

LESSON
25

고급 초6·중2 AI 교실

AI의 편견: 데이터가 치우치면 생기는 일

AI는 배운 데이터를 닮아요. 치우친 데이터는 불공정한 판단을 낳아요.

 **학습 목표** 이번 시간이 끝나면...

- ✓ AI 편향이 생기는 까닭(치우친 데이터, 과거의 차별)을 설명할 수 있어요.
- ✓ 편향, 공정성, 대표성의 뜻을 말할 수 있어요.
- ✓ AI 편향을 줄이는 방법을 제안할 수 있어요.

● 고급 25

AI의 편견: 데이터가 치우치면 생기는 일

AI는 배운 데이터를 닮아요.

데이터가 치우치면 판단도 불공평해져요.

“날아다니는 새만 봤더니
펭귄은 새가 아니래!”

치우친 데이터는 치우친 판단을 낳아요.

✚ 어떤 데이터로 배웠나요?



치우친 데이터

- ✗ 한 무리 자료만 많아요
- ✗ 적은 쪽에 자주 틀려요

VS



골고루 데이터

- ✓ 여러 무리를 고르게
- ✓ 사람이 함께 점검해요

📖 꼭 알아야 할 단어

편향 = 한쪽으로 치우친 판단이나 결과**공정성** = 누구에게나 차별 없이 대하는 것Q. 데이터가 한쪽으로 치우치면
AI의 판단도 치우칠 수 있다?

O / X

정답: O

그림 1 한눈에 보는 25강 — 치우친 데이터와 골고루 데이터의 결과를 비교해 보세요.



쉽게 풀어 보기

데이터의 치우침이 AI의 판단에 어떻게 이어지는지 알아봐요.



AI는 데이터를 닮아요

AI는 데이터에서 규칙을 찾아요. 그래서 데이터가 한쪽으로 치우치면 AI의 판단도 치우쳐요. 한쪽으로 치우친 판단이나 결과를 **편향(bias)**이라고 해요.

데이터가 실제 세상을 고르게 담고 있는 정도를 **대표성(representativeness)**이라고 해요. 대표성이 낮으면, 자료가 적은 무리에게 AI가 자주 틀리거나 불리하게 판단해요.

누구에게나 고르게, 차별 없이 대하는 것을 **공정성(fairness)**이라고 해요. AI를 만들고 쓰는 사람은 공정성을 꾸준히 점검해야 해요.



날아다니는 새 사진만 본 AI로 이해해요

‘새’를 가르치려고 하늘을 나는 참새, 비둘기, 까치 사진만 잔뜩 보여 줬다고 해 봐요. 이 AI에게 펭귄이나 타조 사진을 보여 주면 **“새가 아니에요”**라고 할 수 있어요.

AI가 나빠서가 아니라, **‘새는 하늘을 난다’**는 **치우친 데이터**로 배웠기 때문이에요.

사람을 판단하는 일에서 이런 치우침이 생기면 **훨씬 심각한 문제**가 돼요.



한 줄 요약

AI는 배운 데이터를 닮아요. 데이터가 한쪽으로 치우치거나 과거의 차별이 담겨 있으면, AI도 어떤 사람이나 대상에게 불공평한 판단을 할 수 있어요. 이것을 AI 편향(bias)이라고 해요.

1. 생활 속 편향, 사람에게 생기면

입학, 채용, 대출처럼 사람의 기회를 정하는 일에 AI를 쓴다고 생각해 봐요. 과거 기록에 차별이 담겨 있었다면, AI는 그 차별을 **규칙처럼 배워 되풀이**할 수 있어요.

얼굴 인식, 음성 인식도 마찬가지예요. 어떤 무리의 데이터가 적으면 그 무리에서 더 자주 틀릴 수 있어요.

그래서 데이터를 고르게 모으고, **무리별로 결과를 따로 확인**하고, 중요한 결정은 사람이 최종 검토해요.

완벽하게 공정한 AI는 만들기 어려워서 꾸준히 점검하고 고쳐야 해요.

미리 보기 뒤에 나오는 '직접 해보기 미션'에서 우리 반 '과일' 데이터의 치우침을 직접 확인해 봐요.

미리 생각하기 우리 학교 학생 사진으로만 '학생'을 배운 AI는 어떤 실수를 할 수 있을까요?

다른 나라, 다른 학년, 교복이 없는 학생을 떠올려 봐요.

123 숫자로 보는 25강

사실

2018년

해외 기업의 이력서 평가 AI가 여성 지원자에게 불리했다는 보도가 나왔어요.

예시

3가지

이 강의 핵심 단어: 편향, 공정성, 대표성

예시

25명 중 0명

미션에서 '용과'를 떠올린 친구 수를 예로 든 숫자예요.

