

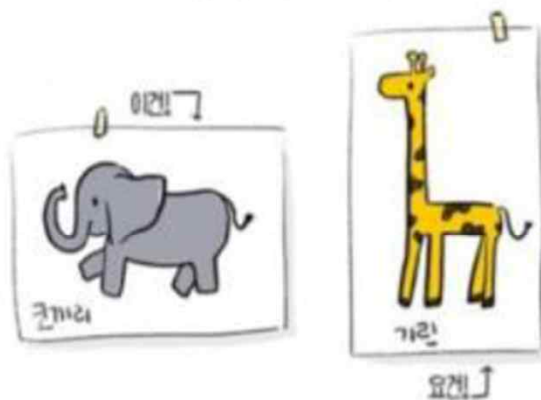
생성형 AI 및 ChatGPT 소개

생성형 AI

머신 러닝

지도 학습 (Supervised Learning)

문제와 정답을 모두 알려주고
공부시키는 방법



예측, 분류

비지도 학습 (Unsupervised Learning)

답을 가르쳐주지 않고
공부시키는 방법

비지도학습은 답을 가르쳐주지 않고 공부를 시키는거야.



연관 규칙, 군집

강화 학습 (Reinforcement Learning)

보상을 통해
상은 최대화, 벌은 최소화하는
방향으로 행위를 강화하는 학습

강화학습은 일종의 게임 같이 보상해주는거야



보상

AI, 머신러닝, 딥러닝



판단 모델

Discriminate model

Discriminative models discriminate between different kinds of data instances.

생성 모델

Generative model

Generative models can generate new data instances.

거대한 모델

엄청난 수의 데이터셋

Prompt Engineering

DALL-E의 학습 데이터셋 중 일부



by Joi Ito

the trail climbs steadily uphill most of the way.



by Danail Nachev

the stars in the night sky.



by Justin Higuchi

musical artist performs on stage during festival.



by Viaggio Routard

popular food market showing the traditional foods from the country.

Conceptual Caption Dataset, Google
3백만개의 이미지-텍스트 pair 데이터셋

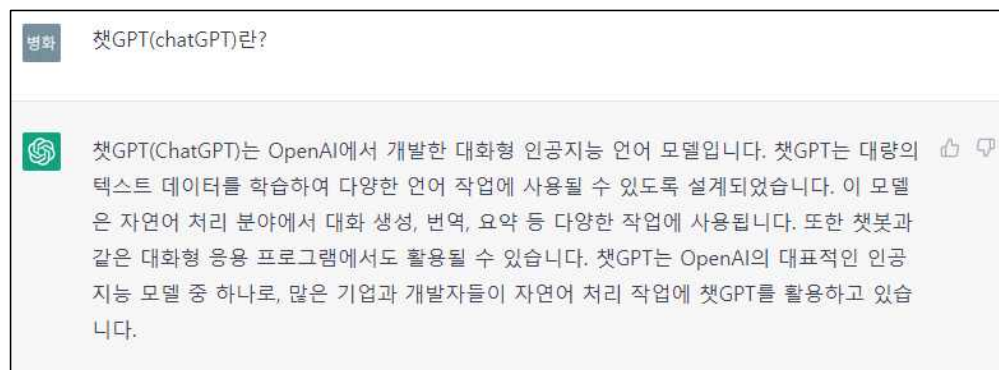
ChatGPT 소개

▣ 챗GPT(ChatGPT)란?

챗GPT(ChatGPT)는 OpenAI에서 개발한 대화형 인공지능 언어 모델입니다.

챗GPT는 대량의 텍스트 데이터를 학습하여 다양한 언어 작업에 사용될 수 있도록 설계되었습니다. 이 모델은 자연어 처리 분야에서 대화 생성, 번역, 요약 등 다양한 작업에 사용됩니다.

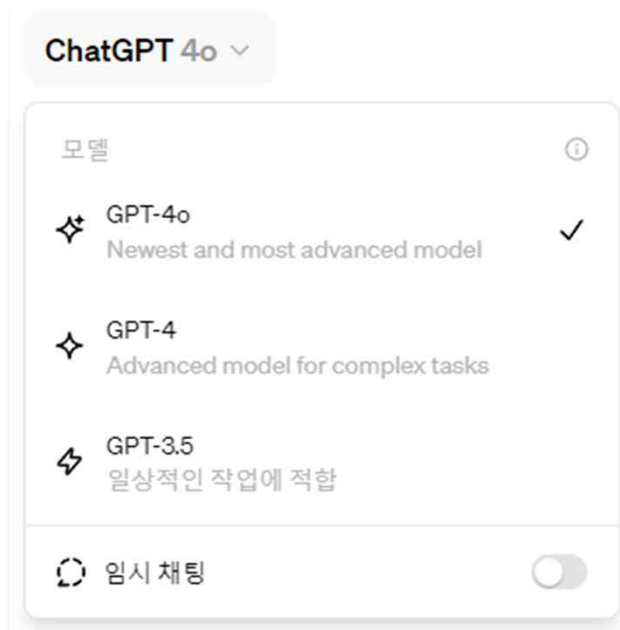
또한 챗봇과 같은 대화형 응용 프로그램에서도 활용될 수 있습니다. 챗GPT는 OpenAI의 대표적인 인공지능 모델 중 하나로, 많은 기업과 개발자들이 자연어 처리 작업에 챗GPT를 활용하고 있습니다. GPT는 Generative Pre-trained Transformer의 약자.



