

LESSON

초급 초6·중2 AI 교실

04

데이터는 AI의 밥이다

AI는 데이터를 먹고 배워요. 넉넉하고, 골고루, 정확해야 해요.



학습 목표 이번 시간이 끝나면...

- ✓ 데이터의 4가지 모습(글, 사진, 소리, 숫자)을 말할 수 있어요.
- ✓ 좋은 데이터의 조건(넉넉하게, 골고루, 정확하게)을 설명할 수 있어요.
- ✓ 치우친 데이터로 배우면 AI가 실수하기 쉬운 까닭을 알 수 있어요.

초급 04

데이터는 AI의 밥이다

AI는 데이터를 먹고 배워요.

넉넉하고 골고루, 정확한 데이터가 좋아요.



“좋아하는 반찬만
먹으면 안 돼!”

AI도 데이터를 골고루 먹어야 잘 배워요.

Q 데이터의 4가지 모습

1



글
책·메시지

2



사진
그림·영상

3



소리
말·음악

4



숫자
점수·기온

꼭 알아야 할 단어

학습 데이터 = AI가 공부할 때 보는 예시 자료

편향 = 한쪽으로 치우쳐 있는 것

Q. 데이터는 많기만 하면
틀린 내용이 섞여도 괜찮다?

O

/

X

정답: X

그림 1 한눈에 보는 04강 — 좋은 데이터의 세 가지 조건을 먼저 찾아보세요.



쉽게 풀어 보기

AI에게 데이터는 밥과 같아요. 어떤 밥이 좋은지 알아봐요.



좋은 데이터는 어떤 데이터일까요?

AI가 공부할 때 보는 예시 자료를 **학습 데이터(training data)**라고 해요. AI는 이 데이터에서 규칙을 찾아요.

좋은 데이터에는 세 가지 조건이 있어요. **넉넉하게**(양이 충분하게), **골고루**(여러 경우를 고르게), **정확하게**(틀린 내용 없이)예요.

데이터가 얼마나 정확하고 쓸모 있는지를 **데이터 품질(data quality)**이라고 해요. 양만 많다고 좋은 데이터는 아니예요.



골고루 먹는 급식으로 이해해요

급식에서 좋아하는 반찬만 먹으면 몸에 필요한 영양소가 모자라요. AI도 비슷해요. 한쪽으로 치우친 데이터만 먹으면 **AI도 한쪽만 아는 편식쟁이**가 돼요.

예를 들어 하얀 강아지 사진으로만 배운 AI는 까만 강아지를 보고 헛갈릴 수 있어요.

또 상한 음식을 먹으면 배탈이 나듯이, **틀린 데이터로 배우면 AI도 틀린 답**을 내놓아요.

배운 사진 1	배운 사진 2	배운 사진 3	새 사진
하얀 강아지	하얀 강아지	하얀 강아지	까만 강아지?

예시 그림 하얀 강아지만 보고 배운 AI는 까만 강아지를 헛갈릴 수 있어요.



한 줄 요약

AI는 데이터를 먹고 배워요. 데이터가 넉넉하고, 골고루 있고, 정확해야 AI가 잘 배울 수 있어요.

데이터는 어디에나 있어요

데이터는 컴퓨터 속에만 있는 게 아니에요. 내가 쓴 **일기와 메시지(글)**, 찍은 **사진**, 녹음한 **목소리(소리)**, 매일 재는 **걸음 수(숫자)**도 모두 데이터예요.

날씨 앱은 기온과 비 온 기록(숫자)으로, 번역기는 아주 많은 문장(글)로, 음성 비서는 많은 말소리(소리)로 배워요.

데이터를 모을 때는 지켜야 할 약속도 있어요. **다른 사람의 이름, 얼굴, 사진**은 허락 없이 모으거나 쓰면 안 돼요.

미리 보기 뒤에 나오는 '직접 해보기 미션'에서 적은 사람과 많은 사람에게 물었을 때의 차이를 직접 느껴 봐요.

미리 생각하기 내가 오늘 만든 데이터를 하나 적어 보세요. 글, 사진, 소리, 숫자 가운데 무엇인가요?

예: 아침에 찍은 사진, 친구에게 보낸 메시지

숫자로 보는 04강

예시

4가지

이 강에서 살펴본 데이터의 모습:
글, 사진, 소리, 숫자

예시

3명

옆자리 친구 3명에게만 물으면
결과가 한쪽으로 치우치기
쉬워요.

