# T크미. CHATBOT 프로젝트

자갈자갈 석은수 손지현 이소정

## Contents 목本

- 1. 프로젝트 구축 개요
- 2. 프로젝트 구축 범위
- 3. 프로젝트 추진 방법론
- 4. 프로젝트 조직 및 역할
- 5. 프로젝트 일정
- 6. 예상이슈

#### 1. 프로젝트 구축 개요

배경

1인가구, 개인주의 증가

→ 우울증 등이 사회문제 대두

- 챗봇을 활용한 고객 응대 사이트 증가
- 기업에서 고객서비스, 교육 등다양한 분야에서 사용 가능

기간

2023.05.01 ~ 2023.06.15

#### 목적

- 가벼운 대화
- 위로, 조언 등 감정적 교류



• 적절한 QA 모델 구축



#### 2. 프로젝트 구축 범위

본 프로젝트는 Al Hub의 "주제별 텍스트 일상 대화 데이터"를 사용합니다.

#### 소스 데이터

- 주제별 텍스트 일상 대화 데이터
- 134,263 건
- 2022년 최신자료

<u>주제별 텍스트 일상 대화 데이터</u>

#### 구축범위

- 기존 Q&A 데이터 활용
- 즉각적 응답 가능

#### 기대효과

- 기존 Q&A 데이터 활용
- 즉각적 응답 가능

## 3. 프로젝트 추진 방법론

	분석 기획	데이터 준비	데이터 분석	시스템 구현	평가 및 전개
주요 Task	<ul> <li>프로젝트 정의</li> <li>계획 수립</li> <li>Kick-off 회의</li> <li>프로젝트 세부 일정 계획 수립</li> </ul>	<ul> <li>필요데이터정의</li> <li>데이터 수집</li> </ul>	<ul> <li>탐색적 분석</li> <li>모델링(딥러닝)</li> </ul>	• 플랫폼 업로드	<ul> <li>모델 발전 계획 수립</li> <li>프로젝트 평가/보고</li> </ul>
주요 산출물	<ul><li>요구사항정의서</li><li>프로젝트계획서</li></ul>		<ul> <li>분석용데이터셋</li> <li>데이터 탐색 보고서</li> <li>데이터 시각화 보고서</li> <li>모델링 결과 보고서</li> </ul>		• 최종보고서
					Ę.

#### 3-1. 프로세스 구성도



## 3-2. 프로젝트 구성도



#### 3-3. 프로젝트 설계도

운영체제 Window 11

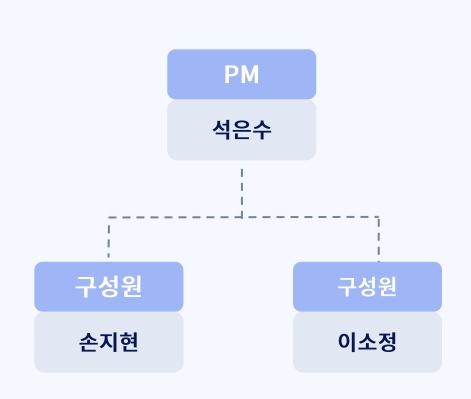
개발툴 Python 3.9

프레임워크 PyTorch, transformer

## 4. 프로젝트 조직 및 역할

프로젝트 조직도

#### 인력별 역할



No.	담당자	역할
프로젝트 관리	석은수	프로젝트 구축 준비
데이터 수집	손지현, 석은수	데이터 수집 데이터셋 구성
분석/설계	이소정, 손지현	데이터 전처리 EDA
개발	이소정	모델링
테스트	이소정	모델 테스트 진행 및 평가
문서작성	전원	회의록, 보고서 등 문서 작성