▣ ChatGPT 역사

- 2015년 12월 11일 샘알트만이 인공지능 연구 회사로 OpenAI 설립 : 마크 주커버그, 엘론 머스크 등의 여러 공동 투자를 받음
- 2017년 구글이 Transformer 발표
- 2018년 OpenAI에서 최초의 GPT(Generative Pre-trained Transformer) 발표 : 15.4억 개 파라미터
- 2019년 GPT-2 모델 발표 : 15.4억 개 파라미터 (1,000~4,000만원 비용 소모)
마이크로 소프트사 1조원 투자
- 2020년 GPT-3 모델 발표 : 1,750억 개 파라미터 (1200만 달러/150억원 비용 소모)
- 2022년 11월 30일 ChatGPT-3.5 발표 : 5일만에 100만명, 3개월만에 1억 사용자 달성
- 2023년 01월 마이크로 소프트사 100억 달러 투자, 파트너십
- 2023년 03월 14일 ChatGPT-4 발표 : 약 1조개 파라미터 사용 , 튜링 테스트 완료 예상
- 2024년 05월 13일 ChatGPT-4o 발표 : 멀티 모달 입력 및 출력, 반응 속도와 성능 향상

▣ ChatGPT-4o : omni



▣ GPT-1 ~ GPT-4 성능 비교



▣ GPT-1 ~ GPT-4 성능 비교

GPT-1	GPT-2	GPT-3	GPT-4
<ul style="list-style-type: none">• 12개 셀프어텐션 계층• 1억 1700만 개 파라미터	<ul style="list-style-type: none">• 48개 셀프어텐션 계층• 15억 개 파라미터	<ul style="list-style-type: none">• 96개 셀프어텐션 계층• 1750억 개 파라미터	<ul style="list-style-type: none">• 192개 셀프어텐션 계층 (추정)• 1조 개 파라미터 이상 (추정)

ChatGPT 데이터셋

- GPT-3에 사용된 데이터셋은 Common Crawl, WebText2, Books1, Books2, Wikipedia 등이며 데이터셋의 총 용량은 735GB 입니다.

Dataset	Tokens (billion)	Assumptions	Tokens per byte (Tokens / bytes)	Ratio	Size (GB)
Common Crawl (filtered)	410B	-	<i>0.71</i>	<i>1:1.9</i>	570
WebText2	19B	<i>25% > WebText</i>	<i>0.38</i>	<i>1:2.6</i>	<i>50</i>
Books1	12B	<i>Gutenberg</i>	<i>0.57</i>	<i>1:1.75</i>	<i>21</i>
Books2	55B	<i>Bibliotik</i>	<i>0.54</i>	<i>1:1.84</i>	<i>101</i>
Wikipedia	3B	<i>See RoBERTa</i>	<i>0.26</i>	<i>1:3.8</i>	<i>11.4</i>
Total	499B				<i>753.4GB</i>

Table 8. GPT-3 Datasets. Disclosed in **bold**. Determined in *italics*.

출처 : lifearchitected.ai

	Wikipedia	Books	Journals	Reddit links	CC	Other	Total
GB	<i>11.4</i>	<i>21</i>	<i>101</i>	<i>50</i>	570		753
Tokens	3	12	55	19	410		499

