

프로젝트 최종 보고서

	사용자 맞춤형 RAG 기반 챗봇 서비스 보고서
서비스명 및 개요	<p>서비스명: 기업 맞춤형 QA 챗봇 서비스 개발</p> <p>서비스 개요: 많은 취업준비생들이 기업과 직무에 대한 정보를 인터넷 검색을 통해 얻는다. 원하는 정보를 찾으려면 해당 기업의 홈페이지에 일일이 찾아, 원하는 직무의 직무기술서를 탐색해야 한다. 이를 해결하기 위해 기업의 직무기술서 데이터를 검색하여 이를 바탕으로 사용자들의 질문에 대한 정확한 답변을 생성하는 QA챗봇 서비스를 설계한다. 이 서비스는 사용자들이 기업과 직무에 관해 탐색하는 비용을 단축하는 데 기여한다. 취업준비생들을 주요 사용자로 하며, 사용자가 직접 챗봇을 통해 정보를 요청하고 원하는 정보를 얻을 수 있다.</p>
목표 및 기대효과	<p>서비스 목표:</p> <ul style="list-style-type: none"> - 취업 준비생들이 직무와 기업에 대해 효과적으로 분석할 수 있음. - 사용자들이 질문을 통해 자신이 원하는 기업과 직무에 필요한 지식 및 경험 등의 정보를 제공받을 수 있음 - 자신이 보유한 역량과 경험이 어떤 직무에 적합한지에 대한 정보를 제공받을 수 있음. <p>기대효과:</p> <ul style="list-style-type: none"> - 정보 탐색에 드는 비용을 줄이고 직무/기업에 대해 분석할 수 있음. - 기업인사팀 입장에서 직무/기업 관련 직접적 문의를 처리하는 비용을 감소시킬 수 있음.
타겟 사용자 및 시장 분석	<p>타겟 사용자:</p> <ul style="list-style-type: none"> - 취업을 준비하는 사람들 중 자신이 가진 역량과 직무 역량이 맞는지 궁금한 사람 - 원하는 기업과 직무가 있지만 뚜렷한 관련 정보가 부족한 사람 <p>시장 분석: 이 서비스는 일반 챗봇과 달리, 검색과 생성 모델을 결합하여 보다 정확하고 유연한 답변을 제공한다는 점에서 차별성을 갖습니다. 취업준비생의 니즈에 적합한 정보를 빠르게 제공하기에 검색 효율성과 답변 생성의 품질을 동시에 충족시킨다는 점에서 강력한 경쟁력을 가집니다.</p>
데이터 구성 및 처리 방법	<p>원천데이터 소스 : 삼성 그룹 직무기술서</p> <p>원천데이터 형식 : PDF 문서 파일</p> <p>원천데이터 처리 방법:</p> <ul style="list-style-type: none"> • 데이터 수집 : 웹 크롤링을 통해 삼성 그룹 직무기술서 데이터 확보 • 데이터 처리 : Document Parse를 활용하여 PDF파일을 HTML 형태의 텍스트 데이터로 변환 테스크 유형에 맞게 기업/직무 관련 질문-답 쌍으로 구성된 QA 데이터셋으로 변환 <p>삼성 그룹 직무기술서 PDF를 구글 페이지에서 크롤링하여, Upstage Document Parser API를 활용해 Markdown형태의 텍스트로 변환하였습니다. 이후, Markdown 문법을 정제하여 순수 텍스트 데이터로 변환하였습니다. 또한, PDF의 파일명을 이용하여 “제품이름”-PDF변환텍스트데이터”의 Pair를</p>

JSONL형식으로 구성하였으며, 이를 기반으로 QA데이터셋 생성에 필요한 기초 데이터를 준비하였습니다.

QA데이터 처리 방법:

- 16개의 변환된 직무기술서 데이터에 Upstage Solar LLM을 사용하여 QA Pair 생성
- QA Pair 데이터 생성 프롬프트:

```
system_prompt = """{document_title} 중 한 페이지를 받게된다. 이를 읽고 사용자가 질문할 만한 질문과 이에 대한 답변을 주어진 정보를 기반으로 만들어라.
단 아래의 JSON format에 맞게 응답해야한다.

(주어진 정보가 질문하기에 충분하지 않다면 빈 list를 반환하라)
질문과 답변은 한국어로 작성해야 한다.

질문과 답변은 다음과 같은 카테고리로 분류된다:
1. '직무분석' - Role, 직무가 하는 일에 관련된 질문
2. '기업분석' - 기업과 관련된 질문
3. '필요역량' - Requirements, 직무 지원을 위해 필요한 역량, 기술과 관련된 질문
4. '우대역량' - Pluses, 직무 채용시 우대되는 역량, 기술과 관련된 질문

