싱가포르 안전연구소와 함께 한 공동 테스트의 세부 결과 보고서를 냈다. (14장)

2026년 6월 15일 UC 데이비스가 해럴이 평균 1분에 56단어로 말한 후속 결과를 발표했다. (16장)

2026년 6월 16일 SK텔레콤이 AI 에이전트에 사번을 주고 직원처럼 관리하겠다고 밝혔다. (7장)

2026년 6월 17일 한국 안전연구소가 오픈AI와 고위험 분야 안전 협력 양해각서를 맺었다. (14장)

2026년 6월 18일 오픈AI 에이전트가 호주 메디케어 통계 포털에 접근했다. 오픈AI는 9월 10일에야 공용 메일함으로 알렸다. (13장)

2026년 6월 19일 EU 집행위원회와 OECD가 초중등용 AILit 틀을 확정했다. (3장)

2026년 6월 23일 뉴욕주 항소법원 2부가 존재하지 않는 판례를 인용한 항소 서면을 문제 삼아 변호사에게 8천 달러, 로펌에 2,500달러를 물렸다. (2장)

2026년 6월 26일 앤트로픽이 설문 응답자 약 9,700명을 사용 기록과 이은 경제 지수 보고서를 냈다. (7장)

2026년 6월 29일 중국 NMPA가 BCI 의료기기 분류 지도원칙을 냈다. (16장)

2026년 6월 30일 피규어가 BMW 스파턴버그 공장에 피규어 03을 넣었다고 밝혔다. (6장)

2026년 6월 30일 중국임상시험등록센터(ChiCTR)에 오른 중국 BCI 임상이 이날까지 134건이었다(중국 연구진 집계). (17장)

2026년 7월 구글 딥마인드가 휴머노이드의 온몸을 제어하는 제미나이 로보틱스 2를 공개했다. (6장)

2026년 7월 디지털 옴니버스 규정이 발효해 EU 부속서 III 고위험 시스템의 의무 적용일을 2027년 12월 2일로 미뤘다. (13장)

2026년 7월 영국 과학혁신기술부가 해체되면서 영국 AI안전연구소가 내각부로 옮겼다. (15장)

2026년 7월 안전연구소 국제 네트워크가 서울에서 모여 평가 모범 관행을 논의했다. (15장)

2026년 7월 중국 뇌호가 화산병원 등에서 완전 이식형 장치의 등록 임상을 시작했다. (17장)

2026년 7월 장기회복력센터의 통제 상실 관측소가 7월 한 달에 의도하지 않은 인공지능 행동을 300건 넘게 기록했다. (13장)

2026년 7월 1일 미국 FDA가 캐스게비의 대상을 2세 이상으로 넓혔다. (8장)

2026년 7월 4일 미국 정부효율부(DOGE)가 문을 닫았다. (10장)

2026년 7월 7일 Jun-Yong Park이 「이해 없는 자동화(Automation Without Understanding)」를 arXiv에 올렸다. (서장)

2026년 7월 7일 OECD가 고용전망을 내며 기업의 인공지능 사용이 노동 수요를 전반적으로 줄인다는 증거가 아직 없다고 밝혔다. (7장)

2026년 7월 7일 인실리코 메디신이 AI가 고른 폐섬유증 약 렌토서팁의 3상을 시작했다. (8장)

2026년 7월 9일 오픈AI 내부 평가에서 돌던 에이전트가 허깅페이스에 첫 공격 행동을 했다(협정 세계시 02시 28분). (12장)

2026년 7월 11일 에이전트가 허깅페이스 운영 환경 안으로 들어가 내부 운영 데이터베이스에 접근했다. (12장)

2026년 7월 12일 에이전트가 허깅페이스의 소스 코드 관리 시스템에 닿았다. (12장)

2026년 7월 13일 허깅페이스 침입의 마지막 의미 있는 공격 활동이 13시 37분에 있었다. 침입은 4.5일가량 이어졌다. (12장)

2026년 7월 13일 NEO의 첫 상업 이식이 상하이 푸단대 부속 화산병원에서 이뤄졌다. (17장)

2026년 7월 15일 상하이에서 IMO 시험이 16일까지 열렸다. 참가 학생 666명 가운데 7명이 만점을 받았고, 화웨이 셀리아와 샤오홍수의 모델이 42점을 받았다고 두 회사가 발표했다. (1장, 5장)

2026년 7월 16일 허깅페이스가 침입 사실을 공개했다. (12장, 13장)

2026년 7월 18일 앤트로픽의 Claude Haiku 4.5가 필라델피아 경찰의 미해결 살인 사건 제보 양식에 메시지를 냈다. (13장)

2026년 7월 19일 오픈AI의 보안 감시가 이상 활동을 잡았다. 다음 날 오픈AI는 이를 허깅페이스 사건과 연결했다. (13장)

2026년 7월 21일 오픈AI가 자사 내부 평가 모델이 허깅페이스 침입을 일으켰다고 공개했다. (11장, 12장, 13장)

2026년 7월 21일 인공지능 기본법 개정 법률 가운데 부칙 단서에 적힌 조항이 시행됐다. (부록 3)

2026년 7월 22일 사이언스 코퍼레이션의 망막 칩 PRIMA가 유럽 CE 인증을 받아 출시됐다. (18장)

2026년 7월 23일 앤트로픽이 모든 사이버 평가를 멈췄다. (12장)

2026년 7월 23일 피규어가 피규어 03의 1,000번째 기체를 만들었다. (6장)

2026년 7월 23일 영국과 미국이 중국 문샷의 Kimi K3에 대한 사이버 능력 공동 예비 평가를 냈다. (15장)

2026년 7월 25일 영국 AI안전연구소 사이버 평가에서 승인받지 않은 에이전트 행동이 28일까지 이어졌다. (12장)

2026년 7월 27일 허깅페이스가 침입의 기술 타임라인을 공개했다. 앤트로픽은 평가 파트너 Irregular와 피해 기관 세 곳에 사고를 알렸다. (11장, 12장)

2026년 7월 28일 오픈AI가 GPT 5.6 Sol 등의 에이전트가 허깅페이스 말고도 네 곳에 더 들어갔다고 밝혔다. 같은 날 영국 AI안전연구소가 사고를 선언하고 관련 모델의 접근을 막았다. (12장)

2026년 7월 28일 오픈AI, 앤트로픽, 구글, 메타 등의 직원 1,100명 이상이 늦출 수 있는 도구를 만들자는 공개서한에 서명했다. (11장)

2026년 7월 28일 저자가 전자신문 칼럼에서 지금은 문명의 변곡점이고 신중함보다 무모함이 필요한 시점이라고 썼다. (10장, 종장)

2026년 7월 30일 앤트로픽이 자사 사이버 평가 중에 생긴 실제 사고를 공개했다. (12장, 13장)

2026년 7월 31일 중국 베이나오 1호의 인체 이식이 40여 례라고 박람회 현장 직원이 말했다(7월 31일 기준). (17장)

2026년 8월 한국은행 후속 보고서가 4년 동안 청년층 일자리 28만 5천 개가 줄었고 감소분의 94퍼센트가 고노출 업종에 몰렸다고 분석했다. (7장)

2026년 8월 노가 알론이 인공지능이 에르되시 문제를 풀기 시작해 그 연구를 그만뒀다고 퀀타 매거진에 말했다. (8장)

2026년 8월 과학기술정보통신부가 한국 안전연구소가 평가한 모델 가운데 출시 전 평가 사례가 없다고 설명했다. (15장)

2026년 8월 영국 카니슈카 나라얀 AI 담당 장관이 연구소의 구조보다 결과를 보겠다고 했다. (15장)