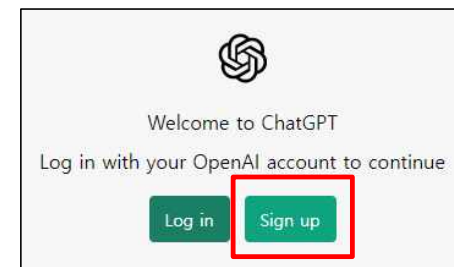
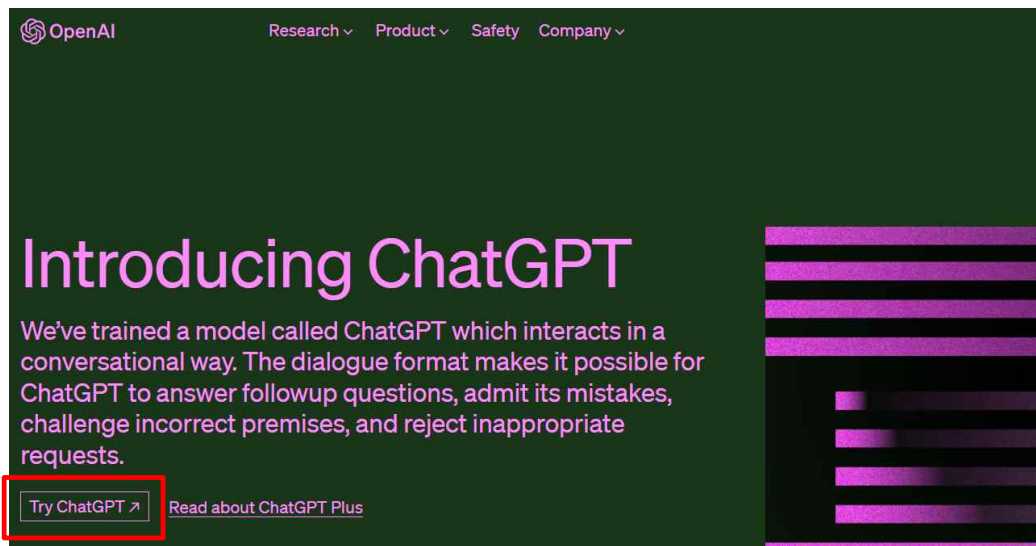
Table 9. GPT-3 Datasets Summary. Disclosed in **bold**. Determined in *italics*.

출처 : lifearchitected.ai

▣ 챗GPT(ChatGPT) 회원 가입

<https://openai.com/blog/chatgpt>

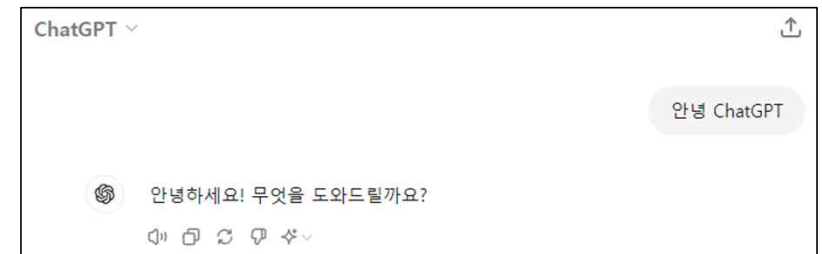
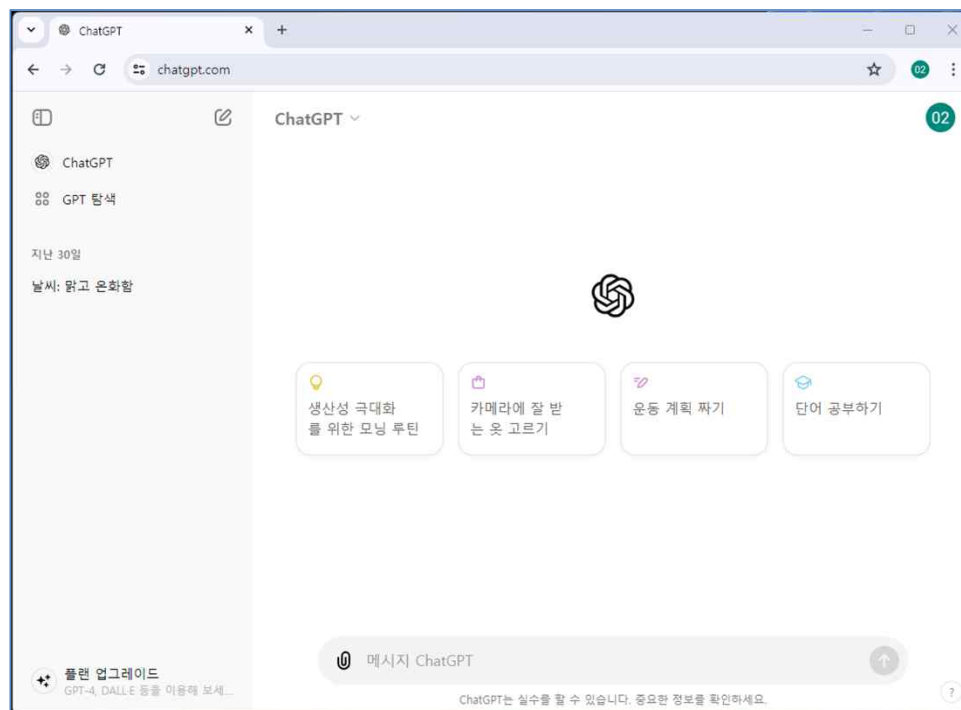
주소로 이동한 후 왼쪽 하단의 Try ChatGPT를 클릭한 sign up 버튼을 클릭한다



구글이나 마이크로소프트
계정으로 로그인 할 수 있
다

❑ 챗GPT(ChatGPT) 회원 가입

화면 하단에 질문을 입력하면 ChatGPT가 대답을 하게 된다 (한글도 지원됨)

