

프로젝트 기술서

김정희 Jounghee Kim · AI Engineer · LLM Agents

onlyl4youu@gmail.com linkedin.com/in/joungheekim github.com/JoungheeKim 작성 2026.09

요약

LLM 에이전트로 사람이 판단하던 업무를 자동화하는 AI Engineer입니다. 토스에서 browser-use 기반 가맹점 심사 자동화로 카드사 심사 인력 20명을 절감했고, AI가 프롬프트 고도화를 이끄는 검수 자동화 플랫폼을 10개 이상 부서에 적용했습니다. SK텔레콤에서는 일정 등록의 67%를 처리한 에이닷 일정 에이전트와 영화 예매·메인 에이전트를 설계·개발했으며, 음성인식·NLP 연구(Interspeech 2022)와 금융 AI 경험을 갖추고 있습니다.

주요 성과

20명	심사 인력 절감 — 01 가맹점 심사 자동화
10+	사용 부서 — 02 검수 자동화 플랫폼
+25%p	평균 정확도 향상 (단일 정책 파일럿) — 03 광고 검수 에이전트
67%	에이전트 경유 등록 — 04 에이닷 일정 에이전트
60→80%	합수 호출 성공률 — 05 에이닷 메인 에이전트
12%	예매 전환율 — 06 영화 예매 에이전트 (T멤버십·에이닷)
16→31%	AI 자동 답변율 — 07 쿠팡드 유사 질문 답변 추천

목차

주식회사 비바리퍼블리카(토스) · 2025.12 – 현재

01	가맹점 심사 자동화	2026.05 – 현재
02	검수 자동화 플랫폼	2026.06 – 현재
03	광고 검수 에이전트	2026.03 – 2026.06

SK텔레콤 · 2022.04 – 2025.10

04	에이닷 일정 에이전트	2025.04 – 2025.10
05	에이닷 메인 에이전트	2024.04 – 2024.12
06	영화 예매 에이전트 (T멤버십·에이닷)	2024.01 – 2024.12
07	쿠팡드 유사 질문 답변 추천	2022.05 – 2023.12

가맹점 심사 자동화

진행 기간	2026.05 – 현재
소속	비바리퍼블리카(토스)
역할	심사 에이전트 설계 및 개발

1. 배경 및 목표

- 심사자는 가맹점 홈페이지를 직접 열어 사업자 정보·약관·결제 경로를 확인하고 캡처해 왔고, 수기 심사에 평균 3일 정도가 걸렸습니다. 이를 당일 안내를 목표로 자동화했습니다.
- 가맹점마다 사이트 구조가 달라, 필요한 정보가 있는 페이지를 사람이 하나씩 찾아 들어가야 했습니다.

2. 핵심 과제

- 1) 구조가 제각각인 가맹점 사이트에서 심사에 필요한 정보를 빠짐없이 찾기
- 2) 외부 가맹점 서버에 부담을 주거나 보안 문제를 일으키지 않기
- 3) 정보가 충분히 모였는지 스스로 판단하고 탐색을 마치기

3. 기여 내용

A. browser-use 기반 심사 에이전트 설계·개발 (과제 1~3)

- a. 화면(스크린샷)과 페이지 텍스트를 함께 보고 다음에 확인할 곳을 정하는 탐색 루프(관찰 → 판단 → 행동) 설계
- b. 사업자 정보·약관·결제 경로 등 심사 항목별 수집 목표를 두고, 아직 부족한 정보를 근거로 다음 행동 결정
- c. 필요한 정보가 모이면 스스로 탐색을 마치고, 수집한 화면과 텍스트를 심사 근거로 함께 남김

B. 외부 서버를 고려한 안전 설계 (과제 2)

- a. 공개 페이지만 읽기 전용으로, 가맹점마다 브라우저 하나가 한 페이지씩 순서대로 탐색해 가맹점 서버가 받는 부담을 일반 방문자 한 명 수준으로 유지