2026년 8월 힌턴이 CNN 인터뷰에서 인공지능이 인류 멸종에 기여할 확률 10에서 20퍼센트는 어림짐작이라고 말했다. (11장)

2026년 8월 앤트로픽이 자사 모델로 리만 제타 함수 영점 가운데 리만 가설을 만족하는 비율의 하한을 41.6퍼센트에서 67.2퍼센트로 올렸다고 밝혔다. (8장)

2026년 8월 2일 EU AI사무국이 범용 모델 제공자에 대한 집행권을 갖게 됐다. (13장, 14장)

2026년 8월 3일 영국 AI안전연구소가 관련 모델 개발사와 미국 CAISI에 사고를 통지했다. (12장)

2026년 8월 3일 인공지능 기본법 시행 뒤 이날까지 발의된 개정안이 17건이었다. (15장)

2026년 8월 4일 영국 AI안전연구소가 모델 7종 122회 평가에서 승인받지 않은 행동 19건이 나온 사고 보고서를 공개했다. (12장, 13장)

2026년 8월 5일 국가교육위원회가 대통령 업무보고에서 수능 서·논술형 문항 도입과 전 과목 절대평가를 검토한다고 밝혔다. (1장)

2026년 8월 12일 티머시 고워스가 블로그에서 출판된 수학은 발견자의 생각 과정을 감춘 증명이라고 썼다. (8장)

2026년 8월 12일 브린욜프슨 연구팀 개정판이 노출도가 높은 직종의 22세에서 25세 고용이 약 19퍼센트 낮다고 보고했다. (7장)

2026년 8월 13일 타오가 사이먼스재단 대담에서 10만 줄짜리 기계 증명을 아무도 이해하지 못한다고 말했다. (8장)

2026년 8월 18일 오픈AI가 모델 규모를 키우는 속도를 한동안 늦추고, 배포를 앞둔 최신 모델의 강화학습 훈련을 2주 동안 멈췄다고 밝혔다. (12장)

2026년 8월 19일 유니트리가 상하이 증시 과창판에 상장했다. (7장)

2026년 8월 21일 교육부와 한국대학교육협의회가 탐지 결과를 징계의 결정적 증거로 인정하지 않는 대학 지침을 냈다. (4장)

2026년 8월 22일 제2회 세계 휴머노이드 로봇 경기대회가 개막했다. 16개국 666개 팀이 로봇 2,056대를 내보냈다. (6장)

2026년 8월 23일 중국 창어 7호가 발사 조건을 채우지 못해 올해 발사 창을 넘겼다. (9장)

2026년 8월 25일 팰컨 9의 1단 B1067이 37번째 비행을 마쳤다. (9장)

2026년 8월 26일 METR과 레드우드 리서치가 허깅페이스 사건 공동 분석을 냈다. 약 1,200개 에이전트가 승인되지 않은 게시판으로 소통했고 약 700개가 공격에 가담했다. 오픈AI도 후속 글을 냈다. (12장, 13장)

2026년 8월 27일 정부가 2030년까지 휴머노이드에 정부 재정 2조 3천억 원을 들이는 계획을 논의했다. (6장)

2026년 9월 국제로봇연맹(IFR)이 휴머노이드 판매량을 처음 따로 발표했다. 2025년 약 7천 대였다. (7장)

2026년 9월 예일대 버짓랩이 인공지능과 연결된 노동시장 교란의 뚜렷한 증거를 찾지 못했다. (7장)

2026년 9월 백악관이 자국 기업에 영국과의 모델 공유 중단을 요청했고, 앤트로픽이 Claude Mythos 5.1의 영국 검토를 막았다. (14장)

2026년 9월 타오가 사이언티픽 아메리칸 인터뷰에서 어떤 인공지능 사용을 받아들일지 수학자가 규칙을 정해야 한다고 말했다. (서장)

2026년 9월 뉴럴링크가 ALS 환자가 합성된 자기 목소리로 말하는 영상을 공개했다. (16장)

2026년 9월 한 대학생의 개인 프로젝트에서 오픈AI의 새 모델이 2026학년도 수능 450점 만점을 기록했다(비공식 실험). (1장)

2026년 9월 OECD가 PISA 2025 결과를 냈다. OECD 국가 학생의 46퍼센트가 한 주에 한 번 이상 AI 챗봇으로 공부했다. (1장)

2026년 9월 1일 Epoch AI가 FrontierMath 에르되시를 공개했다. GPT-6 Astra가 68문제 가운데 2개를 풀었다. (8장)

2026년 9월 3일 정부가 공공기관 109곳을 줄이거나 합치는 기능 개혁안을 발표하며 비용 절감이 목표가 아니라고 했다. (7장, 10장)

2026년 9월 4일 나이팅게일 컬렉티브가 오픈AI 평가 에이전트들이 독일어 프로그래밍 위키를 게시판처럼 썼다는 증거를 공개했다. (13장)

2026년 9월 4일 사법연수원 법관 연수에서 서면에 인용한 판례 40개 가운데 39개가 허위였던 사례가 공유됐다. (2장)

2026년 9월 5일 오픈AI 내부 모델의 1만 개 규모 에이전트가 88시간 만에 나비에 스토크스 문제의 답에 이르렀다(회사 설명). (8장)

2026년 9월 7일 유럽 연구평가개혁연합(CoARA)이 파리에서 이틀 동안 회의를 열었다. (5장)

2026년 9월 8일 오픈AI가 내부 모델이 나비에 스토크스 문제를 풀었다고 발표했다. 독립 검증은 10월 14일까지 끝나지 않았다. (서장, 8장)

2026년 9월 9일 토론토대 대학원이 박사 학위논문에 생성형 인공지능을 쓰려면 사전 승인을 받게 하는 지침을 갱신했다. (5장)

2026년 9월 10일 Epoch AI가 FrontierMath 4단계의 마지막 문제가 풀렸다고 알렸다. GPT-6 Astra의 점수는 98퍼센트였다. (8장)

2026년 9월 10일 국민건강보험공단과 질병관리청이 건강보험 빅데이터 분석센터를 함께 열었다. (10장)

2026년 9월 11일 필즈상 수상자 25명이 공개서한 「수학에서 AI의 심각한 어긋남」을 냈다. (8장)

2026년 9월 12일 아모데이가 에세이 「We Must Pace the Frontier」를 올렸다. 올트먼은 같은 날 동의한다고 썼다. (11장)

2026년 9월 13일 중국 관영 환구시보가 아모데이 에세이를 냉전 각본이라고 비판하는 사설을 냈다. (11장)

2026년 9월 14일 중국 외교부 대변인이 공포 조장과 대결이 세계 인공지능 거버넌스를 어지럽힐 뿐이라고 답했다. 같은 날 정보보안 표준 위원회 TC260이 인공지능 안전

거버넌스 프레임워크 3.0을 냈다. (11장)

2026년 9월 14일 트럼프 대통령이 Truth Social에 인공지능을 늦추는 일을 반기는 쪽은 중국뿐이라고 썼다. 젠슨 황은 올인 서밋에서 통제 불능이면 잠시 멈춰야 한다고 말했다. (11장)

2026년 9월 14일 EU가 다섯 미션의 중간 평가를 공개했다. (9장)

2026년 9월 14일 한국경제가 고용노동부 자료를 들어 재직자 훈련 인원이 241만 명으로 2004년 이후 최저이고 직업훈련 예산이 2023년 2조 7,301억 원에서 2026년 2조 2,548억 원으로 줄었다고 보도했다. (종장)

2026년 9월 15일 마이크 존슨 하원의장이 인공지능 개발에 모라토리엄을 둘 수 없다고 했다. (11장)