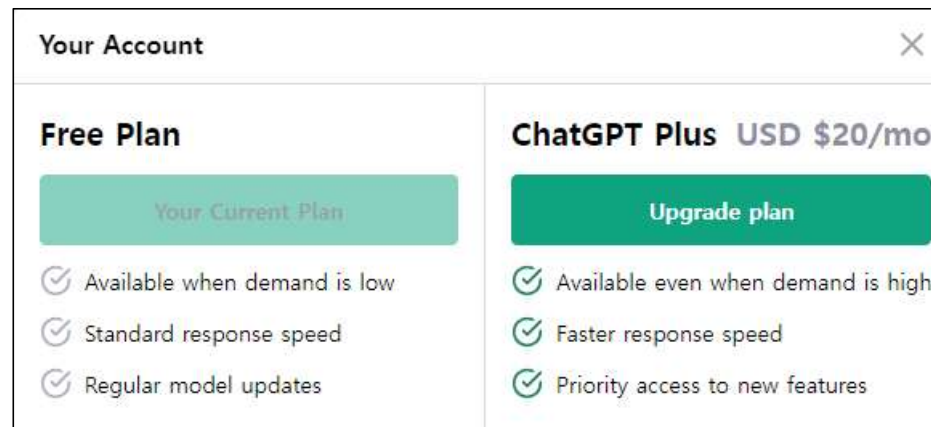
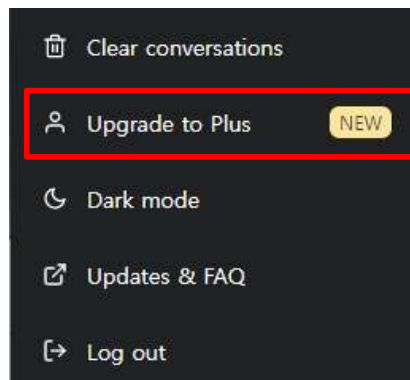


▣ 챗GPT(ChatGPT) 기능과 활용법 : 올바른 답변을 얻기 위한 팁

- 질문할 때 명확하고 구체적으로 질문하라.
- 맞춤법과 문법을 지켜라.
- 가능한 많은 맥락을 제공하라.
- 개인정보를 묻거나 공유하지 말아라.
- 답변을 기다리는 동안 인내심을 가져라.
- 적절한 언어를 사용하라.
- 복잡하거나 설명이 필요한 질문은 여러 문장으로 나누어 제시하라.
- "질문 명확화" 기능을 사용하여 추가 문맥이나 후속 질문을 할 수 있도록 하라.
- 다양한 주제를 살펴보고 필요한 정보를 찾아보라.
- 아이디어를 얻기 위해 ChatGPT를 사용하라.
- 만족스럽지 않은 답변이나 개선점이 있다면 피드백을 제공하라.
- 대화를 친근하고 예의 바르게 유지하라. 비속어나 부적절한 질문은 피하라.

▣ 챗GPT 유료 계정 사용

매월 20달러를 지불하여 **ChatGPT Plus** 유료 계정을 사용하면 좀더 빠른 응답 속도와 새로운 기능 사용과 사용자와 많은 경우에 우선적으로 사용 할 수 있다




무료 계정 사용시에 오류 메시지가 가끔씩 나온다

ChatGPT로 코딩하기

▣ 파이썬 코드 작성

챗GPT로 아래와 같이 입력하면 파이썬 코드를 작성 할 수 있다

“if문과 for문을 사용하여 1에서 1000 까지 짝수를 구하는 파이썬 코드를 만들어 주세요”



The screenshot shows a chat interface with a user prompt and a GPT response. The user prompt is: "if문과 for문을 사용하여 1에서 1000 까지 짝수를 구하는 파이썬 코드를 만들어 주세요". The GPT response includes a code block with the following Python code:

```
python
for i in range(1, 1001):
    if i % 2 == 0:
        print(i)
```


Below the code block, there is an explanation in Korean: "위 코드에서 `range(1, 1001)`은 1부터 1000까지의 수를 나타내며, `if i % 2 == 0`은 i가 짝수인지 검사합니다. 만약 i가 짝수이면 `print(i)`로 출력합니다."

▣ 파이썬 코드 작성

반복 질문 할 경우 아래와 같이 if문을 사용하지 않고 작성된 코드도 알려준다

더 간단하게 코드를 작성할 수도 있습니다. 아래와 같이 `range()`의 세번째 인자를 2로 설정하여 2씩 증가하면서 짝수만 출력하는 방법입니다.

python

 Copy code

```
for i in range(2, 1001, 2):  
    print(i)
```

위 코드를 실행하면 1부터 1000까지의 수 중에서 짝수인 수가 출력됩니다. `range(2, 1001, 2)`는 2부터 1000까지의 정수를 2씩 증가시키는 range 객체를 반환합니다. `print(i)`는 `i`를 출력합니다.

따라서 `for`문에서 `i`는 2부터 1000까지 2씩 증가하면서 실행되기 때문에 1부터 1000까지의 짝수가 출력됩니다.

