

SKN 12기 FINAL

3조

5주차 업무 보고

프로젝트 진행 상황

전주 특이사항

- 중간 발표회 성황리 종료
- 목, 일 멘토링 진행 (5.5시간)

제출 서류

- 전주 제출 서류
 - 화면 설계서 (담당 : 이주영)
 - 시스템 구성도 (담당 : 노명구)
 - 중간 발표 자료
- 금주 제출 서류
 - LLM 활용 소프트웨어 (담당 : 노명구, 황차해)

프로젝트 진행 상황

전주 개발 진행 사항

- 데이터 파트
 - 데이터 전처리 및 후처리 코드 정리
 - 파이프라인 제작을 위한 코드 모듈화
- 프론트 파트
 - 화면 설계서 작성, 프론트엔드 작업 착수
- LLM 파트
 - 데이터 파이프라인 제작
 - 모델 선정을 위한 연구 및 비교작업
 - 회의록 작성용 프롬프트 엔지니어링 착수
- 머신러닝 파트
 - Bert 모델 성능 향상

프로젝트 진행 상황

팀원 별 금주 주요 진행 업무

- 노명구 : sLLM 파인튜닝
- 이준석 : 백엔드 연결
- 이주영 : 프론트엔드 작업 (웹 대시보드)
- 박슬기 : 머신러닝(Bert) 모델 고도화
- 황차해 : sLLM 파인튜닝, 프롬프트 엔지니어링

금주 업무 목표

- 차주 있을 자체 sLLM 모델 및 테스트 자료제출을 위해 sLLM 모델 개발을 금주중으로 끝내는 것을 목표로 함
 - 모델 개발이 1차적으로 완료됨과 동시에 QA 시작 예정
- 차주에 웹 애플리케이션 제출도 있으므로 백엔드 및 프론트엔드 개발 착수
 - 모델 개발 완료 시 바로 QA 병행

프로젝트 진행 상황

화면설계

SKN12-Final-3Team

화면명	메인	화면 ID	SCR-MAIN-01	요구사항 ID	REQ-MAIN-001	작성날짜	2025.07.25
화면 경로	/메인					작성자	이주영
<div> <div>1</div> <div>2</div> <div>3</div> <div>4</div> <div>5</div> </div>	<div> <div> <div>오늘의 프로젝트 현황</div> <div> <div>총 처리 건수</div> <div>12건</div> <div>처리율 100%</div> </div> <div> <div>평균 소요 시간</div> <div>20초</div> <div>20초 미만</div> </div> <div> <div>처리율</div> <div>95%</div> <div>95% 이상</div> </div> <div> <div>활성 사용자</div> <div>0명</div> <div>1명 이상</div> </div> </div> <div> <div>3</div> <div>버튼 액션</div> <div> <div>음성 녹음 업로드</div> <div>문서 분석</div> <div>수동 업무 추가</div> </div> </div> <div> <div>4</div> <div>업무 현황</div> <div> <div>WBS 문서 작성</div> <div>요구사항 정의서(초안)</div> <div>요구사항 협의서</div> </div> <div> <div>프로젝트 기획서</div> <div>수집 데이터</div> <div>데이터 베이스 설계분석</div> <div>데이터 조회 프로그램</div> </div> <div> <div>인공지능 데이터 처리 결과서</div> <div>인공지능 학습 결과서</div> <div>학습된 인공지능 모델</div> <div>수집된 데이터 및 전체 리포트</div> <div>시스템 아키텍처</div> </div> </div> </div>					Description	
						1	각 메뉴 클릭 시 해당 화면으로 이동
						2	각 항목별 아이콘과 수치, 증감률 및 상태 표시
						3	각 버튼 클릭 시 해당 기능 실행
						4	필터/검색으로 업무 리스트를 실시간 확인하고, 상태를 색상으로 시각화
						5	활동 내역을 시간 순으로 보여주며, 클릭 시 상세 확인/화면 이동 가능

