

사용자 맞춤형 RAG 기반 챗봇 서비스 기획(안)

서비스명 및 개요	<p>서비스 명: 제품 맞춤형 QA 챗봇 서비스 개발</p> <p>서비스 개요:</p> <p><u>서비스가 해결하려는 문제점 (Problem Statement)</u></p> <ul style="list-style-type: none">● 많은 정보와 복잡한 구조로 이루어진 제품 매뉴얼은 사용자가 필요로 하는 정보를 찾기 어렵고 이는 고객 만족도 저하와 기업의 서비스 비용 증가로 이어짐● 기존의 FAQ 시스템은 사용자 질문에 대한 맞춤형 답변을 제공하는 데 한계가 있어 즉각적이고 정확한 정보 제공이 어려운 상황 <p><u>제공하고자 하는 주요 기능 및 가치</u></p> <ul style="list-style-type: none">● 사용자 질문 분석 및 답변 제공<ul style="list-style-type: none">○ AI 기반의 자연어 처리 기술을 활용하여 사용자 질문에 맞는 제품 매뉴얼의 정보를 실시간으로 제공● 데이터 학습 및 품질 개선<ul style="list-style-type: none">○ 대규모 QA 데이터셋을 기반으로 지속적인 학습과 개선을 통해 챗봇의 정확도와 신뢰성을 높임● 고객 경험 강화<ul style="list-style-type: none">○ 간단한 질문 입력만으로 문제를 해결할 수 있도록 하여 고객 만족도를 향상● 비용 효율성<ul style="list-style-type: none">○ 고객 서비스팀의 작업 시간을 단축하고 운영 비용을 절감
목표 및 기대효과	<p>서비스 목표:</p> <ul style="list-style-type: none">● 사용자 질문에 신속하고 정확한 답변을 제공하여 고객 문제 해결 시간을 단축합니다.● 제품 매뉴얼의 가치를 극대화하여 사용자와 기업 간의 신뢰를 강화합니다. <p>기대 효과:</p> <ul style="list-style-type: none">● 고객은 간단한 질문 입력만으로 매뉴얼을 탐색하지 않고도 필요한 정보를 얻을 수 있습니다. 이에 따라 고객 만족도 향상으로 제품의 브랜드 충성도가 증가할 수 있습니다.● 고객 서비스 팀은 검색 시간 단축으로 더 많은 고객 요청을 처리할 수 있습니다. 따라서 고객 서비스 비용을 절감할 수 있습니다.

<p>타겟 사용자</p>	<p>타겟 사용자:</p> <ul style="list-style-type: none"> ● 고객: 삼성 제품을 사용하는 고객 <ul style="list-style-type: none"> ○ 빠르고 정확한 제품 관련 정보를 습득 가능 ○ 제품 사용 중 문제가 발생했을 때 챗봇에 질문을 입력하여 실시간으로 해결책을 받음 ● 고객 서비스팀: 고객 요청을 처리하는 기업의 서비스 담당자 <ul style="list-style-type: none"> ○ 빠른 검색 및 답변 생성을 통해 더 많은 요청 처리 가능 ○ 수집된 데이터와 사용자 피드백을 활용하여 제품 개선 및 서비스 전략 수립 <p>시장 분석:</p> <ul style="list-style-type: none"> ● 국내 QA 챗봇 시장 현황 <ul style="list-style-type: none"> ○ 현재 QA 챗봇 서비스는 다양한 산업 분야에서 활발히 도입되며, 고객 서비스 효율성을 높이는 주요 도구로 자리 잡고 있음 ● 주요 동향 및 기회 요인 <ol style="list-style-type: none"> 1. 금융 분야의 도입 확대 <ul style="list-style-type: none"> ○ 국내 주요 은행에서 고객 서비스의 향상과 업무 효율성 향상을 위해 QA 챗봇 서비스를 도입하고 있음 <출처> 2. 기업 서비스 경쟁력 강화 <ul style="list-style-type: none"> ○ AI 챗봇은 고객 경험을 높이고 서비스 경쟁력을 높이는 요소로 평가받고 있음 <출처> 3. 챗봇 시장의 성장 전망 <ul style="list-style-type: none"> ○ 글로벌 챗봇 시장 규모는 2023년에 53억 4천만 달러로 평가되었으며, 2031년까지 263억 7천만 달러에 도달할 것으로 예상 <출처> 4. 다양한 산업 분야에서의 활용 <ul style="list-style-type: none"> ○ QA 챗봇은 금융, 의료, 소매 등 여러 산업 분야에서 데이터 분석, 고객 서비스, 마케팅 등의 다양한 업무를 지원하고 있음 <출처> ● 결론 <ul style="list-style-type: none"> ○ QA 챗봇 시장은 다양한 산업 분야에서 도입되고 있으며 고객 만족도 향상과 기업 효율성 증대에 기여하고 있음
---------------	---