▣ C코드 작성

“C언어로 바꾸어 작성해주세요”라고 요청하면 C로 작성된 코드를 보여준다

병화

C언어로 바꾸어 작성해주세요

다음은 C언어로 작성한 1에서 1000까지의 짝수를 출력하는 코드입니다.

c

Copy code

```
#include <stdio.h>

int main() {
    int i;
    for (i = 1; i <= 1000; i++) {
        if (i % 2 == 0) {
            printf("%d\n", i);
        }
    }
    return 0;
}
```

위 코드에서 `for` 문은 1부터 1000까지의 수를 반복하며, `if (i % 2 == 0)`은 i가 짝수인지 검사합니다. 만약 i가 짝수이면 `printf` 함수를 사용하여 출력합니다.



▣ 웹페이지용 HTML코드 작성

[입력 질문]

아래 정보와 HTML 을 활용해서 로그인 페이지를 만들어 주세요

1. 디자인 컨셉 : 블랙 앤 화이트 모던
2. 구조 : 아이디 입력폼, 비밀번호 입력폼이 있고 그 폼들 아래에는 아이디 찾기, 비밀번호 찾기 하이퍼 텍스트가 있음, 로그인 버튼은 입력폼 두 개의 우측에 있음
3. 배경색 : 그레이, 버튼 스타일: 약간 둥글게
4. 자바스크립트 : 로그인 버튼에 마우스 오버 시 색이 짙게 변함, 아이디는 7자 ~ 15자, 비밀번호는 5 ~ 20자 이내, 특수 문자를 반드시 하나 이상 포함해야 한다

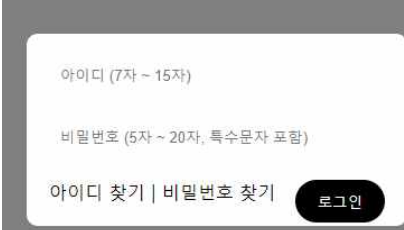
▣ 웹페이지용 HTML코드 작성 (출력 결과)

 아래는 요구사항에 맞는 HTML 코드입니다. 

```
html
Copy code

<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
  <title>Login Page</title>
  <style>
    body {
      background-color: gray;
    }
    #login-form {
      width: 300px;
      margin: 50px auto;
      background-color: white;
      border-radius: 10px;
      padding: 20px;
    }
    input[type="text"], input[type="password"] {
      width: 100px;
    }
  </style>
</head>
<body>
  <div id="login-form">
    <input type="text" value="아이디" />
    <input type="password" value="비밀번호" />
    <button type="button" value="로그인" />
  </div>
</body>
</html>
```

Copy code를 클릭하여 소스코드를 복사하여 메모장에 넣어 test.html 파일로 저장한 다음 웹 브라우저로 열어서 확인해본다



A mockup of a login form. It features a white rectangular box with rounded corners on a gray background. Inside the box, there are two input fields: the first is labeled '아이디 (7자 ~ 15자)' (ID (7 to 15 characters)) and the second is labeled '비밀번호 (5자 ~ 20자, 특수문자 포함)' (Password (5 to 20 characters, including special characters)). Below these fields is a row containing a link '아이디 찾기 | 비밀번호 찾기' (Find ID | Find Password) and a black button with the text '로그인' (Login) in white.

ChatGPT 의 효과적인 질의 전략

Chat-GPT 내부 구조

- OpenAI에서 제작
- GPT-3.5와 GPT-4라는 두 개의 다른 모델 사용
- 대화록을 저장
- 구글 Transformer의 디코더 만으로 구성된 GPT 계열
- 어텐션 메커니즘(자기주의 기제)에 집중

1. 구체적인 질문하기

일반적인 질문: "좋은 책 추천해줘"

구체적인 질문: "2023년에 출간된 역사 관련 좋은 책 추천해줘"

- **효과:** 구체적인 질문은 ChatGPT가 보다 정확한 정보를 제공하는 데 도움이 있다. 첫 번째 질문은 너무 광범위하여 다양한 답변이 나올 수 있지만, 두 번째 질문은 특정 범위를 지정하여 더 정확한 답변을 얻을 수 있다.

2. 문맥 제공하기

일반적인 질문: "프로그래밍 배우기 좋은 언어는?"

문맥 제공 질문: "웹 개발을 위해 프로그래밍 배우기 좋은 언어는?"

- **효과:** 문맥을 제공하면 ChatGPT가 질문자의 의도를 더 잘 이해할 수 있다. 첫 번째 질문은 여러 언어를 제시할 수 있지만, 두 번째 질문은 웹 개발에 적합한 언어를 추천해줄 것이다.

3. 단계별로 질문하기

일반적인 질문: "사업 시작하는 방법 알려줘"

단계별 질문:

1. "사업 아이디어를 찾는 방법은?"
2. "사업 계획서를 작성하는 방법은?"
3. "사업 자금을 조달하는 방법은?"

