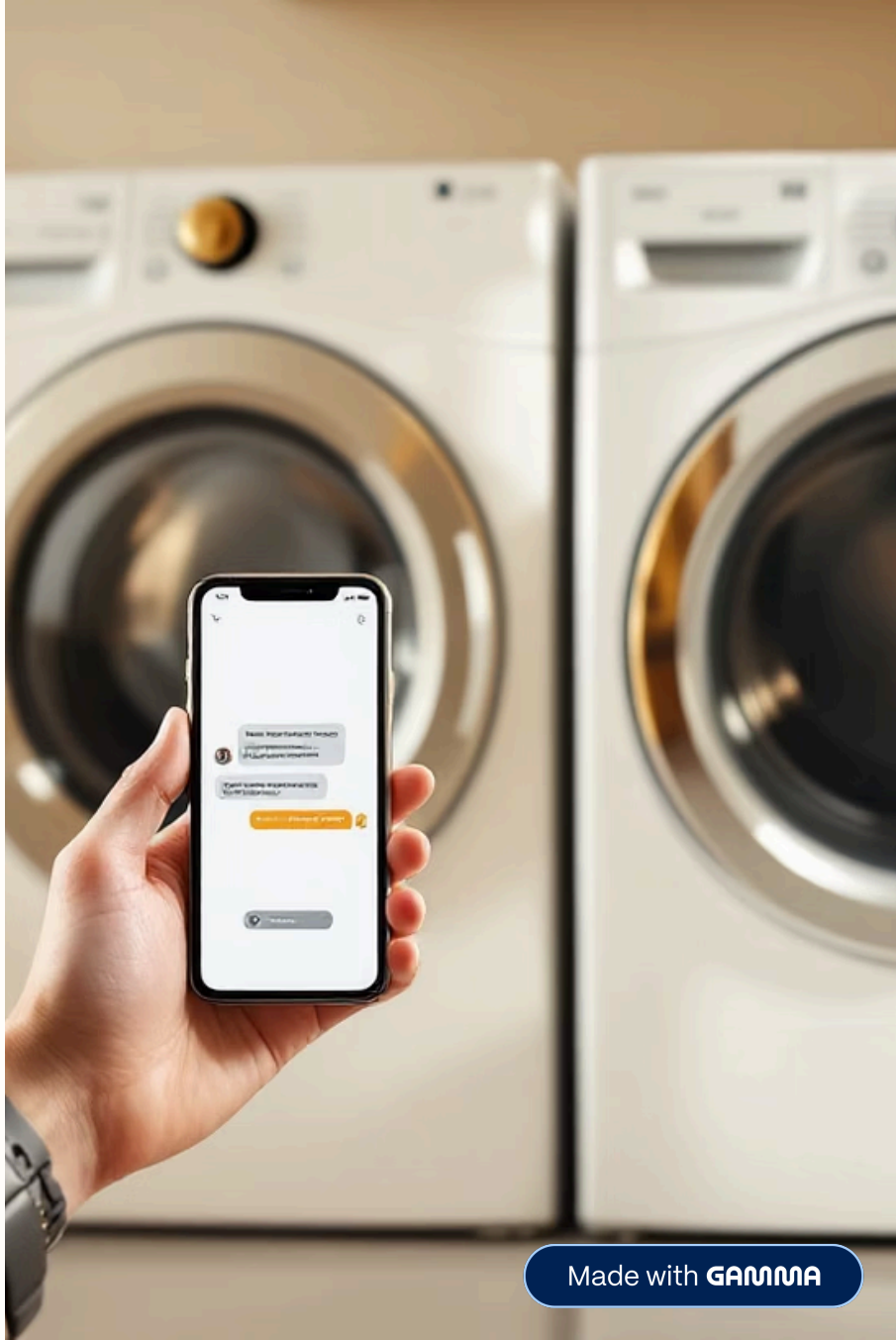
SKN14-3rd-3Team:

세탁기·건조기 매뉴얼 Q&A 챗봇

📅 진행 기간: 2025년 7월



Made with GAMMA

Contents

1. 팀 소개 및 프로젝트 개요
2. 시스템 구조 및 기술 아키텍처
3. 전체 파이프라인 요약
4. 데이터 수집/전처리/크롤링
5. AI 기술 구조: 임베딩, 벡터DB, 이미지 모델
6. LLM / RAG 기반 Q&A 및 Streamlit
7. 기술적 특장점
8. 참고 자료 및 활용코드

Made with GAMMA

팀 소개

1	유용환 삼성 제품 데이터 수집, 매뉴얼/이미지 크롤링, 모델명 자동 추출
2	윤이서 LG 제품 데이터 수집, 상세페이지 크롤링, 전처리
3	이수미 LLM 프롬프트·API 연동, QA 시스템 설계
4	김준기 매뉴얼 임베딩, 벡터DB 구축, 이미지 인식 기능 구현
5	정유진 PPT, README 문서화, 노션-프로젝트 기록
6	김재우 Streamlit UI, 사용자경험·결과 시각화

프로젝트 개요

목표: 이미지 업로드 또는 자연어 질문만으로도 제품 모델을 인식, 해당 모델의 공식 매뉴얼/실제 리뷰 기반 AI Q&A를 실시간 제공하는 지능형 챗봇.