프로젝트 진행 상황

화면설계

SKN12-Final-3Team

화면명	연동	화면 ID	SCR-INTEG-01	요구사항 ID	REQ-INTEG-001	작성날짜	2025.07.25																									
화면 경로	/연동					작성자	이주영																									
<div><div><div><div>ThaIKkak</div><div>AI 교육자료 관리</div></div><div><div>메시지도</div><div>2건의 연동 보기</div></div><div><div>활동 상태</div><div>전체 활동 관리하기</div></div><div><div>업무 관리</div><div>업무 현황 및 관리하기</div></div><div><div>연동</div><div>연동 서비스 등록</div></div><div><div>설정</div><div>시스템 설정</div></div></div></div> <div><div>연동</div><div>1</div><div><div>슬랙(Slack)</div><div>채널: SKN12_FINAL_3TEAM</div><div>최근 접속: 2025.07.23 03:03</div><div>메시지 알림 상태 알림 연결 해제</div></div><div>2</div><div><div>노션(Notion)</div><div>채널: SKN12_FINAL_3TEAM</div><div>최근 접속: 2025.07.23 05:41</div><div>메시지 알림 상태 알림 연결 해제</div></div><div>3</div><div><div>지라(Jira)</div><div>채널: SKN12_FINAL_3TEAM</div><div>최근 접속: 2025.07.23 05:00</div><div>메시지 알림 상태 알림 연결 해제</div></div><div>4</div><div><div>연동 추가 버튼</div><div>Drop Here</div><div>1 서비스 선택</div><div>슬랙, 노션, 지라</div><div>2 계정 연결</div><div>OAuth 인증</div><div>3 권한 대상 선택</div><div>- 노션 페이지</div><div>- 슬랙 채널 등록</div><div>4 기존 알림 조건 재연속</div><div>- 채널이 변경, 수정</div><div>뒤로 가기 취소 완료 & 테스트</div></div></div> <div><table><thead><tr><th>시간</th><th>연동명</th><th>상태</th><th>상세내용</th></tr></thead><tbody><tr><td>2025.07.23 15:30</td><td>프로젝트 알림</td><td>성공</td><td>지라와 페이시 관련 알림</td></tr><tr><td>2025.07.23 15:36</td><td>교과 분석 현황</td><td>실패</td><td>슬랙 채널 최근 관련 오류</td></tr><tr><td>2025.07.23 15:10</td><td>주간 보고서</td><td>성공</td><td>CRM 동기화 완료</td></tr></tbody></table></div>	시간	연동명	상태	상세내용	2025.07.23 15:30	프로젝트 알림	성공	지라와 페이시 관련 알림	2025.07.23 15:36	교과 분석 현황	실패	슬랙 채널 최근 관련 오류	2025.07.23 15:10	주간 보고서	성공	CRM 동기화 완료	<div>Description</div> <table><tr><td>1</td><td>각 서비스별 활성/예러/비활성 상태 아이콘 및 최근 접속 시간 표시, '테스트 알림', '상세 설정', '연결해제' 버튼 제공</td></tr><tr><td>2</td><td>단계별 안내에 따라 서비스 선택, 계정 연결, 대상 선택, 알림 조건 설정 진행, '뒤로 가기', '취소', '완료 & 테스트' 버튼 제공</td></tr><tr><td>3</td><td>간단한 질문과 답변 제공</td></tr><tr><td>4</td><td>성공/실패 상태 색상 태그로 표시, 상세 내용 확인 가능</td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr></table>		1	각 서비스별 활성/예러/비활성 상태 아이콘 및 최근 접속 시간 표시, '테스트 알림', '상세 설정', '연결해제' 버튼 제공	2	단계별 안내에 따라 서비스 선택, 계정 연결, 대상 선택, 알림 조건 설정 진행, '뒤로 가기', '취소', '완료 & 테스트' 버튼 제공	3	간단한 질문과 답변 제공	4	성공/실패 상태 색상 태그로 표시, 상세 내용 확인 가능						
	시간	연동명	상태	상세내용																												
	2025.07.23 15:30	프로젝트 알림	성공	지라와 페이시 관련 알림																												
	2025.07.23 15:36	교과 분석 현황	실패	슬랙 채널 최근 관련 오류																												
	2025.07.23 15:10	주간 보고서	성공	CRM 동기화 완료																												
1	각 서비스별 활성/예러/비활성 상태 아이콘 및 최근 접속 시간 표시, '테스트 알림', '상세 설정', '연결해제' 버튼 제공																															
2	단계별 안내에 따라 서비스 선택, 계정 연결, 대상 선택, 알림 조건 설정 진행, '뒤로 가기', '취소', '완료 & 테스트' 버튼 제공																															
3	간단한 질문과 답변 제공																															
4	성공/실패 상태 색상 태그로 표시, 상세 내용 확인 가능																															

