DAY 09 · 1부 AI가 뭔지

큰 모델이 무조건 좋을까? — 큰 모델과 작은 모델

하루 5분 AI 수업 · 2026년 9월 15일

좋은 크기는 하나가 아니라, 쓸 자리마다 다르다

한 줄로 말하면

1. 모델이 **크다**는 건 5일차에 본 파라미터가 많다는 뜻이야 — 아는 게 많고, 처음 보는 어려운 문제를 잘 버텨.

2. 대신 큰 모델은 **느리고 비싸고**, 네 폰 안에는 아예 안 들어가. 인터넷 너머 서버에서만 돌아.

3. 그래서 큰 모델이 배운 걸 **작은 모델에 옮겨 담아** 쓰기도 해. 좋은 크기는 쓸 자리가 정한다.

그림 한 장으로

같은 일을 시켜도, 사는 곳이 다르다

작은 모델

내 기기 안에서 돈다

빠르다 · 싸다 · 인터넷 없어도 된다

아는 범위는 좁다

오타 고치기 · 사진 속 글자 읽기 · 게임 속 즉답

배운 걸 옮겨 담기

(증류)

큰 모델

서버(데이터센터)에서 돈다

아는 게 많다 · 어려운 문제를 버틴다

느리다 · 비싸다 · 인터넷이 필요하다

긴 글 따져 읽기 · 처음 보는 문제 단계별 풀이

둘은 우열이 아니라 역할이 다르다 — 쓸 자리가 크기를 정한다

직접 해보기

이 일, 어느 쪽에 맡길까?

작은 모델

이런 일에 낫다

- ✓ **비행기 모드에서 사진 속 글자 읽기**
- ✓ **카톡 보내기 전 오타 고치기**
- ✓ **게임 속 NPC가 1초 안에 대답**
- ✕ 처음 보는 경시 문제 풀이

빠르고 싸고, 인터넷 없이도 돈다

큰 모델

이런 일에 낫다

- ✓ **처음 보는 문제를 단계별로 풀기**
- ✓ **30쪽 자료 읽고 따져 묻기**
- ✓ **두 나라 제도를 비교해 정리**
- ✕ 오타 하나 고치기

아는 게 많지만 느리고 비싸다

증류

큰 모델 → 작은 모델

- ✓ **큰 모델이 먼저 답을 잔뜩 만든다**
- ✓ **작은 모델이 그 답을 보고 배운다**
- ✓ **작은 모델이 제 크기보다 잘한다**
- ✕ 그래도 큰 모델을 그대로 대신하진 못한다

2015년 힌턴 팀이 정식화한 방법

큰 모델이 좋은 게 아니라, 자리에 맞는 크기가 좋은 거야.

근거: Hinton·Vinyals·Dean, Distilling the Knowledge in a Neural Network (2015) · arxiv.org/abs/1503.02531

확인해보기

세 문제만 풀어 보자

Q1 모델이 “크다”는 건 주로 무엇이 많다는 뜻일까?

① 화면에 뜨는 글자 수가 많다
② 파라미터(학습으로 맞춰 둔 숫자)가 많다
③ 쓰는 사람이 많다
④ 답하는 데 걸리는 시간이 많다

Q2 작은 모델이 큰 모델보다 확실히 나은 점은?

① 처음 보는 어려운 문제를 더 잘 푼다
② 아는 사실이 더 많다
③ 빠르고 싸고, 인터넷 없이 내 기기에서 돈다
④ 절대 틀린 말을 하지 않는다

Q3 **(O / X)** 큰 모델이 배운 것을 작은 모델에 옮겨 담는 방법이 있다.

왜 그런지까지 알고 가자

Q1

②

크기는 파라미터 수를 말해. 5일차에 감 잡았던 그 숫자야 — 많을수록 담아 둘 수 있는 게 많아진다.

Q2

③

작은 모델은 기기 안에 통째로 들어가서 즉시 답하고, 전기도 돈도 덜 써. 어려운 문제 실력으로 이기는 게 아니라 자리로 이기는 거야.

Q3

0

증류(distillation)라고 불러. 큰 모델이 만든 답을 교재 삼아 작은 모델을 가르치면, 작은 모델이 제 크기보다 훨씬 잘하게 돼.

큰 모델이 좋은 게 아니라.
자리에 맞는 크기가 좋은 거야.



내일은 열린 모델과 닫힌 모델 — 누구나 받아 쓸 수 있는 AI와 그렇지 않은 AI 이야기를 해 볼 거야.