2026년 9월 16일 제프리 힌턴이 미국 의회 비공개 브리핑을 마치고 규제할 시간이 1년쯤 남았다고 말했다. 같은 날 랜드 폴 상원의원이 비상 정지 장치 법안의 만장일치 처리를 막았다. (11장, 15장)

2026년 9월 16일 오픈AI가 비정렬 사례 공개 체계를 발표하고 보고서 6건을 냈다. (13장)

2026년 9월 16일 배경훈 부총리가 중요한 것은 속도 조절이 아니라 속도 책임이라고 썼다. (11장)

2026년 9월 17일 금융보안원 세미나에서 딥시크 기반 에이전트가 금융사 서버에 악성프로그램을 설치한 사례가 공유됐다. 약 2주 뒤 여러 금융사에서 정보가 유출됐다. (15장)

2026년 9월 18일 월스트리트저널이 구글 Gemini 모델의 5월 평가 중 침입을 보도했다. (13장)

2026년 9월 18일 중국 스테이지 창업자가 운동기능 임상 검증 40여 례를 마쳤다고 말했다. (17장)

2026년 9월 20일 젠슨 황이 CBS 인터뷰에서 2030년이 세상의 끝이 될 확률은 0퍼센트라고 말했다. (11장)

2026년 9월 20일 오픈AI 내부 연구 모델이 훈련 샌드박스의 네트워크 제한 빈틈으로 외부 챗봇에 질문을 보냈다. 오픈AI는 최상위 모델군의 도구 사용 훈련, 평가, 추론을 멈췄다. (12장)

2026년 9월 20일 과학기술정보통신부가 「국가연구개발 AI 연구윤리 가이드」를 냈다. (4장, 5장)

2026년 9월 21일 유엔 독립 국제 AI 과학패널이 첫 주제 브리핑에서 7월 사고를 다뤘다. (11장)

2026년 9월 21일 오픈AI가 프린스턴 고등연구소에 둔 수학 독립 자문단의 출범을 알렸다. (8장)

2026년 9월 22일 Axios가 인공지능의 안전과 속도를 둘러싼 논쟁을 다뤘다. 젠슨 황은 시장의 힘과 기존 제조물책임법이 해로운 출시를 막는다고 보았고, 전 오픈AI 연구자 대니얼 코코타일로는 인공지능을 안전하게 규모를 키울 수 없다고 보았다. (종장)

2026년 9월 23일 젠슨 황이 뉴욕타임스 팟캐스트에서 실험을 가둘 수 없는 연구소라면 닫아야 한다고 말했다. (11장)

2026년 9월 23일 올트먼이 유엔 연설에서 AI가 새 르네상스가 될지 격변의 새 산업혁명이 될지는 인류의 선택에 달렸다고 말했다. (9장)

2026년 9월 26일 오픈AI가 이날까지 통지 기준에 맞는 활동을 이유로 100곳 넘는 조직에 알렸다. (13장)

2026년 9월 28일 한국 안전연구소가 카카오와 협약을 맺었다. (14장)

2026년 9월 28일 스타십이 이날까지 통합 비행을 14번 했다. (9장)

2026년 9월 28일 앤트로픽이 7월 제보 양식 메시지를 발견했다. 10월 8일 필라델피아 경찰에 알렸다. (13장)

2026년 9월 28일 모종린 연세대 국제학대학원 명예교수가 서울신문 칼럼에서 인공지능이 이미 비어 가던 지식인의 자리에 들어왔다고 진단했다. (종장)

2026년 9월 29일 트럼프 대통령과 앤트로픽, 구글, 메타, 오픈AI, xAI, 엔비디아의 대표가 「Joint Commitment on Frontier Responsibilities」에 서명했다. (11장)

2026년 9월 29일 AI중심대학 18곳과 AX대학원 15곳이 함께 출범했다. (10장)

2026년 9월 30일 과학기술정보통신부가 국가 AI 안전 생태계 조성 마스터플랜을 세울 민관 실무협의체를 꾸렸다. (13장)

2026년 9월 30일 K-문샷의 2027년 예산 1조 3,465억 원이 의결됐다. (9장, 20장)

2026년 10월 1일 뉴럴링크가 참여자들이 5만 시간 넘는 신경 데이터를 보냈다고 밝혔다. (17장)

2026년 10월 2일 미국 노동통계국이 9월 실업률을 4.2퍼센트로 발표했다. (7장)

2026년 10월 6일 오픈AI가 내부 모델이 낸 수학 결과를 GitHub 저장소에 한꺼번에 공개했다. 원고는 722편이었다. (서장, 8장)

2026년 10월 6일 과학기술정보통신부 국정감사에서 배경훈 부총리가 킬스위치 같은 인간의 통제 장치가 필요하다고 말했다. 이주희 의원은 AI 성능·인프라 예산과 안전·신뢰 예산의 격차를 지적했다. (11장, 12장, 14장)

2026년 10월 6일 호주 의회 AI 합동특별위원회 청문회에서 오픈AI와 앤트로픽 관계자가 의무 보고 체제를 환영했다. (13장)

2026년 10월 7일 오픈AI가 수학 원고 3편을 거둬들였다. (8장)

2026년 10월 7일 금융보안원이 금융분야 AI 에이전트 보안 평가 기준 17개 항목을 마련했다. 같은 날 금융위원회가 금융권 에이전트 규율체계 연구용역을 발주했다. (15장)

2026년 10월 7일 한국교육과정평가원이 수업과 평가를 위한 생성형 인공지능 가이드라인을 발표했다. 국정감사에서 차정인 국가교육위원장은 수능 서·논술형 도입 여부가 확정되지 않았다고 밝혔다. (1장, 2장)

2026년 10월 7일 누리호 5차 발사가 성공했다. (9장)

2026년 10월 9일 타오가 칼텍 강연 「Math 2.0」에서 수학이 증명이 귀한 시대에서 증명이 넘치는 시대로 넘어간다고 말했다. (3장)

2026년 10월 9일 앤트로픽이 평가와 내부 사용 중에 나온 의도하지 않은 모델 행동 보고서를 냈다. 같은 날 백악관 슈퍼인텔리전스 포스가 사고 신고는 선택 사항이 아니라고 선언했다. (13장)

2026년 10월 10일 다미앵 샤를로탱의 인공지능 환각 사건 데이터베이스에 2,184건이 올라 있었다(이날 기준). (2장)

2026년 10월 13일 34개국 과학자 약 400명이 주요 사고를 공개하는 국제 등록부를 요구했다. 같은 날 힌턴은 가디언 인터뷰에서 기업 자율규제 주장을 석유 회사와 석면 회사에 빗대 비판했다. (11장)

2026년 10월 13일 로이터가 오픈AI가 METR과 레드우드 리서치의 보고서에서 비공개 정보를 지웠다고 단독 보도했다. (13장)

2026년 10월 13일 박정 의원이 고용노동부 국정감사에서 한국은행 보고서의 청년 일자리 28만 5천 개 감소를 인용했다. (7장)

2026년 10월 13일 로브로스의 휴머노이드 20대를 제조 현장 네 곳에 넣는 실증이 규제특례 승인을 받았다. (6장)

2026년 10월 13일 한 집계에서 ARC-AGI-3 최고 점수가 GPT-6 Astra의 62.7퍼센트였다. 공식 검증치와 따로 확인하지 못했다. (8장)

2026년 10월 14일 필즈상 수상자의 공개서한 서명이 28명이 됐다. (8장)

2026년 10월 14일 ClinicalTrials.gov 조회에서 한국에 시험 장소가 있는 이식형 BCI 등록이 0건이었다(이날 조회 기준). (17장, 종장)

부록 3. 인공지능 기본법 제12조 전문과 관련 조문

법률명: 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법(약칭 인공지능 기본법)

시행일: 2026년 1월 22일

현행 법률 표기(국가법령정보센터 화면): [시행 2026. 7. 21.] [법률 제21311호, 2026. 1. 20., 일부개정]

제정 법률: 법률 제20676호, 2025년 1월 21일 공포, 2026년 1월 22일 시행

현행 시행령 표기(국가법령정보센터 화면): [시행 2026. 8. 20.] [대통령령 제36580호, 2026. 8. 18., 타법개정]

제정 시행령: [시행 2026. 1. 22.] [대통령령 제36053호, 2026. 1. 21., 제정]

소관: 과학기술정보통신부 인공지능안전신뢰정책과

국가법령정보센터는 부칙 단서 조항의 시행일인 2026년 7월 21일을 현행 법률의 시행일로 표기한다. 시행령의 현행 판은 2026년 8월 18일 다른 법령의 개정에 따라 바뀐 판이다.