사실 = 실제로 있었던 일 · **예시** = 이해를 도우려고 만든 숫자



표로 정리해요

치우친 데이터와 골고루 데이터를 비교해요.



그림으로 보는 비교

데이터	어떤 결과가 생길까요?
치우친 데이터	어떤 무리의 자료만 많아요. → 자료가 적은 무리에게 자주 틀리거나 불리하게 판단해요.
골고루 데이터	여러 무리의 자료를 고르게 모아요. → 더 많은 사람에게 공정하게 작동할 가능성이 커져요. 그래도 사람이 결과를 함께 점검해요.



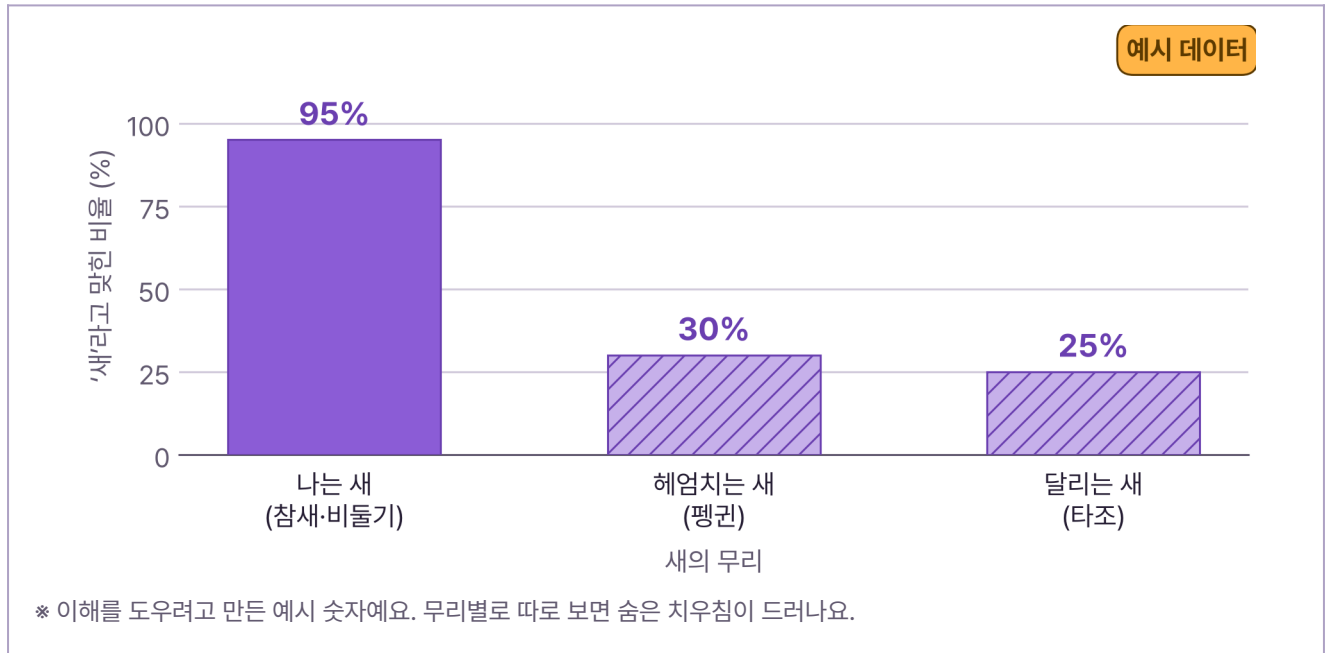
편향이 생기는 곳과 줄이는 방법

생기는 곳	이런 일이 생겨요	이렇게 줄여요
데이터 모으기	어떤 무리의 자료가 적어요.	여러 무리의 데이터를 고르게 모아요.
과거 기록	예전의 차별이 기록에 남아 있어요.	기록을 살펴 차별이 담긴 기준을 빼요.
결과 확인	평균만 보면 문제가 숨어요.	무리별로 결과를 따로 확인해요.
결정	AI 판단을 그대로 써요.	중요한 결정은 사람이 최종 검토해요.

생각해 보기 전체 정확도가 95%인 AI도 불공평할 수 있어요. 어떤 경우일지 적어 보세요.

많은 무리에서는 잘 맞고, 적은 무리에서는 자주 틀리는 경우를 떠올려 봐요.

📊 날아다니는 새 사진만 본 AI의 무리별 정답률 (예시)



그래프 읽기 어느 무리에서 정답률이 낮나요? 이 AI의 정답률을 고르게 하려면 무엇이 필요할까요?

이 그래프는 예시 데이터예요. 데이터의 대표성을 떠올려 봐요.

