

SK네트웍스 Family AI과정 3기

데이터 수집 및 저장 프로젝트 기획서

□ 개요

- 산출물 단계 : 데이터 수집 및 저장
- 평가 산출물 : 프로젝트 기획서
- 제출 일자 : 2024-11-12
- 깃허브 경로 : <https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN03-FINAL-5Team>
- 작성 팀원 : 김종식

프로젝트 주제	LLM 활용 AI 모의면접 시스템
문제정의	<div>1. 취업준비생을 위한 효과적인 면접 준비시스템 필요</div> <div>- 면접경험 등의 부재로 자기 능력표현의 한계 존재</div> <div>- 구직자의 46%, 면접이 가장 어렵다는 조사 결과 (2024.05.08 잡코리아)</div> <div>2. 기존 모의면접 시스템의 한계</div> <div>- 신뢰도 높은 AI 모의 면접 시스템의 부재</div> <div>- 최신 기술을 반영하지 못한 기존 시스템의 한계</div> <div>- 청년 채용역량 강화를 위한 최신 면접 시스템 기술개발 필요</div> <div>- 사용자의 현재 능력을 점검할 수 있는 명확한 평가지표 필요</div> <div>- 직무전문성을 대비한 면접시스템의 부재</div> <div>3. 기업의 인재수집 효율화 필요</div> <div>- 취업준비생의 면접역량을 강화하여 능력있는 인재 발굴</div> <div>- 추후 1차 면접 등 대규모 다수면접을 대체하여 면접비용 감소</div>

<p>시장조사</p>	<p><b>1. 유사시스템 조사(AI 면접관 뷰인터)</b></p> <ul style="list-style-type: none"> <li>- 원하는 검사역량을 선택가능</li> <li>- 기업별 비교데이터 부족</li> <li>- 특정 키워드 중심으로 반응하는 꼬리질문 가능</li> <li>- 종합레포트 제공</li> <li>- 답변내용, 질의에 대한 맞춤형 피드백 부족</li> <li>- 호감도, 집중도 등 자체 지표로 상대평가 수치제공</li> <li>- 영상 데이터 분석</li> <li>- 시선처리, 머리 움직임, 목소리 크기 분석</li> <li>- 다만 분석보다 사용자의 모습을 표현/현황보고에 가까움</li> <li>- 순위평가제(n명중 n위)</li> <li>- 합격가능성(n%), 면접준비상태 판단하여 제공</li> <li>- 고객사 설문 결과 신뢰성이 21%, 비용 대비 효과성이 12%</li> </ul> <p><b>2. 유사시스템 조사(AI 취업플랫폼 에듀스)</b></p> <ul style="list-style-type: none"> <li>- 등급평가제(6개등급 최우수 - 우수 ...)</li> <li>- 사용자 답변을 바탕으로 꼬리질문 생성</li> <li>- 답변별 평가 제공(총평, 개선부분, 개선예시)</li> <li>- 답변내용에 대한 심층적인 피드백은 부족</li> <li>- 긍정어, 부정어, 면접 답변 어휘 감성분석</li> <li>- 역량어휘, 비역량어휘를 구분 해석</li> <li>- 영상, 음성의 경우 분석보다 표현이 핵심</li> <li>- 면접핵심 키워드 워드 클라우드 표시</li> </ul> <p><b>3. 결론</b></p> <ul style="list-style-type: none"> <li>- 답변내용에 대한 심층적인 피드백 제공 필요</li> <li>- 답변어휘의 감성분석, 문장내용에 대한 분석이 필요</li> <li>- 영상, 음성 등의 경우 현황보고 이외의 해석, 분석이 필요</li> <li>- 분야별 전문성있는 질문을 가능할 경우 강점이 될 것으로 추정</li> </ul>
-------------	--

