DAY 56–60 · 6부 앞으로

6부 총정리 — 앞으로

앞으로 — 용어와 원리까지 · DAY 56-60

factnotes

시험 없음 — 읽고 넘어가는 복습 자료

AGI 논쟁 — 자를 어디에 대느냐의 문제

1. AGI는 '사람만큼 두루 잘하는 AI'를 가리키는 말이야. 어디까지 해내야 그렇다고 할지는 합의가 없어서, 같은 AI를 놓고 판정이 갈려.

2. 1950년 튜링은 속을 들여다보지 말고 **겉으로 드러난 대답만 보고** 판정하자고 했어. 기준을 낮게 두는 쪽의 뿌리야.

3. 2019년에는 지능을 '아는 양'이 아니라 **새 기술을 익히는 속도**로 재자는 제안이 나왔어. 기준을 높게 두는 쪽의 뿌리고.

벤치마크 — 정해진 문제집으로 실력을 재는 시험

1. 벤치마크는 모델끼리 같은 조건에서 비교하려고 만든 문제집이야. AGI 논쟁에서 숫자는 대개 여기서 나와.

2. 2026년 2월, 사람 평균이 60%인 퍼즐 시험(ARC-AGI-2)에서 **84.6%**를 받았다는 자사 발표가 나왔어.

3. 설명서 없이 새 규칙을 알아내야 하는 ARC-AGI-3는 2026년 3월 공개 당시 **사람 100% · AI 0.51%**였어.

시험을 통과한 것과 일을 해내는 것

1. 문제집은 언젠가 다 풀려. 그래서 새 시험이 계속 만들어지고, 점수만으로는 논쟁이 끝나지 않아.

2. 실제 사람 일을 통째로 맡겨 본 2025년 평가에서 가장 잘한 AI가 끝까지 해낸 건 **2.5%**였어.

3. 그래서 'AGI'라는 말 자체를 버리고 그냥 '강력한 AI'라고 부르자는 쪽도 있어. 말이 흐릿하면 다툼만 길어지니까.

AI와 직업 — 조각 단위로 봐야 보인다

1. 직업 이름으로 보면 '사라진다, 아니다'밖에 안 나와. 조각으로 쪼개면 모으기 · 만들기 · 고르기 · 책임지기로 갈려.

2. 앞의 두 조각은 **정답 모양이 정해져 있고 반복되는 일**이라 AI가 먼저 가져갔어.

3. 뒤의 두 조각은 무엇이 옳은지 정하고 감당하는 일이라 그대로 남아.

어떻게 쟀나 — AI험 문제 대신 실제 산출물로

1. 2025년에는 9개 분야 **44개 직업의 실제 업무 산출물**로 AI를 채점한 평가가 나왔어. 문제집이 아니라 진짜 일로 잰 거야.

2. 같은 해, 실제 업무를 통째로 맡겼을 때 끝까지 해낸 비율이 **2.5%**라는 평가도 나왔어.

3. 조각은 잘하고 이어 붙이기는 약하다 — 두 숫자를 같이 놓으면 그 모양이 보여.

관찰과 원인은 다른 말이야

1. AI를 많이 쓰는 분야에서 22~25세의 일자리 수가 상대적으로 적더라는 2025년 관찰이 있어.

2. 저자들은 **'원인을 밝힌 게 아니라 관찰'**이라고 못 박았어. 숫자를 읽을 때 꼭 같이 읽어야 하는 문장이야.

3. 조각이 바뀌면 직함도 바뀌어. 2023년에는 '프롬프트 엔지니어'라는 이름이 새로 생기기도 했어.

AI 안전 — 모델 위에 얹는 장치 셋

1. 안전은 모델의 성격이 아니라 바깥에 덧붙이는 장치야. 거절하기 · 하기 전에 물어보기 · 나중에 들여다보기 셋으로 나눠 볼 수 있어.

2. 2022년에는 사람이 일일이 라벨을 붙이는 대신 **'원칙을 적은 글'을 학습 신호로** 쓰는 방법이 나왔어. 거절을 가르치는 뿌리가 된 연구야.

3. 모델은 원래 아는 걸 말해 주는 쪽으로 배워. 거절은 저절로 생기지 않고 따로 가르쳐야 하는 행동이야.

숫자로 보면 — 86%에서 4.4%, 계산은 23.7% 증가

1. 들어오는 질문과 나가는 답을 같이 걸러 주는 장치를 얹었더니, 꼼수 질문의 성공률이 **86%에서 4.4%로** 떨어졌어(2025년).