• **효과:** 복잡한 주제를 여러 단계로 나누면 각 단계마다 구체적인 답변을 받을 수 있다. 이는 정보를 더 체계적으로 이해하는 데 도움이 있다.

4. 키워드 사용하기

일반적인 질문: "기술 발전이 사회에 미치는 영향은?"

키워드 포함 질문: "인공지능 기술 발전이 고용 시장에 미치는 영향은?"

- **효과:** 중요한 키워드를 포함하면 ChatGPT가 질문의 핵심을 더 잘 파악할 수 있다. 첫 번째 질문은 광범위한 답변이 가능하지만, 두 번째 질문은 특정 영역에 대한 답변을 제공한다.

5. 명확하고 간결하게 작성하기

일반적인 질문: "나는 신체 활동을 좋아하지 않는데, 어떤 운동이 적합할까? 추천 좀 해줘"

명확하고 간결한 질문: "신체 활동을 싫어하는 사람에게 적합한 저강도 운동 추천해줘"

- **효과:** 명확하고 간결한 질문은 ChatGPT가 질문을 더 쉽게 이해하고 빠르게 답변을 제공하는 데 도움이 있다.

6. 피드백 제공하기

처음 질문: "한국의 전통 음식 알려줘" **피드백 제공 질문:** "좀 더 구체적으로, 서울에서 쉽게 찾을 수 있는 전통 음식 알려줘"

• **효과:** ChatGPT의 답변에 대해 추가 피드백을 주면 더 구체적이고 원하는 답변을 얻을 수 있다. 처음 질문에서 일반적인 답변을 얻었다면, 피드백을 통해 더 구체적인 정보를 요청할 수 있다.

ChatGPT 정보형 지시와 명령형 지시

ChatGPT에 대한 질의는 정보형 지시와 명령형 지시가 있다. 이 두 가지 유형의 지시는 사용자가 원하는 답변을 효과적으로 얻기 위해 서로 다른 방식으로 질문을 구성하는 방법이다.

정보형 지시 (Informative Prompt)

정보형 지시는 사용자가 특정 정보를 얻고자 할 때 사용된다. 이러한 질문은 보통 설명, 정의, 예시, 역사적 사실 등을 포함한다.

예시:

1. 설명 요청:

1. "인공지능이란 무엇인가?"

1. 답변: "인공지능(AI)은 컴퓨터 시스템이 인간 지능을 모방하여 작업을 수행할 수 있도록 설계된 기술을 의미한다. 이는 머신 러닝, 자연어 처리, 컴퓨터 비전 등을 포함한다."

2. 역사적 사실 요청:

1. "제2차 세계대전의 주요 원인은 무엇인가?"

1.답변: "제2차 세계대전의 주요 원인은 베르사유 조약의 결과로 인한 독일의 경제적, 정치적 불안정, 나치즘의 부상, 영국과 프랑스의 유화 정책, 일본의 아시아 팽창 정책 등이 있다."

3. 정의 요청:

1. "기후 변화란 무엇인가?"

1.답변: "기후 변화는 지구의 기후 패턴이 장기적으로 변하는 현상을 의미하며, 주로 온실가스 배출 증가로 인해 지구 온난화가 주요 원인으로 꼽힌다."

명령형 지시 (Imperative Prompt)

명령형 지시는 사용자가 ChatGPT에게 특정 작업을 수행하도록 지시하는 형태이다. 이러한 질문은 보통 예시 제공, 요약, 리스트 작성, 번역 등의 작업을 포함한다.

예시:

1. 예시 제공:

질의: "효과적인 이메일 작성법 예시를 보여줘"

답변: "안녕하세요 [자신의 이름]님,

저는 [자신의 이름]이며, [회사 이름]에서 [직책]으로 일하고 있습니다.

귀하의 [이메일을 받은 이유]에 대해 문의드리고자 합니다.

감사합니다.

[자신의 이름] 드림."

2. 리스트 작성:

질의: "미국의 대형 기술 회사 5개를 나열해줘"

답변: "1. Apple

2. Microsoft

3. Amazon

4. Google

5. Facebook"

3. 요약 요청:

질의: "최근 기후 변화 보고서를 요약해줘"

답변: "최근 기후 변화 보고서는 지구 온난화가 가속화되고 있으며, 해수면 상승과 극한 기후 현상이 증가하고 있다는 내용을 다루고 있다. 주요 원인은 인간의 활동으로 인한 온실가스 배출이며, 즉각적인 대응이 필요하다는 결론을 내리고 있다."

4. 번역 요청:

질의 : "이 문장을 영어로 번역해줘:

'안녕하세요, 저는 한국에서 왔습니다.'"

답변: "Hello, I am from Korea."

이와 같이 정보형 지시는 정보를 제공받고자 할 때, 명령형 지시는 특정 작업을 수행하도록 할 때 사용된다. 두 가지 형태를 적절히 활용하면 ChatGPT를 보다 효과적으로 사용할 수 있다.

