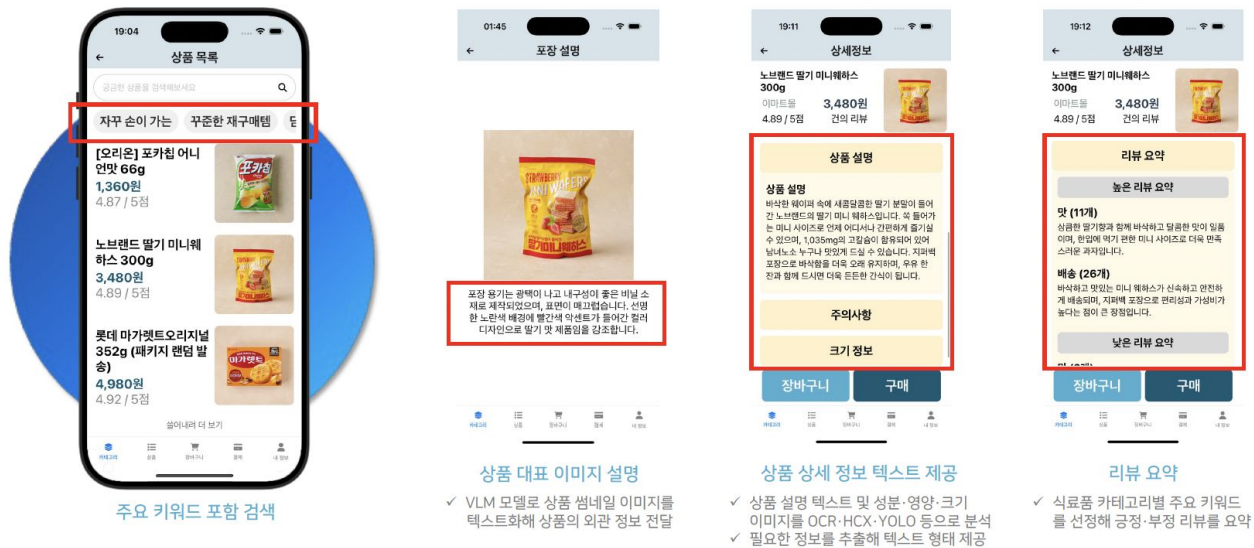


Wrap-Up Report: Company Hackathon

기업 해커톤



시각장애인을 위한 온라인 식료품 쇼핑 서비스 ‘푸드리 (Foodly)’ 제안



‘나야, 자, 연어’팀
곽희준, 김정은, 김진재, 오수현, 윤선웅, 정민지

목차

1. 서비스 소개

- 주제 선정 배경
- 푸드리(Foodly) 소개
- 팀원 소개 및 역할

2. 문제 정의 구체화

- 사용자 요구 분석
- EDA
- 타 서비스와의 차별점

3. 서비스 아키텍처

- 전체 파이프라인 소개
- 5가지 핵심 기능의 기술 소개
 - 1. 대표 이미지 설명
 - 2. 상품 텍스트 요약
 - 3. 크기 정보 설명
 - 4. 성분/영양 정보
 - 5. 리뷰 요약 및 키워드 기반 상품 추천

4. AI Safety

- Hallucination
 - 부정적 내용 필터링
 - 정보 정확성
 - 예측 가능성
- AI Ethics
 - 사람을 위한 AI 개발
 - 다양성 존중
 - 합리적인 설명과 편리성의 조화

5. 결과 및 회고

- 결과물
- 시각장애인의 사용성 평가 분석
- 프로젝트 타임라인

6. 개인 회고

- 곽희준
- 김정은
- 김진재
- 오수현
- 윤선웅
- 정민지

1. 서비스 소개

주제 선정 배경

본 프로젝트는 시각장애인 분들이 식료품 구매 시 겪는 정보 접근의 어려움을 해소하고, 온라인 쇼핑몰의 접근성을 개선하기 위해 기획되었습니다.

시각장애인 분들께서는 영양성분, 알레르기 정보 등 상세한 식료품 정보를 필요로 하나, 온·오프라인 모두에서 이를 독립적으로 확인하기 어렵습니다. 설문 결과, 식료품은 가장 상세한 설명을 원하는 제품으로 나타났으며, 오프라인 구매 시 조사 대상 식품 중 62.3%에서 점자 표기가 미흡하거나 모호하게 제공되고 있는 실정입니다. 또한, 온라인 쇼핑몰에서는 필수 정보가 이미지로만 제공되고 대체 텍스트가 부족하여, 영양 성분 및 원재료 정보를 확인하기 어려운 문제가 있습니다.

식료품의 기본 정보 습득 어려움

✓ 오프라인 구매

- 조사 대상 321개 식품(음료, 컵라면, 우유 등) 중 62.3%가 점자 표기 미제공
- 점자 표기가 있어도 “음료”, “탄산” 등 모호한 표기 제공

✓ 온라인 식료품 쇼핑

- 필수 정보가 이미지 형태로만 제공되고, 대체 텍스트 미흡
(예: “32 페이지 중 1”)



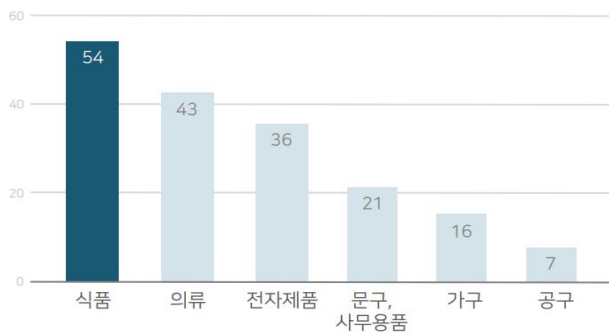
□ 음료는 제품명 대신 ‘음료’ 또는 ‘탄산’으로 점자 표시하는 경우 많아

박정태, (2022). 시각장애인 식품 점자 표기 소비자문제 실태조사. 조사보고서, 1-66.

식료품 정보에 대한 니즈

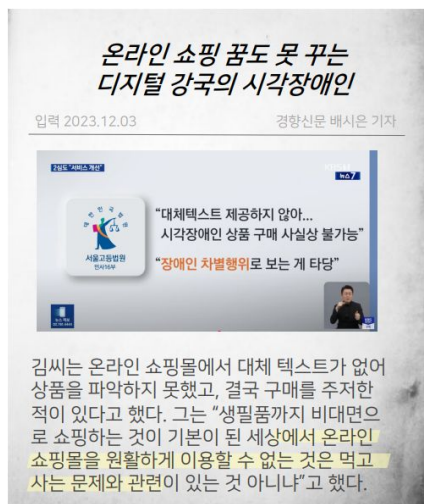
✓ 시각장애인이 가장 자세한 설명을 원하는 제품 유형은 식료품 (78%)

가장 상세 설명을 제공받고 싶은 제품 유형 (시각장애인 73명 대상 설문)

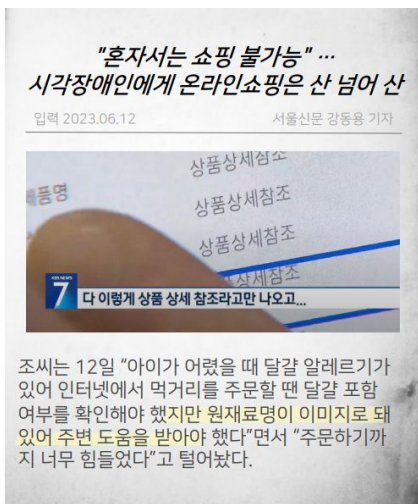


Lee et al. (2021). Designing product descriptions for supporting independent grocery shopping of people with visual impairments. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-6).

더불어, 스크린리더 사용 시 모든 텍스트를 순차적으로 읽어야 하는 불편함과 하단 정보 접근의 어려움으로 인해, 시각장애인 분들은 온라인 쇼핑 이용에 많은 제약을 받고 있습니다. 실제 사례에서도 대체 텍스트 부재로 상품 정보를 파악하지 못해 구매를 주저하거나, 생필품 주문 시 주변 도움에 의존하는 경우가 실제로 존재한다는 것을 설문조사를 통해 파악하였습니다.



<https://www.khan.co.kr/article/202312031347001>



<https://www.seoul.co.kr/news/society/2023/06/12/20230612500255>

스크린리더와의 낮은 호환성

- ✓ 스크린리더를 사용하여 버튼, 제목을 포함한 모든 텍스트를 순차적으로 읽어야 함
- ✓ 원하는 정보 도달까지 많은 시간이 소요
- ✓ 일부 하단에 있는 정보(리뷰, 교환 및 반품 등) 습득 어려움

대체 텍스트 제공 미흡

- ✓ 필수 정보가 이미지로 제공되며, 대체 텍스트 등 설명 부재
- ✓ 식료품의 경우, 영양 성분과 원재료명 등 필수 정보 습득 불가

주변 도움 없이 쇼핑 어려움

- ✓ 도우미, 가족 등 주변인의 도움 필수
- ✓ 독립적으로 상품 구매 불가

따라서 본 프로젝트는 시각장애인 분들이 독립적으로 식료품 정보를 확인할 수 있도록 지원하며, 온라인 쇼핑몰의 접근성을 개선하는 기술 사례를 개발하여 자립적인 소비 생활과 디지털 포용 사회 구현에 기여하고자 합니다.

1. 서비스 소개

푸드리 (Foodly) 소개

푸드리(Foodly)는 시각장애인의 독립적인 쇼핑 활동을 돕기 위한 AI쇼핑 서비스입니다. 본 서비스는 상품의 AtoZ 정보를 자동화하여 제공하며, 사용자는 필요한 제품 정보를 화면 낭독 프로그램을 통해 손쉽게 들을 수 있습니다. VLM을 활용해 상품 대표(썸네일) 이미지를 텍스트로 변환하고, 상품 설명, 상품 크기 정보, 성분/영양 정보 등을 분석하여 텍스트 형태로 제시합니다. 또한, 리뷰 요약을 통해 상품에 대한 긍정적 및 부정적 반응을 제공하고 상품 추천 키워드를 통해 검색 추천 기능을 지원하여 간편한 쇼핑 경험을 제공합니다.

제안 서비스 푸드리(Foodly)

시각장애인의 독립적인 쇼핑 활동을 위해 상품의 A to Z 정보를 자동화해 제공하는 AI 쇼핑 서비스 Foodly 개발



팀원 소개 및 역할

팀원	역할
곽희준	리뷰 요약 및 키워드 추출: ASTE(파이프라인 설계, 데이터 엔지니어링, Metric 선정), Clustering을 통한 추천 키워드 포함 검색과 리뷰 요약
김정은	성분/영양 정보 추출: YOLO11 SFT, CLOVA OCR output 후처리, Rule-based 방식 적용 실험, HCX Fine-Tuning, 평가 Metric 선정
김진재	크기 정보 묘사, 리뷰 요약 및 키워드 추출, 앱 개발: (팀장) 프로젝트 매니징, YOLO11 SFT, Rule-based 후처리, ASTE(HCX/DeepSeek Prompt Engineering, DeepSeek SFT), React Native, React (Chrome Extension) Spring Framework 개발
오수현	리뷰 요약 및 키워드 추출: ASTE(HCX/DeepSeek Prompt Engineering, DeepSeek SFT), Clustering을 통한 추천 키워드 포함 검색과 리뷰 요약
윤선웅	대표 이미지 설명 생성: Janus Pro Fine-Tuning, HCX 후처리(요약, 번역, Hallucination 제거), 1376개 대표 이미지 골드라벨 추출, VLM 평가 metric 설계
정민지	상품 설명 요약, 성분/영양 정보 추출: 상품 설명 요약 HCX Fine-Tuning, 성분/영양 정보 평가 metric 설계 및 golden label 생성, OCR과 LLM을 활용한 성분/영양 정보 추출 로직 설계 및 실험

2. 문제 정의 구체화

사용자 요구 분석

저희는 시각장애인 분들의 의견을 직접 수렴하여, 서비스 필요성을 확인하고 방향성을 확립하였습니다. 온라인 설문조사는 2025년 1월 21일부터 27일까지 시각장애인 온라인 커뮤니티 '넓은마을'에서 진행되어 301명의 응답을 받았으며, 1월 24일에는 실로암시각장애인복지관에서 'Talkback 사용법 안내' 봉사활동을 통해 실제 사용자 경험을 파악하였습니다.