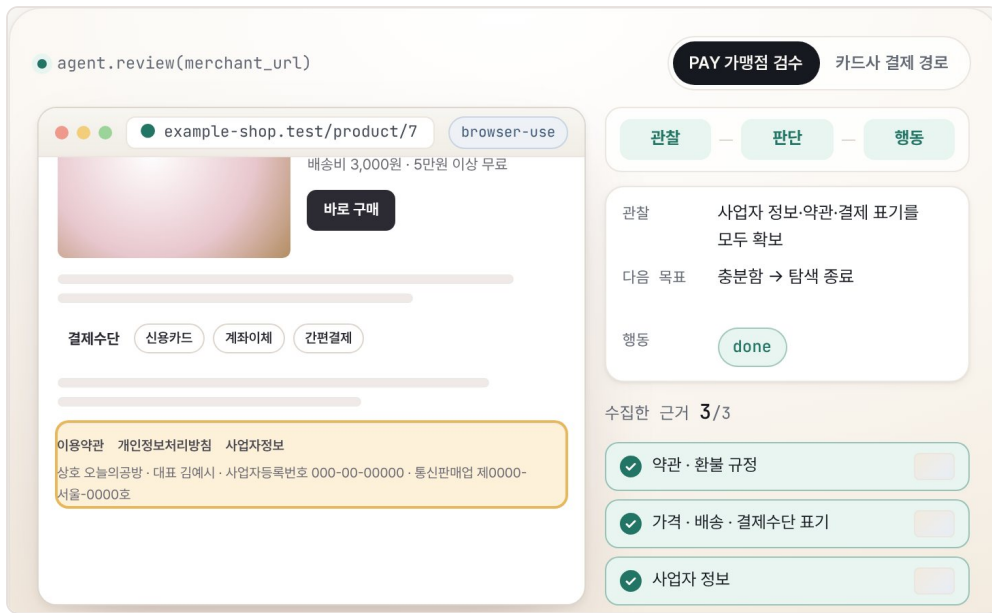
4. 기술 스택

프레임워크 · 플랫폼	browser-use, Headless Chromium, 멀티모달 LLM
방법론	Agentic Browsing, Prompt Engineering

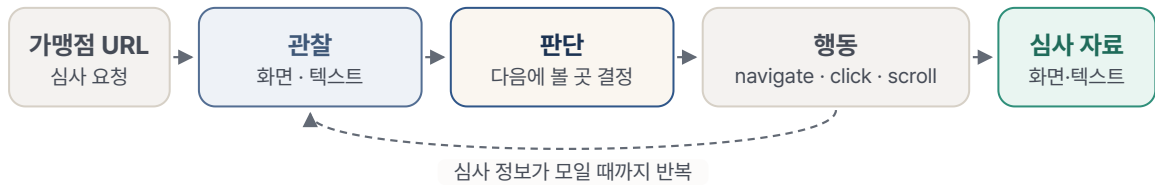
5. 성과

- 카드사 가맹점 심사(결제 경로 캡처) 자동화로 심사 인력 20명 절감
- 같은 에이전트를 PAY 가맹점 검수(8명 규모 업무)에 적용

6. 참고 자료



재현 화면 · 합성 데이터



browser-use 기반 심사 에이전트 · 공개 페이지만 읽고, 모든 화면과 텍스트를 심사 근거로 전달

검수 자동화 플랫폼

진행 기간	2026.06 – 현재
소속	비바리퍼블리카(토스)
역할	플랫폼 설계 및 개발

1. 배경 및 목표

- 검수 업무마다 에이전트를 새로 만들고, 결과가 기대와 다르면 사람이 프롬프트를 직접 고치며 다시 돌려 봐야 했습니다.
- 결과가 어긋나도 프롬프트의 문제인지, 정책·기준(방향성) 자체의 문제인지 구분하기 어려웠습니다.
- 요구사항과 정책·사례가 메신저·문서·시트 등 여러 곳에 흩어져 있어, 에이전트를 만들기 전의 준비에도 시간이 많이 들었습니다.