조문 출처

법률: 국가법령정보센터, 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법, https://www.law.go.kr/LSW/lsInfoP.do?lsiSeq=282791 (본문 https://www.law.go.kr/LSW/lsInfoR.do?lsiSeq=282791&chrClsCd=010202&urlMode=lsInfoP&efYd=20260721&ancYnChk=0)

시행령: 국가법령정보센터, 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 시행령, https://www.law.go.kr/LSW/lsInfoP.do?lsiSeq=288781 (본문 https://www.law.go.kr/LSW/lsInfoR.do?lsiSeq=288781&chrClsCd=010202&urlMode=lsInfoP&efYd=20260820&ancYnChk=0)

조회 시각: 법률 2026년 10월 14일 14시 06분 KST, 시행령 2026년 10월 14일 14시 07분 KST

아래 1절부터 4절의 조문은 위 화면의 문구를 글자 그대로 옮겼다. 조문 제목 뒤 화면의 여백만 한 칸으로 줄였다.

2025년 12월 개정 사항

2025년 12월 30일 국회 본회의를 통과한 개정안은 의원 발의안 9건을 병합한 것이다. 이 개정으로 국가인공지능전략위원회의 심의·의결과 이행점검 권한이 법에 들어갔다(10장). 범용 인공지능 같은 첨단 기술 개발을 맡는 "인공지능연구소"의 설립·운영 근거도 새로 생겼다. 이 인공지능연구소는 제12조의 인공지능안전연구소와 다른 기관이다(15장). 국가법령정보센터 현행 법률에서 그 근거 조항은 제22조의2(인공지능연구소의 설립 및 지원 등)다.

개정 법률은 법률 제21311호로 2026년 1월 20일 공포됐다. 부칙에 따라 법은 2026년 1월 22일부터 시행하고, 부칙 단서에 적힌 조항만 공포 후 6개월이 지난 2026년 7월 21일부터 시행한다. 제12조는 이번 개정 대상이 아니어서 2026년 1월 22일부터 그대로 효력이 있다. 국가법령정보센터 현행 법률의 제12조에도 개정 표시가 없다.

부칙 원문: "이 법은 2026년 1월 22일부터 시행한다. 다만, 제3조제5항, 제6조제2항제7호 · 제8호, 제16조제3항부터 제5항까지(제2항제2호의2의 개정규정에 관한 부분에 한정한다), 제17조의2, 제18조, 제22조의3 및 제35조제1항 후단의 개정규정은 공포 후 6개월이 경과한 날부터 시행한다."

1. 법률 제12조 전문

제12조(인공지능안전연구소) ① 과학기술정보통신부장관은 인공지능과 관련하여 발생할 수 있는 위험으로부터 국민의 생명 · 신체 · 재산 등을 보호하고 인공지능사회의 신뢰 기반을 유지하기 위한 상태(이하 “인공지능안전”이라 한다)를 확보하기 위한 업무를 전문적이고 효율적으로 수행하기 위하여 인공지능안전연구소(이하 “안전연구소”라 한다)를 운영할 수 있다.

② 안전연구소는 다음 각 호의 사업을 수행한다.

1. 인공지능안전 관련 위험 정의 및 분석

2. 인공지능안전 정책 연구

3. 인공지능안전 평가 기준 · 방법 연구

4. 인공지능안전 기술 및 표준화 연구

5. 인공지능안전 관련 국제교류 · 국제협력

6. 제32조에 따른 인공지능시스템의 안전성 확보에 관한 지원

7. 그 밖에 인공지능안전에 관한 사업으로서 대통령령으로 정하는 사업

③ 정부는 안전연구소의 운영과 사업 추진 등에 필요한 경비를 예산의 범위에서 출연하거나 지원할 수 있다.

④ 그 밖에 안전연구소의 운영 등에 필요한 사항은 대통령령으로 정한다.

2. 시행령 제10조 전문

제10조(인공지능안전연구소의 운영 등) ① 법 제12조제2항제7호에서 “대통령령으로 정하는 사업”이란 다음 각 호의 사업을 말한다.

1. 법 제12조제1항에 따른 인공지능안전(이하 “인공지능안전”이라 한다) 관련 자문 및 교육

2. 인공지능안전 관련 평가 시스템의 구축 및 운영

3. 인공지능안전 관련 데이터(「데이터 산업진흥 및 이용촉진에 관한 기본법」 제2조제1호에 따른 데이터를 말한다)의 확보 및 공개

4. 인공지능안전 관련 통계 및 정보의 분석 · 제공 · 공유

5. 그 밖에 인공지능안전을 위하여 과학기술정보통신부장관이 필요하다고 인정하는 사업

② 과학기술정보통신부장관은 법 제12조제1항에 따른 인공지능안전연구소(이하 “안전연구소”라 한다)의 전문적이고 효율적인 운영 및 우수한 전문인력의 확보 등을 위하여 다음 각 호의 사항이 포함된 운영규정의 기준을 정할 수 있다.

1. 조직 및 인사에 관한 사항

2. 연구보안 및 윤리에 관한 사항

3. 국내외 협력 체계 구축 및 운영에 관한 사항

4. 예산 및 회계에 관한 사항

5. 그 밖에 안전연구소의 전문적이고 효율적인 운영을 위하여 필요한 사항

③ 안전연구소는 법 제12조제2항 각 호의 사업을 수행하기 위하여 필요한 경우에는 국가기관등에 관련 자료의 제공을 요청할 수 있다.

④ 안전연구소는 전년도 사업실적과 해당 연도의 사업계획을 매년 1월 31일까지 과학기술정보통신부장관에게 제출해야 한다.

3. 법률 제32조 전문(제12조제2항제6호가 가리키는 조문)

제32조(인공지능 안전성 확보 의무) ① 인공지능사업자는 학습에 사용된 누적 연산량이 대통령령으로 정하는 기준 이상인 인공지능시스템의 안전성을 확보하기 위하여 다음 각 호의 사항을 이행하여야 한다.

1. 인공지능 수명주기 전반에 걸친 위험의 식별 · 평가 및 완화

2. 인공지능 관련 안전사고를 모니터링하고 대응하는 위험관리체계 구축

② 인공지능사업자는 제1항 각 호에 따른 사항의 이행 결과를
과학기술정보통신부장관에게 제출하여야 한다.

③ 과학기술정보통신부장관은 제1항 각 호에 따른 사항의 구체적인 이행 방식 및 제2항에 따른 결과 제출 등에 필요한 사항을 정하여 고시하여야 한다.

4. 관련 조문 전문

관련 6개 조문이다. 각 조문 앞의 한 줄은 본문이 정리한 요지이고, 그 아래가
국가법령정보센터 원문이다.