예시

3가지

좋은 데이터의 조건: 넉넉하게,
골고루, 정확하게

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



표로 정리해요

데이터의 모습과 좋은 데이터의 조건을 정리해요.



데이터의 4가지 모습

모습	어떤 자료일까요?	AI가 배우는 예
① 글	책, 뉴스 기사, 메시지 같은 글자 자료	번역기, 맞춤법 검사기
② 사진	그림, 사진, 영상 같은 눈으로 보는 자료	사진 검색, 얼굴 잠금
③ 소리	말소리, 음악 같은 귀로 듣는 자료	AI 비서, 노래 찾기 앱
④ 숫자	시험 점수, 기온, 걸음 수 같은 숫자 자료	날씨 예보, 걸음 수 앱



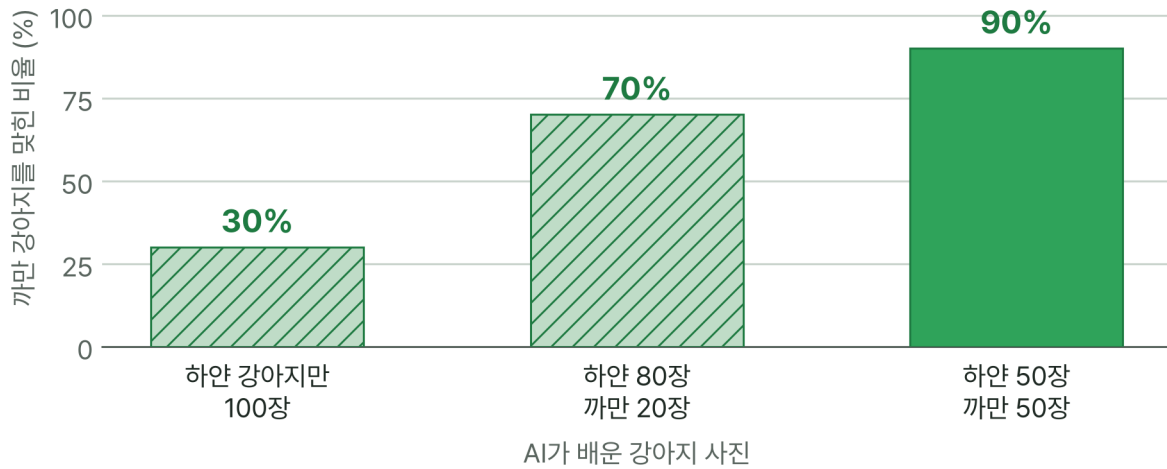
좋은 데이터와 나쁜 데이터

조건	좋은 데이터	나쁜 데이터
양	넉넉해요. 여러 번 볼 수 있어요.	너무 적어요. 규칙을 찾기 어려워요.
고르기	여러 경우를 골고루 모았어요.	한쪽 경우만 잔뜩 모았어요.
정확함	틀린 내용이 없어요.	틀린 이름표, 잘못된 기록이 섞였어요.
지킬 것	허락받고 모았어요.	남의 사진, 개인 정보를 몰래 모았어요.

생각해 보기 고양이 사진 100장 가운데 20장에 '강아지'라는 이름표가 잘못 붙어 있다면, AI는 어떻게 될까요?

틀린 데이터로 배우면 무엇을 배우게 될지 생각해 봐요.

배운 사진에 따라 달라지는 까만 강아지 알아보기 (예시)



※ 이해를 도우려고 만든 예시 숫자예요. 실제 결과는 사진의 질과 AI의 종류에 따라 달라져요.

그래프 읽기 어떤 사진으로 배웠을 때 까만 강아지를 가장 잘 맞췄나요? 그 까닭은 무엇일까요?

이 그래프는 예시 데이터예요. '골고루'라는 말을 떠올려 봐요.

가 꼭 알아야 할 단어

학습 데이터

training data

AI가 공부할 때 보는 예시 자료

편향

bias

한쪽으로 치우쳐 있는 것

데이터 품질

data quality

데이터가 얼마나 정확하고 쓸모 있는지를 나타내는 말

**직접 해보기 미션**

손과 눈으로 직접 확인해 봐요.

