

# GPT-3

 Stakeholders	
 Status	
 Type	
 Created	@2022년 8월 18일 오후 3:19
 Last Edited Time	@2022년 8월 18일 오후 4:39
 Last Edited By	
 날짜	

## GPT-3(Generative Pre-trained Transformer 3)

### 논문 명 Language Models are Few-Shot Learners

전체버전은 1,750억개의 매개변수를 가지고 있어, 2020년 5월 도입된 이전버전 GPT-2 보다 100배 이상의 크기이다.

할 수 있는 것

질문에 답하고, 에세이를 쓰고, 긴 텍스트를 요약하고, 언어를 번역

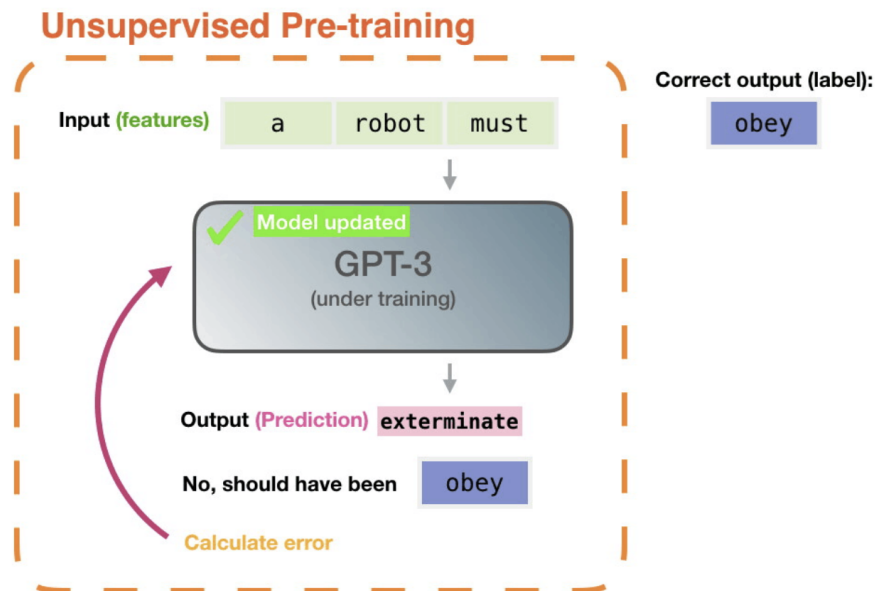
뒤에 올 단어를 예측하는 언어모델

**Text:** Second Law of Robotics: A robot must obey the orders given it by human beings

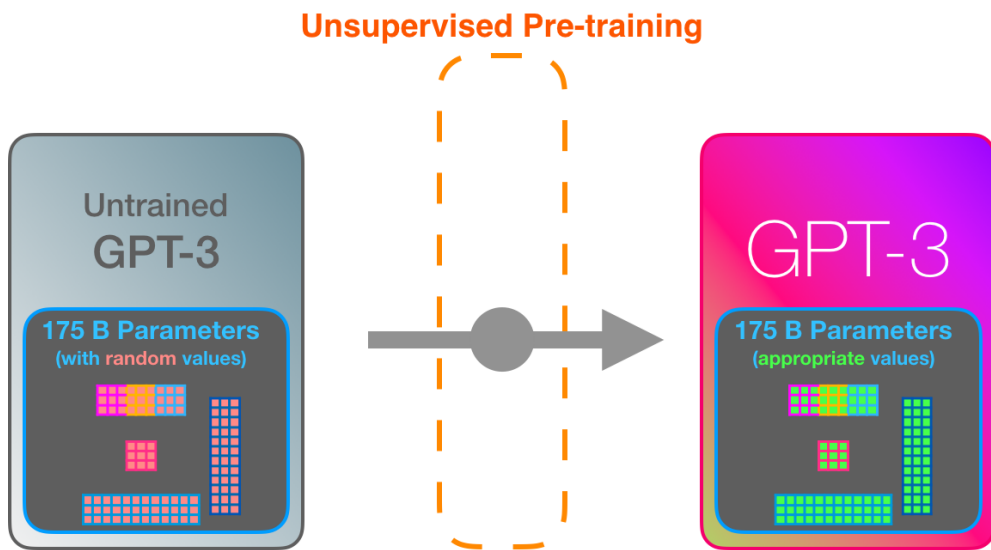


**Generated training examples**

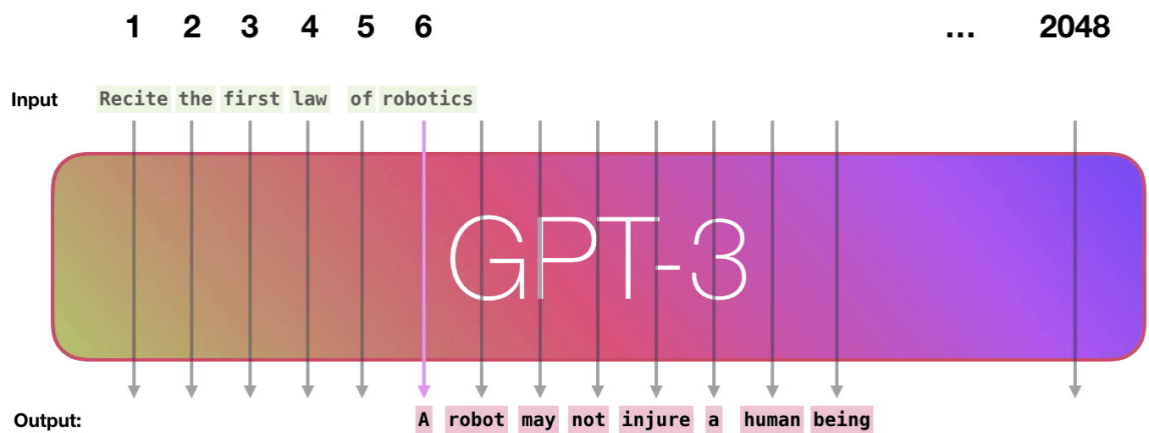
Example #	Input (features)	Correct output (labels)
1	Second law of robotics :	a
2	Second law of robotics : a	robot
3	Second law of robotics : a robot	must
...		