#### 5. 프로젝트 일정



이미지: flaticon.com

## **5-1.** 프로젝트 세부 일정 Gantt Chart

#### **GANTT CHART**

 T3Q
 첫봇

 자연어팀
 설우수, 손지현, 이소정
 날짜
 23년 5월 1일

WBS 번호	작업 제목	시작일	마감일	기간	1주																							
WD3 인모	역합 세국	시작물	4162	기선	월 화	수목	금	월 화	수	목 금	월	한 수	목	금 월	화	수	목 금	월	화 수	누 목	금	월 :	화 수	목	금	릴 화	수	목 금
1	프로젝트 기획																											
1.1	비즈니스 이해 및 범위 설정	23년 5월 3일	23년 5월 10일	6																								
1.2	프로젝트 정의 및 계획 설정	23년 5월 3일	23년 5월 10일	6																								
2	데이터 준비																											
2.1	필요 데이터 정의	23년 5월 11일	23년 5월 12일	2																								
2.2	데이터 수집 및 정합성 검증	23년 5월 12일	23년 5월 23일	8																								
2.3	분석용 데이터 준비	23년 5월 15일	23년 5월 23일	7																								
3	모델 구성 및 학습																											
3.1	모델 탐색	23년 5월 3일	23년 5월 19일	13																								
3.2	모델링	23년 5월 17일	23년 6월 7일	16																								
3.3	모델평가 및 검증	23년 5월 25일	23년 6월 7일	10																								
4	플랫폼 업로드																											
4.1	플랫폼 업로드 교육	23년 6월 5일	23년 6월 9일	5																								
4.2	플랫폼 업로드	23년 6월 7일	23년 6월 13일	5																								
5	평가 및 전개																											
5.1	모델 발전계획 수립	23년 6월 8일	23년 6월 14일	5																								
5.2	프로젝트 평가	23년 6월 8일	23년 6월 14일	5																								
6	발표																											
6.1	기업프로젝트 발표	23년 6월 15일	23년 6월 16일	2																								1

