

ChatGPT 토큰이란?

글자 수와 다른 이유를 직접 관찰하는 초보자 실습

한 문장 결론

토큰은 글자 한 개가 아니라 모델이 텍스트를 처리하도록 나눈 문자 조각의 묶음입니다. 그래서 같은 글자 수라도 언어·숫자·기호·인코딩에 따라 토큰 수가 달라질 수 있습니다.

토큰이 만들어지는 흐름

① 입력 문장	② 작은 조각	③ 모델 처리	④ 답변
배송 지연 주문만 표로 정리해 주세요.	문자 조각 여러 묶음	조각 사이의 관계 계산	정리된 표 초안

먼저 예측해 보기

TXT 실습 파일의 T-01부터 T-08까지 읽고, 글자 수만 보았을 때 토큰이 많을 것 같은 순서를 적어 보세요. 정답을 맞추는 것보다 왜 그렇게 예상했는지 한 문장으로 남기는 것이 중요합니다.

성공 확인

- ☐ 토큰을 글자 한 개와 같은 말로 설명하지 않았다.
- ☐ 숫자·기호·언어가 토큰 나눔에 영향을 줄 수 있다고 예상했다.
- ☐ 관찰값을 실제 ChatGPT 전체 사용량과 구분했다.

정답표: 글자 수만으로는 토큰 수를 알 수 없습니다

측정 조건

tiktoken 0.13.0으로 o200k_base와 cl100k_base 인코딩을 각각 적용했습니다. 이 표는 문장 자체만 측정한 값이며, 실제 대화의 숨은 지시·이전 문맥·도구 사용량은 포함하지 않습니다.

ID	문장	글자	o200k	cl100k
T-01	회의는 오후 3시에 시작합니다.	17	9	13
T-02	The meeting starts at 3 PM.	27	8	8
T-03	보고서를 요약해 주세요.	13	7	11
T-04	Please summarize the report.	28	5	5
T-05	배송 지연 주문만 표로 정리해 주세요.	21	12	19
T-06	order_id, promised_date, shipped_date	37	8	8
T-07	확인 완료 + 체크 이모지	7	4	7
T-08	2026-08-15 06:00 게시 준비 완료	29	15	20

관찰 1 · 글자가 많아도 토큰은 적을 수 있습니다

T-04는 28글자이지만 두 인코딩에서 5토큰입니다. T-01은 17글자인데 9토큰 또는 13토큰입니다. 따라서 '글자 수 ÷ 일정한 숫자'로 모든 언어의 토큰을 정확히 계산할 수 없습니다.

관찰 2 · 숫자와 구분 기호는 조각이 늘어날 수 있습니다

T-08은 날짜, 시각, 세로선 기호가 섞여 15토큰 또는 20토큰으로 측정됐습니다. 사람이 보기에는 짧은 한 줄이어도 모델이 나누는 방식은 다를 수 있습니다.

관찰 3 · 같은 문장도 인코딩에 따라 달라질 수 있습니다

T-01은 o200k_base에서 9토큰, cl100k_base에서 13토큰입니다. 어떤 인코딩을 쓰는지 기록하지 않으면 숫자만 비교하기 어렵습니다.

초보자가 기억할 세 문장

- 1 토큰은 글자 수가 아니라 모델이 처리하는 문자 조각의 묶음입니다.
- 2 정확한 수치는 사용하는 모델과 인코딩을 함께 봐야 합니다.
- 3 짧게 쓰는 것보다 목적·조건·출력 형식을 분명히 쓰는 일이 먼저입니다.

실패 화면과 수정 방법

아래 화면은 실제 ChatGPT 캡처가 아니라 오류 상황을 설명하기 위해 만든 편집 예시입니다.

실패 예시	수정 예시
공식 Tokenizer 입력 칸이 비어 있어 Tokens 0처럼 보임	입력 칸을 클릭하고 문장을 다시 붙여넣은 뒤 선택된 모델 묶음을 확인
내 측정값과 자료의 수치가 다름	모델·인코딩·확인 날짜를 함께 기록하고 같은 조건에서 다시 비교
토큰이 적을수록 항상 좋은 질문이라고 판단	목적·자료·조건·출력 형식을 먼저 지키고 불필요한 반복만 줄임

실전에서 토큰을 대하는 안전한 순서

1	먼저 원하는 결과를 한 문장으로 적습니다.
2	답에 꼭 필요한 자료와 조건만 붙입니다.
3	같은 설명이 반복되는 부분을 한 번만 남깁니다.
4	중요한 작업은 토큰 수보다 원문 대조와 결과 검산을 우선합니다.

마지막 확인

- ☐ 토큰과 글자 수를 구분했다.
- ☐ 측정 조건에 모델 또는 인코딩 이름을 적었다.
- ☐ 관찰값을 요금·계정 한도의 확정값으로 사용하지 않았다.
- ☐ 질문의 명확성을 해치면서까지 문장을 억지로 줄이지 않았다.
- ☐ 중요한 답은 출처와 원문으로 다시 확인했다.

공식 확인 자료

OpenAI Tokenizer: <https://platform.openai.com/tokenizer>

OpenAI API Reference - Chat usage: <https://developers.openai.com/api/reference/resources/chat>

작성·검토 정보

작성: AI 플레이집 편집팀 · 검토: AI 플레이집 운영자 상시 승인 범위 · AI 사용: 공식 문서 확인, 가상 문장 구성, tiktoken 측정, PDF 배치
검사에 AI와 코드를 사용했습니다. 오류 제보: <https://aiplayzip.com/corrections-policy/>