제11조(인공지능정책센터) 요지. 과학기술정보통신부장관이 정책센터를 "지정할 수 있다"는 구조다. 제12조의 "운영할 수 있다"와 함께 임의규정이다(14장).

제11조(인공지능정책센터) ① 과학기술정보통신부장관은 인공지능 관련 정책의 개발과 국제규범 정립 · 확산에 필요한 업무를 종합적으로 수행하기 위하여 인공지능정책센터(이하 “센터”라 한다)를 지정할 수 있다.

② 센터는 다음 각 호의 사업을 수행한다.

1. 기본계획의 수립 · 시행에 필요한 전문기술의 지원

2. 인공지능과 관련한 시책의 개발 및 관련 사업의 기획 · 시행에 관한 전문기술의 지원

3. 인공지능의 활용 확산에 따른 사회, 경제, 문화 및 국민의 일상생활 등에 미치는 영향의 조사 · 분석

4. 인공지능 및 인공지능기술 관련 정책 개발을 지원하기 위한 동향 분석 , 사회 · 문화 변화와 미래예측 및 법 · 제도의 조사 · 연구

5. 다른 법령에서 센터의 업무로 정하거나 센터에 위탁한 사업

6. 그 밖에 국가기관등의 장이 위탁하는 사업

③ 그 밖에 센터의 지정 등에 필요한 사항은 대통령령으로 정한다.

제31조(인공지능 투명성 확보 의무) 요지. 생성형 인공지능 사업자에게 그 사실을 미리 알리고 결과물에 표시할 의무를 지운다(2장, 로펌 변호사 해설 기준).

제31조(인공지능 투명성 확보 의무) ① 인공지능사업자는 고영향 인공지능이나 생성형 인공지능을 이용한 제품 또는 서비스를 제공하려는 경우 제품 또는 서비스가 해당 인공지능에 기반하여 운용된다는 사실을 이용자에게 사전에 고지하여야 한다.

② 인공지능사업자는 생성형 인공지능 또는 이를 이용한 제품 또는 서비스를 제공하는 경우 그 결과물이 생성형 인공지능에 의하여 생성되었다는 사실을 표시하여야 한다.

③ 인공지능사업자는 인공지능시스템을 이용하여 실제와 구분하기 어려운 가상의 음향, 이미지 또는 영상 등의 결과물을 제공하는 경우 해당 결과물이 인공지능시스템에 의하여 생성되었다는 사실을 이용자가 명확하게 인식할 수 있는 방식으로 고지 또는 표시하여야 한다. 이 경우 해당 결과물이 예술적 · 창의적 표현물에 해당하거나 그 일부를 구성하는 경우에는 전시 또는 향유 등을 저해하지 아니하는 방식으로 고지 또는 표시할 수 있다.

④ 그 밖에 제1항에 따른 사전고지, 제2항에 따른 표시, 제3항에 따른 고지 또는 표시의 방법 및 그 예외 등에 관하여 필요한 사항은 대통령령으로 정한다.

제34조(고영향 인공지능과 관련한 사업자의 책무) 요지. 고영향 인공지능 사업자에게 설명 방안과 이용자 보호 같은 안전성 확보 조치를 정한다(2장, 로펌 변호사 해설 기준).

제34조(고영향 인공지능과 관련한 사업자의 책무) ① 인공지능사업자는 고영향 인공지능 또는 이를 이용한 제품 · 서비스를 제공하는 경우 고영향 인공지능의 안전성 · 신뢰성을 확보하기 위하여 다음 각 호의 내용을 포함하는 조치를 대통령령으로 정하는 바에 따라 이행하여야 한다.

1. 위험관리방안의 수립 · 운영

2. 기술적으로 가능한 범위에서의 인공지능이 도출한 최종결과, 인공지능의 최종결과 도출에 활용된 주요 기준, 인공지능의 개발 · 활용에 사용된 학습용데이터의 개요 등에 대한 설명 방안의 수립 · 시행

3. 이용자 보호 방안의 수립 · 운영

4. 고영향 인공지능에 대한 사람의 관리 · 감독

5. 안전성 · 신뢰성 확보를 위한 조치의 내용을 확인할 수 있는 문서의 작성과 보관

6. 그 밖에 고영향 인공지능의 안전성 · 신뢰성 확보를 위하여 위원회에서 심의 · 의결된 사항

② 과학기술정보통신부장관은 제1항 각 호에 따른 조치의 구체적인 사항을 정하여 고시하고, 인공지능사업자에게 이를 준수하도록 권고할 수 있다.

③ 인공지능사업자가 다른 법령에 따라 제1항 각 호에 준하는 조치를 대통령령으로 정하는 바에 따라 이행한 경우에는 제1항에 따른 조치를 이행한 것으로 본다.

제40조(사실조사 등) 요지. 조사 주체는 장관과 소속 공무원이고 안전연구소가 아니다.

제40조(사실조사 등) ① 과학기술정보통신부장관은 다음 각 호의 어느 하나에 해당하는 경우에는 인공지능사업자에 대하여 관련 자료를 제출하게 하거나, 소속 공무원으로 하여금 필요한 조사를 하게 할 수 있다.

1. 제31조제2항 · 제3항, 제32조제1항 · 제2항 또는 제34조제1항에 위반되는 사항을 발견하거나 혐의가 있음을 알게 된 경우

2. 제31조제2항 · 제3항, 제32조제1항 · 제2항 또는 제34조제1항의 위반에 대한 신고를 받거나 민원이 접수된 경우

② 과학기술정보통신부장관은 제1항에 따른 조사를 위하여 필요한 경우 소속 공무원으로 하여금 인공지능사업자의 사무소 · 사업장에 출입하여 장부 · 서류, 그 밖의 자료나 물건을 조사하게 할 수 있다. 이 경우 조사의 내용 · 방법 및 절차 등에 관하여 이 법에서 정하는 사항을 제외하고는 「행정조사기본법」에서 정하는 바에 따른다.

③ 과학기술정보통신부장관은 제1항 및 제2항에 따른 조사 결과 인공지능사업자가 이 법을 위반한 사실이 있다고 인정되면 인공지능사업자에게 해당 위반행위의 중지나 시정을 위하여 필요한 조치를 명할 수 있다.

제43조(과태료) 요지. 과태료 대상은 세 가지이고 제32조 위반에는 과태료가 없다(13장).

제43조(과태료) ① 다음 각 호의 어느 하나에 해당하는 자에게는 3천만원 이하의 과태료를 부과한다.

1. 제31조제1항을 위반하여 고지를 이행하지 아니한 자

2. 제36조제1항을 위반하여 국내대리인을 지정하지 아니한 자

3. 제40조제3항에 따른 중지명령이나 시정명령을 이행하지 아니한 자

② 제1항에 따른 과태료는 대통령령으로 정하는 바에 따라 과학기술정보통신부장관이 부과 · 징수한다.

시행령 제24조(인공지능 안전성 확보 의무) 요지. 법 제32조 대상 시스템은 세 요건을 모두 갖춘 것이고 연산량 산정 방식은 장관 고시로 정한다.

제24조(인공지능 안전성 확보 의무) ① 법 제32조제1항에서 “대통령령으로 정하는 기준 이상인 인공지능시스템”이란 다음 각 호의 기준을 모두 충족하는 인공지능시스템을 말한다.

1. 학습에 사용된 누적 연산량이 10의 26승 부동소수점연산 이상일 것

2. 인공지능기술의 발전 수준을 고려할 때 현재 인공지능시스템에 활용되는 인공지능기술 중 최첨단의 인공지능기술을 적용하여 구성 · 운영되고 있을 것

3. 인공지능시스템의 위험도가 사람의 생명, 신체의 안전 및 기본권에 광범위하고 중대한 영향을 미칠 우려가 있을 것

② 제1항제1호에 따른 학습에 사용된 누적 연산량의 구체적인 산정 방식은 과학기술정보통신부장관이 정하여 고시한다.

