

SK 네트워크스 Family AI 과정 5기

데이터 전처리 인공지능 학습 결과서

□ 개요

- 산출물 단계 : 데이터 전처리
- 평가 산출물 : 인공지능 학습 결과서
- 제출 일자 : 2025-02-13
- 깃허브 경로 : <https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN05-FINAL-4TEAM.git>
- 작성 팀원 : 김지연

개요	<p>이 보고서는 AI 기반 면접 질문 생성 및 답변 평가 모델의 성능을 분석하고, 모델이 실제 면접에서 활용될 수 있도록 적절한 질문과 평가를 수행할 수 있는지 검증하는 것을 목표로 한다.</p> <ul style="list-style-type: none">- 질문 생성 모델은 지원자의 이력서와 채용 공고를 분석하여 적절한 면접 질문을 생성하도록 설계- 답변 평가 모델은 AI가 면접 응답을 정량적으로 평가하고 피드백을 제공하는 역할을 수행
질문 생성 모델	<p>2.1 프롬프트 설계 방법</p> <ul style="list-style-type: none">- 프롬프트 구조 <p>지원자의 이력서 내용 + 기업의 담당 업무 및 자격 요건 + 평가 기준</p> <p>프롬프트 예시</p> <p># 질문 생성 함수</p> <pre>def generate_questions(resume_text, responsibilities, qualifications, evaluation_metrics): prompt = f" 아래는 지원자의 이력서 내용과 회사의 담당 업무 및 지원 자격입니다. 이를 바탕으로 면접 질문을 생성하세요. 각 질문은 실무 관련성과 직무 적합성을 평가할 수 있도록 작성하며, 평가 기준과 담당 업무 및 지원 자격을 반영하여 질문을 생성하세요: 평가 기준: {'', ' '.join(evaluation_metrics)} [지원자의 이력서] {resume_text} [담당 업무] {responsibilities}</pre>

	<p>[지원 자격] {qualifications}</p> <p>질문 유형:</p> <ol style="list-style-type: none"> 1. 주요 프로젝트 경험 관련 질문: 지원자의 주요 프로젝트 경험을 평가하기 위한 질문 2. 문제 해결 사례 기반 질문: 문제 해결 능력을 평가하기 위한 질문 3. 팀워크 및 협업 관련 질문: 팀워크 능력을 평가하기 위한 질문 4. 자기 개발 노력 관련 질문: 자기 개발 능력과 학습 의지를 평가하기 위한 질문 5. 지원 자격과 관련된 질문: 예를 들어, 코드 리뷰, 테스트 코드 작성, RDB 모델링 최적화 등 지원 자격에 명시된 항목과 관련된 질문 <p>각 질문 유형별로 2 개의 구체적인 질문을 작성하세요. 특히 지원 자격과 관련된 항목(예: 코드 리뷰, 테스트 코드 작성, 대규모 트래픽 처리)에 대한 질문을 포함해주세요. 결과를 JSON 형식으로 반환하세요. ""</p> <p>2.2 생성된 질문</p> <pre>{ "질문": { "주요 프로젝트 경험 관련 질문": ["AI 기반 고객 이탈 예측 시스템 프로젝트에서 사용한 기술 스택 중 어떤 부분이 가장 도전적이었고, 어떻게 해결하였나요?", "고객 이탈 예측 정확도를 85%까지 끌어올린 성과를 얻기 위해 어떤 데이터 전처리 및 모델 최적화 기법을 적용하였는지 설명해 주세요."], "문제 해결 사례 기반 질문": ["데이터 분석 도구의 속도를 개선하기 위해 어떤 쿼리 최적화 알고리즘을 도입하였는지 설명해 주세요.", "문제 해결 과정에서 팀원들과 협업하여 새로운 알고리즘을 실행에 옮기기 위해 어떤 커뮤니케이션 전략을 사용하였나요?"], ... } }</pre>
평가 방법	<p>모델이 생성한 질문을 평가하기 위해 다음과 같은 지표를 활용하였다.</p> <p>1. 관련성(Relevane)</p> <ul style="list-style-type: none"> -질문이 지원자의 이력서 및 채용 공고와 얼마나 관련성이 높은지 평가 - SBERT 기반의 문장 임베딩을 활용하여 이력서-질문, 채용공고-질문 간 코사인 유사도를 계산한 후 평균값을 도출 <p>2. 명확성(Clarity)</p> <ul style="list-style-type: none"> -질문의 문법적 오류와 가독성을 평가 -LanguageTool 을 활용하여 문법 오류를 탐지하고, 문장 내 오류 개수를 기준으로 점수화

