DAY 46–55 · 5부 안을 들여다보기

5부 총정리 — 안을 들여다보기

점 하나에서 내 폰까지 · 설명 한 겹 더 · DAY 46-55

시험 없음 — 읽고 넘어가는 복습 자료

점 하나는 곱하고 더하는 것밖에 못 해

1. 신경망은 숫자를 주고받는 점이 층층이 쌓인 구조야.

2. 점 하나는 들어온 숫자에 **무게를 곱해 전부 더하고**, 그 합이 문턱을 넘는지만 봐.

3. 넘으면 다음 층으로 보내고, 못 넘으면 거기서 멈춰.

왜 층을 쌓으면 달라지나

1. 첫 층은 선이나 점 같은 작은 조각만 잡아.

2. 그 조각을 모아 다음 층이 눈과 귀를 만들고, 마지막 층이 고양이라고 답해 — **층이 깊다**는 말이 이 뜻이야.

3. 문턱이 없으면 층을 아무리 쌓아도 한 번 곱해 더한 것과 같아져.

시선은 나눠 갖는 거야

1. 어텐션은 단어가 다른 단어를 얼마나 세게 쳐다볼지 정하는 계산이야.

2. 점수를 그대로 쓰지 않고 **합이 1인 몫**으로 바꿔 나눠 가져.

3. 그래서 한쪽을 세게 보면 나머지는 자동으로 약해져.

시선이 뜻을 정한다

1. '그것'이 누구를 가리키는지는 어울리는 단어에 시선이 몰리면서 정해져.

2. 문장을 앞에서 뒤로 한 줄로 읽는 대신 **한꺼번에 서로를 쳐다보게** 만든 게 큰 변화였어.

3. 2017년에 나온 이 방법이 지금 쓰는 거의 모든 모델의 바탕이 됐어.

뜻을 자리로 바꿔 놓는다

1. 컴퓨터는 글자를 모르니까 단어 하나를 숫자 여러 개가 묶인 값으로 바꿔 둬.

2. 숫자는 사람이 정해 주는 게 아니라 **엄청난 양의 글을 읽으며 스스로** 맞춰 가.

3. 규칙은 하나야 — 같이 쓰이는 말이 비슷하면 숫자도 비슷하게.

자리가 되면 계산이 된다

1. 뜻이 자리로 바뀌면 가깝다와 멀다를 숫자로 잴 수 있어.

2. 글자로만 보면 '사과'와 '귤'은 아무 관계가 없는데, 자리로 보면 **나란히** 놓여.

3. 이게 다음 날 배우는 비슷한 것 찾기의 바탕이야.

글자가 아니라 뜻을 잰다

1. 단어처럼 문장도 숫자 묶음으로 바꿀 수 있어. 그러면 두 문장 사이 거리가 나와.

2. 겹치는 글자가 하나도 없어도 뜻이 같으면 **가까운 자리**에 놓여.

3. 반대로 글자가 거의 같아도 뜻이 반대면 멀어져.

왜 글자 세기로는 안 되나

1. '재밌어'와 '재미없어'는 글자가 거의 같지만 뜻은 정반대야.

2. 글자가 겹치는 정도로만 재면 둘을 **아주 가깝다**고 잘못 세게 돼.

3. 그래서 글 찾기도 글자 맞추기 대신 뜻의 자리를 쓰는 쪽으로 옮겨 갔어.

외우기 대신 찾아서 펴 놓기

1. 외운 것만으로 답하면 모르는 걸 그럴듯하게 지어내기 쉬워.

2. 그래서 질문을 숫자 묶음으로 바꿔 **가까운 자리에 있는 글 조각**을 먼저 꺼내 와.

3. 그 글을 옆에 펴 두고 같이 읽으며 답해.

좋은 점과 약한 점

1. 모델을 다시 가르치지 않아도 돼 — **펴 놓는 글만 갈아 끼우면** 답도 따라서 새로워져.

2. 어디서 나온 말인지 밝힐 수 있어서 확인하기도 쉬워.