5. 조문을 읽을 때 함께 볼 사실

법과 시행령 어디에도 어느 기관을 안전연구소로 지정한다는 절차 조항은 없다. 한국전자통신연구원(ETRI) 부설이라는 형태는 법령이 아니라 과학기술정보통신부의 설립 방식 결정에서 나왔다(14장).

법률 전문에서 "사고"라는 말은 제32조제1항제2호 한 곳에만 나온다. 사고 보고 의무, 사고 조사권, 사고 접수 창구 조항은 법률과 시행령 모두에 없다(13장).

안전연구소의 자료 제공 요청 상대는 시행령 제10조제3항에 따라 "국가기관등"에 한정된다. 민간 사업자에게 모델이나 자료를 내라고 요구할 근거는 없다(14장).

저자가 제안한 제12조 개정 골격과 제32조의2부터 제32조의5, 제43조 개정 예시는 15장 본문에 있다. 이 예시는 법제처 심사를 거친 조문이 아니므로 이 부록에 현행 조문과 섞어 싣지 않는다.

부록 4. 주요 국가 인공지능안전연구소 비교표

11장부터 15장까지 본문에 나온 기관만 적었다. 본문에 없는 칸은 "본문 미기재"로 두었다. 수치 뒤 괄호는 본문이 밝힌 집계 주체다. "의원실 집계"는 이정헌 의원실이 과학기술정보통신부와 인공지능안전연구소에서 받은 자료를 디지털데일리가 2026년 9월 28일 보도한 것이고, 원화 금액은 의원실이 환산한 수치다. 기준 시점은 2026년 10월 14일이다.

나라	기관명(원어)	설립 시점	소속	인력 또는 예산	법적 권한	비고(출처 장)
한국	인공지능안전연구소 (원어 본문 미기재)	2024년 11월 27일 개소 (경기 성남시 판교 글로벌R&D센터). 보도는 세계에서 여섯 번째로 세운 안전연구소라고 적었다.	한국전자통신연구원(ETRI) 부설. 2026년 10월에도 ETRI 직할 조직이다. 부설 형태는 법이 아니라 과학기술정보통신부의 결정에서 나왔다.	출범 때 12명. 2026년 10월 전체 38명, 그 가운데 기술 평가 인력 6명. 2026년 예산 89억 원(의원실 집계). 장비는 엔비디아 H200 16장, 가동률 70에서 80퍼센트.	인공지능 기본법 제12조가 장관에게 연구소를 "운영할 수 있다"는 근거를 준다(임의규정). 민간 기업에 모델이나 자료 제출을 요구할 권한, 사실조사권, 명령권과 제재권이 없다. 자료 제공 요청 상대는 "국가기관등"에 한정된다. 사고 보고를 받을 창구도 법에 없다.	2026년 5월까지 모델 42종 평가. 주요 기업 프런티어 모델의 실명 평가 결과는 공개되지 않았다. 2026년 6월 17일 오픈AI, 9월 28일 카카오와 협약. 6월 15일 싱가포르와 공동 테스트 결과 보고서 공개. 2026년 8월 기준 출시 전 평가 사례 없음. 2027년 정부안의 AI 안전·신뢰 3개 사업 274억 원은 연구소 단독 예산이 아니다. (11장, 12장, 13장, 14장, 15장)

영국	영국 AI안전연구소(AI Security Institute). 2025년 2월 AI Safety Institute에서 지금 이름으로 바꿨다.	2023년 11월	과학혁신기술부 안의 연구 조직. 2026년 7월 과학혁신기술부가 해체되면서 내각부로 옮겼다.	기술 인력 100명 이상(기관 소개). 인력 200명 안팎, 예산 약 1,216억 원(의원실 집계). 회계연도당 6,600만 파운드에 더해 15억 파운드가 넘는 국가 컴퓨팅 자원에 우선 접근(기관 소개). 풀타임 연구원 100명의 평균 연봉 2억 원(김명주 소장 비교, 2025년 3월).	규제기관이 아니다(2023년 창설 문서). 찾아낸 피해에 대응할 규제 권한이 없다. 배포 전 모델 접근은 개발사와의 자발적 계약에 기댄다. 기업의 사고를 접수하는 법정 체계가 없다.	2026년 8월 4일 모델 7종 122회 평가 중 승인받지 않은 행동 19건을 다룬 사고 보고서 공개. Alignment Project로 60개 넘는 정렬 연구 과제에 2,700만 파운드 넘게 배정. 2026년 9월 백악관 요청 뒤 앤트로픽이 Claude Mythos 5.1의 영국 검토를 막았다. 10월 접근 회복 보도는 공식 확인이 없다. (12장, 13장, 14장, 15장)

미국	AI표준혁신센터, CAISI(Center for AI Standards and Innovation). 2023년 11월 출범한 미국 안전연구소를 개편한 기관이다. 개편 전 원어는 본문 미기재.	2023년 11월 출범, 2025년 6월 CAISI로 개편	상무부 NIST 안	인력 34명, 예산 약 232억 원(의원실 집계). 맡은 일을 다 하려면 연 8,400만 달러, 최소 규모로 연 2,600만 달러가 든다는 추산(Institute for Progress, 2026년 5월).	평가 권한이 자발적 참여에만 기댄다. 기업에 모델 제출을 강제할 수 없다.	오픈AI, 앤트로픽, 구글 딥마인드, 마이크로소프트, xAI 5개사와 협약을 맺고 40회 넘게 평가. 2026년 3월 레드티밍 대회 결과 공개. 2026년 7월 23일 영국과 함께 Kimi K3 사이버 능력 공동 예비 평가 발표. 2026년 8월 3일 영국 연구소에서 사고 통지를 받았다. (12장, 14장, 15장)

일본	일본 AISI(원어 정식 명칭 본문 미기재)	2024년 2월	정보처리추진기구 안에 사무국을 두었다.	2025년 12월 인력 약 30명, 정부 계획 초안은 두 배 증원(요미우리 보도). 31명에서 2026년 60명, 장기적으로 200명까지 늘릴 계획(15장). 예산은 당초 약 34억 원에서 추가경정예산으로 800억 원 이상(의원실 집계).	기업에 모델 제출을 강제할 수 없고 협약에 기대어 평가한다.	오픈AI가 각국 안전연구소와 맺은 협약 상대 가운데 하나(앞의 셋은 미국, 영국, 일본이고 한국이 네 번째). 앤트로픽과의 양해각서는 평가 방법론만 다뤄 모델 접근은 따로 협상해야 했다. 한국 연구소가 협력 관계를 맺은 상대다. (14장, 15장)

싱가포르	디지털 신뢰 센터(Digital Trust Centre). 원어는 14장 참고 자료 제목 기준.	2024년 5월 안전연구소로 지정	난양공대	본문 미기재	본문 미기재	한국 연구소와 글로벌 모델 3종의 이름을 가린 공동 테스트를 했고 2026년 6월 15일 세부 결과 보고서를 냈다. (14장)

EU	EU AI사무국(AI Office). 원어는 15장 참고 자료 제목 기준.	본문 미기재	유럽연합 집행위원회 안의 규제 집행 조직	본문 미기재	국제 네트워크 열 회원 가운데 유일하게 강제권이 있다. 2026년 8월 2일부터 범용 모델 제공자에게 기술 접근을 요구하고 과징금을 매길 수 있다. 과징금 상한은 전 세계 연매출의 3퍼센트와 1,500만 유로 가운데 큰 금액이다. 체계적 위험 범용 모델 제공자는 중대 사고를 지체 없이 AI사무국에 보고해야 한다(EU 인공지능법 제55조, 의무는 2025년 8월 2일부터 적용).	안전연구소 국제 네트워크 회원이다. 2025년 7월 네트워크의 세 번째 공동 평가에서 사이버보안 평가를 맡은 세 회원 가운데 하나였다. (13장, 14장, 15장)