시스템 구성	<ol style="list-style-type: none"> <li>1. 웹 <ul style="list-style-type: none"> <li>- 회원가입(카카오톡 소셜 회원가입 기능 제공)</li> <li>- 로그인(카카오톡 소셜 로그인 기능 제공)</li> <li>- 마이페이지(면접 이력, 프로필 정보 조회, 로그아웃, 회원탈퇴)</li> <li>- 모의면접</li> <li>- 면접영상녹화(면접 시 자신의 모습을 확인 및 영상녹화)</li> <li>- 면접결과 레포트 pdf출력</li> </ul> </li> <li>2. 모의면접 <ul style="list-style-type: none"> <li>- 실시간 질문생성(사용자 정보 바탕)</li> <li>- LLM기반 사용자 답변 실시간 분석</li> <li>- 면접결과 분석 레포트 작성(장단점, 총평, 총점 등)</li> </ul> </li> <li>3. 데이터 분석 <ul style="list-style-type: none"> <li>- 파이썬 등 기술 공식데이터 수집</li> <li>- 데이터 난이도별 분류라벨링</li> <li>- 분석 평가지표 설정</li> </ul> </li> </ol>
모델링 방안	<ol style="list-style-type: none"> <li>1. 질문생성 모델 <ul style="list-style-type: none"> <li>- gpt4.o API를 활용하여 질문생성</li> <li>- Prompt Engineering을 통해 질문방향성 조정</li> <li>- FAISS vector DB를 활용하여 데이터 저장 및 정보검색</li> </ul> </li> <li>2. 면접실행 <ul style="list-style-type: none"> <li>- 실시간 면접시스템 구축</li> <li>- STT를 활용하여 사용자 음성데이터를 텍스트로 수집</li> <li>- OpenCV 라이브러리를 활용하여 면접 영상 데이터 수집</li> </ul> </li> <li>3. 평가모델 <ul style="list-style-type: none"> <li>- RAGAS 지표를 바탕으로 답변 정확성 평가</li> <li>- Cosine similarity 계산을 바탕으로 문장의 의미적 유사성 평가</li> <li>- 별도의 평가지표 모델로 독자적인 평가시스템 구축</li> <li>- 실시간 면접시스템 구축</li> </ul> </li> </ol>

사용데이터	<p><b>1. 데이터 수집 목적</b></p> <ul style="list-style-type: none"> <li>- 질의응답 정확성 확보</li> <li>- 질의별 난이도 분리</li> <li>- 면접 피드백의 신뢰도 향상</li> </ul> <p><b>2. 데이터 수집</b></p> <ul style="list-style-type: none"> <li>- 기술 데이터 공식 홈페이지 <b>Crawling</b> 데이터수집</li> </ul> <p><b>3. 데이터 전처리</b></p> <ul style="list-style-type: none"> <li>- 중복데이터 제거</li> <li>- 난이도에 따라 데이터 분리(프로그래머스 기준)</li> </ul>
비즈니스모델	<p><b>1. B2B</b></p> <ul style="list-style-type: none"> <li>- 효율적이고 산업에 맞는 인재를 선발하려는 기업 대상</li> <li>- 산업군에(기업) 맞춰서 전문적인 지식에 대해 질문을 생성</li> <li>- 다양한 산업에 유동적으로 대응가능한 확장성</li> <li>- 전문성있는 피드백을 바탕으로 인재 선발 효율화</li> <li>- 다각도로 인재를 평가하여 기업에 최적화된 인재 선발 가능</li> <li>- 1차 면접과 같은 대규모 면접을 대체하여 면접비용 감소</li> </ul> <p><b>2. B2C</b></p> <ul style="list-style-type: none"> <li>- 취업준비생, 청년층 대상</li> <li>- 취업준비생에게 전문적인 질문을 제공하여 취업역량 제고</li> <li>- 영상, 음성, 답변 등 다방면의 피드백을 제공하여 단점 보완가능</li> <li>- 면접준비의 자신감 향상 및 성공률 증가</li> </ul>
<b>R&amp;R</b>	<p>유혜린(팀장) - 백엔드, AWS</p> <p>김재성 - 데이터분석</p> <p>김원철 - 프론트엔드, 백엔드</p> <p>김종식 - AI</p> <p>하은진 - 데이터분석</p>