## 5-1. 프로젝트 세부 일정 wbs

작업 제목 작업 소유자 결과 시작일 마라일 기간 작업 완료 비율 1구 2주 3 2 3 4 4 3 3 4 4 3 3 5 4 8 3 3 5 4 8 3 3 5 4 8 3 3 5 4 8 3 3 5 4 8 3 3 5 5 8 3 5 5 8 3 5 5 8 3 5 8 5 8	4주 5주 6주 월 화 수 목 금 월 화 수 목 금 월 화 수 목 금
1 사전 공통작업 담당 산출물 일수  1.1 간트차트 작성 손지현 간트차트 23년 5월 9일 23년 5월 9일 1 100%  1.1.1 WBS 작성 석은수 WBS 23년 5월 9일 23년 6월 14일 36 15%  1.2 계획서 작성 손지현 계획서 23년 5월 8일 23년 5월 9일 2 100%  1.3 프로젝트 계획서 ppt작성 이소정 PPT 23년 5월 9일 23년 5월 10일 2 100%  1.4 개발 환경 셋팅 팀원전체 23년 5월 9일 23년 5월 9일 1 100%  1.5 노선 공유페이지 작성 이소정 노선 23년 5월 6일 23년 5월 10일 6 100%  1.6 계획보고서 노선 작성 팀원전체 노선 23년 5월 6일 23년 5월 10일 6 100%  2 기획  2.1 비즈니스 이해 및 범위 설정 석은수 23년 5월 5일 23년 5월 10일 6 100%  3 대이터  3 대이터 준비 석은수 데이터파일 23년 5월 9일 1 100%  3 필요 데이터 전의 석은수 데이터파일 23년 5월 12일 2 0%  3 일요 필요 데이터 전의 이소정 데이터 정의서 23년 5월 11일 2 3년 5월 12일 2 0%  3 일요 데이터 주집 손지현 23년 5월 11일 2 3년 5월 14일 3 0%  3 대이터 주집 삼은수 23년 5월 18일 3 0%	월 화 수 목 금 월 화 수 목 금 월 화 수 목 금
1.1 간트차트 작성 손지현 간트차트 23년 5월 9일 23년 5월 9일 1 100%  1.1.1 WBS 작성 석은수 WBS 23년 5월 9일 23년 6월 14일 36 15%  1.2 계획서 작성 손지현 계획서 23년 5월 8일 23년 5월 9일 2 100%  1.3 프로젝트 계획서 ppt작성 이소정 PPT 23년 5월 9일 23년 5월 10일 2 100%  1.4 개발 환경 셋팅 팀원전체 23년 5월 9일 23년 5월 9일 1 100%  1.5 노선 공유페이지 작성 이소정 노선 23년 5월 9일 23년 6월 15일 40 10%  1.6 계획보고서 노션 작성 팀원전체 노션 23년 5월 5일 23년 6월 15일 40 10%  2 기획  2 기획  2 1 비즈니스 이해 및 범위 설정 석은수 23년 5월 9일 23년 5월 9일 5 100%  3 대이터 전비 석은수 데이터파일 23년 5월 9일 23년 5월 9일 1 100%  3 대이터 준비 석은수 데이터파일 23년 5월 11일 23년 5월 12일 2 0%  3 급인 데이터 전비 석은수 데이터파일 23년 5월 11일 23년 5월 12일 2 0%  3 대이터 준비 석은수 데이터 정의서 23년 5월 11일 23년 5월 12일 2 0%  3 대이터 주비 석은수 데이터 정의서 23년 5월 12일 23년 5월 14일 3 0%  3 대이터 주집 손지현 23년 5월 16일 3 0%  3 대이터 주집 수지현 23년 5월 16일 3 0%	
1.1.1 WBS작성 석은수 WBS 23년 5월 9일 23년 6월 14일 36 15% 12 개획서 작성 손지현 계획서 23년 5월 8일 23년 5월 9일 2 100% 13 프로젝트 계획서 ppt작성 이소정 PPT 23년 5월 9일 23년 5월 9일 2 100% 14 개발 환경 셋팅 팀원전체 23년 5월 9일 23년 5월 9일 1 100% 15 노선 공유테이지 작성 이소정 노선 23년 5월 6일 23년 6월 15일 40 10% 16 계획보고서 노선 작성 팀원전체 노선 23년 5월 6일 23년 6월 15일 40 10% 16 계획보고서 노선 작성 팀원전체 노선 23년 5월 5일 23년 5월 10일 6 100% 17월 17월 17일	
1.2 계획서 작성 손지현 계획서 23년 5월 8일 23년 5월 9일 2 100% 100% 100% 11.3 프로젝트 계획서 ppt 작성 이소정 PPT 23년 5월 9일 23년 5월 10일 2 100% 11.4 개발 환경 셋팅 팀원전체 23년 5월 9일 23년 5월 9일 1 100% 10% 10% 10% 10% 10% 10% 10% 10%	