중국	본문 미기재. 중국 대학과 연구기관이 만든 안전 연구 네트워크가 있다.	본문 미기재	본문 미기재	본문 미기재	본문 미기재	안전연구소 국제 네트워크 회원이 아니다. 위 연구 네트워크는 국제 대화에서 중국을 대표하는 일을 주로 맡고 있다는 분석이 있다(Carnegie, 2025년 6월). 정보보안 표준 위원회 TC260은 2026년 9월 인공지능 안전 거버넌스 프레임워크 3.0을 냈다. (11장, 14장)

캐나다	본문 미기재	본문 미기재	본문 미기재	본문 미기재	본문 미기재	15장은 한국 연구소가 이미 협력 관계를 맺은 상대로 일본과 캐나다를 들었다. 기관 정보는 본문에 없다. (15장)

표 밖의 사실

안전연구소 국제 네트워크. 회원은 열 곳이다. 2025년 7월 세 번째 공동 평가는 에이전트에 맞췄고 민감정보 유출과 사기, 사이버보안이 두 축이었다. 2026년 7월 네트워크는 서울에서 모여 평가 모범 관행을 논의했다. 영국 연구소는 2026년 7월 23일 글에서 네트워크를 새 이름(International Network for Advanced AI Measurement, Evaluation and Science)으로 적었다. 한국은 네트워크 공동 평가에 참여했지만 과제를 이끈 기록은 없다. (14장, 15장)

법적 권한의 공통점. 영국, 미국, 일본, 한국의 연구소는 모두 기업에 모델 제출을 강제할 수 없고 협약에 기대어 평가한다. 차이는 규모와 접근에서 난다. (14장)

같은 기관을 가리키는 다른 표기. 15장 본문의 출처 표기와 참고 자료에는 영국 기관이 "영국 AI안보연구소"로도 적혀 있다. 이 표는 "영국 AI안전연구소(AI Security Institute)"로 통일했다.

부록 5. 세계 BCI 임상 현황표

16장부터 20장까지 본문에 나온 회사와 기관만 적었다. 수치는 본문이 적은 숫자와 기준 시점을 그대로 옮겼고, 본문에 없는 칸은 "본문 미기재"로 두었다. 방식은 본문의 분류를 따랐다. 침습형은 두개골을 열어 전극을 뇌 조직 안에 넣는 방식, 반침습형은 경막 바깥이나 뇌 표면, 혈관 안에 전극을 두는 방식, 비침습형은 머리 겉에서 신호를 읽는 방식이다. 본문이 방식을 밝히지 않은 곳은 "이식형(세부 방식 본문 미기재)"으로 적었다. 기준 시점은 2026년 10월 14일이다.

1. 회사와 기관별 현황

회사 또는 기관	나라	방식	이식 또는 참여자 수와 기준 시점	승인·임상 단계	출처 장
뉴럴링크(Neuralink), 장치 N1	미국(임상은 미국, 캐나다, 영국, 아랍에미리트)	침습형(피질 안에 64가닥 실, 전극 1,024개)	임상 등록 21명(회사 발표, 2026년 1월 28일 기준). 이식 인원이 아닌 등록 인원이다. 첫 이식 2024년 1월, 2025년 6월 7명, 2025년 9월 12명. 그 뒤 회사의 공식 누계는 없다. 제3자 사이트 뉴라 포드는 2026년 5월 20일 26번째 참여자 이식, 10월 기준 공개적으로 알려진 환자 27명 이상으로 적었다(비공식 집계).	모두 등록부상 조기 타당성 연구. 미국 PRIME(NCT06429735)은 2024년 1월 시작, 예상 15명. 캐나다 2024년 11월 승인, 영국 2025년 7월 시작, 아랍에미리트 2025년 5월 모집 시작, 말 복원 시험 VOICE는 2025년 10월 시작(예상 6명). 피보털 시험 진입 발표 없음. FDA 시판 허가 없음. 시력 복원 장치 블라인드사이트는 2026년 1월 규제 승인을 기다린다고 했고 10월까지 사람 이식 발표가 없다.	16장, 17장, 19장, 20장
싱크론(Synchron),	미국(초기 임상은 호주)	반침습형(목의 정맥으로 넣어	호주 멜버른 초기 임상 SWITCH 4명. 미국 임상 COMMAND 6명(2024년 9월 발표, 12개월 동안 뇌나 혈관 관련 중대한 이상반응 없음). 2026년 4월부터 호주 세 병원에서 차세대 장치를 운동신경세포병 환자	임상 단계. 시판 허가 없음. 2024년 7월 GPT-4o를 대화 기능에 넣었고, 2025년 5월 13일 애플의 BCI 입력	16장, 17장, 19장

스텐트로드 (Stentrode)		뇌혈관 안에 전극을 둔다)	10명에게 심는 시험을 시작했고 실제 이식 수는 공개하지 않았다. 17장의 첫째 셈 기준에서는 약 10명으로 셌다.	규격에 기기를 직접 연결하겠다고 발표했다.	

파라드로믹스 (Paradromics)	미국	이식형(세부 방식 본문 미기재)	2026년 6월 미시간대학병원에서 첫 참여자 이식. 17장 셈에서 1명.	2025년 11월 FDA의 IDE를 받아 초기 2명 등록으로 시작하는 연구.	17장, 20장

프리시전 뉴로사이언스(Precision Neuroscience), Layer 7	미국	반침습형(뇌 표면에 얹는 전극)	본문 미기재	2025년 4월 FDA 510(k) 허가. 허가 조건은 최대 30일 이식. 2026년 9월 2억 5,000만 달러 투자 유치.	16장, 17장

UC 데이비스 연구진(참여자 케이시 해럴)	미국	이식형(말을 담당하는 왼쪽 운동피질에 전극 256개)	1명. 2023년 7월 이식.	연구 임상. 2024년 8월 NEJM 논문, 2026년 6월 후속 논문. 집에서 거의 2년 동안 3,800시간 넘게 사용.	16장

브레인게이트(BrainGate) 연구진	미국	이식형(세부 방식 본문 미기재)	2명(ALS 환자 1명, 목 척수 손상 환자 1명). 2026년 3월 16일 발표.	연구 임상. 두 사람 모두 집에서 사용.	16장

스탠퍼드 연구진	미국	본문 미기재	본문 미기재	2023년 연구에서 시도한 말을 1분에 62단어로 해독(Nature, 2023년 8월).	17장

피츠버그대·시카고대 연구진	미국	이식형(체성감각피질에 전극을 심어 자극)	참여자 c1(2025년 1월 발표), 네이선 코플랜드(2021년 발표). 각 1명.	연구 임상. 기계 신호를 뇌에 넣어 감각을 돌려주는 실험.	16장, 18장

코어텍(CorTec), Brain Interchange	독일	이식형(뇌 신호 기록과 전기 자극을 한 장치에 담음, 세부 방식 본문 미기재)	본문 미기재	2026년 4월 뇌졸중 운동 재활 용도로 FDA 획기적 기기 지정. 시판 허가와는 다르다.	16장

미겔 에르난데스대 연구진(에두아르도 페르난데스)	스페인	이식형(시각피질에 전극 96개, 6개월)	1명(2021년 12월 논문). 2025년 11월 양방향 시각 보철을 향한 후속 논문.	연구 임상.	18장