온라인 설문조사

- ✓ 시각장애인 온라인 커뮤니티 '넓은마을'에서 설문조사 진행
- 일주일 (2025.01.21~27) 간 시각장애인 301명 설문 응답
- ✓ 이를 통해 푸드리 서비스 주요 기능 구체화

▲ 설문조사 요청을 위해 '넓은마을' 커뮤니티에 업로드한 글

교육 봉사

- ✓ 1/24, 실로암시각장애인복지관에서 진행된 시각장애인 대상 '음성 안내 서비스 Talkback 사용법 안내' 봉사활동 참여
- ✓ 이를 통해 시각장애인을 위한 앱 구성에 포함되어야 할 항목 파악



▲ 안드로이드 폰 ScreenReader 앱 'Talkback' 사용법 안내 봉사활동에 참여한 모습

주요 평가 항목

온라인 식품 쇼핑 경험

- ✓ 쇼핑 방식 파악
- ✓ e-commerce 이용 시 겪었던 불편 사항 조사

필요 정보 및 기능

- ✓ 습득하고자 하는 정보의 종류·양 파악
- ✓ Foodly에서 제공할 정보·기능 구체화

UX 관점

- ✓ 온라인 쇼핑물 접근성 개선을 위한 의견 수렴

UI 관점

- ✓ 최적의 인터페이스 구성 및 컴포넌트 배치 탐색

설문 결과, 사용자 분들은 상품 이미지 및 상세 정보 확인의 어려움(66%), 리뷰와 평점 정보 부족(38.5%)을 주요 문제로 지적하였습니다. 또한, 상품 구매 시 가격, 상품명, 배송정보와 함께 외관 정보(색상, 디자인 등)와 리뷰 요약 정보가 필요하다는 의견(각각 92% 이상, 93.4%)을 주셨습니다. UX 측면에서는 불필요한 기능을 최소화하고, 간결하고 직관적인 정보 제공(예: 상품 상세 설명은 5~7문장)이 요구되었으며, UI 측면에서는 익숙한 인터페이스와 핵심 정보 우선 배치, 스크린리더를 통한 순차적 탐색이 필요하다는 의견이 많았습니다. 이러한 결과를 바탕으로, 푸드리 서비스는 사용자 요구에 부합하는 기능과 인터페이스를 구현하였습니다.

온라인 식품 쇼핑 경험

✓ 주요 불편 사항

- 상품 이미지, 상세 정보 확인 어려움 (66%)
- 리뷰, 평점 정보 부족 (38.5%)

✓ 서비스 반영 내용

- 상품 이미지 및 상세 정보 텍스트로 상세히 제공
- 리뷰 요약해 제공함으로써 리뷰 접근성 강화

UX 관점

✓ UX 측면 요구 사항

- 불필요한 기능 최소화: 부가 기능 제거, 핵심 정보 중심 설계
- 상품 상세 설명 길이: 5~7문장 정도 길이가 적절 (62.5%)
- 정보 형식: 영양성분은 표보다 줄글 형식이 이해하기 쉽다는 응답 (60.4%)

✓ 서비스 반영 내용

- 직관적이고 간결하게 기능 구현
- 표 대신 줄글 형식으로, 너무 길지 않게 텍스트 정보 제공

필요 정보 및 기능

✓ 필요 정보

- 상품 구매 시 가격, 상품명, 배송정보 중시
- 상품 외관 정보(색상, 디자인 등) 필요 (92%)
- 리뷰 요약 정보 필요 (93.4%)

✓ 서비스 반영 내용

- 화면 최상단에 가격, 상품명, 배송정보 배치
- 상품 외관 정보를 텍스트 설명으로 제공

UI 관점

✓ UI 측면 요구 사항

- 익숙한 인터페이스 유지: 일반적 구조 활용 (예: 제목은 항상 최상단에 배치)
- 핵심 정보 우선 배치: 식품 구매 시 가장 중요한 정보 우선 제공
- 순차 탐색 가능 설계: 스크린리더로 자연스럽게 따라가며 정보 얻는 흐름

✓ 서비스 반영 내용

- 기존 앱 인터페이스와 유사한 구조 채택
- 단계별, 순차적 정보 제공으로 직관적으로 상품 정보 탐색 지원

2. 문제 정의 구체화

EDA

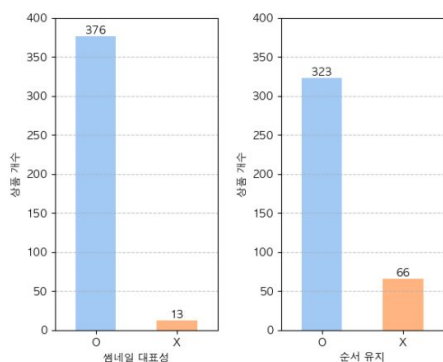
본 연구에서는 설문 조사와 봉사활동에서 얻은 인사이트를 바탕으로, 온라인 쇼핑몰의 정보 제공 방식을 분석하기 위해 데이터 수동 레이블링과 EDA를 수행하였습니다. Google Sheets를 활용하여 네이버 쇼핑의 주요 쇼핑몰인 이마트몰에서 389개 식료품, 총 1,364개 이미지에 대해 수동 레이블링을 진행하였습니다.

R256																	
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	ID	img-ID	카테고리	설명	상문 상세 URL	연재/개방	이러 URL	손서 표시	한글판 내로	상문명 개수	상문명 개수	조문 필요 여부	OCR 필요 여부	상문 설명	조기 정보 제공 여부	조기 정보 제공 여부	
298	emart-01	emart-01-0	국자명	노자본, 백자본, 흑자본 등 11개	https://chinese.net	연재	https://chinese.net	O	O	단행	단행						
299	emart-01	emart-01-1					https://chinese.net	O				X	X	X	X	X	
300	emart-01	emart-01-2					https://chinese.net	O				X	X	X	X	X	
301	emart-01	emart-01-3					https://chinese.net	O				X	X	X	X	X	
302	emart-01	emart-01-4					https://chinese.net	O				X	X	X	X	O	
303	emart-01	emart-01-5					https://chinese.net	O									
304	emart-01	emart-01-6					https://chinese.net	O									
305	emart-02	emart-02-0	국자명	노자본, 지, 백자본 등 11개	https://chinese.net	연재	https://chinese.net	O	O	단행	단행						
306	emart-02	emart-02-1					https://chinese.net	O				X	X	X	X	X	
307	emart-02	emart-02-2					https://chinese.net	O				X	X	X	X	X	
308	emart-02	emart-02-3					https://chinese.net	O				X	X	X	X	O	
309	emart-02	emart-02-4					https://chinese.net	O									
310	emart-02	emart-02-5					https://chinese.net	O									
311	emart-02	emart-02-6	국자명	노자본, 지, 백자본 등 11개	https://chinese.net	연재	https://chinese.net	X	O	단행	단행						
312	emart-03	emart-03-1					https://chinese.net	X				X	X	X	X	X	
313	emart-03	emart-03-2					https://chinese.net	X				X	X	X	X	X	
314	emart-03	emart-03-3					https://chinese.net	X				X	X	X	X	O	
315	emart-03	emart-03-4					https://chinese.net	X									
316	emart-03	emart-03-5					https://chinese.net	X									
317	emart-03	emart-03-6					https://chinese.net	X									
318	emart-04	emart-04-0	국자명	백자본, 노자본 등 11개	https://chinese.net	연재	https://chinese.net	O	O	단행	단행						
319	emart-04	emart-04-1					https://chinese.net	O				X	X	X	X	X	
320	emart-04	emart-04-2					https://chinese.net	O				X	X	X	X	X	
321	emart-04	emart-04-3					https://chinese.net	O				X	X	X	X	O	
322	emart-04	emart-04-4					https://chinese.net	O									
323	emart-04	emart-04-5					https://chinese.net	O									
324	emart-05	emart-05-0	국자명	노자본, 지, 백자본 등 11개	https://chinese.net	연재	https://chinese.net	O	O	단행	단행						
325	emart-05	emart-05-1					https://chinese.net	O				X	X	X	X	X	
326	emart-05	emart-05-2					https://chinese.net	O				X	X	X	X	X	
327	emart-05	emart-05-3					https://chinese.net	O				X	X	X	X	O	
328	emart-05	emart-05-4					https://chinese.net	O									
329	emart-05	emart-05-5					https://chinese.net	O									
330	emart-06	emart-06-0	국자명	노자본, 지, 백자본 등 11개	https://chinese.net	연재	https://chinese.net	O	O	단행	단행						

분석 대상은 제공 정보의 순서 유지, 썸네일 대표성, 크기 정보, 그리고 영양/성분 정보의 제공 방식 등이며, 이를 토대로 적절한 AI 기술 매칭 가이드를 구축하였습니다. 구체적으로, 크기 정보는 비교군 제공율 97.83%를 확인하였고, 보관 정보는 77.71%의 상품에서 성분 표기가 이루어짐을 파악하였습니다. 또한, 전체 389개 상품 중 83.03%(323개)에서 상세 정보의 정형성이 확인되어, “상세 이미지 → 크기 정보 → 영양/성분 정보” 순서로 제공되는 구조와 썸네일 대표성이 평가되었습니다.

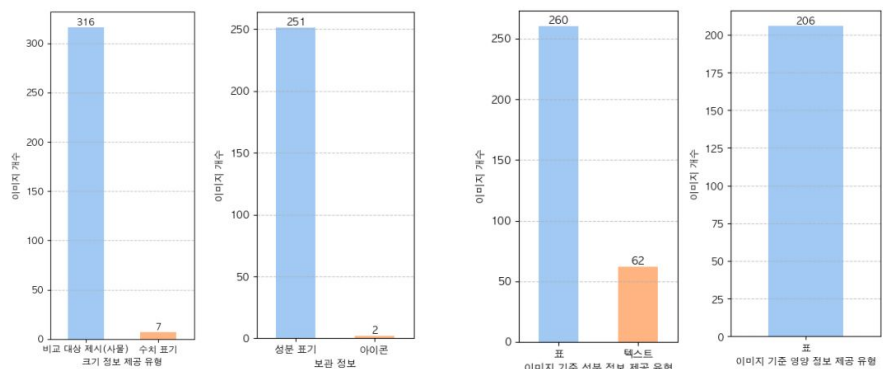
이와 같이 확보된 기초 데이터를 바탕으로, 성분/영양 정보는 이미지, 텍스트, 표 방식으로 제공되며, 노이즈(없음, 가려짐, 왜곡, 반사, 회전) 유형에 따른 모델 성능을 추가로 분석하였습니다. 이를 통해 향후 시스템 설계와 AI 기술 적용에 필요한 구체적인 가이드를 마련하였습니다.