	<p>3. 전문성(Professionalism)</p> <p>- 질문이 직무 관련 키워드와 얼마나 밀접하게 연결되어 있는지 평가</p> <p>- SBERT 임베딩을 사용하여 질문과 직무 관련 키워드 간 유사도를 계산하여 점수화</p> <p>이 세 가지 지표를 조합하여 최종 점수를 산출하며, 모델이 생성한 질문이 채용 평가에 적절한지 정량적으로 분석할 수 있도록 하였다.</p>
평가 점수 및 결과	<p>모델이 생성한 질문을 평가한 결과, 다음과 같은 평가 점수가 도출되었다.</p> <ul style="list-style-type: none"> - 관련성(Relevance): 0.82 - 명확성(Clarity): 0.91 - 전문성(Professionalism): 0.78 - 최종 점수(Final Score): 0.83 <ul style="list-style-type: none"> ● 생성된 질문은 이력서 및 채용 공고와 높은 관련성을 가진다 ● 문법적으로 명확성이 뛰어난 수준을 보인다 ● 다만, 전문성 부분에서 일부 직무 관련 키워드와의 연결성이 다소 부족한 사례가 발견되었다 <p>양질의 인터뷰 질문을 생성하는 능력을 갖추었지만, 특정 직무에 대한 전문성을 더욱 반영할수록 추가적인 프롬프트 최적화 및 데이터 보강이 필요함을 확인할 수 있다.</p>
답변 평가 모델	<p>3.1 프롬프트 설계 방법</p> <p>- 프롬프트 구조</p> <p>지원자의 이력서 내용 + 기업의 담당 업무 및 자격 요건 + 평가 기준</p> <p>프롬프트 예시</p> <pre># 답변 평가 함수 def evaluate_response(response, question, evaluation_metrics, responsibilities, qualifications): prompt = f""" 아래는 면접 질문과 지원자의 응답입니다. 이 응답을 평가 기준에 따라 점수화하고, 각 항목별로 개선사항을 제시하세요. 점수 부여 시 다음과 같은 기준을 일관되게 적용하세요. === 면접 질문 === {question} === 지원자 답변 === {response} === 평가 기준 및 점수 부여 기준 === 1. 질문 이해도 (0~10 점): - 10 점: 질문을 완벽히 이해하고, 논리적으로 답변함. - 8~9 점: 질문을 이해했으나 일부 부족한 설명이 있음.</pre>

