DAY 39 · 4부 조심할 것

이 답에 누가 빠져 있진 않을까 — AI의 편견

하루 5분 AI 수업 · 2026년 10월 15일

AI는 미워하지 않아. 다만 많이 본 쪽으로 기울 뿐이야. 사람을 고르는 질문이 나오면 '다른 경우는?' 하고 한 번 더 물어.

한 줄로 말하면

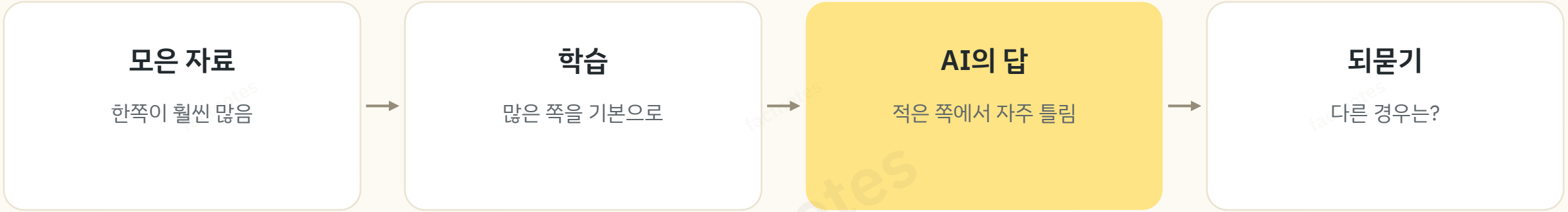
1. AI는 마음먹고 누구를 미워하는 게 아니야. **배운 자료가 한쪽으로 치우쳐 있으면 답도 그쪽으로 치우치는** 것뿐이야.

2. 2018년에 얼굴 사진을 맞히는 프로그램들을 검사한 연구가 있었어. **피부색이 어두운 여성 얼굴에서 틀리는 비율이 가장 높았고, 피부색이 밝은 남성 얼굴에서 가장 낮았어**.

3. 그러니까 겁낼 게 아니라 습관을 하나 만들면 돼. **사람을 나누거나 고르는 질문에는 '다른 경우는?' 하고 한 번 더 묻기**.

그림 한 장으로

치우침이 흘러가는 길



맨 앞 상자가 기울어져 있으면 뒤 상자는 저절로 같이 기울어. 마지막 상자는 네가 넣을 수 있는 칸이야.

핵심: 치우친 자료에서 시작하면 치우친 답이 나온다 — 마지막에 되묻는 건 네 몫

이 답에 편견이 섞였을까

편견이 섞였을 수 있어

이쪽이 맞는 상황

- ✓ '의사가 진료하는 그림 그려 줘'라고 열...
- ✓ 우리 반 이름 목록을 주고 '반장으로 누가...
- ✓ '밤에 조심해야 할 사람이 누구야?'라고...

그럴 일은 별로 없어

이쪽이 맞는 상황

- ✓ '3 곱하기 7은?' 하고 물었더니 21이...
- ✓ 영어 단어 하나의 뜻을 물었더니, 사전에...
- ✓ '집 와이파이 비밀번호 바꾸는 법'을 물었...

규칙

고르는 기준

- ✓ 사람을 나누거나 고르는 질문일수록 치우치기...

상황을 하나씩 읽고 둘 중 하나를 골라 봐. 어떤 질문이 잘 치우치고 어떤 질문이 덜 치우치는지 감이 잡힐 거야.

웹 수업에서 직접 해 볼 수 있어요 · next-token.net/learn/day-39/

확인해보기

세 문제만 풀어 보자

Q1

AI에 편견이 생기는 가장 큰 이유는?

① AI가 사람을 싫어해서
② 인터넷이 느려서
③ 계산을 잘 못해서
④ 배운 자료가 한쪽으로 치우쳐서

Q2

2018년 얼굴 사진 검사 연구에서 드러난 건?

① 피부색이 어두운 여성 얼굴에서 가장 많이 틀렸다
② 모든 얼굴에서 똑같이 틀렸다
③ 사진이 흐릴 때만 틀렸다
④ 사람보다 늘 정확했다

Q3

다음 중 편견이 섞이기 쉬운 질문은?

① 3 곱하기 7은?
② 누가 반장에 어울려?
③ 물의 끓는점은?
④ 와이파이 바꾸는 법은?

왜 그런지까지 알고 가자

Q1

④

AI에는 미워하는 마음이 없어. 많이 본 쪽을 기본값처럼 답할 뿐이고, 그 '많이 본 쪽'이 이미 기울어져 있었던 거야.

Q2

①

같은 프로그램인데 누구 얼굴이냐에 따라 틀리는 비율이 달랐어. 학습에 쓴 사진이 한쪽으로 쏠려 있었기 때문이야.

Q3

②

사람을 고르거나 나누는 질문이 제일 위험해. 답이 하나로 정해진 질문은 치우칠 자리가 거의 없어.

AI는 미워하지 않아. 다만 많이 본 쪽으로 기울 뿐이야.
사람을 고르는 질문이 나오면 '다른 경우는?' 하고 한 번 더 물어.



내일은 AI가 만들었다고 밝히기 — 과제에서 어디까지 AI가 했는지 표시하는 법.