

SK네트웍스 Family AI 과정 20기

데이터 전처리 인공지능 학습 결과서

산출물 단계	데이터 전처리
평가 산출물	데이터 전처리 인공지능 학습 결과서
제출 일자	2026년 2월 2일
깃허브 경로	SKN20-FINAL-5TEAM
작성 팀원	김황현, 박찬, 최소영

1. 모델 선정 및 활용 전략

- 본 프로젝트는 실무적인 엔지니어링 문제를 생성하고 정교한 코칭 피드백을 제공하기 위해, 직접적인 모델 학습 대신 최신 대규모 언어 모델(LLM)을 활용한 프롬프트 엔지니어링 방식을 채택하였습니다.

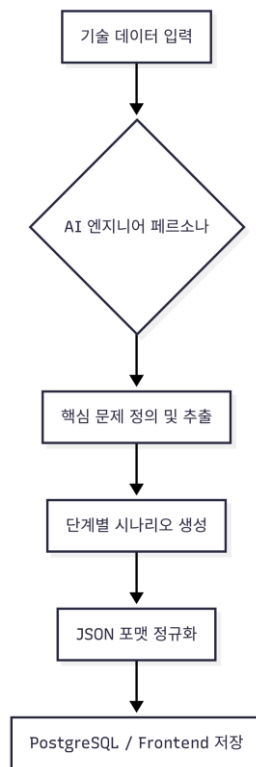
모델명	종류	선정이유
Claude 3.5 Sonnet	LLM	최종 선정. 코딩 능력 및 논리적 추론이 매우 뛰어나며, 시스템 설계 시나리오 생성 시 정교한 JSON 구조 유지 능력이 탁월함
GPT-4o	LLM	다양한 지식 범위를 가졌으나, 복잡한 기술 블로그 분석 및 특정 규격(JSONB) 출력의 안정성 측면에서 비교군으로 검토함

2. 모델 구조 및 아키텍처 (프롬프트 파이프라인)

2.1.1 1, 2번 트랙

직접적인 레이어 설계 대신, **프롬프트 기반의 문제 추출 아키텍처**를 구축했습니다.

2.1.2 프롬프트 워크플로우

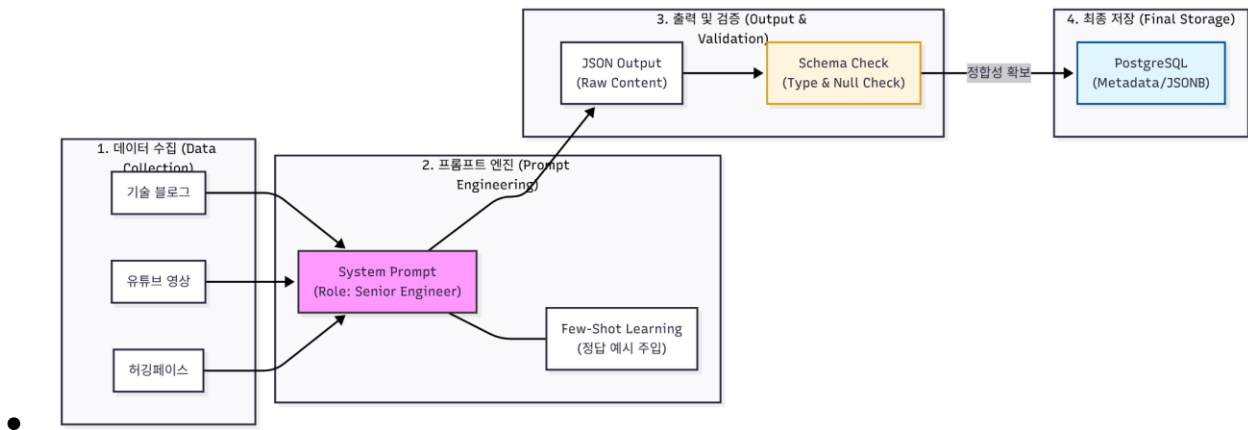


2.1.3 핵심 구성 요소

- **페르소나 설계:** "너는 AI 엔지니어다"라는 지침을 통해 모델이 실무에서 자주 발생하는 예외 상황이나 성능 병목 지점을 핵심 문제로 뽑아내도록 유도.
- **Context Injection:** 기술 블로그나 영상의 로우 데이터를 입력값으로 주어 모델이 근거 없는 답변(Hallucination)을 하지 않도록 제어.

2.2.1 3번 트랙

아래는 비정형 기술 데이터를 구조화된 학습 콘텐츠로 변환하는 LLM 프롬프트 파이프라인입니다.