1.3     프로젝트 계획서 ppt 작성     이소정     PPT     23년 5월 9일     23년 5월 9일     23년 5월 9일     1     100%       1.4     개발 환경 셋팅     팀원전체     23년 5월 9일     23년 5월 9일     1     100%       1.5     노선 공유페이지 작성     이소정     노선     23년 5월 6일     23년 5월 15일     40     10%       1.6     계획보고서 노션 작성     팀원전체     노선     23년 5월 5일     23년 5월 10일     6     100%       2     기획       2.1     비즈니스 이해 및 범위 설정 석은수     23년 5월 5일     23년 5월 9일     5     100%       2.2     프로젝트 정의 및 계획 설정 석은수, 손지현     23년 5월 9일     23년 5월 9일     1     100%       3.1     데이터 준비     석은수     데이터파일     23년 5월 11일     23년 5월 12일     2     0%       3.2     필요 데이터 정의     이소정     데이터 정의서     23년 5월 14일     23년 5월 16일     3     0%       3.2.1     데이터 작합성 검증     이소정, 석은수     23년 5월 16일     23년 5월 16일     3     0%       3.2.2     데이터 적합성 검증     이소정, 석은수     23년 5월 16일     23년 5월 16일     3     0%	
1.4 개발 환경 셋팅 팀원전체 23년 5월 9일 23년 5월 9일 1 100% 10% 155 노선 공유페이지 작성 이소정 노선 23년 5월 6일 23년 6월 15일 40 10% 10% 166 계획보고서 노선 작성 팀원전체 노선 23년 5월 5일 23년 5월 10일 6 100% 10% 10% 10% 10% 10% 10% 10% 10% 10	
1.5 노선 공유페이지 작성 이소정 노선 23년 5월 6일 23년 6월 15일 40 10% 10% 166 계획보고서 노선 작성 팀원전체 노선 23년 5월 5일 23년 5월 10일 6 100% 2 2 기획 2.1 비즈니스 이해 및 범위 설정 석은수 23년 5월 5일 23년 5월 9일 5 100% 2 2 프로젝트 정의 및 계획 설정 석은수, 손지현 23년 5월 9일 23년 5월 9일 1 100% 2 2 2 프로젝트 전의 및 계획 설정 석은수, 손지현 23년 5월 9일 23년 5월 9일 1 100% 2 2 3년 5월 9일 2 3년 5월 9일 1 100% 2 2 3년 5월 9일 2 3년 5월 9일 1 100% 2 2 3년 5월 9일 1 100% 2 2 3년 5월 9일 1 100% 2 3년 5월 12일 2 3년 5월 12일 2 2 0% 2 2 3년 5월 12일 2 3년 5월 12일 2 3년 5월 14일 3 0% 2 2 3년 5월 14일 3 3 2 3년 5월 14일 3 2 3년 5월 14일 3 3 2 3년 5월 14일	
1.6 계획보고서 노션 작성 팀원전체 노션 23년 5월 5일 23년 5월 10일 6 100%  2 기획  2.1 비즈니스 이해 및 범위 설정 석은수 23년 5월 5일 23년 5월 9일 5 100%  2.2 프로젝트 정의 및 계획 설정 석은수, 손지현 23년 5월 9일 1 100%  3 데이터  3.1 데이터 준비 석은수 데이터파일 23년 5월 11일 23년 5월 12일 2 0%  3.2 필요 데이터 정의 이소정 데이터 정의서 23년 5월 12일 23년 5월 14일 3 0%  3.1 데이터 수집 손지현 23년 5월 14일 23년 5월 16일 3 0%  3.2.1 데이터 주집 수지현 23년 5월 14일 23년 5월 16일 3 0%  3.2.2 데이터 적합성 검증 이소정, 석은수 23년 5월 16일 23년 5월 18일 3 0%	
2     기획       2.1     비즈니스 이해 및 범위 설정 석은수     23년 5월 5일     23년 5월 9일     5     100%       2.2     프로젝트 정의 및 계획 설정 석은수, 손지현     23년 5월 9일     23년 5월 9일     1     100%       3     데이터       3.1     데이터 준비     석은수     데이터파일     23년 5월 11일     23년 5월 12일     2     0%       3.2     필요 데이터 정의     이소정     데이터 정의서     23년 5월 12일     23년 5월 14일     3     0%       3.2.1     데이터 주집     손지현     23년 5월 14일     23년 5월 16일     3     0%       3.2.2     데이터 적합성 검증     이소정, 석은수     23년 5월 16일     23년 5월 18일     3     0%	