2. 3,000시간 넘게 꼼수를 던졌는데 통째로 뚫는 방법은 나오지 않았다고 해. 그래도 0은 아니야.

3. 값이 같이 붙어 와. 같은 연구에서 계산이 **23.7% 더** 들었어.

과잉 거절 — 멀쩡한 부탁까지 거절하는 일

1. 장치를 세게 조이면 위험한 질문만 막는 게 아니라 평범한 부탁도 막혀. 이걸 과잉 거절이라고 불러.

2. 그래서 요즘은 거절 실력과 과잉 거절을 한 쌍으로 같이 재. 한쪽만 좋아져도 쓸 만한 AI가 아니거든.

3. 나중에 들여다보는 쪽도 사람 손에서 도구로 옮겨 갔어. 2025년에는 **AI가 AI를 자동으로 감사하게** 만든 도구가 무료로 공개됐어.

로봇과 AI — 앞쪽 세 조각과 뒤쪽 한 조각

1. 로봇이 한 번 움직이려면 보기 · 알아듣기 · 정하기 · 움직이기를 거쳐. 앞 세 조각이 최근 몇 년에 AI로 바뀐 부분이야.

2. 2023년에는 인터넷 사진과 글로 배운 지식을 **로봇 팔 제어에 그대로 이어 붙인** 연구가 나왔어.

3. 2025년에는 말을 알아듣는 대형 모델을 로봇 몸에 연결한 모델이 공개됐어. 원래 뿌리가 다른 두 분야가 이때부터 본격적으로 붙었어.

손과 다리는 여전히 기계 기술이야

1. 손으로 잡고 다리로 걷는 조각은 수십 년 쌓인 제어 기술이 맡아. AI가 대신한 게 아니야.

2. 예전 로봇은 정해 준 좌표대로만 움직여서 물건이 1cm만 삐뚤어져도 허공을 집었어. 손과 다리는 정확했지만 **무엇을 할지 정하는 조각이 없었던** 거야.

3. 그래서 '로봇이 AI다'가 아니라 '로봇 안에 AI가 들어갔다'고 말하는 게 정확해.

빠른 회로와 느린 회로를 층으로 나눈다

1. 균형은 1초에 수백에서 수천 번 잡아야 하는데 생각은 그 속도로 못 해. 한 회로로 다 하려면 둘 중 하나가 늦어져.

2. 2026년 공개된 한 로봇은 전신 균형을 **1초에 1,000번**, 관절을 200번 고쳐 잡고, 의미를 따지는 부분은 훨씬 천천히 돌린다고 밝혔어.

3. 그래서 진짜 어려운 쪽은 생각이 아니라 몸이야. 속도가 다른 일을 같은 몸 안에서 겹쳐 돌려야 하거든.

배우는 순서 — 60일을 한 고리로 묶으면

1. 60일 내내 반복해서 나온 답은 하나였어 — 묻고, 확인하고, 직접 해 보고, 내 말로 설명하기.

2. 새 도구를 만났을 때, AI 답을 받았을 때, 혼자 새로 배울 때 **세 상황 다 순서가 같아**.

3. 한 바퀴 돌면 다시 첫 칸이야. 이 고리를 혼자 돌릴 수 있으면 가르쳐 주는 사람 없이도 계속 배울 수 있어.

이 순서가 답인 이유

1. AI가 가져가는 조각과 사람한테 남는 조각이 여기서 갈려. **초안은 AI가 주고, 맞는지 정하는 건 너**야.

2. 그래서 '확인하고' 칸을 빼면 순서가 무너져. AI는 아주 자신 있게 틀리기도 하니까.

3. 도구는 몇 달마다 바뀌지만 배우는 방법은 안 바뀌어. 이게 새 도구 이름을 외우지 않아도 되는 이유야.

혼자 따라가 보고 싶다면

1. 사람은 쉽게 푸는데 AI는 못 푸는 퍼즐 대회가 매년 결과를 공개해. **AI가 어디까지 왔는지 직접 따라갈 수** 있어.

2. 'AI가 배울 글이 언제 바닥날까'를 숫자로 따져 본 글도 있어. 답보다 따지는 방법을 보는 게 더 쓸모 있어.

3. 출처는 아래에 모아 뒀어. 한 줄 요약을 믿지 말고 직접 열어 보는 것도 순서의 한 칸이야.

앞으로를 묻는 질문은
대부분 우리 쪽을 묻고 있어.



다음 주부터는 매주 한 편, 그 주의 AI를 본다.