- 고객센터/기술지원의 '비효율'과 '반복적 문의' 해소
- "내가 쓰는 기계, 이름을 몰라도" 사진 한 장이면 해결
- 삼성/LG 공식 정보 + 사용자 리뷰를 융합한 "현장 중심 Q&A"

Made with GAMMA

시스템 구조 및 기술 아키텍처

📐 아키텍처 다이어그램

데이터 수집/정제

Selenium/BeautifulSoup: 삼성/LG 공식몰 모든 모델, 옵션별 이미지·PDF·스펙·리뷰 자동 크롤링

파일명/폴더/CSV/메타데이터 표준화

이미지/텍스트 전처리

PIL/OpenCV: 이미지 리사이즈, 증강

pdfminer: 매뉴얼 PDF → 문단/문서화

OCR: 실사진/라벨에서 모델명 인식

임베딩 & 벡터DB

Chroma: VectorDB , 모델-문단-FAQ-리뷰 매핑

AI Q&A/LLM 구조

OpenAI GPT-4o-mini: Retrieval-Augmented Generation(RAG)

프롬프트: "제품명 + 질문 + 근거 매뉴얼/FAQ/리뷰" → 자연어 답변

답변의 근거 문단/리뷰/링크 자동 표시

Frontend/대시보드

Streamlit: 이미지 업로드, 질문 입력, 답변/FAQ/리뷰/링크 화면 구현

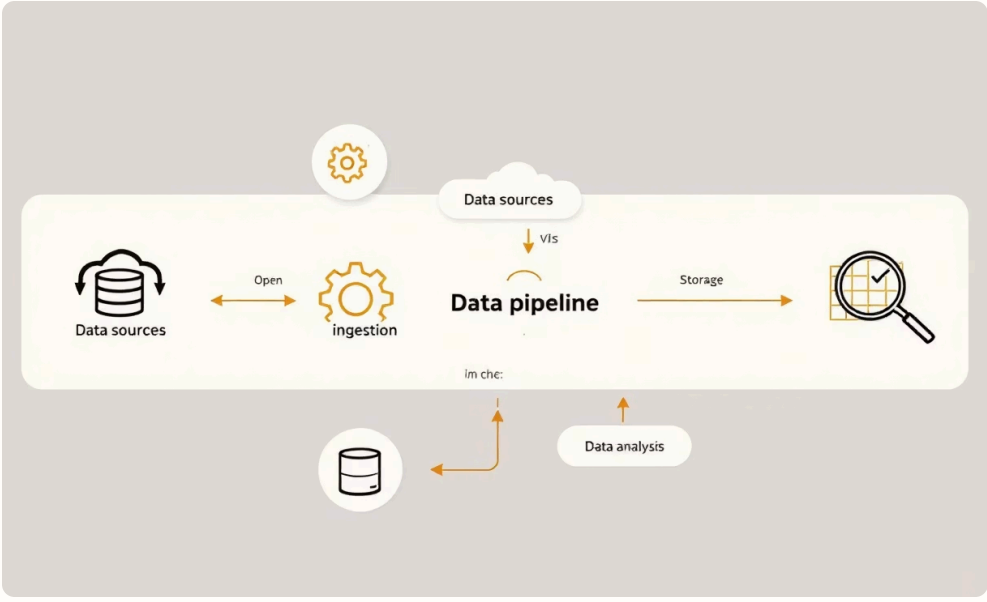
피드백(만족/불만/기타) 버튼, 운영자 관리 UI

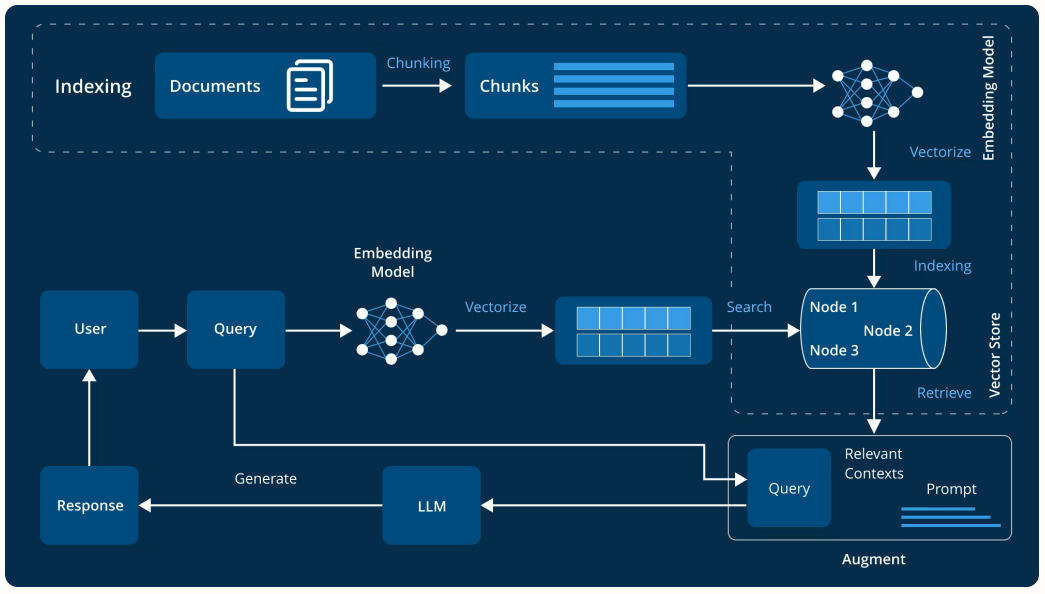
자동화/피드백/재학습

사용자의 피드백/미해결 문의 → 신규 DB 자동 갱신

Made with GAMMA

전체 파이프라인 요약





1. 크롬 웹드라이버 설정 및 사이트 접속
2. 품절상품 제외(옵션 해제)
3. “더보기” 반복 클릭해 모든 제품 로드
4. 모든 제품 한글/영문 모델명, 상세 링크 추출
5. 각 제품별 이미지 모두 다운로드(최대화질, 폴더/파일 자동생성)
6. 전체 메타데이터를 CSV로 저장
7. 최종 통계/에러/진행상황 출력
