| **(1) 프로젝트명** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 자연어 데이터를 활용한 능동형 챗봇 만들기 |  |  |  |  |  |  |  |  |
| **(2) 프로젝트 목표** | | | | | | | | |
| 제공된 데이터를 모델에 알맞은 형태로 변환, 적용하여 chatbot service program을 구성한다. |  |  |  |  |  |  |  |  |
| **(3) 프로젝트 개발스펙** | | | | | | | | |
| - pythonproject\_data\_카카오톡채널.txt - gpt-3.5-turbo - pycharm |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  | | | |  |  |  |  |
| **(4) 학습 콘텐츠 (학습할 기술)** | | | | | | | | |

**[1단계]**

1. 데이터 파일(project\_data\_카카오톡채널.txt)을 로딩하여 데이터를 생성한다

2. 수집된 데이터를 정형화 시킨다.

3. chatGPT api를 이용하여 질의 응답이 가능한 모델을 구성한다.

4. 질의는 prompt Engineering을 이용하여 효율성을 높인다.



**[2단계]**

1. 제공된 데이터(카카오싱크.txt)는 LLM library에 이용한다.

2. prompt engineering 기법을 이용하여 chatGPT api에 해당 데이터를 적용한다.

3. LLM + prompt engineering은 LangChain library를 이용하여 구성한다.

4. 여러 반복을 통해 카카오싱크 api 사용법을 잘 설명 해주는 챗봇을 완성한다.



예시)

User : 카카오싱크 기능이 무엇이 있는지 설명해주세요

Assistance :

1. 간편 가입: 동의 화면에서 서비스 약관까지 한 번에 동의 받을 수 있습니다.
2. 더 다양한 사용자 정보 활용: 서비스 회원 가입 시 필요한 다양한 사용자 정보를 제공받을 수 있습니다. 이름, 이메일, 전화번호, 연령대, 생일, 성별, 출생연도, 배송지 등 정보를 제공합니다.

**[3단계]**

1. 비정형 데이터를 정형화 한다.

2. 정형화된 데이터를 token화 한다.

3. 사용할 데이터를 preprocessing을 한다.

4. Embedding 기법을 사용하여 데이터를 벡터화 한다.

5. 벡터 데이터를 이용하여 모델을 학습한다.

6. 생성된 모델을 서비스 형태로 구성한다.

7. 서비스 형태는 web application(next.js, node.js, django, spring, asp등등)과 view programming을 이용하여 구현한다.



5. UI 구현 사항

예시를 참고로 자유롭게 구현한다.

1. 화면 구성
   1. User와 Assistance 부분으로 구분한다. 이때 User와 Assistances는 출력에 구분이 되도록 아래의 예시를 참고하여 구성한다.
   2. 처음 시행 시 챗봇 서비스 인사를 기본으로 구성한다.
   3. 실행은 User와 질문과 Assistance을 대답으로 진행한다.