프로젝트 진행 상황

회의록 생성 - 1차 테스트

```
{
  "timestamp": 1753056019000,
  "timestamp_order": "4-1",
  "speaker": "SPEAKER_05",
  "text": "[TGT] 세  
계부지 모델 전철이가 이번 주까지예요? [/TGT]",
  "label": 0
}
{"timestamp": 1753056024000, "timestamp_order": "5-1", "speaker": "SPEAKER_01", "text": "[TGT] 전  
철이 이번 주까지가 맞습니다. [/TGT]", "label": 0}
{"timestamp": 1753056027000, "timestamp_order": "6-1", "speaker": "SPEAKER_05", "text": "[TGT] 이  
공식 모델 습도 이번 주까지네요? [/TGT]", "label": 0}
{"timestamp": 1753056029000, "timestamp_order": "7-1", "speaker": "SPEAKER_01", "text": "[TGT] 네  
맞습니다 일단은 저희 금요일까지로 알고 있는데 네 조금 시간이 빠듯한 부분이 있습니다 그래서 좀  
제가 지금 보내드린 거가 저희 산출물 관리하는 구글 스프레드 시트 인데요 [/TGT]", "label": 0}
{"timestamp": 1753056054000, "timestamp_order": "8-1", "speaker": "SPEAKER_01", "text": "[TGT] 여  
기 보면 주차라고 되어 있는 게 저희 제출하는 주차고 이번 주차 3주차여서 지금 5개 있습니다.  
[/TGT]", "label": 0}
{"timestamp": 1753056076000, "timestamp_order": "10-1", "speaker": "SPEAKER_01", "text": "[TGT] 다  
른 조우들도 저희랑 조금 비슷한 상황이긴 한데요. [/TGT]", "label": 0}
{"timestamp": 1753056080000, "timestamp_order": "11-1", "speaker": "SPEAKER_01", "text": "[TGT] 인  
공지능을 전혀 생각 안 하고 그냥 채집팀 API를 생각하고 있다가 멘토님들이 이런 이런 부분들 넣는  
게 좋겠다 해가지고 지금 오늘부터 들어간 팀들도 있고 몇몇 군데에서 그런 얘기를 들었는데 거의 다  
버튼 모델을 기반으로 생각을 하고 있더라고요. [/TGT]", "label": 0}
{"timestamp": 1753056106000, "timestamp_order": "13-1", "speaker": "SPEAKER_05", "text": "[TGT] 이  
번에 설명드린 대로만 얘기한다면 버튼 모델이 되겠네요. [/TGT]", "label": 0}
{"timestamp": 1753056115000, "timestamp_order": "17-1", "speaker": "SPEAKER_01", "text": "[TGT]  
어, 일단 저 인공지능 데이터 전설의 결과서는 오늘 제가 초안을 작성을 했는데, 이거 일단 제가 보여  
드리면서 같이 말씀을 나눌까요? [/TGT]", "label": 0}
```



요약

- 회의는 인공지능 모델을 사용하여 회의록에서 필요한 문장을 분류하는 작업에 대한 논의였다. 현재 사용 중인 데이터는 2,000개의 문장으로 구성되어 있으며, 이 데이터는 회의와 관련된 문장(1)과 관련되지 않은 문장(0)으로 라벨링되었다. 그러나 이 데이터는 실제 회의록과는 차이가 있으며, 라벨링 기준이 명확하지 않아 모델 성능이 낮은 것으로 보인다. 다양한 모델(캐시버트, TF-IDF, 로지스틱 회귀, 클로버트, 모놀로그 페이워 일렉트라)을 비교했지만, 정확도 차이가 크지 않았고, 데이터셋의 문제로 인해 모델 선택이 어려운 상황이다. 또한, 인퍼런스 시 실제 데이터의 노이즈 처리가 필요하며, 레이블링 작업을 GPT를 활용해 개선할 계획이다.