8. 리뷰 DB 구축 → 벡터DB 임베딩/LLM 질문응답 컨텍스트로 활용
9. 고객 질문 → 동일모델/증상 리뷰 본문/별점/키워드를 RAG로 노출, LLM에 근거 제공

기술적 의의

1. 실제 서비스에 바로 적용 가능한 **대량 웹 데이터 수집/정제 자동화** 예시
2. 리뷰/제품 데이터가 LLM·챗봇의 ‘지식베이스’가 되는 AI 파이프라인
3. 모든 단계가 반복/재사용/확장 가능하게 설계된 자동화 모듈

Made with GAMMA

데이터 수집/전처리/크롤링

all_image_downloader.py:

- "더보기" 자동클릭, 전체 제품/색상/슬라이드 이미지/모델명 동적 크롤링
- 폴더/파일명/CSV 표준화 (예: 제품명_모델코드_색상명_슬라이드번호.png)

all_pdf_downloader.ipynb:

- 각 모델별 매뉴얼/스펙 PDF 자동수집/폴더정리
- PDF→텍스트 변환, 제품/모델/문단/키워드 메타데이터화

lg_review_summary.js / lg_review_detail.js:

- LG 공식몰에서 요약/상세 리뷰 전체 크롤링, txt/zip 자동 저장
- 키워드/별점/장단점/이미지/동영상까지 구조화

🔽 [삼성/LG 공식몰 이미지·PDF·스펙·리뷰 자동화]

🗂️ 데이터 표준화/정제

- 모델명/코드 자동추출, 중복/오류/이상치 제거, 파일구조 및 CSV 자동 생성
- 이미지/문단/리뷰 모두 같은 Key값으로 매칭

AI 기술 구조: 임베딩, 벡터DB, Q&A

임베딩

- 최신 SOTA 모델 사용
- 각 매뉴얼 문단, 리뷰, 질문 모두 768 차원 이상 임베딩 벡터화
- Chroma에 VectorDB 저장

문서 검색(벡터DB)

- 사용자의 질문/이미지를 임베딩
- 벡터DB에서 '유사도' 기준 Top-N 문단/리뷰 즉시 추출
- 검색시 모델명 자동 태깅

이미지 모델

- 이미지 인식 기능 구현
- OCR(모델명 라벨 직접 추출)