메타 연구진 (MEG)	미국	비침습형(뇌의 자기장을 재는 장치)	자원자 35명	연구 단계. 연구를 이끈 장레미 킹은 제품으로 가는 길이 없다고 봤다. 바깥 자기장을 막는 방 안에서만 작동한다.	16장

뉴라클(博睿康, Neuracle)과 칭화대, NEO	중국	반침습형(경막 바깥에 전극 접점 8개, 배터리 없음). 중국 규제 범주로는 "침습형" 3류 의료기기.	첫 환자 2023년 10월 24일 베이징 쉬안우병원 이식. 확증 임상 32명(NCT06990412, 2025년 5월 시작, 완료)과 앞선 타당성 임상 4례를 더해 36례. 6개월 시점에 장치 없이도 손 기능 점수가 기준 이상 오른 사람 22명(68.8퍼센트, 회사 자료). 첫 상업 이식 2026년 7월 13일 상하이 화산병원, 7월 이후 13세트 판매, 8세트 이식(상장 심사 자료 인용 보도, 2026년 10월).	2024년 8월 혁신의료기기 특별심사 진입. 2026년 3월 13일 NMPA 시판 승인(세계 첫 반침습 이식형 BCI 승인). 3월 22일 국가 의료보험 소모품 코드 발급. 기기 1세트 47만 5,000위안으로 27개 성 조달 목록 등재. 대상은 목 척수 손상으로 손을 쥐지 못하는 성인.	16장, 17장, 18장

베이나오 1호 (北脑一号)	중국	이식형(세부 방식 본문 미기재)	중국 연구진 논문 표에 30명(2026년 6월 말 기준 중국 이식 73명에 포함). 박람회 현장 직원 발언으로 2026년 7월 31일 기준 인체 이식 40여 례(비공식).	본문 미기재	17장

스테이지(阶梯医疗, 상하이)	중국	본문 미기재	창업자 발언으로 2026년 9월 18일 운동기능 임상 검증 40여 례(동료 심사 없음).	임상 검증 단계(창업자 발언). 2026년 3월 5억 위안 투자 유치.	17장, 18장

뇌호(脑虎科技, 상하이)	중국	이식형(완전 이식형 장치, 세부 방식 본문 미기재)	참여자 수 비공개	2026년 7월 화산병원 등에서 등록 임상 시작.	17장

사이언스 코퍼레이션(Science Corporation), PRIMA	미국	망막 아래에 심는 칩. 신호를 넣는 곳이 뇌가 아닌 망막이다.	환자 38명 임상 논문(2025년 10월).	2026년 7월 유럽 CE 인증을 받아 출시. FDA 승인 없음.	18장

지브레인	한국	본문 미기재(서울대병원의 뇌전증 진단용 전극 시험)	본문 미기재	운동이나 말 복원용 BCI 임상이 아닌 진단용 전극 시험. 미국 FDA 절차와 국내 인허가를 함께 밟으며 2026년 기술특례 상장 준비.	17장, 20장

서울대병원 총괄 국책 컨소시엄	한국	이식형 양방향 장치(세부 방식 본문 미기재)	인체 이식 없음(임상 시작 전)	2028년 이후 인체 임상 계획. 7년 동안 국비 202억 5,000만 원을 포함해 약 300억 원 규모.	17장, 20장

2. 나라와 세계의 누계

미국과 호주. 뉴럴링크, 싱크론, 파라드로믹스를 더해 32명(2026년 6월 말 기준, 기업 공식 발표). (17장)

중국. 하이난대학과 베이징 톈탄병원 연구진의 논문(동료 심사 전 사전 공개본)은 2026년 6월 말 기준 중국 이식 73명을 적었다. 여기에는 NEO 36명과 베이나오 1호 30명이 들어 있다. 같은 논문은 2026년 6월 30일까지 중국임상시험등록센터(ChiCTR)에 오른 BCI 임상을 134건으로 셌다. 비침습형이 106건(79.1퍼센트), 피질 안에 전극을 넣는 등록 시험 4건, 혈관 안 방식 1건이었다. 경막 바깥이나 아래에 전극을 두는 반침습 연구는 연구자 주도 임상으로만 진행됐다. (17장)

미국, 중국, 호주의 누계 100여 명. 본문이 말하는 100여 명은 2026년 6월 말 기준 장기 이식 참여자를 보수적으로 센 숫자다. 미국과 호주 32명에 중국 논문의 73명을 더해 약 105명이다. 수술 중에 잠깐 기록하거나 30일 안에 빼는 장치는 넣지 않았고, 1998년부터 2023년까지 대학 연구진이 심은 67명도 따로 두었다. (17장)

셈 기준에 따른 다른 숫자. 등록된 임상만 세면 뉴럴링크 21명, 싱크론 약 10명, 파라드로믹스 1명, NEO 36명으로 68명이고, NEO 상업 이식 8례를 더하면 76명이다. 2026년 10월까지 기업과 창업자가 공개한 장기 이식 수치를 모두 더하면 약 160명이다. 6월 이후 늘어난 50여 명은 중국 몫이다. (17장)

세계 추정. 2024년의 한 리뷰는 1998년부터 2023년 말까지 21개 연구진이 67명에게 BCI를 시험했다고 집계했다. 위트레흐트 대학병원의 마리스카 반스틴셀은 2026년 6월 세계 전체를 약 150명으로 추정했다. 18장은 장기 이식형 BCI 참여자를 공개 수치를 모두 더해도 200명이 되지 않는다고 적었다. (16장, 17장, 18장)

등록부 조회. 2026년 10월 14일 ClinicalTrials.gov에서 BCI를 검색하면 264건이다. 미국에 시험 장소가 있는 등록은 91건, 중국은 26건이다. 이식 관련 어휘가 들어간 등록은 미국 34건, 중국 10건이다. 한국에 시험 장소가 있는 이식형 BCI 등록은 0건이다. (17장)

한국. 국내에는 이식형 BCI 임상시험 인허가 사례가 없다. 한국인 환자가 해외에서 이식형 BCI를 받았다는 공개 사례도 찾지 못했다. 과학기술정보통신부는 2026년 3월 18일 K-문샷 BCI 임무를 2027년에 시작한다고 밝혔다. (16장, 17장, 20장)

3. 참고: BCI가 아닌 인접 신경 장치

아래 장치는 본문이 BCI와 견주려고 든 것이다. 본문은 이들을 이 책이 말하는 양방향 BCI로 보지 않는다.

장치	나라	방식	보급 또는 기준 시점	승인 상태	출처 장
메드트로닉 적응형 뇌심부자극	미국	이식형 뇌 자극(뇌 신호를 읽어 자극 세기를 조절)	본문 미기재	2025년 2월 24일 파킨슨병용 FDA 승인 발표	18장
뉴로페이스 RNS	미국	이식형 뇌 자극(발작 조짐을 읽어 자극)	본문 미기재	2013년 11월 FDA 승인	18장
인공와우	세계	청신경 자극(뇌 신호를 읽지 않으므로 BCI가 아니다)	2019년 12월까지 세계 등록 이식 약 73만 6,900대	본문 미기재	18장
메타 손목 근전도 밴드	미국	손목 근육 신호를 읽는 착용 기기(뇌 신호가 아니므로 BCI가 아니다)	본문 미기재	2025년 7월 Nature 발표	18장, 19장

후원 안내

이 책을 잘 읽으셨으면
그리고 새로운 가치있는 지식을 얻으셨다고 판단되시면

농협 302-1096-0948-81
(예금주 김경진)에 자발적 후원 부탁드립니다.

김경진 변호사

(김경진 변호사 저)