

LESSON
23

고급 초6·중2 AI 교실

거대언어모델(LLM)은 어떻게
문장을 만들까?

글을 토큰으로 나누고, 앞 내용을 보고 다음 토큰을 확률로 골라요.

**학습 목표** 이번 시간이 끝나면...

- ✓ LLM의 핵심 4가지(토큰, 사전 학습, 어텐션, 확률로 고르기)를 설명할 수 있어요.
- ✓ 어텐션이 문장 속 관계를 찾는 방법을 예로 들 수 있어요.
- ✓ LLM이 패턴을 계산할 뿐 세상을 직접 이해하는 것은 아니라는 점을 알 수 있어요.

● 고급 23

거대언어모델(LLM)은
어떻게 문장을 만들까?

LLM은 아주 많은 글로 미리 배운 AI예요.

다음에 올 토큰을 확률로 골라 이어 가요.

**“주변 조각을 보면
이 조각 자리가 보여!”**

LLM도 토큰끼리의 관계를 살피요.

LLM의 4가지 핵심

1

토큰으로
쪼개기

2

많은 글로
사전 학습

3

어텐션으로
관계 보기

4

확률로
다음 고르기

꼭 알아야 할 단어

토큰

= AI가 글을 나눈 작은 조각

어텐션

= 토큰끼리 관련을 따져 집중하는 방법

Q. LLM은 문장을 토큰이라는
작은 조각으로 나눠서 처리한다?

O

/

X

정답: O

그림 1 한눈에 보는 23강 — 문장이 토큰 조각으로 나뉘는 모습부터 보세요.



쉽게 풀어 보기

LLM이 글을 조각내고, 관계를 찾고, 다음 말을 고르는 방법을 알아봐요.



글을 조각내고, 다음 조각을 골라요

LLM은 글을 그대로 읽지 않고 **토큰(token)**, 곧 AI가 글을 처리하려고 나눈 작은 조각으로 나뉘요. 토큰은 단어일 수도, 단어보다 작은 조각일 수도 있어요.

먼저 책, 웹 문서처럼 엄청난 양의 글로 '다음 토큰 맞히기'를 연습해요. 많은 글로 언어의 패턴을 미리 배우는 이 단계를 **사전 학습(pre-training)**이라고 해요.

문장을 만들 때는 토큰끼리 얼마나 관련 있는지 따져 집중하는 **어텐션(attention)**을 써요. 그리고 다음 토큰 후보마다 확률을 매겨 하나를 골라 붙여요.



직소 퍼즐 조각 맞추기로 이해해요

퍼즐 조각 하나만 들고 있으면 어디에 들어갈지 몰라요. 하지만 **주변 조각의 색과 무늬를 함께 보면** 자리가 보이지요.

LLM도 문장을 작은 조각(토큰)으로 나누고, 각 조각이 다른 조각들과 어떤 관계인지 살펴요.

"비가 와서 우산을 폈는데, 그것은 파란색이었다"에서 '**그것**'이 '**우산**'과 **이어진다**는 걸 찾아내는 식이에요.



한 줄 요약

거대언어모델(LLM, large language model)은 아주 많은 글로 미리 배운 커다란 인공지능망이에요. 글을 '토큰'이라는 조각으로 나누고, 앞 내용을 보고 다음에 올 토큰을 확률로 골라 문장을 이어 가요.

생활 속 LLM과 그 한계

챗봇, 글 요약, 이메일 초안 쓰기, 코드 작성 도우미가 LLM을 써요. 질문에 맞춰 자연스러운 문장을 빠르게 만들어 줘요.

하지만 LLM은 **글의 패턴을 계산**할 뿐, 사람처럼 세상을 직접 겪으며 이해하는 것은 아니에요. 그래서 문장은 그럴듯해도 내용은 틀릴 수 있어요.

또 배운 글에 치우침이 있으면 답에도 치우침이 생길 수 있어요. 중요한 내용은 출처를 확인하고, 개인 정보는 넣지 않아요.

미리 보기 뒤에 나오는 '직접 해보기 미션'에서 문장을 조각내고 어텐션처럼 선을 그어 봐요.

미리 생각하기 “민수는 지우에게 책을 빌려줬다. 그녀는 고맙다고 했다.”에서 ‘그녀’는 누구일까요?
어떻게 알았나요?

‘그녀’ 주변의 어떤 조각을 보고 판단했는지 생각해 봐요.

숫자로 보는 23강

사실

2017년

요즘 LLM 대부분이 바탕으로 삼는 ‘트랜스포머’ 구조가 발표됐어요.

사실

수십억 개

거대언어모델의 가중치 수는 보통 수십억 개가 넘어요.

예시

5문장

미션에서 낱말 조각으로 잘라 볼 문장 수

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

✓

표로 정리해요

LLM의 핵심 4가지를 퍼즐 비유와 함께 정리해요.

