**[프로젝트 기획안]  
RAG 기반 챗봇 서비스 구현**

**| 마일스톤 1**

Step 1: 프로젝트 문제 정의   
Step 2: 기획안 작성

Step 3: 데이터 선정

**본 문서의 무단 사용 및 불법 배포 시 법적 조치를 받을 수 있습니다**

**프로젝트 개요**

**프로젝트명 : RAG 기반 챗봇 서비스 구현**

**프로젝트 개요**:  
 본 프로젝트는 RAG(Retrieval-Augmented Generation) 기술을 활용하여 여러분들만의 챗봇 서비스를 설계하고 구현하는 것을 목표로 합니다. 먼저 설계하고자 하는 서비스의 주제를 기획한 뒤, 해당 서비스에 필요한 데이터를 선정합니다. 이후, 수집된 데이터를 활용해 RAG 파이프라인을 설계하고, 이를 배포하여 챗봇 서비스를 개발하게 됩니다. 이 과정을 통해 RAG의 개념을 경험해보고 실무 적용 능력을 강화합니다.

**프로젝트 목표**

* 주제에 맞는 데이터를 선정, 정제, 활용하는 경험을 쌓습니다.
* 사용자 요구에 부합하는 AI 서비스를 기획하고, 이를 기술적으로 구현하는 과정을 체득합니다.
* RAG 기반 챗봇 설계를 통해 AI 서비스 개발 실무 경험을 배양합니다.

**프로젝트 결과물**:

* 서비스 기획안
* RAG 파이프라인
* RAG 기반 챗봇 서비스 데모

**프로젝트 로드맵**

* **마일스톤 1 (권장 작성 기간 : ~12/24 화)**Step 1: 프로젝트 문제 정의   
  Step 2: 기획안 작성  
  Step 3: 데이터 선정
* **마일스톤 2 (권장 작성 기간 : 12/24 화 ~12/30 월)**  
  Step 4 : RAG 서비스 아키텍처 설계   
  Step 5: 서비스 기능 요건 정의  
  Step 6: 서비스 배포

**Step 1: 프로젝트 문제 정의**

안녕하세요, 프로젝트를 시작하기에 앞서 **AI 서비스를 기획할 때 고려해야 할 핵심 요소**들을 함께 살펴보겠습니다. 이번 과정을 통해 여러분들이 AI 서비스를 설계하는 기본 흐름을 이해하고, 이를 바탕으로 여러분만의 AI 서비스를 기획하는 능력을 키우는 것을 목표로 합니다.

앞으로 여러분들은 **체크리스트**와 **프로젝트 예시**를 활용해 문제 정의와 이의 해결 방안을 구체화하는 방법을 알아보고, 프로젝트 과정에서 아이디어를 구체화하는 데 도움을 드리고자 합니다.

* **✅ 체크리스트**여러분이 서비스를 설계하는 과정에서 반드시 생각해보아야 할 질문들을 제공합니다. 이를 바탕으로 해결하고자 하는 문제를 명확히 정의하고, 서비스의 가치를 구체화할 수 있습니다.
* **🔷 프로젝트 예시**해당 프로젝트 예시는 여러분들께서 3주차에 진행하시게 될 프로젝트의 예시입니다.   
  프로젝트 예시를 바탕으로 문제를 정의하고 AI 기술을 활용하여 해결하는 과정을 구체적으로 보여줍니다. 예시를 참고하여 본 프로젝트에서 여러분의 아이디어를 더욱 구체화해보세요.

1) Pain Point(문제점) 찾기  
2) Expected Solution(예상 해결 방안) 도출 - AI의 역할 정의   
3) Task(태스크) 유형 정의   
4) 데이터 요구사항 구체화   
5) 결론

1. **Pain Point(문제점) 찾기**

서비스 아이디어를 구체화하기 위해서 해결하고자 하는 문제를 명확히 정의하는 것이 중요합니다.  
아래 질문들을 바탕으로 고민해보세요!

**✅ 체크리스트**

* (문제 상황 파악) 어떤 불편함이나 비효율을 개선하고 싶은가? 혹은 AI를 활용하여 도움이 되었던 경험은 무엇인가?
* (타겟 사용자) 해결하고자 하는 문제는 누구에게 영향을 미치는가?
* (서비스) 무엇을 만들 것인가?

1. **Expected Solution(예상 해결 방안) 도출 - AI의 역할 정의**

**AI의 역할을 정의하고 문제 해결 방향을 구체화합니다.**해당 문제가 AI로 해결 가능한 영역인지, AI를 활용하면 더 효율적인지 판단합니다.

**2-1/ AI가 잘 할 수 있는 영역**

**✅ 체크리스트   
(본 프로젝트에서는 아래 질문을 RAG로 바꾸어 보세요!)**

* 해결하고자 하는 문제가 AI가 해결할 수 있는 영역인가? (e.g. 자동화, 예측, 패턴 인식, 자연어 처리 등)   
   *> 해당 문제는 자연어 처리에 능숙한 LLM이 처리할 수 있는 영역인가?   
  > 해당 문제는 외부 문서/정보 검색을 바탕으로 답변하는 것이 유리한 영역인가?*
* 해당 문제에 적용가능한 AI 알고리즘, 모델, 기술이 있는가?
* 해당 AI 기술이 현재 문제를 해결하는데 적합한가?