```JSON
[
 {{
 "question": str, # 사용자의 질문
 "answer": str, # 질문에 대한 응답
 "category": str # 질문의 카테고리 ('직무분석', '기업분석', '필요역량', '우대역량', ...)
 }}
]
```"""

user_prompt = """직무 공고 : {page_content}"""

prompt_template = ChatPromptTemplate([
    ("system", system_prompt),
    ("user", user_prompt)
])
```

- QA 데이터셋의 카테고리 분류(제품 사용법 이해, 문제 해결, 정보 탐색 및 선택, 제품 사용 방법 이해, 정보 검색 및 선택) 과정에 Few-shot기법 사용
- 평가 Metric을 사용하여 부적격 판단이 되는 데이터셋에 한해서 Upstage Solar LLM이 생성한 부적격 이유, Question 및 Context를 참고하여 직접 Ground Truth 생성

데이터 생성 과정에서 LLM은 다음과 같은 경우에 활용되었습니다.

첫째, PDF 메타데이터에 제품 이름이 명시되지 않은 경우 이를 보완하기 위해 사용하였습니다.

둘째, 기초 데이터를 기반으로 QA Pair데이터셋을 생성하는 데 적용되었습니다.

셋째, 생성된 질문을 카테고리별로 분류하는 작업에 활용되었습니다.

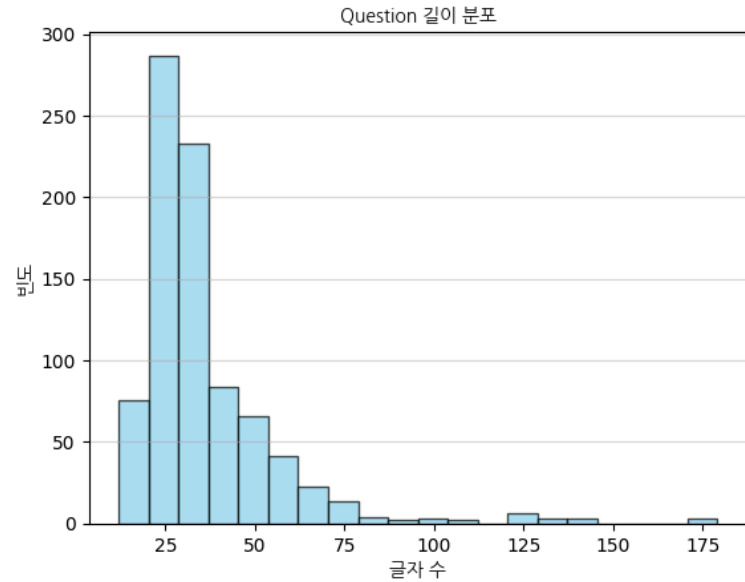
마지막으로, 데이터의 품질을 평가하고 부적격 데이터를 판별하는 과정에서도 LLM을 활용하여 데이터의 정확성과 적합성을 확보하였습니다.

데이터 통계 :

- 데이터 개수 :
 - Chunk 길이를 1024로, overlap은 10으로 세팅
 - 총 데이터 개수 : 850개
- 데이터 길이 통계 :