 LLM의 핵심 4가지

핵심	하는 일	퍼즐 비유
① 토큰 (token)	글을 단어나 단어보다 작은 조각으로 나누고, 조각마다 번호를 붙여요.	그림을 퍼즐 조각으로 나뉘요.
② 사전 학습 (pre-training)	책, 웹 문서처럼 엄청난 양의 글로 '다음 토큰 맞추기'를 연습해요.	퍼즐을 아주 많이 맞춰 봐요.
③ 어텐션 (attention)	문장 속 토큰들 가운데 서로 관련이 깊은 것을 찾아 더 집중해요.	주변 조각의 색과 무늬를 함께 봐요.
④ 확률로 고르기	다음 토큰 후보마다 확률을 매기고 하나를 골라 붙여요. 이것을 되풀이하면 문장이 돼요.	가장 잘 맞을 조각을 골라 끼워요.

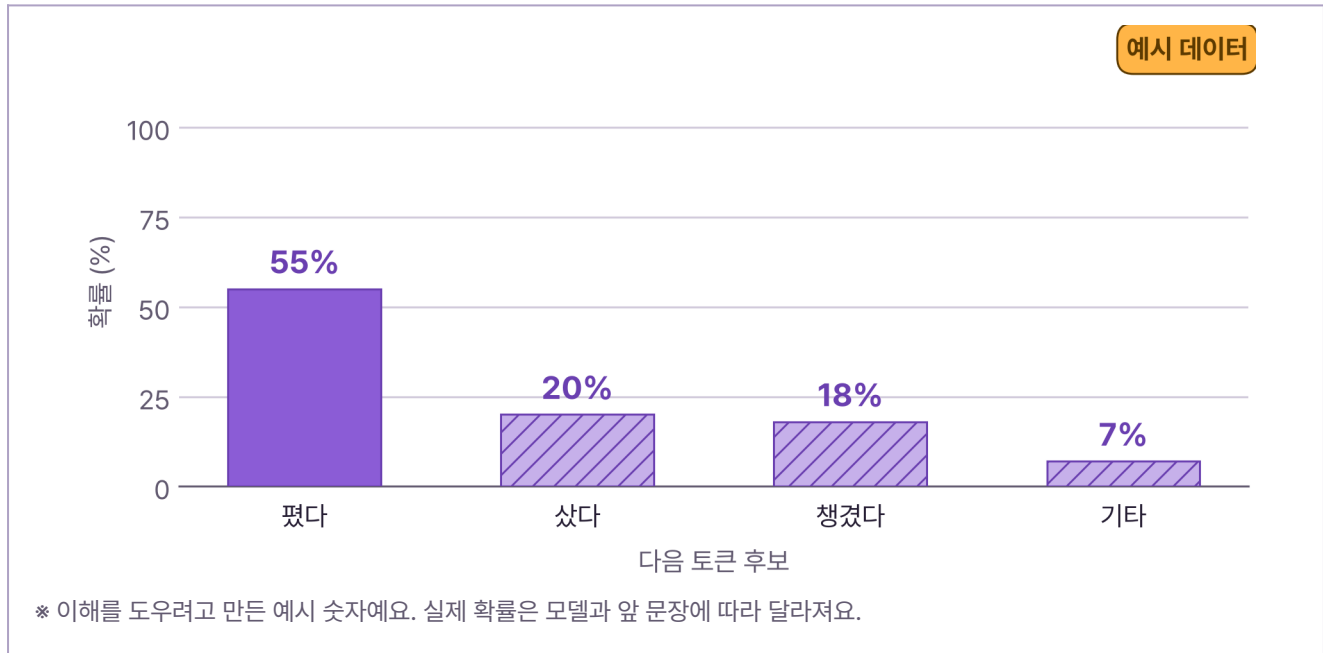
 LLM이 하는 것과 하지 못하는 것

살펴봐요	잘하는 것	하지 못하거나 약한 것
문장	자연스러운 문장 이어 가기	내용이 사실인지 스스로 보장하기
관계	멀리 떨어진 단어끼리의 관계 찾기	세상을 직접 겪으며 이해하기
지식	배운 글의 패턴 활용	배운 뒤에 생긴 일 알기

생각해 보기 LLM이 “그렇듯하지만 틀린 문장”을 만들 수 있는 까닭을 표를 보고 적어 보세요.

‘패턴 계산’과 ‘사실 확인’의 차이를 생각해 봐요.

“비가 와서 우산을 ___” 다음 토큰 후보와 확률 (예시)



그래프 읽기 가장 확률이 높은 후보는 무엇인가요? 앞 문장이 “비가 그쳐서”로 바뀌면 확률은 어떻게 달라질까요?

이 그래프는 예시 데이터예요. 앞 내용이 다음 토큰의 확률을 바꿔요.