2.1 비즈니스 이해 및 범위 설정 석은수 23년 5월 5일 23년 5월 9일 5 100% 2.2 프로젝트 정의 및 계획 설정 석은수, 손지현 23년 5월 9일 1 100% 100% 3.1 데이터 준비 석은수 데이터파일 23년 5월 11일 23년 5월 12일 2 0% 3.2 필요 데이터 정의 이소정 데이터 정의서 23년 5월 12일 23년 5월 14일 3 0% 3.2.1 데이터 수집 손지현 23년 5월 14일 23년 5월 16일 3 0% 3.2.2 데이터 적합성 검증 이소정, 석은수 23년 5월 16일 23년 5월 18일 3 0% 3.2.2	
2.2     프로젝트 정의 및 계획 설정 석은수, 손지현     23년 5월 9일     23년 5월 9일     1     100%     100%       3     데이터       3.1     데이터 준비     석은수     데이터파일     23년 5월 11일     23년 5월 12일     2     0%     0%     0%       3.2     필요 데이터 정의     이소정     데이터 정의서     23년 5월 14일     23년 5월 14일     3     0%     0%     0%       3.2.1     데이터 주집     손지현     23년 5월 16일     23년 5월 16일     3     0%     0%     0%       3.2.2     데이터 적합성 검증     이소정, 석은수     23년 5월 16일     23년 5월 16일     3     0%     0%	
3.1 데이터 준비     석은수     데이터파일     23년 5월 11일     23년 5월 12일     2     0%     0%       3.2 필요 데이터 정의     이소정     데이터 정의서     23년 5월 12일     23년 5월 14일     3     0%       3.2.1 데이터 수집     손지현     23년 5월 14일     23년 5월 16일     3     0%       3.2.2 데이터 적합성 검증     이소정, 석은수     23년 5월 16일     23년 5월 18일     3     0%	
3.1 데이터 준비 석은수 데이터파일 23년 5월 11일 23년 5월 12일 2 0% 3.2 필요 데이터 정의 이소정 데이터 정의서 23년 5월 12일 23년 5월 14일 3 0% 3.2.1 데이터 수집 손지현 23년 5월 14일 23년 5월 16일 3 0% 3.2.2 데이터 적합성 검증 이소정, 석은수 23년 5월 16일 23년 5월 18일 3 0%	
3.2     필요 데이터 정의     이소정     데이터 정의서     23년 5월 12일     23년 5월 14일     3     0%       3.2.1     데이터 수집     손지현     23년 5월 14일     23년 5월 16일     3     0%       3.2.2     데이터 적합성 검증     이소정, 석은수     23년 5월 16일     23년 5월 18일     3     0%	
3.2.1     데이터 수집     손지현     23년 5월 14일     23년 5월 16일     3     0%       3.2.2     데이터 적합성 검증     이소정, 석은수     23년 5월 16일     23년 5월 18일     3     0%	
3.2.2 데이터 적합성 검증 이소정, 석은수 23년 5월 16일 23년 5월 18일 3 0%	
3.3 분석용 데이터 선정 손지현 데이터파일 23년 5월 18일 23년 5월 19일 2 0%	
3.4 데이터 전처리 팀원전체 23년 5월 19일 23년 5월 23일 5 0%	
4 모델	
4.1 모델 선정 손지현 모델파일 23년 5월 17일 23년 5월 20일 4 0%	
4.2 모델 학습 이소정 23년 5월 21일 23년 6월 1일 13 0%	
4.3 검증 및 평가 석은수 23년 6월 2일 23년 6월 7일 6 0%	
5 평가 및 보고서 작성	
5.1 프로젝트 평가 팀원전체 23년 6월 8일 23년 6월 14일 7 0%	
5.2 보고서 작성 석은수 <u>완료 보고서,</u> 23년 6월 7일 23년 6월 13일 7 0%	
6 플랫폼 및 발표	
6.1 플랫폼 업로드 팀원전체 h5파일 23년 6월 7일 23년 6월 13일 7 0%	
6.2 발표 석은수 23년 6월 15일 23년 6월 15일 1 0%	12

## 6. 예상 이슈

시스템 구축 과정에서 아래와 같은 3가지 이슈가 예상되며, 대응방안은 다음과 같습니다.

No.	예상이슈	대응방안
1	데이터 수집 시 저작권을 침해	라이센스 검토 및 구입
2	챗봇의 인종, 성차별적인 윤리적 문제	양질의 데이터 수집, 데이터의 다양성 증가
3	낮은 성능	모델변경, 데이터 전처리 재검토, 파인튜닝

#