상품 메타 정보의 정형성 확인



- 총 389개 상품 중 83.03%(323개)의 상품 상세 정보 정형성 확인
- 순서 유지: 상세 이미지 → 크기 정보 → 영양/성분 정보
- 썸네일 대표성: VLM, IC, VQA 성능 비교
- 이마트 상품 설명 생성에 대한 정형화된 파이프라인 구축

상품 상세 설명 유형별 결론 도출



크기 정보: 비교군 제공 (97.83%)

- 상품과 비교군의 상대적 크기 추출
- VLM 모델 성능 평가, YOLO, Bbox 크기 비교

보관 정보: 성분 표기에 제공 (77.71%)
- 영양/성분 표기의 OCR + LLM 성능과 연관

성분/영양 정보 제공 유형

- 이미지(100%)
- 텍스트: OCR + LLM
- 표: YOLO + OCR + LLM

성분/영양 정보 노이즈 유형

- 없음, 가려짐, 왜곡, 반사, 회전
- 개별 노이즈 유형에 대한 모델의 성능 파악

2. 문제 정의 구체화

타 서비스와의 차별점

타 서비스와 차별화되는 점은 이미지 대체 텍스트 제공, 리뷰 요약 기능, 그리고 성분·영양 및 크기 정보를 상세하게 텍스트로 제공한다는 점입니다. 일반 쇼핑몰(예, 쿠팡)은 임의 탐색 방식을 채택하고 있으며 대표 이미지 설명, 리뷰 요약, 성분·영양, 크기 정보 제공이 미흡합니다. 반면, 시각장애인 쇼핑몰인 픽포미는 순차 탐색 및 대표 이미지 설명, 리뷰 요약 기능을 지원하지만 성분·영양 및 크기 정보는 제공하지 않습니다. 이에 저희 서비스는 시각장애인 접근성을 고려한 순차 탐색 외에도, 추가 정보를 제공함으로써 더욱 차별화된 서비스를 구현하고 있습니다.



	쿠팡 등 (일반 쇼핑몰)	픽포미 (시각장애인 쇼핑몰)	푸드리
시각장애인 접근성	임의 탐색	순차 탐색	순차 탐색
대표 이미지 설명	X	O	O
리뷰 요약	X	O	O
성분, 영양 정보 텍스트 제공	X	X	O
크기 정보 제공	X	X	O

아래는 기존 서비스와 푸드리와 차별점을 정리해둔 예상 Q&A의 일부입니다.

2. 기존의 시각장애인 대상 쇼핑 서비스와 푸드리의 차별점은 무엇인가요?

기존 서비스로는 지마켓에서 제공하던 OCR 서비스, 소리마켓, 픽포미 등이 있습니다. 그러나 지마켓의 OCR 서비스는 상품 정보를 정확히 읽지 못하는 경우가 많아 사용자 불편이 크다는 피드백이 있었고, 소리마켓은 이미 운영을 중단했습니다. 픽포미의 경우 주요 원재료나 영양성분 같은 핵심 정보를 유료로 별도 문의해야만 확인할 수 있고, 리뷰 또한 긍정·부정 정도에만 제한적으로 요약해 제공한다는 한계가 있습니다.

반면 저희 서비스 푸드리는 **다양한 측면의 상품 상세 정보를 기본적으로 자세히** 제공합니다. 이에 추가로 리뷰를 분석해 상품의 주요 특징을 추출하고, 이를 바탕으로 **상품을 재정렬**해주는 기능을 추가함으로써 사용자 맥락을 반영한 맞춤형 추천도 같이 제공합니다 (본 해커톤에서는 과자, 라면/간편식 카테고리에 한정). 소비자로서 알아야 하는 다양한 정보를 풍부하게 제공한다는 점이 저희 푸드리만의 차별점입니다.

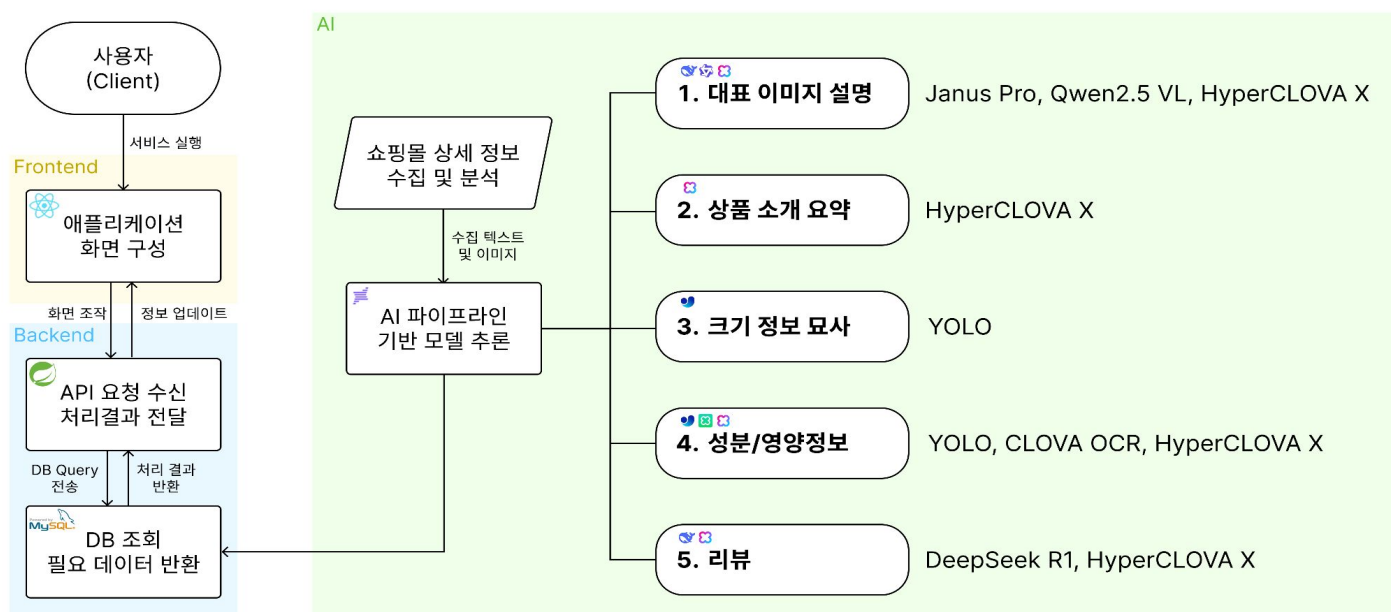
3. 서비스 아키텍처

3.1 전체 파이프라인 소개

본 절에서는 푸드리 서비스의 전체 파이프라인을 소개하고, AI 활용 아키텍처를 설명합니다. AI 파이프라인은 크게 다섯 가지 주요 태스크(Task)로 구성되며, 대부분의 과정에서 한국어에 최적화된 네이버 HyperCLOVA X를 활용하여 데이터를 가공합니다. 이를 통해 상세 정보를 소비자 친화적인 형태로 변환하는 후처리 과정을 수행하였습니다. 각 태스크별 AI 모델 활용 방식은 다음과 같습니다.

- **대표 이미지(썸네일) 설명**: Janus Pro 및 Qwen2.5 VL을 활용하여 이미지를 텍스트로 변환
- **크기 정보 묘사**: YOLO 모델을 활용하여 제품의 크기 정보를 분석
- **성분 및 영양 정보 추출**: YOLO와 CLOVA OCR을 조합하여 텍스트 정보를 정확하게 인식 및 추출
- **리뷰 요약 및 키워드 기반 상품 추천 시스템**: DeepSeek R1을 활용하여 리뷰 데이터를 분석하고 핵심 내용을 요약

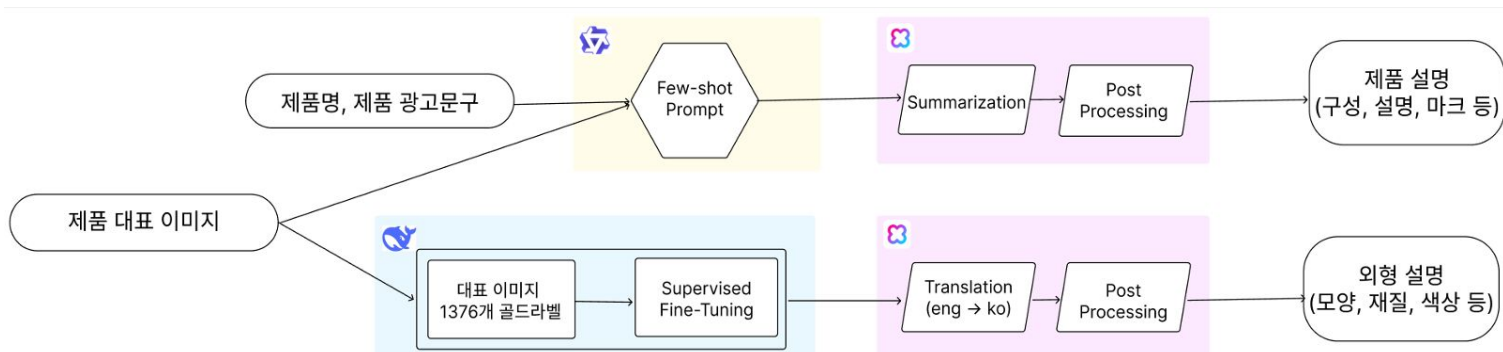
이와 같은 AI 기반 파이프라인을 통해 푸드리 서비스는 사용자에게 보다 직관적이고 유용한 정보를 제공할 수 있도록 설계되었습니다.



전체 파이프라인 이미지

3.2 대표 이미지(썸네일) 설명

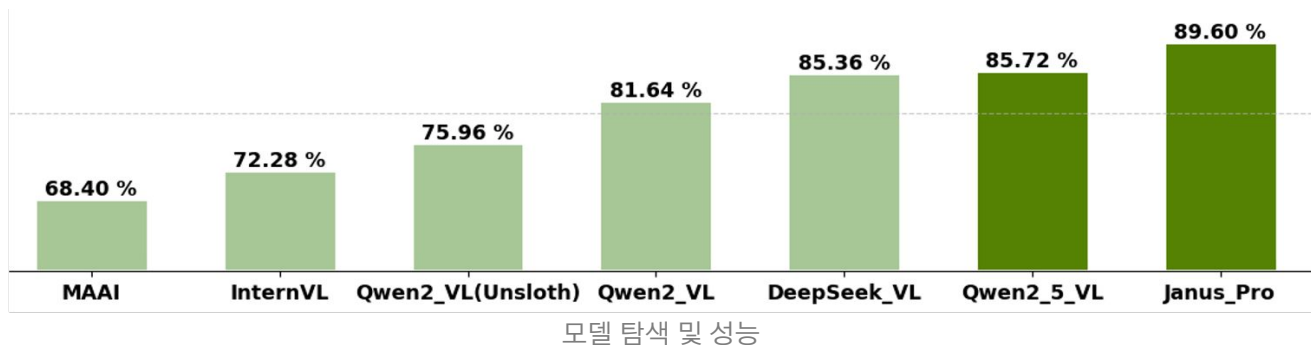
푸드리 서비스의 AI 파이프라인에서 가장 먼저 수행되는 과정은 상품의 대표적인 상태를 확인할 수 있는 대표 이미지 설명입니다. 대표 이미지에서 추출할 정보는 인증 마크, 구성품 및 설명 정보에 대응하는 제품 자체의 설명과 모양, 재질 색상 등 외관 관련 정보로 구분됩니다. 아래는 대표 이미지(썸네일) 설명에 대한 파이프라인입니다.



대표 이미지 설명 아키텍처

3. 서비스 아키텍처

NCSOFT/Varco-Vision의 벤치마크인 NCSOFT/K-MMBench와 NCSOFT/K-LLaVA-W를 차용하여 각각 이미지 이해능력 평가와 자연스러운 한국어 생성능력을 자체 평가 기준표를 구성하였습니다. GPT-4o를 활용한 LLM-as-a-Judge 방식으로 최종 7개의 VLM을 평가하고, 분석하였습니다. 그 결과, deepseek-ai/Janus-Pro-7B와 Qwen/Qwen2.5-VL-Instruct-7B가 각각 1위와 2위를 기록하였습니다, Janus Pro는 제품 설명에서, Qwen2.5 VL은 외형 설명에서 높은 점수를 보였습니다..