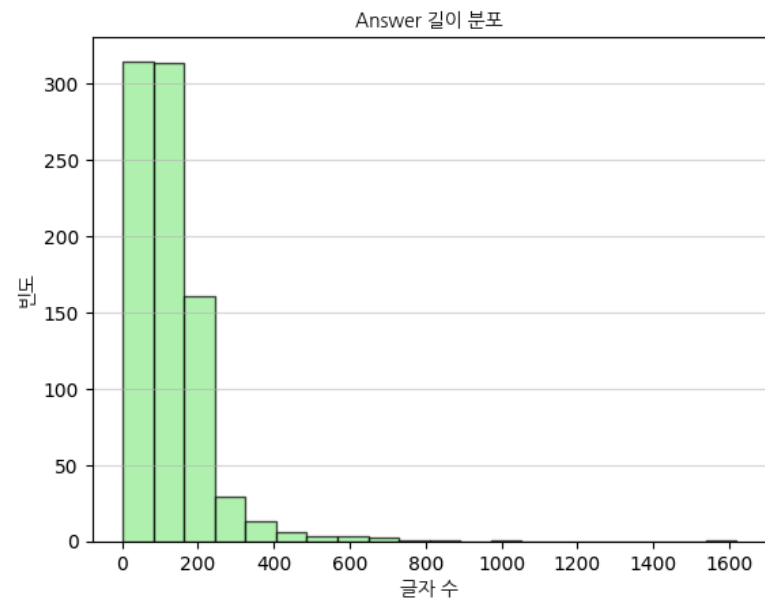
■ Question

◆ 평균길이 : 36.07자



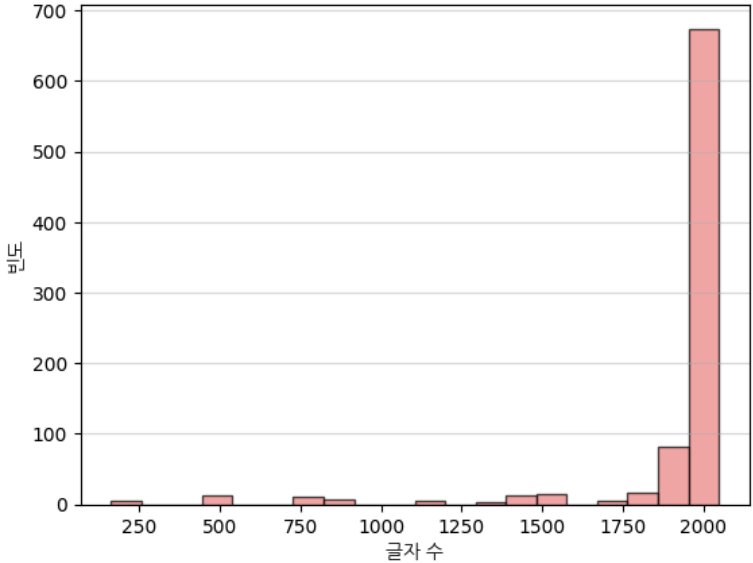
■ Answer

◆ 평균길이 : 128.07자



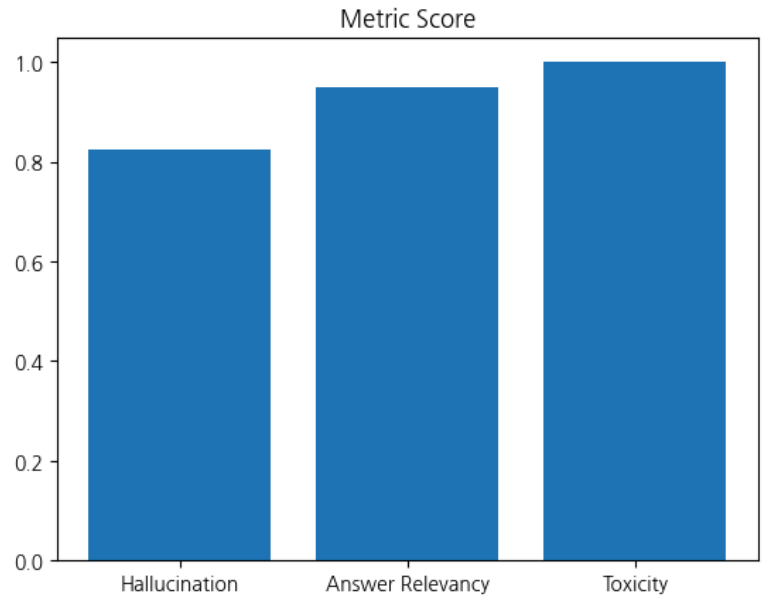
■ Context

◆ 평균길이 : 1913.72자

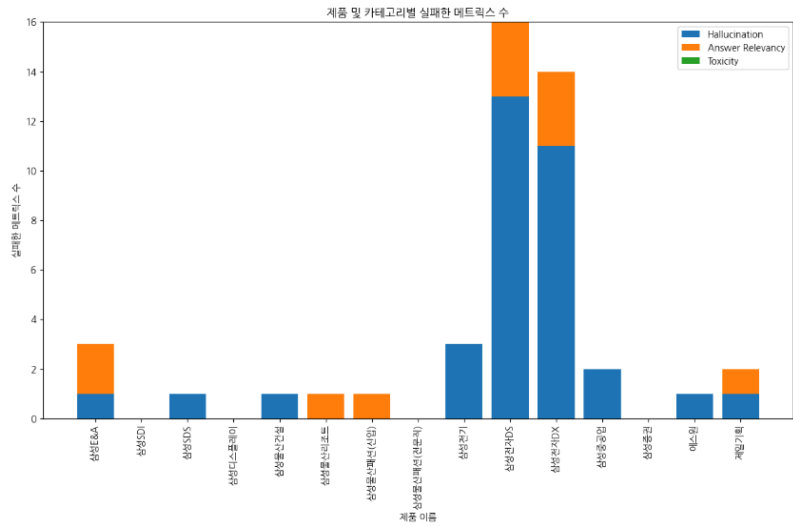
	<p style="text-align: center;">Context 길이 분포</p>  <table border="1"> <caption>Context 길이 분포 데이터 (추정)</caption> <thead> <tr> <th>글자 수</th> <th>빈도</th> </tr> </thead> <tbody> <tr><td>250</td><td>10</td></tr> <tr><td>500</td><td>20</td></tr> <tr><td>750</td><td>15</td></tr> <tr><td>1000</td><td>10</td></tr> <tr><td>1250</td><td>5</td></tr> <tr><td>1500</td><td>15</td></tr> <tr><td>1750</td><td>10</td></tr> <tr><td>1900</td><td>80</td></tr> <tr><td>2000</td><td>680</td></tr> </tbody> </table>	글자 수	빈도	250	10	500	20	750	15	1000	10	1250	5	1500	15	1750	10	1900	80	2000	680
글자 수	빈도																				
250	10																				
500	20																				