4990억개 데이터셋 중에서 가중치 샘플링하여 3000억개로 구성된 데이터셋으로 사전학습

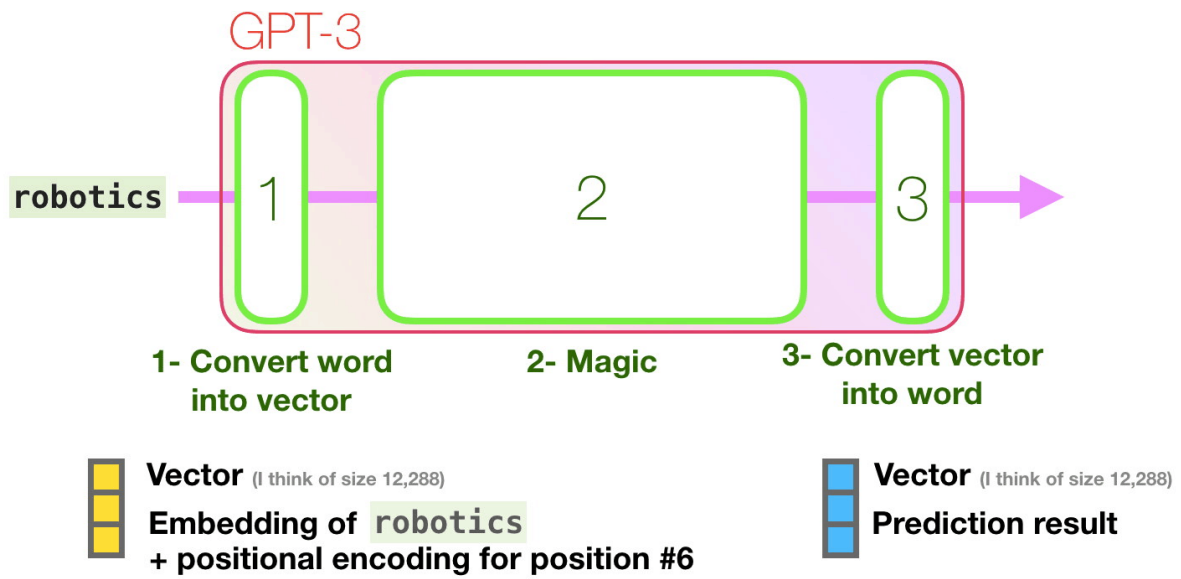


비지도학습을 통해 파라미터 학습



2048개의 토큰 너비 = context window

<http://jalammar.github.io/images/gpt3/06-gpt3-embedding.gif>

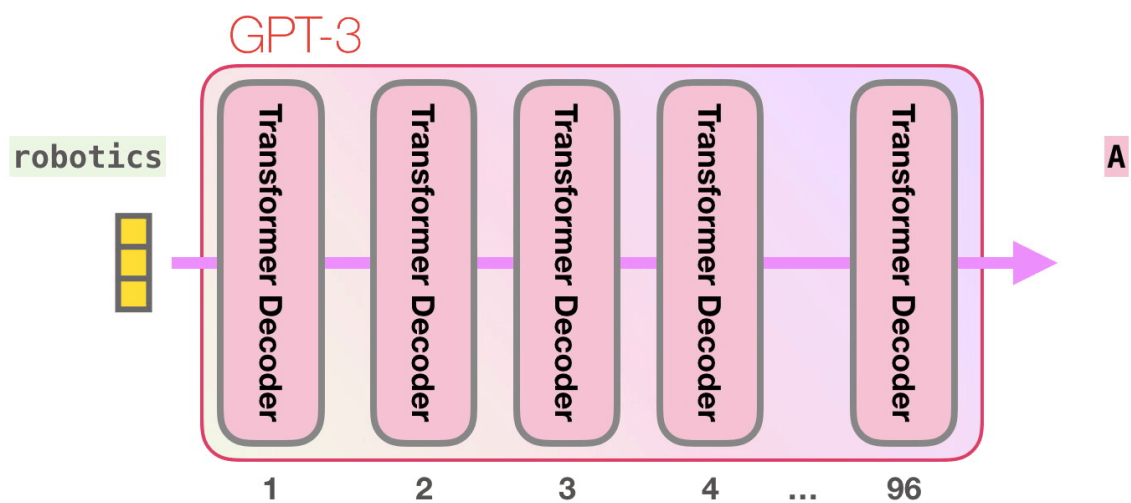


1단계: 단어를 단어에 해당하는 벡터로 변환

2단계: 벡터를 통해 예측값을 연산

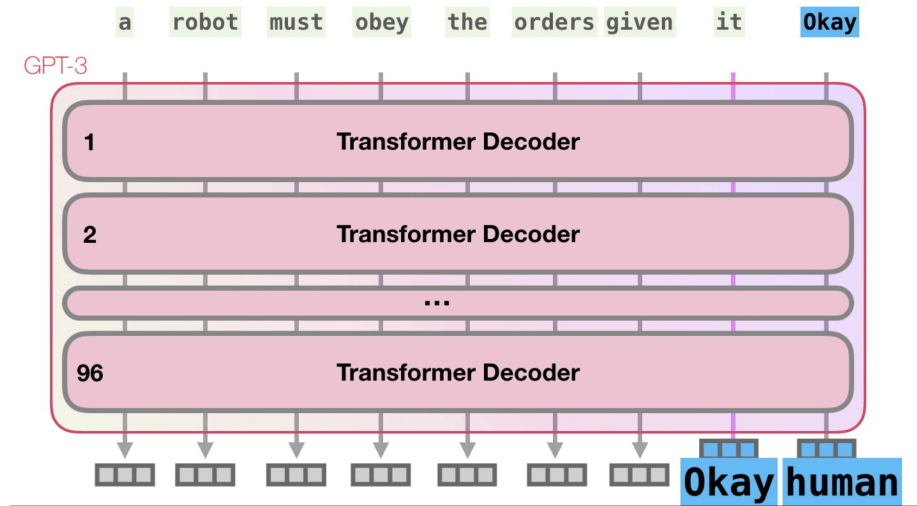
3단계: 연산된 결과 벡터를 다시 단어로 변환

2단계 → 96개의 트랜스포머 디코더 레이어로 이뤄져있음



