

● DAY 14 · 2 부 여기까지 온 길

문장 속 단어들이 서로를 쳐다보기 시작했다 — 2017 트랜스포머

하루 5 분 , AI 이야기 · 2026 년 9 월 20 일



단어들이 서로를 쳐다보게 했더니 , AI 가 문장을 통째로 이해하기 시작했어 .

한 줄로 말하면

- 1 2017 년 구글 연구팀이 낸 논문 한 편이 지금 쓰는 AI 의 뼈대를 정했어 . 이름은 **트랜스포머** , 논문 제목은 '**Attention Is All You Need**' 야 .
- 2 그 전까지는 문장을 앞에서부터 한 단어씩 차례로 읽었어 . 트랜스포머는 **문장 속 모든 단어가 서로를 동시에 쳐다보게** 했어 — 이걸 '어텐션' 이라고 불러 .
- 3 동시에 본다는 건 **한꺼번에 계산할 수 있다**는 뜻이야 . 그래서 그래픽카드에 딱 맞았고 , 이 뼈대 위에서 GPT · 클로드 · 제미니가 전부 자랐어 .

차례로 읽기 vs 한꺼번에 쳐다보기

2017 년 이전 · 차례로

- 단어를 앞에서부터 하나씩
- 앞 단어의 기억이 점점 흐려짐
- 한 번에 하나 — 기다려야 함

트랜스포머 · 동시에

- 모든 단어가 모든 단어를 쳐다봄
- 멀리 있는 단어도 바로 짐작
- 한꺼번에 계산 — 그래픽카드에 딱

같은 문장을 읽어도 , 오른쪽은 단어 사이의 관계를 표로 한 번에 계산해 .

핵심 : 순서대로 읽기를 버리고 , 서로 쳐다보게 하니 빨라지고 똑똑해졌다

'그것' 이 뭘 가리키는지 , 단어들이 서로 쳐다보기

'그것' 을 보면

동물이 길을 건너지 못했다 . 그것은 너 ...

동물이 길을 건너지 못했다 . 그것은
너무 피곤했다 .

'피곤했다' 가 옆에 있으니 '그것' 은 동물이야 .
길은 피곤할 수 없잖아 . 트랜스포머는 이걸 문장
전체를 한 번에 훑어서 알아...

문장을 바꾸면

동물이 길을 건너지 못했다 . 그것은 너 ...

동물이 길을 건너지 못했다 . 그것은
너무 넓었다 .

단어 하나 (' 넓었다 ') 만 바뀌었는데 '그것' 이
쳐다보는 곳이 길로 옮겨 갔어 . 앞에서부터
차례로만 읽었다면 이 차이를 잡기...

예전 방식이라면

앞에서부터 한 단어씩 — 앞 내용은 점 ...

동물이 길을 건너지 못했다 . 그것은
너무 넓었다 .

차례로 읽는 옛 방식은 문장이 길어질수록 앞
단어가 흐려졌어 . 그래서 '그것' 이 뭘지 자주
놓쳤지 . 어텐션은 거리와 상관없이...

단어 하나를 고르면 , 그 단어가 문장 속 어느 단어를 얼마나 세게 쳐다보는지 색으로 보여줘 . 진할수록 세게 본다는 뜻이야 .

학습 .html 에서 직접 만져 볼 수 있어

세 문제만 풀어 보자

Q1

트랜스포머가 그 전 방식과 가장 다른 점은 ?

- ① 문장 속 단어들이 서로를 동시에 쳐다본다
- ② 단어를 더 빨리 외운다
- ③ 인터넷을 검색해서 답한다
- ④ 사람이 규칙을 손으로 넣어 준다

Q2

' 동시에 본다 ' 가 왜 중요했을까 ?

- ① 문장이 더 예뻐져서
- ② 한꺼번에 계산할 수 있어서 그래픽카드에 딱 맞았기 때문에
- ③ 메모리가 덜 들어서
- ④ 사람이 읽기 편해서

Q3

' 그것은 너무 피곤했다 ' 에서 ' 그것 ' 이 ' 동물 ' 인 걸 어텐션은 어떻게 알까 ?

- ① 제일 앞 단어라서
- ② 사전에 적혀 있어서
- ③ 주변 단어와 어울림 점수로
- ④ 무작위로 골라서

왜 그런지까지 알고 가자

Q1

①

앞에서부터 차례로 읽는 대신, 모든 단어가 모든 단어를 한 번에 본다. 이게 '어텐션' 이고, 논문 제목이 '어텐션만 있으면 된다' 였던 이유야.

Q2

②

차례로 읽으면 앞 단어가 끝나야 다음을 볼 수 있어. 동시에 보면 한 방에 계산할 수 있고, 어제 배운 그래픽카드가 바로 그런 계산을 잘하지.

Q3

③

단어마다 '누구를 봐야 하나' 점수를 매겨. '피곤' 과 어울리는 건 '길' 이 아니라 '동물' 이라 점수가 그쪽으로 몰리는 거야.

**단어들이 서로를 쳐다보게 했더니,
AI 가 문장을 통째로 이해하기 시작했어 .**



내일은 2020 년 GPT-3 — 이 뼈대를 1,750 억 개 숫자로 키우면 무슨 일이 생기는지 .