Made with GAMMA

LLM/RAG 기반 Q&A 및 Streamlit

LLM/RAG 기반 Q&A

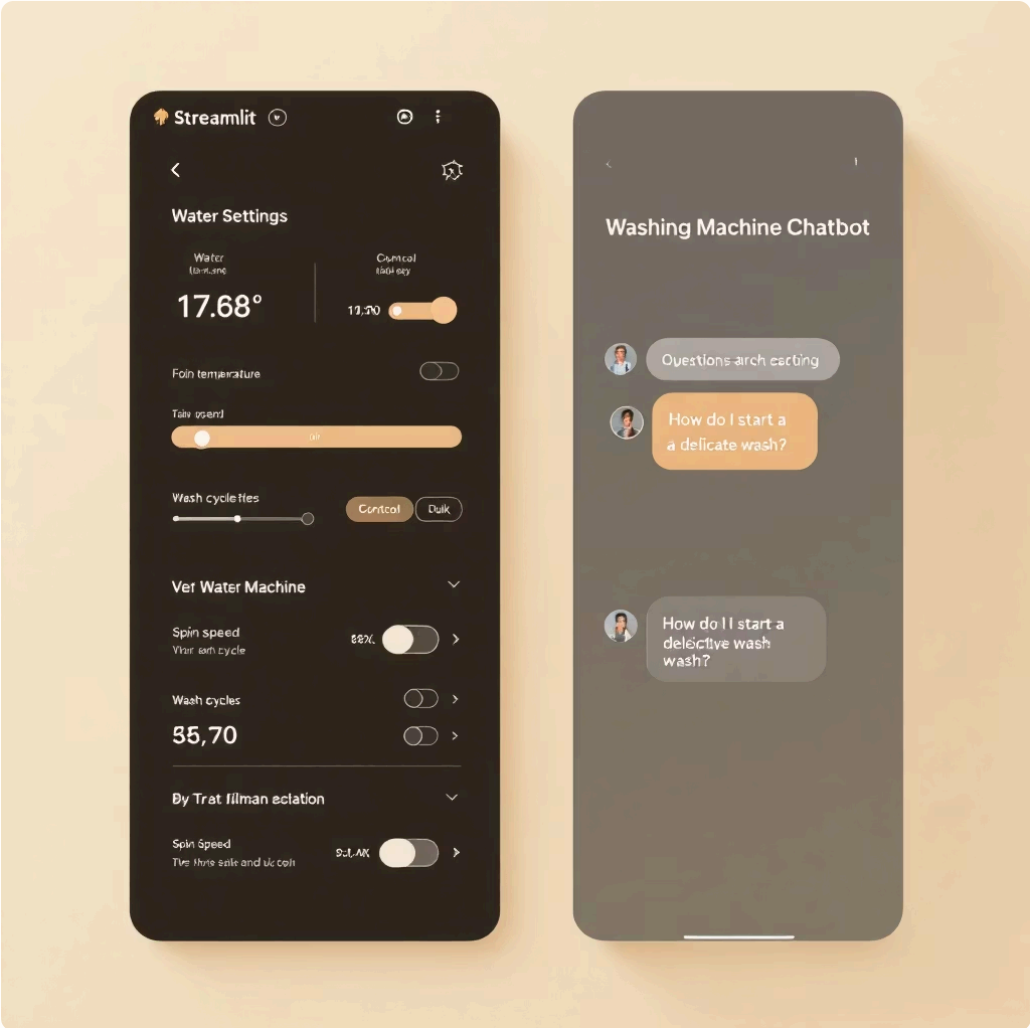
RAG 구조 Prompt

> [제품 모델명] [질문] [근거 매뉴얼 문단/실제 리뷰]
> → 답변 생성, 원문/링크 첨부

GPT-4o-mini(OpenAI API)

자동 피드백/지식 업데이트

- Streamlit에서 사용자 피드백 DB 자동 수집



🖼️ Streamlit 대시보드

- 이미지 업로드, 모델명 자동출력, 질문 입력, 답변·근거·리뷰·PDF/링크 등 자동표시
- 관리자: 신규 모델·PDF·리뷰 한 번에 업로드/재임베딩

