DAY 56–60 · 6부 앞으로

6부 총정리 — 앞으로

앞으로를 한 겹 더 · 설명과 이유까지 · DAY 56-60

factnotes

시험 없음 — 읽고 넘어가는 복습 자료

AGI 논쟁은 실력 다툼이 아니라 기준선 다툼이야

1. AGI는 '사람만큼 두루 잘하는 AI'를 가리키는 말인데, 어디까지 해내야 그렇다고 할지는 아무도 정해 놓지 않았어.

2. 기준을 낮게 — '사람처럼 말이 통하면 AGI' — 잡으면 **이미 통과**야. 1950년 튜링이 낸 기준이 그쪽에 가까웠어.

3. 기준을 높게 — '처음 보는 일을 스스로 배워 끝까지 해내면 AGI' — 잡으면 **아직**이야.

그래서 '언제 오냐'는 답도 같이 바뀐다

1. 시험으로 재면 답이 갈려. 2026년 2월, 사람 평균이 60%인 퍼즐 시험에서 84.6%를 받은 모델이 나왔어.

2. 설명서 없이 새 규칙을 알아내는 시험에서는 2026년 3월 공개 당시 **사람 100% · AI 0.51%**였어.

3. 같은 말을 서로 다른 뜻으로 쓰고 있으니 논쟁이 시끄러운 거야. 그래서 'AGI'라는 말을 그냥 버리자는 쪽도 있어.

AI가 가져가는 건 직업이 아니라 조각이야

1. 일 하나를 쪼개면 네 조각이 나와. 모으기 · 만들기 · 고르기 · 책임지기.

2. 앞의 두 조각은 AI가 빠르게 가져갔어. **정답 모양이 정해져 있고 여러 번 반복되는 쪽**이거든.

3. 뒤의 두 조각은 남아. 무엇이 맞는지 정하고 그걸 내보내도 된다고 결정하는 일이야.

통째로 넘어가지 않는 이유 — 끝까지 해내는 게 약해

1. AI는 한 조각은 엄청 잘하는데 처음부터 끝까지 혼자 이어서 해내는 건 아직 약해.

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

3. 그래서 무게중심이 뒤로 밀려. 확인하고 책임지는 쪽으로 옮겨 가고, 없던 직함이 새로 생기기도 해.

안전은 모델의 성격이 아니라 위에 얹는 장치야

1. AI 안전은 모델을 착하게 만드는 마법이 아니야. 모델 위에 하나씩 덧붙이는 장치야.

2. 장치는 크게 셋이야 — **거절하기 · 하기 전에 물어보기 · 나중에 들여다보기**.

3. 거절은 저절로 생기지 않아. 모델은 원래 아는 걸 말해 주는 쪽으로 배우니까, 따로 가르쳐야 하는 행동이야.

값을 치르고 사는 거야

1. 2025년 한 연구에서는 규칙을 어기게 만드는 꼼수 질문의 성공률이 **86%에서 4.4%로** 떨어졌어.

2. 대신 계산이 약 24% 더 들었어. 그리고 멀쩡한 부탁까지 덜컥 거절해 버리는 '과잉 거절'이 생겨.

3. 그래서 요즘은 거절 실력과 과잉 거절을 같이 재. '이 AI 안전해요?'보다 좋은 질문은 '무슨 장치가 붙어 있어요?'야.

로봇의 앞쪽 세 조각에 AI가 들어갔어

1. 예전 로봇은 정해 준 좌표대로만 움직였어. 물건이 1cm만 삐뚤어져 있어도 허공을 집었지.

2. 2023년부터 보는 AI와 말을 알아듣는 AI를 로봇 몸에 그대로 연결했어. 그러니까 **'빨간 컵 집어'가 통하기 시작**했어.

3. 눈 · 말귀 · 판단 세 조각이 AI로 바뀐 거고, 손과 다리는 여전히 기계 기술이야.

어려운 쪽은 생각이 아니라 몸이야

1. 균형은 1초에 수백에서 수천 번 잡아야 하는데, 생각은 그 속도로 못 해.

2. 그래서 요즘 로봇은 **빠른 몸 회로와 느린 생각 회로를 층으로 나눠** 돌려.

3. 2026년 공개된 한 로봇은 전신 균형을 1초에 1,000번, 관절을 200번 고쳐 잡는다고 밝혔어.

준비물은 새 지식이 아니라 네 칸의 순서야

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

2. 새 도구를 만났을 때도, AI 답을 받았을 때도, 혼자 배우고 싶을 때도 **순서가 똑같아**.

3. 막히는 건 실패가 아니라 순서의 한 칸이야. 막힌 데를 다시 묻는 게 다음 칸이거든.

이 순서가 답인 이유

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

2. 도구는 몇 달마다 바뀌지만 배우는 방법은 안 바뀌어. 그래서 오늘 익힌 게 내년에도 쓸모 있어.

3. 같은 AI를 써도 이 네 칸을 도느냐 마느냐로 갈려. 더 잘난 사람이라서가 아니라 한 바퀴 더 돌았을 뿐이야.

앞으로를 묻는 질문은
대부분 우리 쪽을 묻고 있어.



다음 주부터는 매주 한 편, 그 주의 AI를 본다.