2.2.2 구성 요소 설명

- **System Prompt:** "10년 차 시니어 엔지니어" 페르소나를 부여하여 전문적인 문제 시나리오 생성 유도.
- **Few-shot Learning:** 모델이 logic_type, scenario 등 정해진 JSON 필드 규격을 엄격히 준수하도록 실제 예시 데이터를 프롬프트에 포함.
- **Schema Validator:** 생성된 데이터가 시스템에서 정의한 데이터 타입(UUID, JSONB 등)과 일치하는지 검증.

3. 학습 설정 및 하이퍼파라미터 (프롬프트 설정)

직접 학습 대신 모델의 응답 품질을 결정짓는 핵심 설정값입니다.

항목	설정 내용	설명
Instruction	"핵심 문제 뽑아줘"	데이터에서 학습 가치가 높은 핵심 요소를 추출하도록 명령

Persona	AI 엔지니어 / 시니어 개발자	전문적이고 실무적인 톤앤매너 유지
Temperature	0.3	일관성 있는 JSON 구조와 논리적 정답을 얻기 위해 낮게 설정
Few-shot	1~2개 예시 주입	모델이 출력 형식을 어기지 않도록 가이드 제공

4. 학습 결과 및 성능 평가 (품질 평가)

모델 학습 대신 생성된 콘텐츠의 유효성을 평가했습니다.

- **평가 방법:** 생성된 50여 개의 연습 문제 샘플 중 무작위 10건을 추출하여 기술적 오류 확인.
- **성능 지표:**
 - **문제 적합성:** AI 엔지니어의 관점에서 실제 면접이나 실무에 나올 법한 문제인가?
 - **형식 준수율:** 시스템(JSON)에 즉시 반영 가능한 형태로 출력되었는가?

5. 과적합/과소적합 대응 (프롬프트 튜닝 결과)

- **과적합 대응:** 모델이 특정 블로그 내용만 그대로 베끼는 현상을 방지하기 위해 "다양한 실무 케이스를 조합하라"는 지침 추가.
- **과소적합 대응:** 문제가 너무 평이하게 나오는 경우 "난이도 별 5단계 기준을 명확히 정의"하여 문제의 깊이를 확보.

6. LLM 평가 시스템 성능 분석

생성된 문제의 품질을 자동으로 평가하기 위해 4가지 평가 시스템을 비교 분석했습니다.

6.1 평가 시스템 비교

평가 방식	유형	장점	단점
NFR 기반 평가	표준형	5가지 비기능 요구사항으로 체계적 평가	문제별 특성 반영 제한적
기능/비기능 분리	심화형	설계와 답변을 독립적으로 평가	평가 구조가 다소 복잡
신속 평가	속도 우선	단일 호출로 빠른 피드백	동적 기준 조정 어려움
2단계 평가	정확도 우선	모범답안 생성 후 비교하여 투명성 확보	API 호출 2회 필요

6.2 테스트 결과

동일한 학생 답변에 대해 각 평가 시스템의 채점 결과를 비교했습니다.

문제	NFR 기반	기능/비기능 분리	신속 평가	2단계 평가
축제 홍보용 단축 URL 생성기	85	85	90	90
오늘의 학식 점심 알림	80	85	90	90
인기 교양 과목 수강 신청	85	85	90	90

6.3 질문 생성 품질 분석

GPT-4o-mini 모델을 활용하여 고급 버전(압박 면접)과 기본 버전(유도 심문) 질문을 각각 생성하고 품질을 분석했습니다.

- 10개 시나리오에 대해 각각 6개 질문 생성(고급 3개 + 기본 3개)
- 고급 버전: 시스템 장애 대응, 비즈니스 제약, 기술적 타당성 등 심화 주제 포함
- 기본 버전: 데이터 흐름 분석, 확장성 검토, 설계 보완점 유도 등 기초 개념 확인
- 테스트 시점에 따라 질문 스타일에 약간의 변화 관찰됨

6.4 핵심 발견사항

- **점수 일관성:** 신속 평가와 2단계 평가 방식이 3개 문제 모두에서 90점으로 가장 높고 일관된 평가 결과를 보였습니다.
- **평가 차별성:** NFR 기반과 기능/비기능 분리 방식은 문제에 따라 80-85점 범위에서 더 세밀한 차등 평가를 수행했습니다.
- **피드백 품질:** 기능/비기능 분리와 2단계 평가 방식이 점수 분해 및 개선방안 제시에서 가장 상세한 피드백을 제공했습니다.
- **모델 안정성:** 질문 생성 시 동일 시점 테스트에서 유사한 패턴이 나타났으며, 프로덕션 환경에서는 일관성 확보를 위한 추가 설정이 필요합니다.

7. 결론

모델을 새로 학습시키지는 않았지만, LLM의 역할을 'AI 엔지니어'로 정의하고 정교한 프롬프트 설계를 통해 고품질의 실무 학습 데이터셋을 구축하였습니다. 이는 전통적인 학습 방식보다 적은 비용으로 전문적인 데이터를 확보하는 효율적인 전략이었습니다.