안건

1. 제목: 인공지능 모델 학습 데이터 개선
 - 현재 사용 중인 학습 데이터는 실제 회의록과 차이가 있으며, 라벨링 기준이 명확하지 않아 모델 성능이 낮다. 따라서 데이터셋을 실제 회의록 기반으로 재구성하고, GPT를 활용해 레이블링 작업을 개선해야 한다.
 - 관련 발언자: SPEAKER_01, SPEAKER_05
2. 제목: 모델 성능 평가 기준 명확화
 - 현재 모델 평가 기준이 명확하지 않아 성능 비교가 어렵다. 정확도, 리콜, F1 스코어 등 다양한 지표를 고려해 평가 기준을 설정해야 한다.
 - 관련 발언자: SPEAKER_01, SPEAKER_05
3. 제목: 인퍼런스 시 노이즈 처리 방안 마련
 - 학습 데이터는 깨끗하지만, 실제 인퍼런스 데이터에는 노이즈가 포함되어 있어 모델이 헛갈릴 수

기존 유효 발화 데이터

생성된 회의록
청킹 겹침으로 인한 **중복 요약 문제**
발생

프로젝트 진행 상황

프롬프트 엔지니어링

회의는 제427회 국회 제2차
농림축산식품해양수산위원회에서 진행되었으며,
전재수 해양수산부장관 후보자에 대한 인사청문회가
주요 안건이었습니다.

주요 내용:

- 농업재해보험법 개정안 상정 보류 요청 (전종덕 위원)
- 해수부 부산 이전에 대한 논의
- 북극항로 개척의 중요성 강조
- 어민 부채 문제와 수산업 현안 논의
- 해양환경 보호 및 외국인 노동자 문제 다뤄짐

기본 프롬프트 결과

회의 요약:

전재수 해양수산부장관 후보자 인사청문회가 진행되었으며,
해수부 부산 이전과 북극항로 개척에 대한 논의가 중심이 되었다.

생성된 5개 액션 아이템:

1. 해수부 부산 이전 관련 직원들과의 공식 소통 기구 구성 (담당: 전재수 후보자, 마감: **2023-12-31**, 우선순위: **high**)
2. 어가 부채 문제 해결을 위한 정책 검토 (담당: 해양수산부, 마감: **2024-03-31**, 우선순위: **high**)
3. 북극항로 개척을 위한 연구용역 발주 (담당: 해양수산부, 마감: **2024-01-31**, 우선순위: **medium**)
4. 해양수산부의 재생에너지 정책 강화 (담당: 해양수산부, 마감: **2024-06-30**, 우선순위: **medium**)
5. CITES 관련 뱀장어와 관련하여 대응 방안 마련 (담당: 해양수산부, 마감: **2024-02-28**, 우선순위: **medium**)

주요 결정사항:

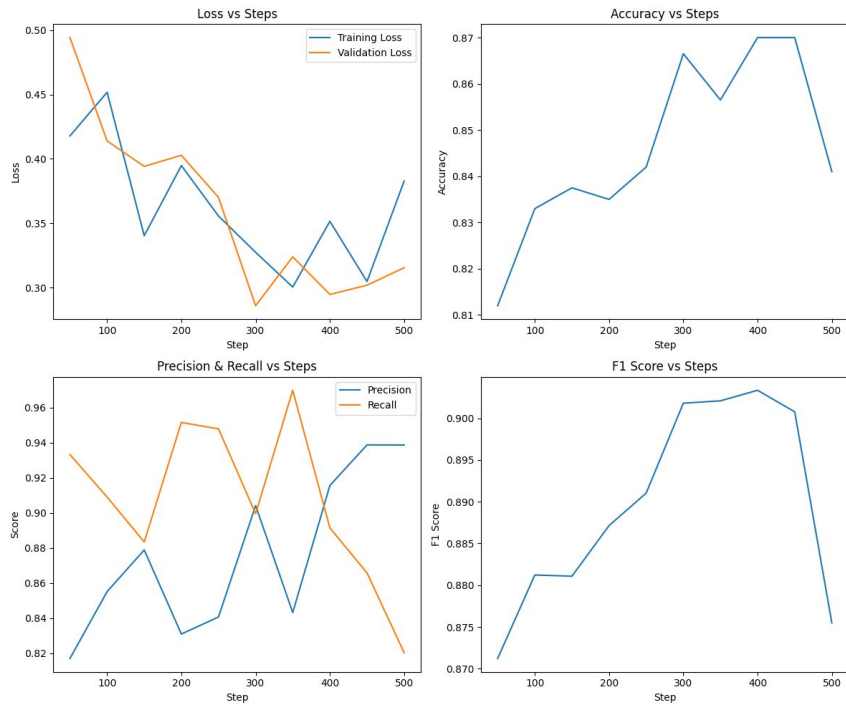
- 해수부 부산 이전을 추진하기로 결정
- 북극항로 개척을 위한 연구용역 발주 결정