750	15																				
1000	10																				
1250	5																				
1500	15																				
1750	10																				
1900	80																				
2000	680																				
<p>데이터 품질 평가 및 결과</p>	<p>데이터 품질 평가 방식:</p> <ul style="list-style-type: none"> ● QA 평가셋 구성 방식 : <ul style="list-style-type: none"> ■ 전체 데이터셋 중 Random Sampling 방식을 사용해 500건 추출 ● 정량 평가 방식(평가 지표) : <ul style="list-style-type: none"> ■ Hallucination 생성된 답변에서 허위 정보를 평가하는 메트릭 ■ Answer Relevancy 생성된 답변의 관련성을 평가하는 메트릭 ■ Toxicity 생성된 답변의 유해성을 평가하는 메트릭 ● 정성 평가 방식 : <p>전체 데이터 셋 중 일부를 추출하여 매뉴얼의 내용 연관성과 불용어, 답변의 정확성 등 기준으로 평가 실시</p> <p>QA 평가셋은 전체 데이터 중 무작위 추출(Random Sampling) 방식을 통해 500개의 데이터를 선정하여 구성하였습니다. 선정된 데이터는 Deepeval의 HallucinationMetric, AnswerRelevancyMetric, ToxicityMetric을 활용해 정량적으로 평가하였습니다.</p> <p>추가로, 정성적 평가를 위해 각 팀원이 데이터를 직접 분석하여 평가하여 데이터의 품질을 다각도로 검증하였습니다.</p> 																				

데이터 품질 결과 보고:

