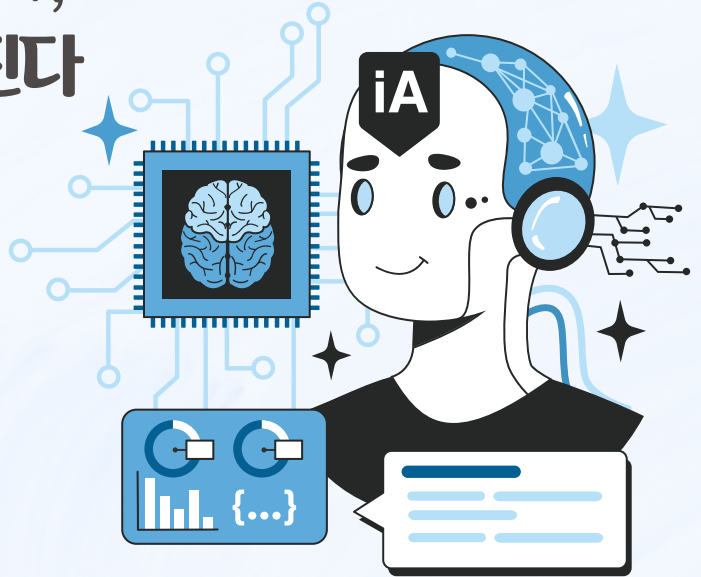


# 친구 같은 ChatGPT, 자세히 봐야 친해진다

강지희 객원기자



인공지능 챗봇 ‘챗GPT’(ChatGPT)가 2022년 11월 30일 출시되자마자 일주일 만에 100만 사용자를 돌파했다. ChatGPT는 친구처럼 자연스럽게 대화할 수 있으며, 구글 같은 검색 서비스를 대체할 가능성이 있다고 평가받는다. 너도나도 뛰어드는 ChatGPT 시장에 우리는 어떻게 대비해야 할까? 지피지기 백전백승이라 했다. 상대를 알아야 대비 방안을 찾을 수 있다. ChatGPT가 어떻게 작동하는지 알아보자.

## 거대 언어 모델과 인간의 피드백으로 자연스럽게 말한다

ChatGPT를 포함한 챗봇은 모두 자연어 처리를 거쳐야 한다. 자연어 처리는 우리가 일상에서 사용하는 자연어의 의미를 분석해 컴퓨터가 처리할 수 있도록 하는 일을 말하며, 대화 및 자동 질의응답 시스템이 필요한 챗봇에서 꼭 필요한 과정이다.

자연어 처리에는 텍스트 전처리가 꼭 필요하다. 텍스트 전처리를 하지 않으면 뒤에서 배울 자연어 처리 기법들이 제대로 동작하지 않

기 때문이다. ChatGPT 같은 챗봇에서는 ‘단어 토큰화’를 전처리로 많이 활용한다. 주어진 코퍼스에서 토큰이라 불리는 단위로 나누는 작업을 ‘토큰화’(tokenization)라 부르며, 토큰의 기준을 단어로 하는 경우 ‘단어 토큰화’(word tokeniazation)라고 한다.

또 텍스트를 컴퓨터가 이해하고, 효율적으로 처리하려면 텍스트를 적절히 숫자로 변환하는 벡터화가 필요하다. ChatGPT는 각 단어를 인공신경망 학습을 통해 벡터화하는 워드 임베딩을 사용한다.

간단한 대화나 질문 이상으로 복잡한 문장을 만들려면, 해당 내용에 관한 지식이나 상식이 어느 정도 필요하다. GPT(Generative Pre-trained Transformer)에서 트랜스포머(Transformer)는 구글이 2017년에 처음 발표한 논문 'Attention Is All You Need'에서 등장한 신경망 모델로, 신경망 알고리즘 ‘어텐션(Attention)’을 이용해 자연어를 이해하고 활용하는 데 큰 강점을 보인다. 영어로 ‘주의’를 의미하는 어텐션은 각 ‘단어’나 ‘문장’ 중 어디가 중요하고, 어떤 관계성에 중점을

뒤야 하는지 살핀다. 덕분에 ChatGPT는 AI 챗봇 중 가장 사람에 가까운 대화를 할 수 있다.