참석자: 전재수 후보자, 어기구 위원장, 정희용 간사, 서삼석 위원,
조경태 위원, 이원택 위원, 이병진 위원, 문대림 위원, 조승환 위원,
임미애 위원,
김선교 위원, 전종덕 위원, 참고인 이상국

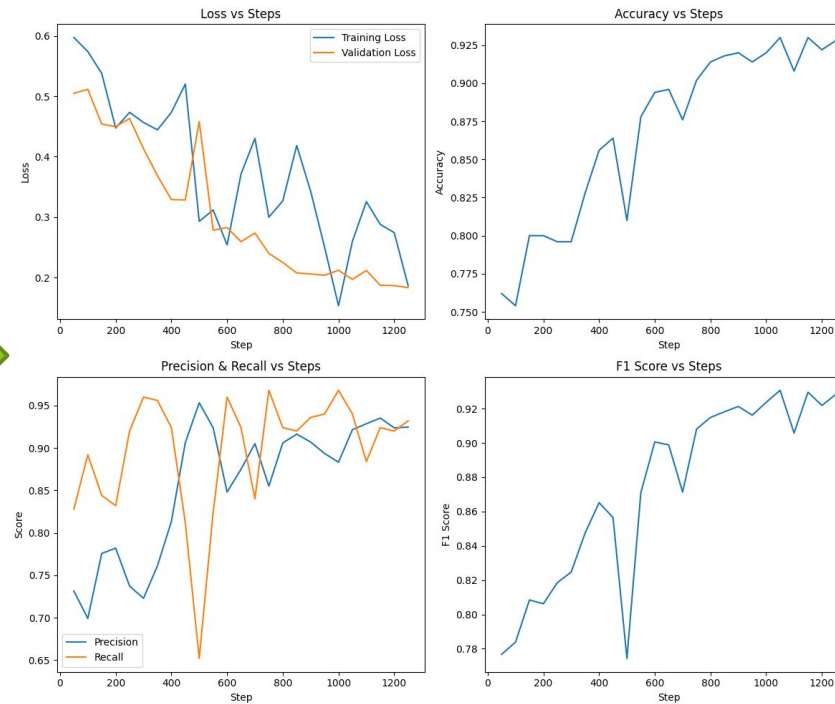
구조화된 프롬프트 결과

프로젝트 진행 상황

머신러닝 파트 - 튜닝 고도화



기존
최고 F1 score 0.90



데이터 정제 후
최고 F1 score 0.93

프로젝트 진행 상황

머신러닝 파트 - 테스트 데이터 리포트

	precision	recall	f1-score	support
유효 발화(0)	0.7871	0.7694	0.7781	980
무효 발화(1)	0.8433	0.8563	0.8498	1420
accuracy			0.8208	2400
macro avg	0.8152	0.8129	0.8139	2400
weighted avg	0.8203	0.8208	0.8205	2400

기준



	precision	recall	f1-score	support
유효 발화(0)	0.7956	0.7745	0.7849	980
무효 발화(1)	0.8472	0.8627	0.8548	1420
accuracy			0.8267	2400
macro avg	0.8214	0.8186	0.8199	2400
weighted avg	0.8261	0.8267	0.8263	2400

데이터 정제 후
0.6~1% 성능 향상 확인

프로젝트 진행 상황

전체 데이터 처리 Pipe Line 구축 완료