소비자에게 보다 정확한 제품 정보를 제공하기 위해 두 모델을 함께 활용하기로 결정하였으며, 고도화 과정에서 Hallucination 방지를 위해 Few-Shot 프롬프팅, SFT, HyperCLOVA X 후처리를 수행하였습니다.

	제품 설명 (Qwen2.5-VL-7B-Instruct)	외형 설명 (itsmenlp/finetuned-Janus-Pro-7B)
모델 수행 작업	Few-Shot, 상품명과 상품 홍보문구 입력 - 상품 명과 상품 대표 홍보 문구 요약의 5가지의 Few shot 제공하여 제품에 대한 사실 기반 설명과 홍보성 문구를 매력적으로 제시	deepseek-ai/Janus-Pro-7B 파인튜닝 - SSG, GS The Fresh 등 상품의 대표이미지 1376개에 대한 재질, 모양, 색상, 투명성, 디자인 속성 추출 (GPT-4o) - 수동 검토 120개 진행하여 개선 예시 제공하여 속성 재추출 후 학습데이터로 사용
HCX 후처리	- 긴 문장 요약, 반말 및 불완전 서술어 후처리 - 조리방법 및 활용법 제거(Hallucination)	- 한국어 번역 진행, 추정의 말투 제거 - 높이 및 너비 정보 제거 (Hallucination)

제품 설명, 외형 설명 고도화 수행 과정

제품명 (기린 옥수수 모닝롤)

제품 설명 (제품정보, 인증마크, 추가 정보)



고소함이 가득한 옥수수 모닝롤입니다.



한 손에 쏘옥 들어오는 크기이며 고소함이 가득한 [기린] 옥수수 모닝롤 360g입니다.



외형 설명 (포장 모양, 재질, 색상, 디자인)

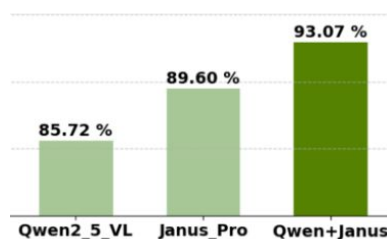
포장 용기는 하얀색 손잡이가 달린 투명한 플라스틱 가방이며, 제품 이름과 기타 세부 정보가 담긴 노란색 라벨이 붙어 있습니다. 이 봉지 안에는 황금색 갈색 빵 롤이 가득 차 있으며, 전반적인 디자인은 심플하면서도 시각적으로 매력적입니다.



포장 용기는 상단에 하얀색 트위스트 타이가 있는 노란색 부분이 달린 투명한 비닐 봉지이며, 제품 이름과 기타 세부 정보가 담긴 하얀색 텍스트가 적혀 있습니다. 이 봉지 안에는 황금색 빵 롤이 가득 차 있습니다.



최종적으로 제품 설명과 외형 설명을 합친 대표 이미지 설명 평가를 진행하였고 약 **0.93**의 점수를 얻었습니다.



최종 성능

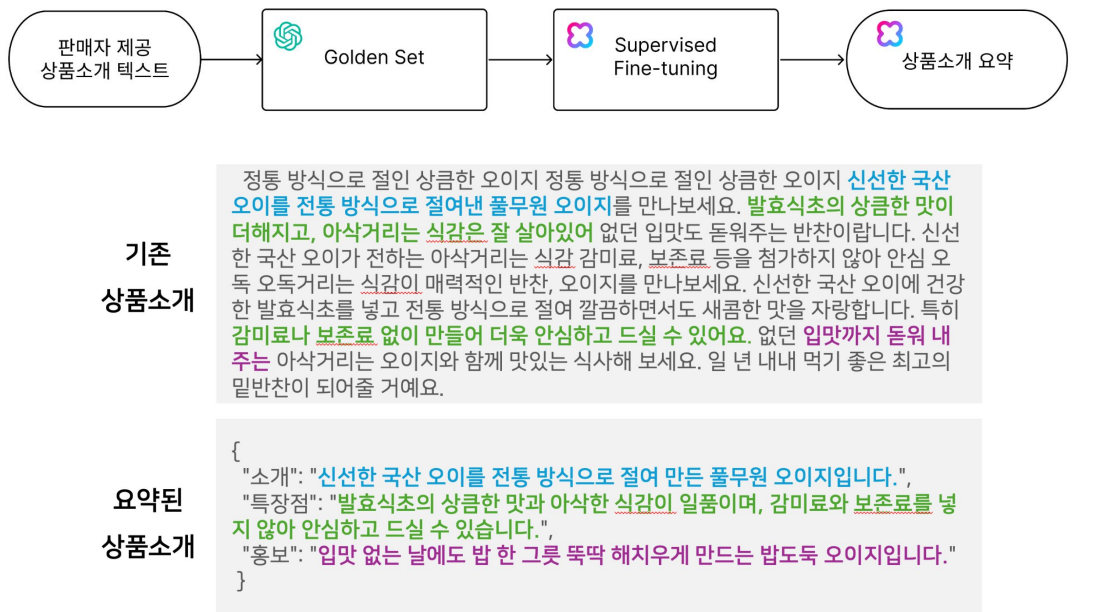
고도화 진행과정과 예시

3. 서비스 아키텍처

3.3 상품 텍스트 요약

시각장애인은 장황한 상품 설명을 모두 청취해야 하므로 핵심 정보 파악이 어렵습니다. 이를 해결하기 위해 상품 요약을 최상단에 배치하였습니다. 이러한 상품 상세 설명 전에 제공되는 요약문은 설명의 일관성을 유지하고 소비자의 구매 욕구를 자극하기 위해 중요한 역할을 합니다. 이를 위해 요약문 작성 지침을 구축하고, Zero-Shot 및 Few-Shot 학습을 적용하였습니다. 그러나 초기 실험에서 지침 미준수 및 Hallucination 문제가 발생하는 한계가 확인되었습니다.

이러한 문제를 해결하기 위해 수동으로 Golden Set을 구축한 후, SFT(Supervised Fine-Tuning)를 수행하였습니다. 그 결과, 모델이 Hallucination 없이 기존 설명에서 추출된 형태로 문장을 생성하는 것을 확인하였습니다.



3.4 크기 정보 설명

상품 상세 설명에서 가장 먼저 확인할 수 있는 정보는 상품의 크기입니다. 설문 조사 결과, 온라인 쇼핑 시 가장 불편했던 점으로 상품 이미지 및 상세 정보 습득의 어려움(보이스오버 호환성 포함)이 1위로 꼽혔다. 또한, 필요한 정보 및 기능으로 포장 상태, 외관 및 크기 정보가 공동 1위를 차지했습니다. EDA 결과, 상품 목록 및 상품 상세 페이지에서 크기 이미지를 제공하는 상품이 약 87.40%이며, 특히, 순서 유지 O 상품 323개는 모두 크기 정보를 포함하고 있었습니다. 이를 바탕으로, 상품 크기에 대한 정보를 더욱 효과적으로 요약 제공할 필요성을 느꼈습니다.



이를 해결하기 위해 먼저 VQA, Image Captioning, Vision-Language Model의 성능을 비교해 보았습니다. 그러나 대조군과 제품을 명확히 구별하지 못해 정확도가 낮은 문제가 있었습니다. 이에 YOLO 모델을 활용해 제품과 대조군의 bbox를 생성하고, SFT를 수행하여 정확도를 개선했습니다. 최종적으로 Rule-based 방식을 적용해 크기 정보를 보다 명확하게 설명하도록 구성했습니다.



3. 서비스 아키텍처

3.5 성분/영양 정보

시장장애인은 식품 패키지에 기재된 성분/영양 정보를 직접 확인하기 어려우며, 기존의 온라인 쇼핑 플랫폼에서도 텍스트 기반 정보 제공이 부족하여 접근성이 낮습니다. 기존 방식으로는 OCR 기술을 활용하여 성분 및 영양 정보를 추출하는 방법이 있지만, 이미지의 품질과 배경 복잡도에 따라 정확도가 달라지는 한계가 존재합니다. 또한, 성분 및 영양 정보는 다양한 포맷으로 존재하기 때문에 일괄적인 정보 추출이 어렵습니다.

문제상황

[illegible]

OCR 결과

원재료명 밀가루[밀(미국산)], 쇼트닝[가공
유지(팜스테아린에스테르화유(팜스테오·
본 제품용 공정거래위원회 기준치에 대한 비율
원칙 고시에 의거 피해보상 나트
름 100 mg 5 % 340 mg 17 %
아린유(말레이시아산)], 유화제, 레시틴, d-토
코페롤(혼합형), 포도당, 기타첨가물(설탕,
전 · 객관축살 · 수산화 요금 탄수화
물 21 g 6 % 69 g 21 %

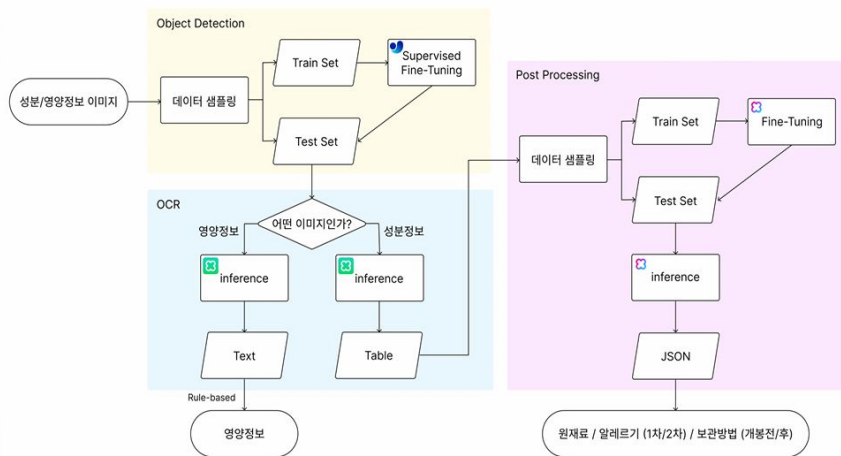
OCR 오류 보정 필요

LLM 결과

총 내용량 204 g / 1회(30 g)당 155 kcal
나트륨 : 340 mg (17 %)
탄수화물 : 69 g (21 %)
당류 : 25 g (25 %)
지방 : 10 g (18 %)
트랜스지방 : 0 g
포화지방 : 4.2 g (28 %)
콜레스테롤 : 5 mg (1 %)
단백질 : 2 g (3 %)

LLM 기반 후처리 안정화 필요

파이프라인



이러한 문제를 해결하기 위해 LLM 후처리와 Rule-based 방식을 결합하여 성분/영양 정보 추출 성능을 개선였습니다. 이미지 내에서 표 형태의 데이터를 구조적으로 인식하여 분석하는 방법을 검토하고, 데이터 정제 및 보정 과정을 추가하여 보다 높은 신뢰도를 확보했습니다.

성분 정보 추출 단계에서는 Supervised Fine-Tuning된 YOLO 모델을 활용하여 성분 정보가 포함된 영역을 검출하고, CLOVA OCR을 통해 해당 영역의 텍스트 데이터를 표 형태로 변환하였습니다. 이후 Fine-Tuning한 HyperCLOVA X 모델을 활용한 후처리를 적용하여 OCR 단계에서 발생한 불완전한 텍스트를 보완하고, 자연어 처리 기법을 통해 식품 성분명을 표준화하여 JSON 형태로 변환하였습니다. 마지막으로 Levenshtein Distance를 기반으로 원본 텍스트와의 유사도를 측정하여 데이터의 일관성과 정확성을 평가하였습니다.