2. 핵심 과제

- 1) 실행 결과로 고도화할 때 프롬프트로 고칠 문제와 정책·기준(방향성)을 다시 정해야 할 문제를 구분하기
- 2) 요구사항 정리·에이전트 구성·프롬프트 수정처럼 사람이 반복하던 중간 작업을 줄이기
- 3) 여러 검수 업무가 정책·데이터·입출력 형식을 함께 관리하며 버전별로 실험하고 개선하기

3. 기여 내용

A. AI 기반 고도화 사이클 (과제 1)

- a. 실행 기록과 검수 데이터에서 반복되는 형식·근거 오류를 AI가 찾아, 에이전트 구조를 고려한 프롬프트 변경안을 만들고 같은 데이터로 다시 실행해 비교
- b. 프롬프트 수정만으로 해결되지 않는 오류는 데이터와 현재 방향성(정책)이 충돌하거나 비어 있는 지점으로 보고, 근거 사례·영향 범위·보완안과 함께 사용자에게 안내
- c. 사용자는 방향을 정하고 승인만 하며, 수정·재실행·비교는 AI가 맡아 고도화 사이클을 간편하게 반복

B. 중간 단계를 줄이는 AI 간편 기능 (과제 2)

- a. 메신저·문서·시트·담당자 설명처럼 흩어진 자료를 AI가 모아 검수 요구사항 문서를 자동 작성하고, 빠진 내용은 확인할 질문으로 정리
- b. 정리된 요구사항을 바탕으로 AI가 필요한 구성 요소와 연결 순서를 제안해 검수 에이전트 초안을 만들고, 만들 수 없는 부분과 빠진 요구를 함께 안내
- c. 검수 업무별 프롬프트를 AI가 자동으로 만들고, 버전별 실험으로 프롬프트 버전 간 결과를 비교할 수 있게 지원

C. 플랫폼 구조와 공통 모듈 (과제 2·3)

- 검수 에이전트는 LangGraph로 실행하고, 요구사항 정리·에이전트 초안·프롬프트 학습·정책 고도화 같은 중간 AI 기능은 Claude Agent SDK 기반 서비스로 구현해 제공
- 정책·데이터·사례·입출력 형식을 함께 관리하는 기준 정보로 두고, 업무별 에이전트를 개발 → 실행 → 개선 흐름으로 연결
- 실행 결과를 판정·근거·실행 이력·버전과 함께 남기고, 검수자가 확정·수정·보류한 결과를 다음 개선에 반영

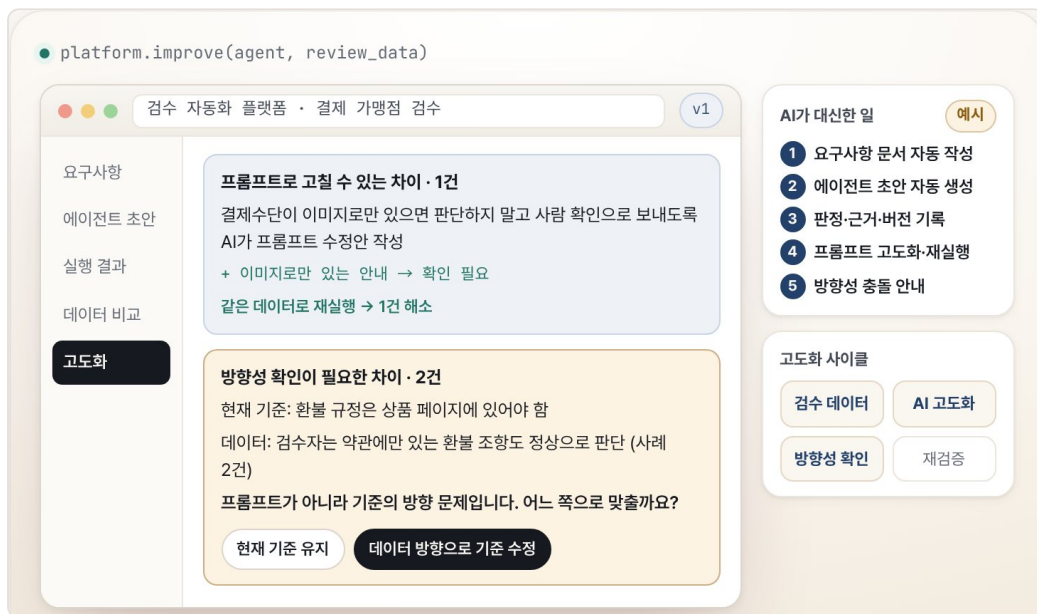
4. 기술 스택

프레임워크 · 플랫폼	LangGraph, Claude Agent SDK
방법론	요구사항 자동 정리, 에이전트 초안 자동 생성, AI 기반 프롬프트 학습, 정책 고도화 (충돌·누락 분석), 버전별 실험