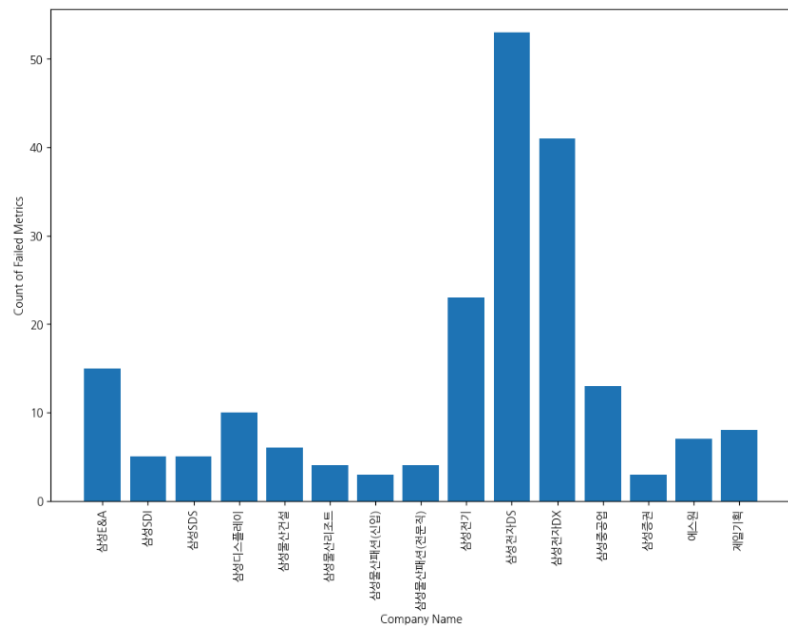
- 각 Metric Score



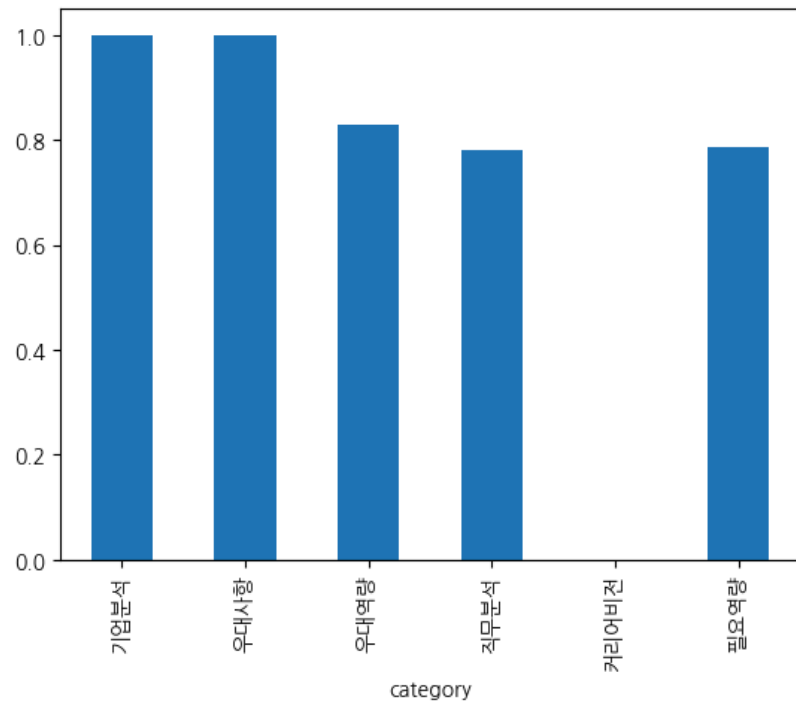
- 제품별 평가 Metrics 별 부적격 데이터 수



- 제품 별 부적격 데이터



• 카테고리 별 데이터 수



• 평가 결과 :

- 커리어비전을 제외한 나머지 카테고리의 전체적인 성공률이 80% ~ 100% 범위로 기본적인 품질 수준 확보
- 기업별 특성
 - ◆ 삼성물산 패션(신입)과 삼성 증권의 품질이 가장 높음

- ◆ 삼성전자DS 부문의 품질이 상대적으로 낮음
- 카테고리별 현황
 - ◆ 기업분석과 우대사항 : 가장 높은 신뢰도
 - ◆ 우대역량과 직무분석, 필요역량 : 상대적으로 낮은 성과
 - ◆ 커리어비전 : 개선필요
- 품질 개선 사항 및 개선 방법 :
 - 단기 개선 과제
 - ◆ 커리어비전 카테고리의 데이터 품질 개선 혹은 기업분석, 직무분석, 필요역량, 우대역량 이외의 카테고리를 생성하지 않도록 프롬프트를 조정
 - 중장기 발전 방향
 - ◆ 삼성전자 DS부문의 우수사례를 타 기업에 확대 적용
 - ◆ 카테고리별 특화된 품질 관리 기준 수립
 - ◆ 자동화된 품질 모니터링 시스템 구축
 - 품질 관리 체계 개선
 - ◆ 정기적인 품질 평가 및 피드백 시스템 구축
 - ◆ 제품군별/카테고리별 차별화된 품질 관리 전략 수립
 - ◆ 데이터 생성부터 검증까지 전 과정의 표준화

이러한 개선 전략을 체계적으로 실행함으로써, QA데이터셋의 전반적인 품질 향상과 함께 보다 신뢰성 있는 사용자 지원 서비스를 제공할 수 있을 것으로 기대됩니다.