영양 정보 추출 단계에서는 YOLO 모델을 이용하여 영양 정보가 포함된 영역을 검출하고, CLOVA OCR을 통해 표 형식의 영양 정보를 텍스트로 변환하였습니다. 이후 Rule-based 후처리를 적용하여 OCR로 추출된 영양 정보를 정리하고, 표준 단위 변환 및 수치 일관성을 유지하기 위한 규칙을 적용하였습니다. 마지막으로 Exact Match 방식을 기반으로 OCR 결과와 실제 데이터를 비교하여 정확도를 검증하였습니다.

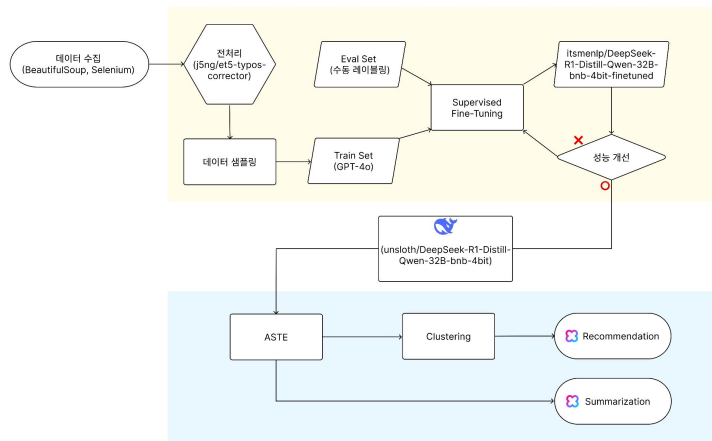
최종적으로 개발된 파이프라인을 통해 성분 정보와 영양 정보를 효과적으로 추출하고 정확도를 검증하였습니다. 성분 정보의 유사도를 평가하기 위해 Levenshtein Distance를 활용하였으며, 총 50개 항목을 비교한 결과 원재료 평균 유사도가 96.70%, 알레르기 정보 평균 유사도가 90.62%, 보관 방법 평균 유사도가 83.62%로 나타났습니다. 또한, 영양 정보의 경우 Exact Match 방식을 기반으로 평가한 결과, 전체 1,640개 중 1,571개가 일치하여 총 일치 확률이 95.79%로 확인되었습니다.

Accuracy		
데이터셋		Metric
		EM *
영양정보		95.79%
		LD *
성분정보	원재료	96.70%
	알레르기 (1차/2차)	90.62%
	보관방법 (개봉전/후)	83.62%

Exact Match *: OCR 결과와 정답셋이 완전히 일치하는 경우를 확인
Levenshtein Distance *: 두 문자열 간의 최소 편집 횟수(삽입, 삭제, 변경)

3. 서비스 아키텍처

3.6 리뷰 요약 및 키워드 기반 상품 추천 시스템



리뷰 정보를 기반으로 한 추천 키워드별 상품 재정렬과 상품 속성에 따른 긍정/부정 감정 리뷰 요약 구현했습니다. 대량의 리뷰 데이터를 처리하기 위해 SFT한 DeepSeek-R1-Distill 모델을 사용하여 Aspect Sentiment Triplet Extraction(ASTE)을 수행했습니다. 이 과정을 통해 맛, 배송 등 속성과 속성에 대한 사용자의 평가, 긍정, 중립, 부정 감정을 Triplet 형태로 추출했습니다. 모델의 ASTE 출력에 있어서 정형성과 일관성을 유지하기 위해 SFT를 수행했는데, 이 때 두 가지 데이터셋을 마련했습니다.

- **Golden Set 구축:** GPT-4o를 사용하여 실험적으로 ASTE를 수행한 Silver Label 데이터를 기반으로, Stratified Random Sampling 기법을 활용하여 100개의 데이터셋을 구축했습니다. 다양한 리뷰를 포함하기 위해 속성, 감정, 평가 항목, 상품 종류, 리뷰 길이 등을 고려하여 구성했습니다. 이후 수동 검수로 2차 레이블링을 진행하여 라벨의 정확도와 일관성을 확보했습니다.
- **Train Set 구축:** GPT-4o를 사용해 ASTE를 수행한 Silver Label 데이터를 기반으로, Sentence-BERT 임베딩을 활용해 리뷰 데이터를 벡터화한 후, K-Means 클러스터링을 통해 900개의 클러스터를 만들었습니다. 각 클러스터의 중심에 위치한 샘플을 선택하여, DeepSeek R1 스타일의 CoT 데이터와 레이블 데이터로 가공했습니다.

또한 모델 성능을 평가하기 위해 라벨(GL)과 예측 Triplet 세트의 일치 여부를 확인하는 방식으로 진행했습니다. 주요 평가 절차는 다음과 같습니다.

1. **세트 개수 확인:** 라벨(GL)과 예측된 Triplet 세트의 개수가 동일한지 확인합니다. 이때 매칭 기준으로 BERTScore를 사용하여, 유사도가 0.8 이상인 후보 쌍을 선정한 후, Hungarian Algorithm을 활용해 최적의 1:1 매칭을 수행합니다.
2. **항목 일치 확인:** 매칭된 후보 쌍에서 ‘속성’과 ‘감정’ 항목이 정확히 일치하는지 확인합니다. 일치하면 True Positive(TP)로 처리하고, 일치하지 않으면 False Negative(FN)로 기록합니다. 매칭되지 않은 세트는 라벨에만 존재하면 FN으로, 예측에만 존재하면 FP로 처리합니다.
3. **정량적 평가:** 최종적으로, 추출된 각 항목에 대한 TP, FP, FN 값을 바탕으로 Precision, Recall, F1-score를 계산하여 모델 성능을 정량적으로 평가합니다.

위 과정으로 추출된 Triplet을 활용하여 추천 키워드별 상품 재정렬과 리뷰 요약 작업을 진행했습니다.

1. **추천 키워드 재정렬:** 유사 리뷰를 Clustering하여 각 클러스터에서 HCX를 통해 마케팅적인 요소가 담긴 키워드를 추출했습니다. 이후 사용자가 선택한 키워드(예: ‘갯성비템’)에 따라, 리뷰 수가 많은 순으로 상품을 재정렬했습니다.
2. **리뷰 요약:** HCX를 사용하여 상품의 속성에 따른 감정을 반영한 리뷰 요약을 생성했습니다. 이를 통해 사용자가 상품의 속성에 따른 긍정적/부정적 리뷰를 한눈에 파악할 수 있도록 했습니다.

4. AI Safety

4.1 Hallucination

부정적 내용 필터링

1. 리뷰 요약: Hate-Speech 탐지 모델인 j5ng/et5-typos-corrector를 사용합니다. 낮은 리뷰를 요약할 때 중립적인 의견을 포함하여 감성 편향을 완화합니다. 또한, HCX 프롬프팅 시 혐오 표현에 대한 절제를 요구합니다.

리뷰 원본	정제 후
존*맛있음 계속시켜먹을듯	hate-speech 맛있음 계속 시켜 먹을 듯.
이런 ***의 것 이런 것을 * *같은 것을 팔아 말도 되지 않는 *할 놈의 맛	이런 hate-speech의 것 이런 것을 hate-speech 같은 것을 팔아 말도 되지 않는 hate-speech의 맛.

< Hate-Speech 탐지 예시 >

정보 정확성

1. 영양/성분 정보: OCR을 활용하여 전체 정보를 제공하는 것이 아니라, 필요한 정보만을 정확하게 추출하는 것을 목표로 합니다. 또한, Rule-based 방식과 LLM 후처리 방식을 비교 분석하여 높은 정확도를 달성합니다.

2. 크기 정보: VLM, Image Captioning, VQA에서 발생하는 Hallucination 문제를 인식하고 대응합니다. 또한, YOLO Bbox 비교를 기반으로 정확한 크기 정보를 제공합니다.



VLM (Hallucination)	YOLO + Rule-based
봉지의 높이는 16.5cm, 너비는 6.5cm입니다. 캔의 부피는 500ml입니다.	배송받는 제품의 실제 너비는 16cm 정도로 500ml 캔 너비의 2배보다 조금 크고, 실제 높이는 21cm 정도로 500ml 캔 높이보다 조금 큼니다.

< 크기 정보 Hallucination 개선 예시 >

예측 가능성

1. 상품 텍스트 요약: HCX에 대해 SFT를 수행합니다. 소비자의 구매 욕구를 자극하는 매력적인 문구를 생성하되, 상세 정보에 제공되지 않은 거짓 정보가 생성되지 않도록 제한합니다.

2. 대표 이미지 설명: 제품 텍스트에 대한 Few-shot 데이터를 생성합니다. 또한, VLM이 생성하는 제품 설명이 상세 정보에 제공되지 않은 거짓 정보인지 검토하여 정확성을 확보합니다.

3. 리뷰 요약: DeepSeek-R1-Distill-Qwen에 대해 SFT를 수행합니다. 또한, 직접 구축한 Golden Set의 스타일과 일관된 구조를 따르도록 유도합니다.

4.2 AI Ethics

사람을 위한 AI 개발

AI를 적극적으로 활용하여 시각장애인이 직접 쇼핑할 수 있도록 합니다. 단순한 정보 제공을 넘어, 리뷰에서 추출한 키워드를 활용하여 사용자와의 인터랙션 기회를 제공합니다.

다양성 존중

중증 시각장애인의 접근성을 개선하기 위해 스크린리더와의 호환성을 강화합니다. 또한, 경증 시각장애인을 위해 단순하고 직관적인 인터페이스를 구성합니다.

합리적인 설명과 편리성의 조화

주변인의 도움 없이 스스로 쇼핑할 수 있도록 제품에 대한 자세한 설명을 제공합니다. 또한, 정보 과부하를 방지하기 위해 상품 설명과 리뷰를 요약하여 제공합니다.

5. 결과 및 회고

5.1 결과물

‘푸드리’ 이용 흐름

애플리케이션에 접속하면 카테고리 메뉴를 확인할 수 있으며, 카테고리를 선택하거나 원하는 식품을 검색하여 상품을 정렬할 수 있습니다. 상품 리스트를 순차적으로 탐색하고 더 자세히 알고 싶은 상품의 내부로 들어가면 상품의 상세 설명 요약글, 주의사항, 크기 정보, 영양 및 성분정보, 리뷰 요약을 제공합니다.

시각장애인 특화 UI/UX 구성

시각장애인이 온라인 쇼핑을 할 때 웹과 비교하여 활용도가 높은 어플리케이션으로 ‘푸드리’를 구현하였습니다. 중증 및 경증 시각장애인을 모두 고려한 직관적이고 단순한 UI 디자인을 반영하였으며 화면 낭독 프로그램(VoiceOver, TalkBack 등)과 호환되도록 설계하여 접근성을 향상시켰습니다. 최종적으로 시각장애인이 겪는 정보 격차 문제를 해결하기 위해 상품구매에 필요한 정보를 시각장애인 특화 UI/UX를 통해 제공합니다.

데모 영상 : https://www.youtube.com/shorts/KuRIOKePB_w







5.2 시각장애인의 사용성 평가 분석

서비스의 사용성은 사용하시는 소비자 분들의 평가가 가장 객관적이고 중요하다고 생각하여, 문제 정의와 마찬가지로 시각장애인분들과 함께 서비스 시연을 진행했습니다. 사용성 평가를 진행하며 쿠팡, 배민 B마트와 비교했을 때, 저희 서비스 푸드리는 상품 탐색과 장바구니 담기 과정이 훨씬 빠르고 편리했습니다. 탐색 편의성은 5점 만점에 4.6점을 기록하며 접근성 개선 효과로 사용자 기대감이 높음을 확인하였습니다.