**(TIP!) RAG 사용 시 도움이 되는 경우**

***(Recap.)* RAG(Retrieval Augmented Generation) :** 검색 엔진과 LLM을 결합한 파이프라인. 사용자의 질문과 관련된 외부 문서/지식을 검색하여 이를 바탕으로 LLM이 답변을 생성함.

1. LLM이 거짓 정보를 생성하는 경우 (Hallucination) : General한 지식을 학습한 LLM이 알지 못하는 전문적인 의료/법률 등의 도메인 지식들을 주입하여 정확한 답변 생성할 수 있음.
2. 사전학습 기간이 오래 걸리기에 최신 정보를 반영하지 못하는 경우(Knowledge Cutoff) : 실시간 검색을 통해 최신 정보를 반영하여 보다 정확한 답변 생성할 수 있음.

**2-2/ 비즈니스 임팩트 측정**실제로 AI 기술을 현업에서 적용하고자 할 때, AI 도입이 실제 서비스, 비즈니스에 큰 임팩트가 있는 지 판단해야 합니다.

**✅ 체크리스트**

* AI 기술을 적용하기 위해 필요한 자원과 역량이 있는가?
* AI 기술을 적용하여 얻을 수 있는 이익이 비용과 시간에 비해 충분한가?   
  (One Step More!) 실제 현업 비지니스 기준
  + Make money : AI 기술을 적용하여 이익을 창출할 수 있는가?
  + Save money : AI 기술을 적용하여 비용을 절약할 수 있는가?
* 해당 서비스를 통해 사용자/기업/사회 등은 어떤 가치/이득을 얻을 수 있는가?

1. **Task(태스크) 유형 정의**

위 모든 과정을 반복하며 해당 문제가 AI로 해결이 가능하고, 비즈니스 임팩트가 있다고 판단했다면 이제 구체적인 AI 서비스의 기능에 대해서 고민해봅시다. AI 서비스의 주요 기능을 구체화하고, 이를 구현하기 위한 적합한 태스크 유형을 선정하는 파트입니다.

이번 프로젝트에서는 다양한 자연어 처리 task 들을 알아보고, 그 중 우리 서비스는 **“어떤 유형들”**로 서비스를 설계해나가면 좋을 지 고민해봅시다.

**✅ 체크리스트   
단, 우리는 챗봇 서비스를 만들기에 최종 형태는 “챗봇” 형태여야합니다.**

* 해결하고자 하는 문제가 풀어야하는 업무 프로세스를 최대한 구체적으로 정의하였는가?
* 위의 업무 프로세스가 위의 자연어처리 태스크 중 어떤 타입을 통해 효과적으로 해결될 수 있는가?
* 해당 태스크 유형이 서비스에서 제공하는 주요 기능과 일치하는가?
* 해당 태스크 유형이 사용자 경험을 효과적으로 개선할 수 있는가?

1. **데이터 요구사항 구체화**

본격적으로 서비스를 설계하기 이전에, 서비스 구현 가능성을 점검해야 합니다. 특히, 필요한 **데이터**가 준비되었거나 구할 수 있는 상태인지, 서비스와 태스크 유형에 적합한 지를 우선 확인해야 합니다.

또한, 일반적으로 기술 및 도구의 활용 가능 여부도 점검해야 하지만, 이번 프로젝트에서는 정해진 API와 도구를 활용하기 때문에 **데이터 준비 상태**에 초점을 맞춰 확인합니다.

**✅ 체크리스트**

**원천데이터 유형 확인**

* 필요한 데이터는 무엇인가? (e.g. 제품 매뉴얼 문서, FAQ 문서 등)
* 데이터의 도메인과 특징은 무엇인가? (e.g. 의료, 제조업, 법률 등)

**원천데이터 형식(format) 확인**

* 원천데이터의 형식은 무엇인가? (e.g. PDF, HTML, txt 등)
* 원천데이터 형식이 태스크 유형과 맞지 않다면, 적합하게 변환할 수 있는가?

**데이터 소스 선정**

* 데이터셋을 구할 수 있는가? (e.g. 오픈 소스 활용 가능 여부 / 웹 크롤링 가능 여부)
* 데이터를 구할 수 있는 출처가 명확한가? (e.g. AI Hub, Kaggle 등)
* 데이터셋의 저작권이 활용가능한 라이센스인가? (e.g. MIT License, CC 라이선스 등)

**데이터 품질 점검**

* 데이터셋이 충분히 크고 다양한가?
* 데이터의 품질을 신뢰할 수 있는가?