가 꼭 알아야 할 단어

편향

bias

한쪽으로 치우친 판단이나 결과

공정성

fairness

누구에게나 고르게, 차별 없이 대하는 것

대표성

representativeness

데이터가 실제 세상을 고르게 담고 있는 정도

**직접 해보기 미션**

손과 눈으로 직접 확인해 봐요.

우리 반 '과일' 데이터 살펴보기

반 친구들에게 '과일' 하면 떠오르는 것을 하나씩 적어 달라고 한 뒤 표로 정리해요. 아마 사과, 바나나, 딸기가 많고 용과나 두리안은 거의 없을 거예요.

1 친구들의 답을 모아 많이 나온 순서대로 표에 적어요. 이름은 적지 않고 답만 모아요.

2 이 표로만 배운 AI라면 용과 사진을 보고 과일이라고 알아볼지 생각해요.

3 데이터를 골고루 만들려면 무엇을 더 모아야 할지 목록을 만들어요.

※ 친구의 답을 놀리지 않아요. 적게 나온 답일수록 데이터를 골고루 만드는 데 소중한 거예요.

기록표

	과일 이름	사람 수
1위		
2위		
3위		
4위		
5위		

- 데이터를 골고루 만들려면 무엇을 더 모아야 할까요? 목록을 적어 보세요.

우리 반 데이터에서 빠진 과일

사람을 판단하는 AI라면 생길 문제

🔑 정리 AI는 배운 데이터를 닮아요. 골고루 모으고, 무리별로 점검해야 공정해져요.

**확인 퀴즈**

스스로 풀고, 정답은 마지막 쪽에서 확인해요.

1. **O/X** 데이터가 한쪽으로 치우치면 AI의 판단도 치우칠 수 있다.☐

O

☐

X

2. **고르기** AI 편향을 줄이는 방법으로 알맞지 않은 것은?☐

① 여러 무리의 데이터를 고르게 모으기

☐

② 사람이 결과를 점검하기

☐

③ 이미 많은 무리의 데이터만 더 모으기

3. **빈칸** 데이터가 실제 세상을 고르게 담고 있는 정도를 _____ (이)라고 해요.4. **쓰기** 무리별로 결과를 따로 확인해야 하는 까닭을 써 보세요.

**중2 한 걸음 더****실제로 드러난 AI 편향**

실제로 한 해외 기업이 이력서를 평가하는 AI를 시험 삼아 만들었다가, 과거 지원자 자료에 남성이 훨씬 많았던 탓에 여성 지원자에게 불리하게 점수를 매기는 문제가 드러나 쓰지 않기로 했다는 보도가 있었어요. 얼굴 인식 기술의 정확도가 피부색이나 성별에 따라 다르다는 연구도 있었어요.

그래서 개발자는 데이터를 다양하게 모으고, 무리별로 결과를 따로 확인하고, 중요한 결정은 사람이 최종 검토해야 해요. **'완벽하게 공정한 AI'는 만들기 어려워서 꾸준히 점검하고 고쳐야** 해요.

- 생각해 봐요: 학교에서 AI로 동아리 회원을 뽑는다면, 공정하게 하려고 어떤 점을 점검해야 할까요?

**용어 정리**

이번 시간에 나온 어려운 말을 모았어요.

용어	영어	쉽게 말하면	예시
편향	bias	한쪽으로 치우친 판단이나 결과	나는 새만 새라고 함
공정성	fairness	누구에게나 고르게, 차별 없이 대하기	모든 무리에서 비슷한 정답률
대표성	representativeness	데이터가 세상을 고르게 담은 정도	여러 종류의 새 사진
차별	discrimination	정당한 까닭 없이 다르게 대하기	성별로 점수 깎기
무리	group	공통점으로 묶인 집단	헤엄치는 새
정답률	accuracy	맞힌 비율	95%, 30%
최종 검토	final review	마지막에 사람이 살펴보고 결정하기	채용 결정
이력서	résumé	학력과 경력을 적은 문서	지원자 서류
과거 기록	historical data	예전에 쌓인 자료	예전 지원자 자료
점검	audit	문제가 없는지 꼼꼼히 살피기	무리별 결과 확인

**정답과 풀이**

퀴즈를 다 푼 다음에 확인해요

1. ① — AI는 배운 데이터의 모습을 그대로 닮기 쉬워요.
2. ③ — 치우침이 더 커질 수 있어요.
3. 대표성(representativeness) — 대표성이 낮으면 편향이 생기기 쉬워요.
4. (예시 답) 전체 평균이 높아도 자료가 적은 무리에서는 자주 틀릴 수 있어서, 따로 봐야 숨은 불공정이 드러나기 때문이에요.