	평균 쇼핑시간*	만족도**
	67초 • 장바구니까지 접근 과정 복잡	0.7점 • 불필요한 정보가 많음 • 임의 탐색
	41초 • 장바구니까지 접근 과정 복잡	1.5점 • 커서 튕김 현상이 심함 • 임의 탐색
	40초 • 장바구니까지 접근 과정 단순	4.6점 • 다양한 정보 접근이 가능 • 순차 탐색

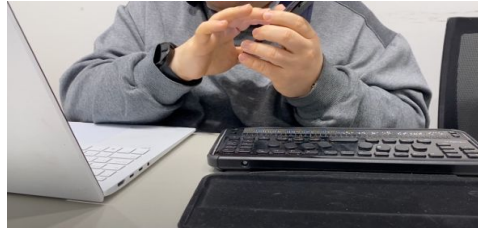
평균 쇼핑 시간*: 진라면 순한맛 제품을 장바구니에 담는 시간
만족도**: 7개 항목에 대한 종합 평점 (5점 만점)

	사용자 후기
 김*경님 (40대, 여, 경증)	출시된다면 식료품은 여기에서 구매하고 싶습니다.
 이*혜님 (30대, 여, 중증)	시각장애인이 독립적으로 쇼핑을 하게끔 서비스를 만들어주셔서 감사합니다.
 곽*희님 (40대, 남, 중증)	소리마켓과 비교했을 때, 푸드리가 더 자세히 알려 주는 것 같아 좋습니다.
 김*성님 (40대, 남, 중증)	전체적으로 정보 접근성이 편리합니다.

5. 결과 및 회고



김*경님 [2025년 2월 3일] <https://youtu.be/f3zdA3h91ro>



김*섭님 [2025년 2월 4일] <https://youtu.be/vQHpcyR101c>



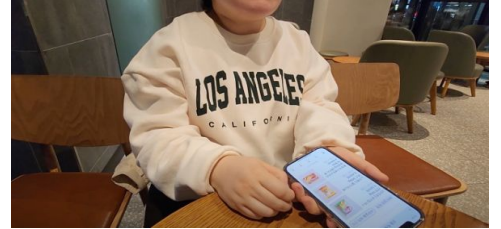
현*수님 [2025년 2월 5일] https://youtu.be/F_3NkGswAag



곽*희님 [2025년 2월 3일] <https://youtu.be/U3fCvXRK0fE>



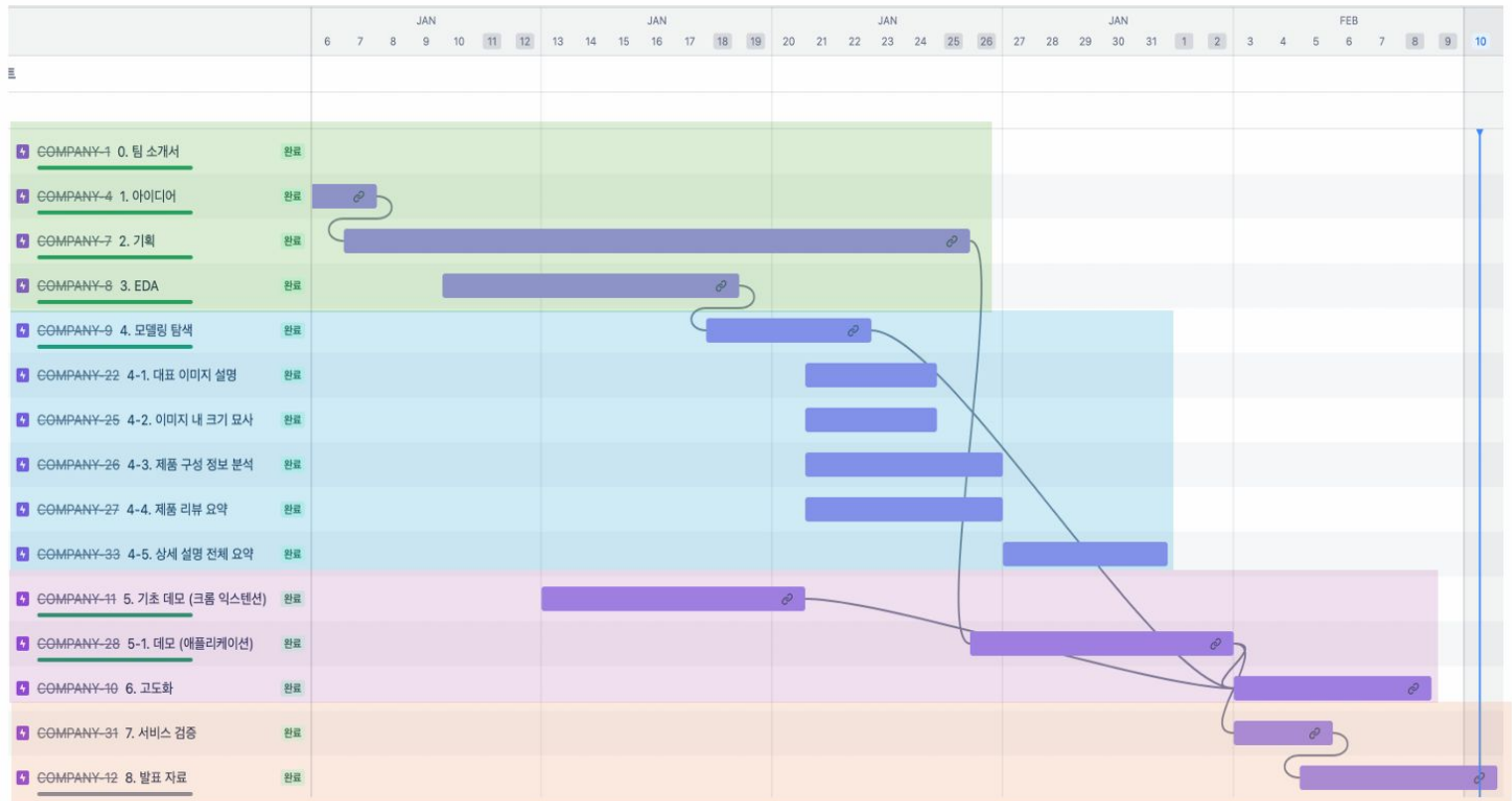
서*선님 [2025년 2월 4일] <https://youtu.be/ivIXUYE2inM>



이*혜님 [2025년 2월 5일] <https://youtu.be/2VxAxmquvuo>

5.3 프로젝트 타임라인

프로젝트를 진행하는 한 달 동안 기획된 내용을 바탕으로 기능별로 필요로 하는 모델을 각각 개발하였으며, 개발된 결과물을 디자인하여 보여줄 수 있도록 프론트 백엔드를 개발하였고, 완료된 이후 서비스를 사용하며 검증 및 평가하는 흐름으로 진행하였습니다.



6. 개인 회고 - 곽희준

팀에서의 나의 역할

초기 아이디어 한 줄에서부터 전반적인 프로젝트 진행 방향을 구체화하였으며, 베이스 모델 선정과 모델링을 통해 서비스 1차 프로토타입을 완성하였습니다.

이후 ASTE Task를 정의하고 클러스터링을 통해 추천 키워드 포함 상품 검색 및 리뷰 요약 기능을 구현하고, 전체 파이프라인 코드 정리를 통해 모듈 연계와 코드 가독성 및 유지보수성을 향상시켰습니다.

나는 내 학습 목표를 달성하기 위해 무엇을 어떻게 했는가?

아이디어 구체화에서는 시각장애인의 정보 접근성 문제(대체 텍스트 미제공, 비효율적 정보 제공)를 정의하고, 판결 사례 및 장애인소비자 모바일거래 실태조사를 참고하여, 상품 상세 페이지 내 이미지 대체 텍스트 생성과 핵심 정보 요약 제공 등 서비스 기획 및 아키텍처를 구체화하였습니다.

VLM 베이스 모델 선정에서는 리더보드 기준 상위 5개 후보 모델을 테스트하고, Human Evaluation을 통해 처리 속도와 성능을 평가하였으며, 결과에 따른 1차 서비스 파이프라인 설계를 구체화했습니다.

OCR+LLM 파이프라인 구현에서는 CLOVA OCR을 활용하여 표 형식 이미지의 텍스트 분할 이슈(큰 이미지, 작은 ROI, 노이즈 문제)를 분석하고, HyperCLOVA X 후처리 모듈을 추가하여 성분/영양정보의 OCR 결과를 복구, 정성 평가와 평균 처리 시간을 확인하였습니다.

1차 프로토타입 완성에서는 VLM 모델과 OCR+LLM 모듈을 통합하여 1차 프로토타입을 완성하였습니다.

ASTE Task 정의에서는 리뷰의 속성, 평가, 감정을 세분화하여, 이를 바탕으로 리뷰 요약 및 추천 키워드 검색 시스템을 제안하였습니다.

ASTE를 위한 텍스트 전처리에서는 특수문자, 중복 표현 제거 등 전처리 작업을 수행하였습니다.

ASTE Task의 학습, 평가 데이터 생성에서는 GPT-4o를 활용한 학습 데이터와 수동 레이블링을 통한 평가 데이터를 생성하였습니다.

ASTE 평가 메트릭 정의에서는 BERTScore와 Hungarian Algorithm을 활용하여 Precision, Recall, F1 평가 메트릭을 자체 설계하였습니다.

ASTE 모델 SFT 최종 결과로 F1-score가 15.07%p(56.06 → 70.13)향상되었습니다.

ASTE의 ‘평가’ 항목 클러스터링 구현에서는 CounterVectorizer, TF-IDF, Word2Vec, FastText, Sentence BERT 등 텍스트 임베딩 기법을 비교하고, UMAP 차원 축소와 HDBSCAN 구현 및 파라미터 최적화를 진행하였으며, T-SNE 시각화, Silhouette Score, DBI를 통해 정량 평가하였습니다.

클러스터링 최종 결과로 SilhouetteScore 0.8과 DBI 0.2를 달성하는 성과를 얻었습니다.

추천 키워드 포함 상품 검색에서는 HyperCLOVA X를 활용하여 각 ‘평가’ 항목 클러스터를 대표하는 한 줄 키워드를 추출하였습니다.

리뷰 요약에서는 ‘속성’과 ‘감정’ 항목 별로 HyperCLOVA X를 통해 원본 리뷰에서 요약문을 추출하였습니다.

전체 파이프라인 코드 정리에서는 썸네일, 크기 정보, 상세정보 요약, 성분/영양 정보, 리뷰 관련 모델들의 전체 파이프라인 코드를 정리하여 모듈 간 연계를 최적화하였습니다.

마주한 한계는 무엇이며 아쉬웠던 점은 무엇인가?

실시간 처리를 구현하고자 하였으나, 처리 시간이 오래 걸려 부득이하게 DB에 저장하는 방식으로 전환하였습니다. ASTE Task가 이미 논문으로 존재함을 늦게 인지하였고, 클러스터링 방법론에 대한 지식 부족으로 다양한 기법을 적용하지 못한 점도 한계로 남았습니다.

한계/교훈을 바탕으로 다음 프로젝트에서 시도해보고 싶은 점은 무엇인가?

ASTE Task에 대해 관련 논문 조사를 통해 고도화를 진행하겠습니다. 클러스터링 방법론을 보다 심도 있게 학습하여 다양한 기법을 적용해 보겠습니다.