디코더 안에서 이루어지는 과정은 GPT-2와 동일하다고함

GPT-3에서 차이점은 dense and sparse self-attention layers를 사용



각각의 토큰들이 모든 레이어를 통과하면 첫번째 단어 출력 okay

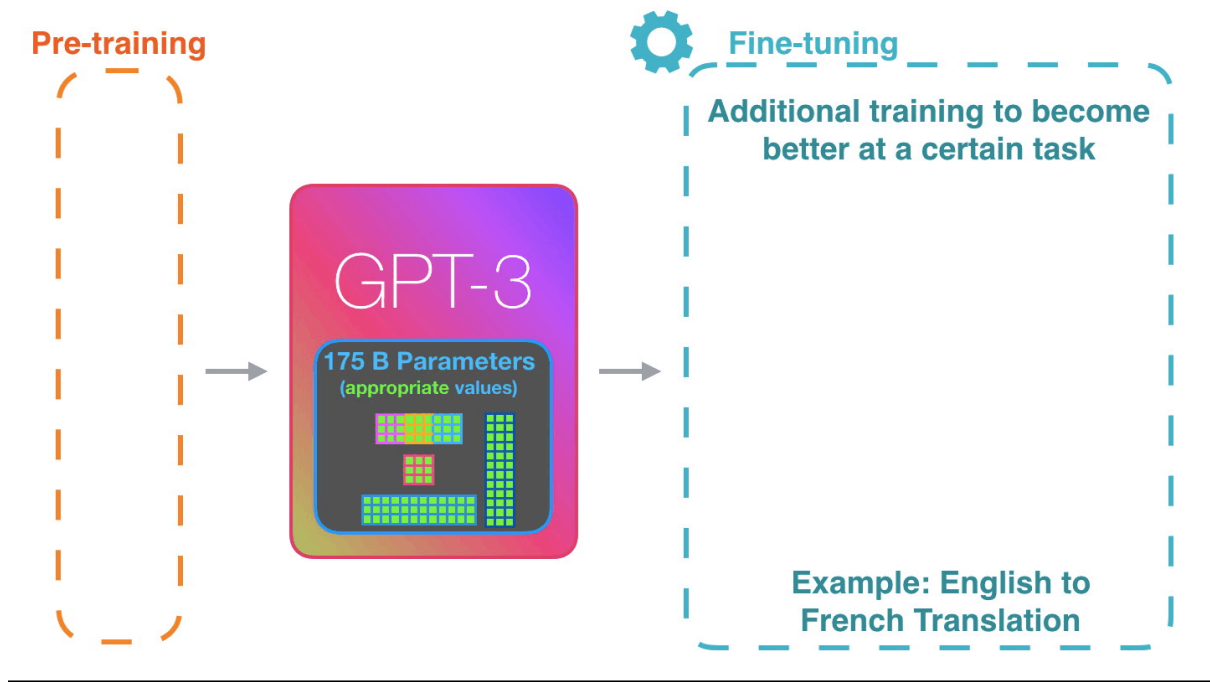
단순히 입력이 완료되었다는 의미로 그 다음부터 정답 출력

[example] an input that says "search" [toCode] Class App extends React Component... </div> } } }  
[example] a button that says "I'm feeling lucky" [toCode] Class App extends React Component...  
[example] an input that says "enter a todo" [toCode]



Class App extends React Component { ...

자연어로 리액트 코드를 설명하면 리액트 코드를 생성할 수도 있음



task에 따라서 fine-tuning하면 모델의 weight를 업데이트 시켜서 특정 task에서 더 좋은 성능을 발휘하게 함

한계점 5가지

1. 하나의 모델이 다양한 문제를 범용적으로 잘 풀지만 모두 최고수준으로 푸는 것이 아니다.
2. 현실의 물리 법칙 잘 모른다.

“치즈를 냉장고에 넣으면 녹을까?” 같은 문제에 사람이라면 당연히 안 녹는다고 답하겠지만 GPT-3는 녹는다고 답한다.

3. 모델이 너무 크다.
4. '기억'이라는 개념이 없다. 최근의 2048 토큰만 보고 생성
5. 이해하고 학습하는게 아니라 통계적으로 그럴듯한 단어를 생성?