데이터
구성 및
활용

원천데이터 소스: 삼성 갤럭시 제품 매뉴얼 문서

원천 데이터 형식: PDF 문서 파일

데이터 처리 방법:

- 데이터 수집: usermanual.wiki 페이지에서 Selenium을 사용한 웹 크롤링을 통해 제품 매뉴얼 데이터 확보.
- 데이터 전처리:
 - Document Parser를 활용하여 PDF 파일을 HTML, Markdown 형태의 텍스트 데이터로 변환.
 - 변환한 데이터에서 HTML 및 Markdown 태그와 문법을 정제하여 순수 텍스트 데이터로 변환.
 - 텍스트 데이터와 제품 이름 Pair를 JSONL 형식의 데이터로 변환하여 QA 데이터셋 생성 준비.

데이터 처리 방법(Detail):

1. 삼성 갤럭시 제품 매뉴얼 PDF를 usermanual.wiki 페이지에서 크롤링
2. Upstage Document Parser API를 활용해 Markdown 형태의 텍스트로 변환
3. Markdown 문법을 정제하여 순수 텍스트 데이터로 변환
4. PDF의 메타데이터를 활용하여 "제품이름"- "PDF 변환 텍스트 데이터"의 Pair를 JSONL 형식으로 구성
5. 이를 기반으로 QA 데이터셋 생성에 필요한 기초 데이터를 준비

QA 데이터 생성 방법:

- Upstage Solar LLM을 사용해 20개의 매뉴얼 데이터의 QA Pair 생성
- QA Pair 데이터 생성 프롬프트:

```
system_prompt = """{document_title} 중 한 페이지를 받게된다. 이를 읽고 사용자들이 질문할 만한 질문과 이에 대한 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. '문제 해결' - 제품 사용 중 발생할 수 있는 문제를 해결하는 질문
2. '제품 사용법 이해' - 제품 사용법과 관련된 질문
3. '정보 탐색 및 선택' - 제품에 대한 정보나 선택과 관련된 질문

```
```JSON
```

```
[
 {
 "question": str, # 사용자의 질문
 "answer": str, # 질문에 대한 응답
 "category": str # 질문의 카테고리 ('문제 해결', '제품 사용법 이해', '정보 탐색 및 선택', ...)
 }
]
```

```
```
```

```
user_prompt = """제품 매뉴얼 : {page_content}"""
```

| | |
|-----------------------------------|---|
| <p>데이터
구성 및
활용</p> | <ul style="list-style-type: none"> ● QA 데이터셋의 카테고리 분류(제품 사용법 이해, 문제 해결, 정보 탐색 및 선택, 제품 사용 방법 이해, 정보 검색 및 선택)와, 매뉴얼에서 제품명 추출하는 과정에 Few-shot기법 사용 ● 평가 Metric을 사용해 부적격 판단이 되는 데이터셋에 한해서 Upstage Solar LLM이 생성한 부적격 이유, Question 및 Context를 참고해 직접 Ground Truth 생성 <p>데이터 생성 과정에서 LLM은 다음과 같은 경우 활용되었습니다.</p> <ol style="list-style-type: none"> 1. PDF 메타데이터 제품 이름이 명시되지 않은 경우 보완 2. 기초 데이터 기반 QA Pair 데이터셋 생성 3. 생성된 질문을 카테고리별로 분류 4. 데이터의 품질을 평가하고 부적격 데이터를 판별하는 과정에서 LLM 활용해 데이터 정확성과 적합성 확보, 또한, 부적격 데이터의 경우 Ground Truth 생성. |
|-----------------------------------|---|