ChatGPT에 다음과 같은 정보를 요청할 때는 주의해야 한다

1. 불법 정보를 요구하기

예시 요청: "해킹하는 방법을 알려줘."

관련 설명: ChatGPT는 불법 활동이나 윤리적으로 문제가 되는 행위에 대한 정보를 제공하지 않도록 설계되어 있다. 해킹, 사기, 불법 다운로드 등과 같은 불법적인 활동에 대한 정보 제공은 법적으로 문제가 될 수 있으며, 사회적으로도 바람직하지 않다.

주의사항: 불법적인 행위에 대한 정보는 절대 요구하지 않도록 한다. 이러한 정보는 법적 처벌을 받을 수 있으며, 사회적 책임을 다하는 것이 중요하다.

2. 개인 정보를 요구하기

예시 요청: "나의 은행 계좌 정보를 알려줘."

관련 설명: ChatGPT는 개인 정보 보호와 사용자의 안전을 최우선으로 한다. 따라서 사용자나 다른 사람의 개인 정보를 요구하는 질문에 대해서는 답변하지 않도록 설계되어 있다. 이러한 요청은 프라이버시 침해와 관련된 법적 문제를 일으킬 수 있으며, 사용자 데이터를 보호하는데 중요한 역할을 한다.

주의사항: 개인 정보(예: 이름, 주소, 전화번호, 은행 계좌 정보 등)는 절대 온라인 플랫폼을 통해 공유하거나 요구하지 않아야 한다. 개인 정보를 안전하게 관리하는 것이 중요하다.

3. 법률 전문 정보를 요구하기

예시 요청: "이 계약서가 법적으로 유효한지 알려줘."

관련 설명: ChatGPT는 법률 자문을 제공할 자격이 없으며, 법적 문제에 대해 구체적이고 정확한 정보를 제공할 수 없다. 법률 문제는 변호사와 같은 법률 전문가의 도움을 받아야 한다.

주의사항: 법적 문제나 계약에 관한 상담은 반드시 변호사와 같은 법률 전문가에게 문의해야 한다. ChatGPT의 답변을 법률 자문으로 사용해서는 안 된다.

4. 의료 전문 정보를 요구하기

예시 요청: "내 증상을 기반으로 질병을 진단해줘."

관련 설명: ChatGPT는 의료 전문 지식이 부족하며, 사용자에게 정확한 의료 조언을 제공할 수 없다. 의료 정보는 항상 인증된 의료 전문가에게 상담받아야 하며, ChatGPT는 기본적인 정보 제공만 가능하다.

주의사항: 건강 문제나 질병에 대한 상담은 항상 의사, 약사 등의 의료 전문가와 상의해야 한다. ChatGPT의 답변을 의료 조언으로 간주해서는 안 된다.

요약

ChatGPT는 사용자의 안전과 법적 책임을 최우선으로 하며, 다음과 같은 정보 요청에는 응답하지 않거나 제한된 정보를 제공한다:

- **불법 정보:** 불법 활동에 대한 정보 제공은 금지된다.
- **개인 정보:** 개인 정보는 보호되어야 하며, 요청하지 않아야 한다.
- **법률 정보:** 법률 자문은 법률 전문가에게 받아야 한다.
- **의료 정보:** 의료 상담은 항상 의료 전문가에게 받아야 한다.

이러한 가이드라인을 준수함으로써 사용자는 ChatGPT를 보다 안전하고 책임감 있게 사용할 수 있다.

ChatGPT의 페르소나

사용자가 원하는 특정 역할이나 스타일로 답변을 제공하도록 설정하는 것이다. 이를 통해 사용자 경험을 향상시키고, 더 적절한 답변을 얻을 수 있다.

1. 교육적인 페르소나

예시: 초등학교 교사 페르소나

사용 방법: 초등학생이 이해하기 쉬운 방식으로 설명이 필요할 때

질문: "초등학생이 알아들을 수 있게 중력에 대해 설명해줘."

페르소나 설정: "당신은 초등학교 교사입니다. 중력을 초등학생이 이해하기 쉽게 설명해 주세요."

답변: "안녕, 어린이들! 중력은 지구가 모든 물체를 끌어당기는 힘이야. 우리가 뛰어오르면 다시 땅으로 떨어지는 이유가 바로 중력 때문이란단다. 중력 덕분에 우리가 땅에 서 있을 수 있고, 물건들이 공중에 떠다니지 않게 되는 거야."

2. 전문가 페르소나

예시: IT 전문가 페르소나

사용 방법: 기술적인 질문에 대한 상세한 답변이 필요할 때

질문: "클라우드 컴퓨팅이란 무엇인가요?"

페르소나 설정: "당신은 IT 전문가입니다. 클라우드 컴퓨팅에 대해 자세히 설명해 주세요."

답변: "클라우드 컴퓨팅은 인터넷을 통해 서버, 저장소, 데이터베이스, 네트워킹, 소프트웨어 등의 컴퓨팅 자원을 제공하는 서비스입니다. 이는 사용자가 물리적인 장비를 직접 구매하거나 관리할 필요 없이, 필요할 때마다 리소스를 사용할 수 있게 해줍니다. 클라우드 컴퓨팅의 주요 장점은 확장성, 유연성, 비용 절감입니다."