6. 개인 회고 - 김정은

팀에서의 나의 역할

이번 프로젝트에서 저는 시각장애인용 온라인 식료품 쇼핑 서비스의 성분/영양 정보 추출 및 처리를 담당하였으며, 모델의 성능을 최적화하는 역할을 수행했습니다.

YOLO SFT: 식료품 이미지 내에서 성분/영양 정보를 각각 검출하도록 YOLO 모델을 Fine-Tuning

OCR output 후처리: CLOVA OCR을 통해 추출된 텍스트/표 형식의 데이터를 후처리 진행

Rule-based 방식 적용 실험: OCR 성능 보완을 위해 영양 정보에 대해 규칙 기반 필터링 및 보정 기법 실험

HCX Fine-Tuning: HyperCLOVA X 모델의 성능을 향상시키기 위해 데이터 특성에 맞춘 파인튜닝 진행

평가 Metric 선정: 정보 추출 정확도를 정량적으로 평가할 수 있도록 적절한 평가 지표를 설계하고 검증

이 외에도, 대표 이미지와 크기 이미지에 대한 Image Captioning/VQA/VLM 테스트를 진행하였으며, 최종 결과물들을 종합하여 데이터베이스를 구축하였습니다.

나는 내 학습 목표를 달성하기 위해 무엇을 어떻게 했는가?

이번 프로젝트를 통해 저는 정보 추출 정확도를 높이고, OCR 및 LLM 기반 시스템의 한계를 극복하는 방법을 탐구하는 것을 목표로 하였습니다.

Object Detection: 성분 정보와 영양 정보를 개별적으로 검출할 수 있도록 YOLO11 모델을 SFT 수행

Rule-based 방식 vs LLM 후처리 비교: 규칙 기반 필터링과 HCX Fine-Tuning을 병행하여 OCR의 오차를 최소화할 수 있는 방법을 모색

마주한 한계는 무엇이며 아쉬웠던 점은 무엇인가?

LLM의 할루시네이션 문제: LLM을 활용한 정보 추출 과정에서 모델이 실제 존재하지 않는 정보를 생성하는 문제가 발생했습니다. 특히 HCX Fine-Tuning을 수행했음에도 할루시네이션을 완전히 억제하기 어려웠습니다. 이를 해결하기 위해 Rule-based 방식과 LLM을 혼합한 하이브리드 접근 방식을 실험했지만, 완전한 해결에는 이르지 못했습니다.

OCR 성능의 근본적 한계: CLOVA OCR의 기본 성능이 좋지만, 특정 폰트나 배경색이 있는 경우 인식 오류가 빈번했습니다. 모델 후처리로 일부 해결했지만, OCR 자체의 성능을 근본적으로 향상시키는 것이 쉽지 않았습니다.

전체 파이프라인의 최적화 부족: 정보 추출을 위한 다양한 실험을 진행했지만, 결과적으로 파이프라인이 복잡해졌습니다. 여러 모델과 후처리 기법을 조합하다 보니 연산량이 증가했고, 실제 운영 시 최적화가 필요해 보였습니다.

한계/교훈을 바탕으로 다음 프로젝트에서 시도해보고 싶은 점은 무엇인가?

특정 기술에 대해 깊이 있는 탐구: 이번 프로젝트에서는 여러 기술을 폭넓게 실험했지만, 다음 프로젝트에서는 하나의 기술(예: OCR, Fine-Tuning)에 대한 깊이 있는 연구를 진행해보고 싶습니다. 특히 OCR 및 LLM 후처리의 한계를 극복하기 위해, 특정 태스크에 최적화된 모델 설계에 더 집중할 계획입니다.

LLM의 신뢰성 강화: 할루시네이션 문제를 해결하기 위해, Fact-checking Mechanism을 도입하는 실험을 진행하고 싶습니다. Rule-based 방식과 LLM을 하이브리드로 조합하는 방법을 더욱 정교하게 설계하고, Fine-Tuning을 더 세밀하게 조정해볼 예정입니다. Knowledge Distillation 같은 기법을 활용해 더 경량화된 모델을 만들고 싶은 목표도 있습니다.

효율적인 파이프라인 설계: 처음부터 최적화된 경량 모델과 후처리 구조를 설계하여, 복잡한 파이프라인을 단순화하는 방향을 고려할 것입니다. 실시간 처리가 중요한 서비스의 경우, 지나치게 무거운 모델을 배제하고 효율적인 경량 모델을 중심으로 실험을 진행할 계획입니다.

이번 프로젝트에서 많은 성과를 거두었지만, 동시에 해결해야 할 과제도 많이 보였습니다. 다음 프로젝트에서는 더 깊이 있는 연구, 신뢰성 높은 AI 시스템 구축, 최적화된 파이프라인 설계를 목표로 삼아 발전해 나가고 싶습니다.

6. 개인 회고 - 김진재

팀에서의 나의 역할

본 프로젝트의 팀장으로서 전체적인 프로세스의 일정과 마일스톤을 관리하고, 각 태스크별 필요 일정 조절 및 프로젝트 관련 자료 기초 문서화를 담당하였습니다. 서버 기초 환경을 세팅하거나 회의를 진행하였으며, 일부 Task에서 아이디어나 방법론을 고안하여 제안하는 등의 역할을 수행하였습니다.

프로젝트 내에서는 이미지 내 크기 정보 묘사 Task를 수행하기 위한 효과적인 방법론을 고안하여, 대조군 제품과 실제 제품을 Object Detection 할 수 있는 YOLO 모델을 SFT하여, Rule-based로 자연어를 생성하도록 후처리하였습니다. 리뷰 요약 Task에서는 Aspect Sentiment Triplet Extraction Task를 수행하기 위한 DeepSeek로 Distillation된 Qwen 모델의 Fine-Tuning 방법론을 제안하고, Unsloth 기반으로 32B 파라미터 모델을 SFT하였습니다.

또한 보여지는 Demo 구현을 담당하였는데 기획에서 프로토타이핑 전반을 총괄하였으며, 초기 1차 프로토타입에서는 Spring Framework와 stages.ai 서버를 연결하여 서버를 활용할 수 있는 백엔드를 구축한 후 Chrome Extension을 이용하여 쇼핑몰 사이트 이미지 데이터를 전송하면 조합하여 반환하도록 구현하였습니다. 최종 구현물에서는 React Native를 이용한 Cross-Platform 애플리케이션으로 프론트엔드 및 백엔드 API를 디자인하였습니다.

나는 내 학습 목표를 달성하기 위해 무엇을 어떻게 했는가?

이미지 내 크기 정보 묘사의 Hallucination 문제를 해결하기 위해 최대한 Rule-based 방식으로 접근할 것을 제안하고, BBox를 수동 라벨링하여 SFT를 하였고 그 결과 이미지 형식(대조군, 제품)이 일치하는 경우 95%에 근접하는 성능에 도달할 수 있었습니다.

리뷰의 Aspect Sentiment Triplet Extraction에서 높은 성능을 내기 위해 Reasoning 모델을 탐색하였고, Deepseek R1 으로부터 Distillation을 수행한 모델을 탐색하였습니다. ASTE 관련해서 Fine-Tuning을 수행할 수 있는 방법론을 제안하여 unsloth 라이브러리를 이용한 메모리 Efficient한 학습을 수행하였고, 높은 성능을 달성할 수 있었습니다. (Zero-shot Baseline 0.075 -> **Fine Tuning 0.7508로 획기적인 성능 향상 달성**)

프로토타입 구현에서는 Figma를 이용하여 개발하는 뷰를 디자인하였으며, 디자인한 내용을 바탕으로 프로젝트 초반 1~2주는 Backend를 Spring Framework, Frontend Chrome Extension을 React 라이브러리에서 구현하였습니다. Stages AI 서버와 연동하여 백엔드에 요청을 보내면 서버에 가서 인퍼런스를 수행하고 결과를 받아오는 일련의 과정을 구현하였고, 배포는 AWS EC2와 RDS를 이용하였습니다. 최종 프로토타입에서는 React Native 라이브러리를 이용한 크로스 플랫폼을 설계하였으며, Spring Framework API를 추가로 개발하여 완성된 서비스를 구현하였습니다.

마주한 한계는 무엇이며 아쉬웠던 점은 무엇인가?

프로젝트의 기간이 짧아 차근차근 고도화를 하기에 벅찬 시간이었습니다. ASTE Task에서도 밑바닥부터 보다 다양한 방법론을 채택해보고, Rule-based 방식부터 하나씩 구현하며 차근차근 성능을 향상해보고 싶었습니다. 또한, 애플리케이션 설계에서도 다양한 디바이스에 맞는 반응형 디자인 설계를 강화하는 코드를 작성해보고 싶었지만, 한 달이라는 프로젝트가 지나치게 짧았던 점이 아쉬웠습니다. 그러나 프로젝트 자체의 기간이 워낙 짧았던 것이고, 한정된 기간 내에서 모든 팀원이 노력했기에 이미 충분히 한계를 넘어서 퀄리티와 분량의 프로젝트를 소화해내었다고 생각합니다.

한계/교훈을 바탕으로 다음 프로젝트에서 시도해보고 싶은 점은 무엇인가?

조금 더 시간을 가질 수 있다면, 많은 것들을 체계화해서 하나씩 진행해보고 싶습니다. 우선, 체계적인 문서화 기반을 다지고 형태에 맞는 문서를 최대한 많이 작성하는 것을 목표로 하고 싶습니다. 또한 코드도 최대한 작성한 만큼 올리면서 맞추어가는 시간을 가져보고 싶습니다. 이러한 체계를 하나씩 갖추어 진행하다보면, 약간의 진행 과정에서 지연은 있겠지만, 프로젝트 중후반부터 발표를 위해 차근차근 준비할 수 있는 시간을 가질 수 있었다고 생각합니다.

6. 개인 회고 - 오수현

팀에서의 나의 역할

봉사활동과 사용성 평가를 위한 대면 인터뷰를 통한 현장 경험을 바탕으로 프로젝트의 방향성을 구체화하는 데 기여하였습니다. 식료품에 대한 상세 설명 제공을 위해 실제 e-commerce 사이트에서 제공하는 설명의 종류와 형태를 두 번의 EDA를 통해 파악하였습니다. 결과적으로 프로젝트의 방법론, 파이프라인 구성에 대한 인사이트를 도출했습니다. 또한 리뷰 요약 및 키워드 기반 상품 추천 시스템 개발에서는 전반적인 파이프라인 설계와 구현에 참여하였습니다.

나는 내 학습 목표를 달성하기 위해 무엇을 어떻게 했는가?

EDA: 실제 온라인 쇼핑몰의 데이터를 수집하여 총 2889개(1차 2493개+2차 389개)의 상품에 대한 EDA 결과 도출된 케이스 분석을 통해 프로젝트의 전체 파이프라인 기획에 인사이트를 제공했습니다.

리뷰 요약 및 키워드 기반 상품 추천 시스템:

● 파이프라인 구성

- 리뷰 요약, 키워드 기반 상품 추천 시스템을 수행할 수 있는 전반적인 파이프라인 구성
- 서비스의 가능성 검증과 고도화를 위한 1차 프로토타입 구현

● 전처리

- 수집한 리뷰 데이터에 대해 길이별, 내용별 분석 후 필터링 수행
- 정형화된 형태를 위해 맞춤법 교정 방법론(Papago 역번역, Py-Hanspell, ET5 기반 모델) 비교
- ET5 기반의 모델을 활용하여 hate-speech 탐색

● ASTE(Aspect-based Sentiment Analysis)

- Rule-based, 형태소 분석 기반의 접근법 탐색
- HyperCLOVA X와 DeepSeek R1 모델의 성능 실험 및 비교 분석
- 프롬프팅: Zero-Shot, Few-Shot(2, 5, 10) 프롬프팅으로 HyperCLOVA X와 DeepSeek R1 모델 비교
- SFT: 최종 선정된 프롬프트를 사용하여 DeepSeek R1 모델에 대해 14B 모델과 quantized unsloth 32B 모델의 비교 분석을 통해 최적의 모델을 도출
- SFT를 통해 DeepSeek R1 모델의 F1-score를 15.07%p(56.06 → 70.13) 향상하여, HyperCLOVA X 대비 우수한 성능을 달성

● Clustering

- 다양한 Clustering 방법론을 조사하여 본 프로젝트와 적합한 거리 기반의 Hierarchical 알고리즘 비교
- 하이퍼파라미터 튜닝과 실험을 통해 Agglomerative 방식 선정
- 프로젝트에 적합한 평가 메트릭을 선정하여 이를 바탕으로 Cluster 기반 리뷰 요약 및 추천 키워드 추출 시스템을 구현

마주한 한계는 무엇이며 아쉬웠던 점은 무엇인가?