실무 활용/확장

고객지원

제품 이름을 몰라도 사진/질문만으로 고장, 사용법, 에러코드, 청소법 등 바로 안내

고객센터/마케팅

자주 묻는 질문 자동 DB화, 신규 에러/이슈 빠른 대응

AI/LLM 음성 지원 기능

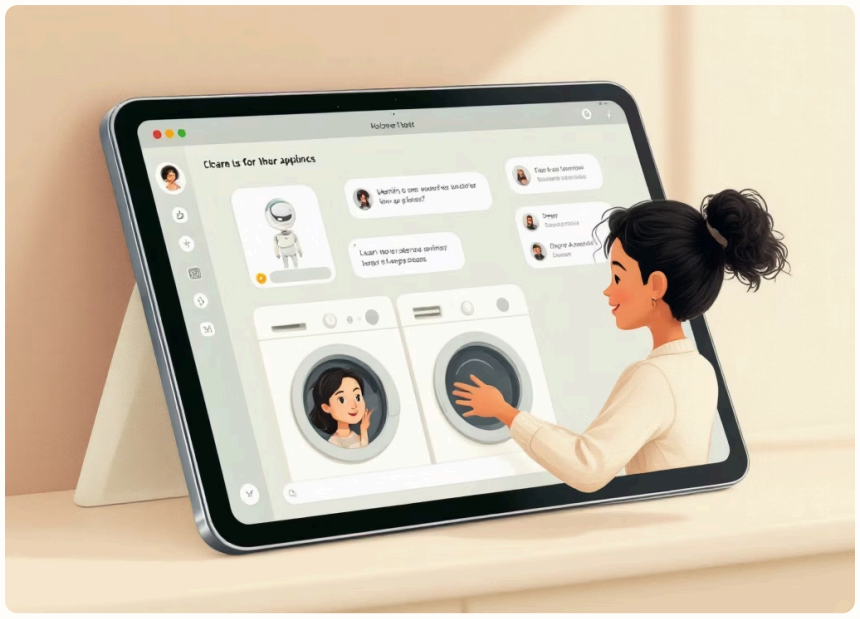
실제 리뷰 기반 RAG → 자체 챗봇/"음성"/AR 연동까지 확장

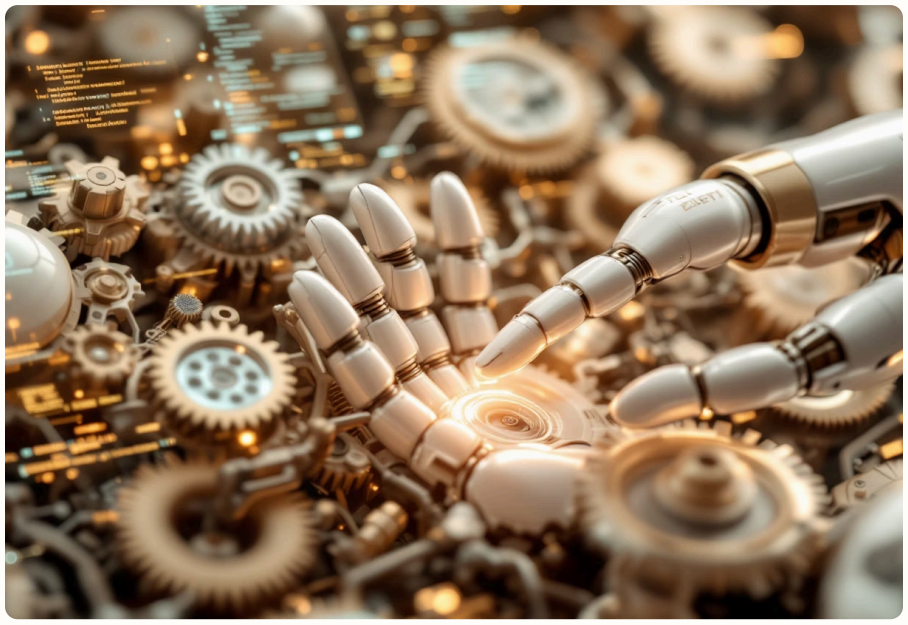
확장성

다국어, 타 브랜드, 추가 카테고리 적용

Made with GAMMA

기술적 특장점





기술적 특장점/구현 상세

- 동적 크롤링/이미지/텍스트/리뷰/스펙 완전 자동화
- CSV, 폴더, 파일명, DB까지 표준화
- CNN, OCR, OpenAI LLM, RAG 등 최신 AI기술 융합
- 예외처리/유지보수성/재사용성 극대화
- 클래스·모듈 구조화로 유지·확장성 매우 우수

참고 자료 및 활용 코드

- 프로젝트 노션 : https://www.notion.so/shqkel/SKN14-3rd-3-22d9cb46e5e2803f805ece660342eae9
- GitHub 저장소: https://github.com/skn-ai14-250409/SKN14-3rd-3Team
- [삼성/LG 공식몰, 다나와 등 크롤러/리뷰/매뉴얼/PDF 자동화 코드]
- [OpenAI, Huggingface, Pinecone, Streamlit, Selenium, BeautifulSoup, Figma, Tavily 등]

Made with GAMMA