1. **결론**

지금까지 AI 서비스 기획에서 고려해야 할 주요 포인트를 함께 살펴보았습니다. **체크리스트와 프로젝트 예시**를 활용해 문제를 명확히 정의하고, AI를 활용한 해결 방안을 구체화해보세요. **🚀 여러분들만의 기획안을 위한 문제 정의를 시작해보세요!**

* 문제 정의 :
* 기대 솔루션 :
* 태스크 유형 :
* 데이터셋 :

**Step 2. 서비스 기획안 구체화**

|  |  |
| --- | --- |
| **프로젝트 기획 (안)** | |
| **서비스명 및 개요** | **서비스명**: (서비스 이름)  **서비스 개요**:  - 서비스가 해결하려는 문제점(Problem Statement)  - 제공하고자 하는 주요 기능 및 가치  - 타겟 사용자와 예상 사용 시나리오 |
| **목표 및 기대효과** | **서비스 목표**:  - 구체적인 목표 설정  **기대효과**:  - 사용자에게 제공되는 직접적인 이점  - 조직/팀/사회 등 발전에 기여할 부분 |
| **타겟 사용자** | **타겟 사용자**:  - 예상 사용자 유형 (예: 고객 서비스 팀, 일반 사용자, 개발자 등)  - 주요 요구사항 및 사용 목적 |
| **데이터 구성 및 처리 방법** | **원천데이터 소스**:  **원천데이터 형식 :**  **원천데이터 처리 방법**:   * 데이터 수집 * 데이터 처리 |

**Step 3: 데이터 선정 및 전처리 가이드**

이번 단계에서는 위 기획안과 강좌에서 배운 내용을 바탕으로 데이터를 수집 /선정하고 처리하는 과정을 배웁니다. 데이터 선정은 서비스 구현의 출발점으로, 목표와 요구사항에 부합하는 데이터를 확보하고 정제하는 과정이 중요합니다.

위 기획안을 바탕으로 데이터의 유형/형식/출처 확인이 되었다는 가정하에 본격적으로 데이터를 선정 및 수집하고 전처리해봅니다.

1) 기획안 - 데이터셋 내용 정리  
2) 데이터 소스 선정  
3) 데이터 처리  
4) 데이터 활용 계획 점검

1. **기획안 - 데이터셋 내용 정리**

* **원천데이터 소스:**
* **도메인**:
* **원천 데이터 형식**
* **출처**
* **저작권** :

1. **데이터 소스 선정**

본격적으로 데이터를 수집하기 이전에 고려해야할 체크리스트를 다시 한 번 되짚어봅니다.

**2-1. 데이터 소스 선정 방식**

* **사용 가능한 오픈소스 데이터셋**:
* **웹 크롤링을 통한 데이터 수집**:

**✅ 체크리스트**

* **오픈소스 데이터 사용 시**:
  + 데이터 출처와 저작권이 명확한가?
  + 데이터가 사용 조건에 부합하며, 라이선스 문제가 없는가?
* **데이터 수집 시**:
  + 개인정보 보호와 관련된 규정을 준수했는가?
  + robots.txt 규정 및 크롤링 제한 조건을 확인했는가?

1. **데이터 처리**

이제 수집한 원천 데이터를 원하는 방식대로 변환 및 처리하는 방법에 대해서 알아보겠습니다.

**3-1. (데이터 변환 필요 시,) 데이터 형식 변환**

**3-2. 데이터 정제**

**3-3. 데이터 통합**

**3-4. 데이터 품질 점검**

**✅ 체크리스트**

* 데이터 형식을 모델과 서비스 요구사항에 맞게 변환하였는가?
* 이상치, 결측치, 중복 항목 등을 제거하여 데이터의 일관성을 확보하였는가?
* 데이터가 서비스에 적합하고, **정확성**, **일관성**, **포괄성 등**을 갖추었는가?

1. **데이터 활용 계획 점검**

**4-1. 데이터와 모델 연계**

* 데이터가 **RAG 파이프라인**과 원활히 연계되도록 설계합니다.   
  데이터가 검색 엔진에 인덱싱되고, 생성 모델이 검색된 데이터를 활용하여 답변을 생성하는 프로세스 입니다.

**Conclusions**

마일스톤 1에서는, 서비스 목표와 데이터 요구사항을 명확히 정의하고, 필요한 데이터를 빠르게 선정합니다.   
오픈소스 데이터를 적극 활용하되, 시간이 충분한 경우 웹 크롤링을 통해 고유 데이터를 확보하는 방향으로 진행하시길 권장드립니다. 서비스 기획 및 데이터 선정을 마무리하고 나면, **2주차 RAG 파이프라인 구현 프로젝트를** 진행하기 위한 준비가 완료됩니다!

구체적인 개발환경 및 프로젝트 구조는 2주차에 RAG 파이프라인의 구성요소, 설계 방식을 배우는 마일스톤2에 해당 강좌와 함께 제공될 예정입니다.