실시간 처리를 구현해보고 싶었으나, 예상보다 처리 시간이 길어지면서 부득이하게 DB에 저장하는 방식으로 전환할 수밖에 없었습니다. 리뷰 분석 및 추천 키워드 생성에 있어서 생성형 AI 모델을 활용했는데, 이를 현업에서는 실제로 어떻게 수행하는지에 대한 의문이 남아 아쉬웠습니다. 또한, ASTE를 수행에 있어서 보다 학문적으로 접근하지 못했습니다. 더불어, 클러스터링 방법론에 대한 지식이 부족하여 다양한 기법을 적용하고 표준화된 평가를 수행하지 못한 것도 아쉬운 점으로 남습니다. 이 부분은 향후 더 깊이 있는 학습을 통해 보완해야겠다고 느꼈습니다.

한계/교훈을 바탕으로 다음 프로젝트에서 시도해보고 싶은 점은 무엇인가?

이번 프로젝트를 바탕으로 현업에서는 애자일하게 어떤 방식으로 진행을 하고, 고도화, AI Safety 관점에서는 어떤 점을 집중적으로 살펴보는지를 경험해보고 싶습니다.

6. 개인 회고 - 윤선웅

팀에서의 나의 역할

저는 대표 이미지(썸네일)를 바탕으로 제품 정보를 심도 있게 분석·요약하는 전 과정을 주도하였습니다. 구체적으로, 상품의 구성 요소, 인증 마크, 색상, 재질 및 외형적 특징 등 세부 요소를 누락 없이 파악하여 체계적인 텍스트 데이터로 정리하였으며, 이를 토대로 소비자에게 명확한 제품 정보를 제공할 수 있도록 하였습니다. 또한, 다양한 VLM(Vision Language Model)을 면밀히 검증한 결과, Qwen2.5-VL-Instruct와 Janus-Pro-7B를 최적의 조합으로 활용하는 전략을 수립하여 모델의 성능 극대화를 도모하였습니다. 더불어, Hallucination 문제를 효과적으로 방지하기 위해 Few-Shot 프롬프팅, SFT 및 HyperCLOVA X 후처리 기법을 설계·적용함으로써, 신뢰도 높은 이미지 설명 시스템 구현에 기여하였습니다. 이러한 노력은 소비자에게 보다 정확하고 신뢰성 높은 제품 정보를 전달하는 데 핵심적인 역할을 하였습니다.

나는 내 학습 목표를 달성하기 위해 무엇을 어떻게 했는가?

최신 이미지 이해 및 자연어 생성 모델의 동향을 파악하기 위하여 관련 논문과 주요 벤치마크(NCSOFT/K-MMBench, K-LLaVA-W 등)를 심도 있게 분석하였습니다. 실제 프로젝트 환경에서 Qwen2.5-VL-Instruct와 Janus-Pro-7B 모델을 적용하여 성능 차이를 면밀히 비교·분석함으로써 이론적 지식과 실무 경험을 효과적으로 융합하였습니다. 대표 이미지 설명 태스크 수행 과정에서는 모델이 발생시키는 Hallucination, 정보 누락, 중복 표현 등 다양한 오류 유형을 체계적으로 점검하여, 개선 가능한 요소들을 도출하는 데 주력하였습니다. 또한, “모델 최적화 기법 이해”를 목표로 Few-Shot 프롬프팅, SFT, HyperCLOVA X 등 다양한 후처리 기법을 직접 설계하고 적용함으로써 모델 성능 향상에 기여하였습니다. 이와 같은 전 과정을 철저히 문서화하여 팀원들과 공유함으로써 개인의 역량은 물론, 팀 전체의 기술적 성장에 중요한 기여를 하였습니다.

마주한 한계는 무엇이며 아쉬웠던 점은 무엇인가?

Qwen2.5-VL-Instruct와 Janus-Pro-7B 모델은 자연어 생성에서는 우수했으나, 이미지 내 세부 구성 요소 인지에 한계를 보였습니다. 복잡한 텍스처, 미세한 아이콘, 인증 마크 등 세부 요소가 완전하게 포착되지 않아, 시각적 피처 추출 알고리즘과 데이터셋 보완이 필요함을 확인하였습니다.

Few-Shot 예시 구성 또한 실제 상황의 다양성과 복잡성을 충분히 반영하지 못해, 모델이 부정확하거나 왜곡된 정보를 생성하는 문제가 발생하였습니다. 이에 따라 보다 정교한 예시 데이터셋 구축과 프롬프트 엔지니어링 심화가 요구됩니다.

또한, 시간과 리소스 제약으로 HyperCLOVA X 후처리 기법에 대한 폭넓은 실험(파인튜닝 범위 세분화 및 후처리 스크립트 최적화)이 제한되어, 후처리 단계의 변수 조정 및 검증 범위가 축소된 점은 아쉬움으로 남으며, 향후 추가 자원 확보를 통한 심도 있는 연구가 필요합니다.

한계/교훈을 바탕으로 다음 프로젝트에서 시도해보고 싶은 점은 무엇인가?

다중 스케일 피처 추출 및 고해상도 이미지 활용 등 첨단 이미지 처리 기법과 데이터셋 보강 전략을 도입하여, 복잡한 텍스처와 미세 디테일의 정확한 인지를 실현하고자 합니다. 실제 상황의 다양성을 반영하는 정교한 Few-Shot 예시 구성과 프롬프트 엔지니어링 기법을 개발함으로써, Hallucination 및 부정확한 정보 생성을 효과적으로 완화할 예정입니다. HyperCLOVA X 후처리 기법에 대해 파인튜닝 범위의 세분화 조정과 스크립트 최적화를 포함한 폭넓은 실험을 진행, 최적의 성능 도출을 위한 변수 조정 및 검증 범위를 확장할 계획입니다.

6. 개인 회고 - 정민지

팀에서의 나의 역할

크게 두 가지 측면에서 기여했습니다. 첫 번째는 **기획 및 문제정의** 단계로, EDA와 설문조사를 통해 우리가 정의한 ‘시각장애인을 위한 온라인 식료품 쇼핑 서비스’가 해결이 시급하고 의미 있는 문제임을 구체적으로 도출하는 데 주로 참여했습니다. 이를 통해 “이 서비스가 정말 필요한가?”라는 의문을 실제 사용자 의견에 기반한 논리적 근거로 설명하고, 어떤 기능이 왜 필요한지 구체화하는 데 기여했습니다. 두 번째는 **모델링** 단계로, 상품 설명을 요약하고 상품의 주요 원재료/영양정보/알레르기 정보 등을 추출하는 모델 시스템 구축에 참여했습니다. 주로 LLM을 활용한 텍스트 정제를 수행했고, 이 성능을 높이기 위해 프롬프트 엔지니어링과 파인튜닝을 주도적으로 진행했습니다.

나는 내 학습 목표를 달성하기 위해 무엇을 어떻게 했는가?

[기획 및 문제정의]

- **EDA:** 프로젝트 초기에 수동 평가 기준을 구체화하고 (평가 항목 및 레이블 정의), 직접 수동평가를 진행했습니다. 이를 통해 데이터의 주요 특성을 파악하고 모델링 과정에 반영할 수 있었습니다.
- **설문 및 인터뷰:** 시각장애인에게 실제로 필요한 기능을 파악하기 위해 접촉 창구를 리스트업하고 설문 참여자를 모집하는 한편 대면 인터뷰도 준비했습니다. 인터뷰 시 필요한 질문 항목과 스크립트를 작성해, 팀원 모두가 일관된 환경에서 인터뷰를 진행할 수 있도록 세팅했습니다.

[모델링]

- **평가 기준 탐색:** 이미지 캡셔닝, 텍스트 요약, 할루시네이션 정량화 방법 등을 조사하고 정리했습니다. 이를 바탕으로 상품 영양정보 및 성분정보 추출의 정확도를 평가하는 평가 기준을 세우고, 평가용 데이터셋의 정답지를 구성했습니다.
- **상품 텍스트 요약:** 시각장애인이 이해하기 쉽게 상품 설명을 요약하는 모델을 설계하고, 프롬프트 엔지니어링 및 파인튜닝을 통해 성능을 높였습니다.
- **상품 상세 정보 추출:** 상품 상세 정보 이미지에서 OCR 모델로 텍스트를 표 형식으로 추출하고, 이를 LLM으로 재구성하도록 하는 방식의 고도화를 담당했습니다. GPT-4 등의 모델에서 이 흐름의 가능성을 확인해 파인튜닝용 골든셋을 직접 생성하며 고도화를 수행했으나, HCX 모델에서의 정확도가 낮아 결국 rule-base 방식과 혼합해 최종 결과물을 도출했습니다.

마주한 한계는 무엇이며 아쉬웠던 점은 무엇인가?

기획과 문제 정의 측면에서는 충분히 만족스러웠지만, 상대적으로 **모델 고도화**에는 많은 시간을 들이지 못해 아쉬움이 남았습니다. 프롬프트 엔지니어링에서 빠른 구현에 집중한 탓에 익숙한 전략과 스타일 위주로 사용했습니다. 프로젝트가 마무리된 후 다른 팀들의 다양한 프롬프트 접근법을 보며, 상품 상세 정보 추출에서 파인튜닝 성능이 낮았던 원인과 더 나은 전략을 적용할 기회를 놓친 것이 아쉬웠습니다.

한편 **문제 전달** 부분도 아쉬웠습니다. 시각장애인의 실제 온라인 쇼핑 환경에 대한 경험을 듣고 자료를 조사하며, 내부 팀원끼리는 충분히 공감과 이해를 끌어냈습니다. 그러나 이 분야에 배경지식이 전혀 없는 사람들에게는 얼마나 와닿았을지 확신하기 어려웠습니다. PPT를 만들 때 수치와 사례를 활용해 최대한 이해하기 쉽게 전달하려 노력했으나, 발표가 마무리된 지금 “과연 이 문제가 제대로 전달되었을까?”라는 고민이 남습니다. 만약 이 부분이 부족했다면, 어떻게 전달하면 더 잘 와닿을 수 있을 지 고민이 남습니다.

한계/교훈을 바탕으로 다음 프로젝트에서 시도해보고 싶은 점은 무엇인가?

앞으로는 이미 익숙한 방법론에만 머무르지 않고, 문제에 가장 적합한 다양한 접근 방식을 적극적으로 모색하며 모델링을 진행하고 싶습니다. 프롬프트 엔지니어링 및 파인튜닝 기법을 한층 더 정교하게 다루고 싶고, LLM을 더 다양하게, 잘 활용해보고 싶습니다. 여기에 덧붙여 생성형 AI 모델을 평가하는 기준(G-EVAL 등)에 대해 더 깊이 파고들며, 객관적이고 신뢰도 높은 평가 방법을 더 찾아보고 싶습니다.