- 6~7 점: 질문의 핵심을 이해했으나, 논리적 흐름에서 약간의 오류가 있음.
 - 5 점 이하: 질문의 핵심을 충분히 이해하지 못했거나, 명확하지 않은 답변을 제공함.
2. 논리적 전개 (0~10 점):
- 10 점: 서론-본론-결론이 명확하며, 논리적 흐름이 자연스러움.
 - 8~9 점: 전체적인 논리는 좋으나 일부 구성이 부족함.
 - 6~7 점: 논리적 흐름이 다소 불완전하며, 연결이 부자연스러움.
 - 5 점 이하: 논리적 전개가 부족하여 이해하기 어려운 답변임.
3. 내용의 구체성 (0~10 점):
- 10 점: 구체적인 예시와 데이터를 활용하여 풍부한 설명을 제공함.
 - 8~9 점: 구체적인 내용이 포함되어 있으나 추가적인 설명이 필요함.
 - 6~7 점: 일부 구체성이 부족하며, 답변이 다소 일반적임.
 - 5 점 이하: 추상적인 답변으로 인해 이해하기 어려움.
4. 문제 해결 접근 방식 (0~10 점):
- 10 점: 문제 해결 프로세스를 체계적으로 설명하고, 적절한 해결책을 제시함.
 - 8~9 점: 문제 해결 과정이 명확하지만, 구체적인 예시가 부족함.
 - 6~7 점: 문제 해결 과정이 부분적으로 설명되었으며, 흐름이 다소 불완전함.
 - 5 점 이하: 문제 해결 과정이 명확하게 설명되지 않음.
5. 핵심 기술 및 직무 수행 능력 평가 (0~10 점):
- 10 점: 지원자의 기술 역량과 실무 경험이 담당 업무 및 지원 자격 요건과 완벽히 부합함.
 - 8~9 점: 기술적 역량과 실무 경험이 대부분 부합하지만, 일부 추가 설명이 필요함.
 - 6~7 점: 기술적 역량은 있지만, 실무 경험이 다소 부족하거나 직접적인 연관성이 적음.
 - 5 점 이하: 기술적 역량이 담당 업무 및 지원 자격 요건과 크게 부합하지 않음.

=== 회사의 담당 업무 및 지원 자격 ===

[담당 업무]

{responsibilities}

[지원 자격]

{qualifications}

평가 결과를 개선사항을 포함하여 **json 형식**으로 반환하세요.

3.2 생성된 평가 예시

```
{
  "질문 이해도": {
    "점수": 10,
    "개선 사항": "질문의 의도를 정확히 파악하고 그에 맞는 답변을
제공하였습니다. 개선 사항은 없습니다."
  },
}
```

	<p>"논리적 전개": { "점수": 9, "개선 사항": "전체적으로 논리적인 흐름을 가지고 있지만, 각 알고리즘을 선택한 이유와 그에 따른 결과를 좀 더 연결시키는 설명이 필요합니다." }, "내용의 구체성": { "점수": 10, "개선 사항": "구체적인 알고리즘 선택 이유와 그 결과를 명확히 제시하였습니다. 개선 사항은 없습니다." ... }</p>
평가 방법	<p>답변 평가 모델의 성능을 검증하기 위해 다음과 같은 평가 지표를 사용하였다.</p> <p>1. 일관성(Consistency) -동일한 답변을 여러 번(총 3 번 실행) 평가하여 점수 산출 후 비교</p> <p>2. 응답의 다양성(Diversity) -AI 평가 피드백이 특정 문장을 반복하는지 확인하기 위해 텍스트 엔트로피 값 계산 -엔트로피 값이 3.5~4.0 수준에서 적절한 범위, 5 이상일 경우 더욱 풍부한 어휘가 사용되었음을 의미</p> <p>3. 설득력(Persuasiveness) -AI 평가 피드백이 쉽게 이해될 수 있는지를 확인하기 위해 Flesh-Kincaid Readability Score(가독성 점수) 계산 - 값이 60~70 이상이면 가독성이 높음을 의미</p>
평가 점수 및 결과	<p>분석 결과</p> <ul style="list-style-type: none"> - 일관성: 같은 질문으로 평가를 총 3 번 진행하였지만 점수 변동은 거의 일어나지 않음 - AI 피드백 1 엔트로피: 3.9069 - AI 피드백 2 엔트로피: 3.5850 - 가독성 점수: 98.25 <ul style="list-style-type: none"> ● 엔트로피 값이 3.5~4.0 사이 이므로 적절한 범위로, 다양한 표현을 사용하고 있지만, 추가적인 어휘 확장도 가능하다 ● 가독성 점수는 90 점 이상으로 평가 피드백 내용은 가독성이 매우 높다