3. 친근한 대화 페르소나

예시: 친구 페르소나

사용 방법: 친근하고 편안한 대화가 필요할 때

질문: "오늘 하루가 힘들었어. 위로가 필요해."

페르소나 설정: "당신은 친구입니다. 친구에게 위로의 말을 해 주세요."

답변: "오, 정말 힘든 하루였구나. 그런 날도 있는 법이야. 그래도 이렇게 잘 버텨줘서 정말 대단해! 오늘은 맛있는 거 먹고 푹 쉬어. 내일은 더 나은 하루가 될 거야. 언제든지 내게 털어놔도 돼."

4. 코치 페르소나

예시: 생활 코치 페르소나

사용 방법: 동기부여나 생활 습관 개선에 대한 조언이 필요할 때

질문: "새로운 습관을 어떻게 형성할 수 있을까?"

페르소나 설정: "당신은 생활 코치입니다. 새로운 습관 형성에 대한 조언을 해 주세요."

답변: "새로운 습관을 형성하는 첫 번째 단계는 작은 목표를 세우는 거야. 예를 들어, 매일 10분 동안 걷는 것으로 시작해봐. 그런 다음 이를 매일 반복하면서 조금씩 시간을 늘려가면 돼. 또한, 습관 형성에 있어 일관성이 매우 중요해. 매일 같은 시간에 같은 행동을 반복하면 점차 습관이 형성될 거야. 마지막으로, 자신에게 보상을 주는 것도 좋은 방법이야. 목표를 달성했을 때 스스로에게 작은 보상을 주면 동기 부여가 될 수 있어."

ChatGPT의 환각현상(hallucination)

모델이 실제로는 존재하지 않거나 부정확한 정보를 생성하는 현상을 말한다. 이러한 현상은 자연어 처리 모델에서 발생할 수 있는 중요한 문제 중 하나이다.

정의

환각현상(hallucination): 언어 모델이 실제로는 존재하지 않거나 부정확한 정보를 생성하는 현상이다. 예를 들어, 특정 사실이나 데이터를 질문했을 때, 모델이 자신 있게 잘못된 정보를 제공하는 경우가 이에 해당한다.

환각 현상의 원인

- 1. 훈련 데이터의 한계:** 모델이 학습한 데이터가 불완전하거나 부정확할 경우, 잘못된 정보를 생성할 가능성이 높다.
- 2. 모델의 확률적 특성:** 언어 모델은 입력에 대한 확률적인 응답을 생성하기 때문에 항상 정확한 정보를 제공하지 않을 수 있다.
- 3. 정보의 최신성 부족:** 모델이 훈련된 후 업데이트되지 않은 정보에 대해 질문을 받을 경우, 부정확한 답변을 생성할 수 있다.
- 4. 복잡한 질문:** 너무 복잡하거나 명확하지 않은 질문에 대해 모델이 적절한 답변을 찾지 못하고 추측성 답변을 할 수 있다.

대응 전략

1. 질문 구체화: 구체적이고 명확한 질문을 통해 모델이 정확한 정보를 제공할 수 있도록 한다.

예: "세계에서 가장 높은 산은?" 대신 "2023년 기준 세계에서 가장 높은 산은 무엇인가?"

2. 출처 확인: 모델이 제공한 정보의 출처를 확인하고, 중요한 정보일 경우 추가 검증을 한다.

예: 모델이 제시한 사실을 신뢰할 수 있는 웹사이트나 출판물을 통해 확인한다.

3. 단계적 질문: 복잡한 질문은 단계별로 나누어 묻는다.

예: "한국의 경제 성장과 관련된 최근 뉴스는?" 대신 "2023년 한국의 경제 성장률은 얼마인가?"와 "최근 한국 경제에 영향을 미친 주요 사건은 무엇인가?"로 나누어 질문한다.

4. 모델 업데이트: 최신 정보를 반영할 수 있도록 모델을 정기적으로 업데이트한다.

이는 OpenAI와 같은 모델 개발사에서 주로 수행하며, 사용자는 최신 버전을 사용하도록 한다.

5. 피드백 제공: 모델의 응답에 대해 피드백을 제공하여 개선에 기여한다.
예: "이 정보가 부정확합니다" 또는 "더 구체적인 답변이 필요합니다"
와 같은 피드백을 통해 모델 개선에 도움을 준다.

6. 외부 검증 도구 활용: 중요한 정보는 외부 검증 도구나 전문가의 의견을 통해 확인한다.

예: 중요한 의학 정보나 법률적 조언은 전문가의 검토를 받는다.

이와 같은 대응 전략을 활용하면 ChatGPT의 환각현상을 줄이고 보다 정확한 정보를 얻을 수 있다.

감사합니다