



세종대학교
SEJONG UNIVERSITY



세종대학교 QA 데이터 수집

20년 5월	다인 네트 가드 네 마들 되레전	0
20년 5월	다인 네트 가드 네 마들 되레전	0
bbt 위하네	핀웨이 피하네이조	0
3년 4월 5일 10시 30분	여름을 되레전 올조	0
3년 4월 10시 30분	여름을 되레전 올조	0
4년 5월 10시 30분	여름을 되레전 올조	0
4년 5월 10시 30분	여름을 되레전 올조	0
0	4	1994

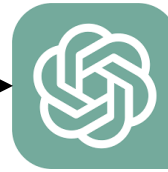
CusiporData



일상 대화 데이터 수집

Prompt Tuning

Text
Data



GPT3.5 turbo api 활용 데이터 생성

- GPT-3.5의 Q&A 생성 능력
- GPT 프롬프트 튜닝을 통해 수집한 텍스트에 대한 질문 + 답변 쌍 생성
- GPT의 hallucination 문제, 생성된 Q&A 쌍 직접 검수

동아리
질문

~어떤 활동을 해
~어떤 동아리야?
~정보 소개 해줘

질문 포맷 지정

‘세종대학교’
‘교양선택’
‘컴퓨터공학과’

‘세종대’
‘교선’
‘컴공’

유사어 / 준말 어휘 증강

22학번 영어졸업
인증 기준

영어졸업인증
22학번 기준

문장 내 단어 랜덤 스왑

최종 데이터

SejongQA data	39609 rows
Chit-chat data	81986 rows