가 꼭 알아야 할 단어

토큰

token

AI가 글을 처리하려고 나눈 작은 조각

사전 학습

pre-training

많은 글로 언어의 패턴을 미리 배우는 단계

어텐션

attention

토큰끼리 얼마나 관련 있는지 따져 집중하는 방법

**직접 해보기 미션**

손과 눈으로 직접 확인해 봐요.

문장 조각내고 어텐션 선 긋기

문장 5개를 종이에 쓰고 가위로 낱말 조각으로 잘라 보세요. (예: “비가 와서 / 나는 / 우산을 / 썼는데, / 그것은 / 파란색이었다.”)

1 문장 5개를 쓰고 낱말 조각으로 잘라요.

2 ‘그것’, ‘거기’, ‘그때’ 같은 말이 어떤 조각을 가리키는지 색연필로 선을 그어 이어요. 선이 모이는 조각이 어텐션이 ‘집중’하는 곳이라고 생각하면 돼요.

3 조각을 섞어 짝에게 주고, 앞 조각만 보고 다음 조각을 맞춰 보게 해요.

※ 가위를 쓸 때는 조심해요. 문장에는 실제 친구 이름 대신 지어낸 이름을 써요.

기록표

	가리키는 말	이어진 조각	짝이 맞았나요?
문장 1			
문장 2			
문장 3			
문장 4			
문장 5			

- 짝이 다음 조각을 맞추기 쉬웠던 문장과 어려웠던 문장은 무엇인가요? 왜 그럴까요?

선이 가장 많이 모인 조각

LLM과 닮은 점, 다른 점



정리 LLM은 토큰 사이의 관계를 따져 집중하고, 다음 토큰을 확률로 골라 문장을 이어 가요.

**확인 퀴즈**

스스로 풀고, 정답은 마지막 쪽에서 확인해요.

1. **O/X** LLM은 문장을 토큰이라는 작은 조각으로 나눠서 처리한다.☐

O

☐

X

2. **고르기** LLM이 다음 말을 고르는 방식으로 알맞은 것은?☐

① 확률이 높은 후보들 가운데서 골라요

☐

② 사전에서 아무 단어나 골라요

☐

③ 매번 사람에게 물어봐요

3. **빈칸** 많은 글로 언어의 패턴을 미리 배우는 단계를 _____(이)라고 해요.4. **쓰기** LLM의 답을 그대로 믿지 않고 확인해야 하는 까닭을 '패턴'이라는 말을 넣어 써 보세요.

**중2 한 걸음 더**

트랜스포머와 '거대'의 뜻

요즘 LLM 대부분은 2017년에 발표된 '**트랜스포머(transformer)**' 구조를 바탕으로 해요. 챗GPT의 'GPT'에서 T도 트랜스포머를 뜻해요. 트랜스포머는 어텐션 덕분에 긴 글에서 멀리 떨어진 단어끼리의 관계도 잘 잡아요.

'거대'라는 말은 가중치의 수가 보통 수십억 개가 넘고, 배운 글도 엄청나게 많다는 뜻이에요. 하지만 LLM은 **글의 패턴을 계산할 뿐**, 사람처럼 세상을 직접 겪으며 이해하는 것은 아니에요. 그래서 그럴듯하지만 틀린 문장을 만들 수도 있어요.

- **생각해 봐요:** 사람이 문장을 이해하는 방식과 LLM이 문장을 만드는 방식은 어떻게 다를까요?



용어 정리

이번 시간에 나온 어려운 말을 모았어요.

용어	영어	쉽게 말하면	예시
거대언어모델	LLM	아주 많은 글로 배운 커다란 인공신경망	대화형 AI
토큰	token	AI가 글을 처리하려고 나눈 작은 조각	"우산/을"
사전 학습	pre-training	많은 글로 언어의 패턴을 미리 배우기	다음 토큰 맞추기 연습
어텐션	attention	관련 깊은 토큰에 더 집중하는 계산	'그것' → '우산'
트랜스포머	transformer	어텐션을 바탕으로 한 신경망 구조	GPT의 T
확률	probability	그럴 것 같은 정도	'떨다' 55%
가중치	weight	신경망 속 중요도 숫자	수십억 개
문맥(맥락)	context	앞뒤 내용과 상황	"비가 와서"
지시어	reference word	앞에 나온 것을 가리키는 말	그것, 거기, 그때
패턴	pattern	글 속에서 되풀이되는 특징	"비" 다음에 "우산"



정답과 풀이

퀴즈를 다 풀 다음에 확인해요

1. 0 — 토큰은 단어일 수도 있고, 단어보다 작은 조각일 수도 있어요.
2. ① — 앞 내용을 보고 후보마다 확률을 계산해 골라요.
3. 사전 학습(pre-training) — '다음 토큰 맞추기'를 엄청나게 연습해요.
4. (예시 답) LLM은 글의 패턴으로 그럴듯한 다음 말을 고를 뿐 사실을 확인하지 않아서, 틀린 내용도 자연스럽게 말할 수 있기 때문이에요.