우리 반 인기 급식 메뉴 조사하기

‘우리 반 인기 급식 메뉴’를 두 가지 방법으로 조사해 보세요. 두 결과가 같나요, 다른가요? 적은 사람에게만 물으면 결과가 한쪽으로 치우칠 수 있다는 걸 느껴 보세요.

1 내 옆자리 친구 3명에게만 물어보기

2 반 친구 모두에게 손을 들어 답하게 하기

3 두 결과를 아래 표에 적고 비교해요.

※ 이름은 적지 말고 메뉴와 숫자만 기록해요.

기록표

	1등 메뉴	2등 메뉴	3등 메뉴
친구 3명			
반 전체			

- 두 결과가 같았나요, 달랐나요? 왜 그렇게 되었을까요?

- AI가 친구 3명의 답만 보고 배운다면 어떤 일이 생길까요?

적은 사람에게 물었을 때 좋은 점

적은 사람에게 물었을 때 걱정되는 점

🔑 **정리** 데이터가 적거나 한쪽으로 치우치면, 그 데이터로 배운 AI도 한쪽으로 치우쳐요.

**확인 퀴즈**

스스로 풀고, 정답은 마지막 쪽에서 확인해요.

1. **O/X** 데이터는 많기만 하면 틀린 내용이 섞여도 괜찮다.☐

O

☐

X

2. **고르기** AI에게 좋은 데이터가 아닌 것은?☐

① 양이 넉넉한 데이터

☐

② 여러 경우를 골고루 모은 데이터

☐

③ 한쪽 경우만 잔뜩 모은 데이터

3. **빈칸** 한쪽으로 치우쳐 있는 것을 _____(이)라고 해요.4. **쓰기** 데이터를 모을 때 지켜야 할 약속을 한 가지 써 보세요.

**중2 한 걸음 더**

쓰레기가 들어가면 쓰레기가 나온다

컴퓨터 과학에는 “쓰레기가 들어가면 쓰레기가 나온다(Garbage In, Garbage Out)”라는 말이 있어요. 넣은 데이터가 나쁘면 결과도 나쁘다는 뜻이에요.

실제로 몇몇 얼굴 인식 프로그램이 피부색이 어두운 사람의 얼굴을 더 자주 잘못 알아본다는 연구 결과가 나온 적이 있어요. 배운 사진이 한쪽으로 치우쳐 있었던 것이 원인 중 하나로 꼽혀요.

그래서 AI를 만드는 사람은 데이터가 골고루인지 꼼꼼히 살피고, 데이터를 모을 때 **개인 정보와 저작권**도 지켜야 해요.

- **생각해 봐요:** 우리 학교 학생들을 돕는 AI를 만든다면, 데이터를 골고루 모으려고 어떤 점을 챙겨야 할까요?



용어 정리

이번 시간에 나온 어려운 말을 모았어요.

용어	영어	쉽게 말하면	예시
데이터	data	AI가 배우는 재료가 되는 기록	사진, 글, 소리, 숫자
학습 데이터	training data	AI가 공부할 때 보는 예시 자료	강아지 사진 1,000장
편향	bias	한쪽으로 치우쳐 있는 것	하얀 강아지 사진만 모으기
데이터 품질	data quality	데이터가 얼마나 정확하고 쓸모 있는지	틀린 이름표가 없는 사진
이름표(레이블)	label	데이터에 붙인 정답 이름	사진에 붙인 '강아지'
조사	survey	사람들에게 물어 자료를 모으기	인기 급식 메뉴 손들기
표본	sample	전체 가운데 골라 살펴본 일부	친구 3명
개인 정보	personal information	누구인지 알 수 있는 정보	이름, 얼굴 사진, 주소
저작권	copyright	만든 사람이 자기 작품에 대해 갖는 권리	그림, 글, 노래
쓰레기 입력, 쓰레기 출력	GIGO (Garbage In, Garbage Out)	나쁜 데이터를 넣으면 나쁜 결과가 나온다는 말	틀린 기록 → 틀린 예측



정답과 풀이

퀴즈를 다 푼 다음에 확인해요

1. X — 틀린 데이터로 배우면 AI도 틀린 것을 배워요.
2. ㉠ — 한쪽으로 치우친 데이터로 배우면 다른 경우에서 실수하기 쉬워요.
3. 편향 — 하얀 강아지 사진만 모은 데이터도 편향된 데이터예요.
4. (예시 답) 다른 사람의 이름이나 사진 같은 개인 정보는 허락 없이 모으지 않아요.