3. 대신 찾아온 글이 틀리면 답도 같이 틀려. 찾기가 절반이야.

다 큰 모델을 조금 다듬는다

1. 크게 한 번 배운 기본 모델 위에 필요한 일만 더 학습시키는 걸 파인튜닝이라고 해.

2. 안에 든 숫자는 거의 그대로 두고 **아주 일부만** 고쳐.

3. LoRA라는 방법은 고치는 숫자를 만분의 일까지 줄였어.

어느 쪽을 써야 하나

1. 따로 더 가르치기는 **말투와 형식**을 바꾸는 데 세.

2. 새 사실을 넣고 싶으면 앞에서 본 책 펴놓고 답하기가 맞아.

3. 크게 키우는 단계는 몇 달이 걸리지만, 다듬는 단계는 그에 비하면 훨씬 짧고 싸.

정답 대신 어느 쪽이 더 나은지

1. 같은 질문에 답을 두 개 내게 하고 사람이 더 나은 쪽을 고르는 일을 수만 번 되풀이해.

2. 모은 건 맞고 틀림이 아니라 **사람의 취향**이야.

3. 답마다 일일이 점수를 매기면 끝이 없으니까, 고른 기록으로 채점 모델을 먼저 만들어.

효과가 얼마나 컸나

1. 그다음부터는 그 채점 모델이 대신 점수를 주고, AI는 그 점수를 따라 다듬어져.

2. 2022년에 공개된 연구에서, 이렇게 다듬은 작은 모델의 답을 사람들은 **130배 넘게 큰 모델**의 답보다 더 좋아했어.

3. 크기보다 다듬기가 더 큰 차이를 만든 드문 사례야.

계산의 모양이 원래 같았다

1. 그래픽카드(GPU)는 쉬운 계산을 한꺼번에 아주 많이 하려고 만든 부품이야.

2. 화면의 점 수백만 개를 동시에 칠하려면 **똑똑한 몇 명보다 단순한 수천 명**이 필요했거든.

3. AI 학습도 어려운 문제 하나가 아니라 쉬운 곱셈 수조 개를 동시에 푸는 일이야.

숫자로 보면

1. 학습에 많이 쓰는 카드 한 장 안에는 아주 작은 스위치가 **약 800억 개** 들어 있어.

2. 그 한 장이 1초에 약 1,000조 번 계산해.

3. 그래도 모자라서 라마 3는 1만 6천 장을 한 덩어리로 묶어 54일을 돌렸어.

한 번은 작고, 다 더하면 커

1. 2025년에 공개된 측정에서 글로 한 번 물었을 때 중간값은 **0.24와트시**였어.

2. TV를 9초 켜 두는 정도라, 한 번만 보면 걱정할 양이 아니야.

3. 문제는 그런 질문이 쉬지 않고 쌓인다는 거지.

건물 전체로 보면

1. 전 세계 데이터센터가 2024년에 쓴 전기는 **415테라와트시**, 전 세계 전기의 약 1.5%야.

2. 2030년엔 945테라와트시로 두 배가 넘을 거라고 국제에너지기구가 봤어.

3. 작아 보이는 1.5%지만, 6년 만에 두 배가 되는 항목은 흔하지 않아.

조각으로 나눠 쓴다

1. 요즘 폰 안에는 숫자 **30억 개쯤 되는 작은 모델**이 들어 있어.

2. 인터넷이 끊겨도 요약하고 다듬고 알림을 정리해.

3. 폰 안 모델과 서버 모델은 크기가 수십에서 수백 배 차이 나.

왜 나눠 쓰나

1. 폰이 먼저 해 보고 버거운 조각만 골라 밖으로 내보내.

2. 안에서 끝내면 **빠르고, 전기도 덜 들고, 내 글이 밖으로 안 나가**거든.

3. 긴 글이나 어려운 질문은 앞에서 본 그 건물까지 다녀와야 해.

안은 마법이 아니라 계산이었어.
단순한 일이 엄청나게 쌓인 거지.



6부에서는 앞으로를 본다.