챗봇은 언어 모델을 활용해 문장을 구성한다. 예를 들어 '나는 밥을' 뒤에 '먹는다'나 '한다'를 넣는다고 예측하는 것이다. 기계가 뒤에 따르는 단어나 문장이 적절하거나 말이 안 된다고 사람처럼 정확히 판단할 수 있다면, 기계의 자연어 처리 성능이 뛰어나다고 할 수 있다.

ChatGPT는 거대 언어 모델(LLM, Large Language Model)을 활용해 파라미터를 증가시키고 방대한 양의 데이터를 학습했다. 이 기능을 통해 사용자들은 ChatGPT에서 단락뿐만 아니라 여러 페이지의 콘텐츠를 작성할 수 있다.

또 사람의 피드백을 이용한 인간 피드백형 강화학습(RLHF)으로 ChatGPT의 인공지능을 고도화했다. 이 학습은 크게 3단계로 구성된다. 첫 번째 단계에서 OpenAI는 인터넷에서 떠도는 데이터 전체를 신뢰하지 않았다. 따라서 신뢰할 수 있는 질문과 답변만을 가지고 지도 학습(Supervised fine-tuning)을 시켰다.

두 번째 단계에서는 첫 단계의 언어 모델에 질문을 하고 여러 답변을 받는다. 그렇게 생성된 여러 답변을 놓고 질문자의 의도와 부합하는지 평가해 순위를 매긴다. 또 이를 별도로 데이터베이스로 만들고, 이 데이터를 이용해 학습을 반복하면서 답변 우선순위를 예측하도록 한다.

세 번째 단계에서는 학습에 사용한 데이터가 아닌 새로운 질문으로 생성된 답변을 보고 문제점을 미세하게 조정해 정확도를 높인다. 정확히는 PPO(Proximal Policy Optimization) 알고리즘을 사용해, 언어 모델의 파라미터 일부 또는 전체를 파인튜닝(fine-tuning, 미세 조정) 한다. 유용성과 진실성 같은 최종 성능평가도 사람이 직접 한다. ChatGPT는 단순히 답변을 예상해 답을

내놓는 수준을 넘어, 지식을 다루는 영역에서 다방면에 활용될 수 있다.

### BERT와의 차이점은?

파인튜닝을 사용하는 트랜스포머 기반 언어 모델로는 ChatGPT 말고도 다른 접근방식이 있다. 구글에서 2018년에 개발한 자연어 처리 사전 훈련 기술 BERT(Bidirectional Encoder Representations from Transformers)다. 차이점으로 BERT는 트랜스포머의 인코더, GPT는 디코더를 사용했다. 인코더는 어떤 대상을 목적을 갖고 처리하기 위해 처리를 위한 방식으로 변환한 것이다. 디코더는 인코딩된 대상을 원래 형태로 되돌리는 것을 말한다.

또 GPT는 generative라는 이름과 같이 생성형 언어 모델을 대표한다. 단어 하나가 주어지면 AI가 수백만 개의 웹페이지로 구성된 방대한 데이터베이스를 통해 인간과 비슷한 대화를 생성한다. 반면, BERT는 다양한 문장력에서는 GPT보다 떨어지지만, 특정 주제에 대해서는 더 적은 비용과 리소스로 대응할 수 있다.

### GPT, 꾸준히 성장 중

2020년 6월 나온 GPT-3는 비약적인 발전을 이룬다. GPT-3는 GPT-2보다 약 12배 많은 1,750억 개의 파라미터를 갖고, 필터 전 45TB, 필터 후 570GB 크기의 방대한 데이터셋을 학습했다. GPT-3.5와 GPT-4는 사람 수준의 농담을 하거나 개발자가 만든 코드의 오류를 순식간에 잡아내는 수준에 이르렀다. 사실관계와 달라도 내용을 사실인 것처럼 표현하는 할루시네이션 예러가 단점이지만, 사용자가 명확한 디렉션을 전달한다면 충분히 유용하게 사용할 수 있을 것으로 전망된다. 