

SK네트웍스 Family AI과정 3기

데이터 전처리 학습된 인공지능 모델

□ 개요

- 산출물 단계 : 데이터 전처리
- 평가 산출물 : 학습된 인공지능 모델
- 제출 일자 : 2024-12-23
- 깃허브 경로 : <https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN03-FINAL-5Team>
- 작성 팀원 : 김종식, 하은진

사용 데이터 및 기술	<ol style="list-style-type: none">1. 개요<ol style="list-style-type: none">1-1. 본 프로젝트는 모의면접 시스템 개발을 목적으로 한다. 이를 위해 질문 생성과 평가의 근거가 될 데이터를 수집한다2. 사용데이터<ol style="list-style-type: none">1-1. 파이썬 공식데이터1-2. geeks for geeks – Programming knowledge data1-3. interview bit – interview data
----------------	---

<p>상황</p>	<ol style="list-style-type: none"> 1. 개요 <ol style="list-style-type: none"> 1-1. 본 프로젝트는 모의면접 시스템 개발을 목적으로 한다. 이를위해 인공지능 모델을 바탕으로 질문을 생성하고 사용자의 답변을 받아 평가하는 로직을 구현한다. 2. 사용모델 및 역할 <ol style="list-style-type: none"> 2-1. Open AI Chat gpt (질문생성용) <p>vector DB의 내용과 사용자 이력서의 키워드를 바탕으로 질문을 생성한다. 프롬프트의 미세조정으로 바탕으로 질문의 방향성과 내용을 조정한다.</p> 2-2 Open AI Chat gpt (평가용) <p>생성된 질문을 기반으로 프롬프트를 미세 조정하여 답변 정확성, 직무 및 역량, 논리성 평가 측면에서 사용자 답변 평가를 진행한다.</p> 2-3. FAISS <p>vector DB를 구축하는 프레임워크로써 사용한다. Open AI의 text-embedding-3-small 임베딩 모델을 채택하여 임베딩을 수행하고 RAG 기반 질문생성 시스템 개발의 바탕으로 되는 vector DB를 구축한다.</p> 2-4. Hugging Face Seq2Seq model <p>이력서의 키워드를 추출하는데 사용한다. Hugging Face는 각종모델과 데이터셋 등이 모여진 HUB로써 그 중 ilsilverskiold/tech-keywords-extractor 모델을 채용한다. 해당 모델은 키워드 추출에 특화되어 있어 사용자의 이력서 정보에서 유의미한 정보(키워드)를 추출하기 위해 사용한다.</p>
<p>수행 결과</p>	<ol style="list-style-type: none"> 1. 개요 <ol style="list-style-type: none"> 1-1. 본 문단에서는 모의 면접모델의 작동로직과 결과에 대해 설명한다. 2. 수행로직 <ol style="list-style-type: none"> 2-1 이력서에서 기술적 핵심키워드(python, Vue.js, GitHub, Stripe API, Authentication, Firestore, React) 등을 추출 2-2 Open AI Chat gpt모델 기반 RAG로 질문생성 후 프론트엔드에 전달 2-3. 사용자의 답변을 평가모델에 인입 <p>RAGas 지표 answer correctness, answer_relevancy 계산</p> 2-4. 평가모델 <p>답변 정확성 평가 - 우수, 보통, 미흡으로 평가 후 세부 피드백 제시 직무 및 역량 평가 - 우수, 보통, 미흡으로 평가 후 세부 피드백제시 논리성 평가 - 우수, 보통, 미흡으로 평가 후 세부 피드백 제시 보고서 산출 -강점, 약점, 한줄평, 총점 계산 후 최종 보고서산출물</p>

보완점

1. 데이터셋의 정확도가 질문의 품질에 영향을 미침 데이터셋의 품질강화가 필요
2. langgraph적용 시 속도가 느려져 빠른 질의응답이 요구되는 모의면접에는langgraph를 적용하기 어려움
3. 기술키워드의 한계가 존재 프로젝트 경험 등 문장 맥락에따라 자동 판단 시스템이 있는 것도 보완
4. 면접의 일방적인 텍스트화로 인해 집중력 감소 우려, 말로해주는 면접 시스템으로 보완필요