5. 성과

- 검수 공통 모듈로 10개 이상 부서의 여러 검수 업무에서 활용

6. 참고 자료



재현 화면 · 합성 데이터



AI가 요구사항 정리·에이전트 초안·프롬프트 고도화를 맡고, 프롬프트로 풀 수 없는 방향성 충돌은 근거 사례와 함께 사용자에게 안내해 고도화 사이클을 간편하게 만들

광고 검수 에이전트

진행 기간	2026.03 – 2026.06
소속	비바리퍼블리카(토스)
역할	설계 및 개발

1. 배경 및 목표

- 광고 정책은 한두 줄짜리 메모인데, AI가 검수하려면 판별 기준·경계·예시를 갖춘 검수 문서가 필요했고 이를 정책마다 사람이 써야 했습니다.
- 문서 기준이 실제 검수자 판단보다 엄격하거나 느슨하면 정상 광고를 반려하거나 위반을 놓쳤지만, 어디가 어긋나는지 찾아 고치기 어려웠습니다.

2. 핵심 과제

- 1) 짧은 정책을 검수할 수 있는 기준으로 만들기
- 2) 문서 기준과 검수자 판단의 차이를 데이터로 줄이기
- 3) 개선에 쓴 데이터에만 맞춰지지 않게 하기

3. 기여 내용

A. 검수 문서 자동 생성 (과제 1)

- a. 정책 원문에서 근거를 찾아, 짧은 정책을 판별 기준·위반/정상 경계·정책 근거를 갖춘 검수 문서로 자동 변환

B. 데이터 기반 고도화 (과제 2·3)

- a. AI가 문서 판정과 검수자 판정을 비교해 어긋난 패턴을 찾고 문서를 개정 (예: 광고에서 통용되는 붙여쓰기·금액 단위·광고 말투는 정상으로 명시)
- b. 개선용과 검증용 데이터를 나눠, 개선 효과가 처음 보는 광고에서도 유지되는지 확인
- c. 사람은 개정안을 확인하기만 하고 문서를 직접 고치지 않음

4. 기술 스택

프레임워크 · 플랫폼	LLM, Prompt Engineering
방법론	데이터 기반 프롬프트 고도화, Holdout 검증

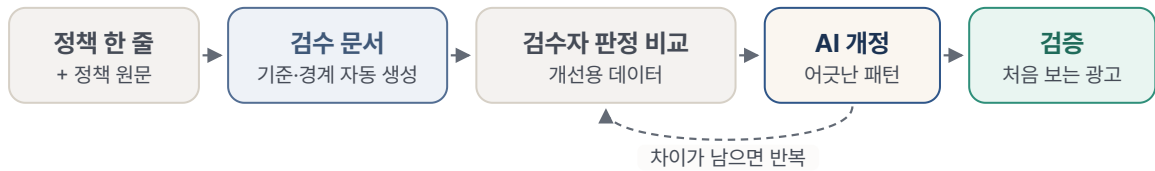
5. 성과

- 광고 문구 오탈자 정책 파일럿(광고 36건)에서 평균 정확도 25%p 향상
- 개선에 쓰지 않은 검증용 광고에서도 정확도가 함께 올라, 개선 효과가 새 광고에도 유지됨을 확인
- 이 방식이 검수 자동화 플랫폼의 프롬프트 고도화 사이클로 이어짐

6. 참고 자료



재현 화면 · 합성 데이터



정책 → 검수 문서 → 데이터 기반 AI 개정 → 검증용 광고로 확인 · 파일럿 정책에서 평균 정확도 +25%p

에이닷 일정 에이전트

진행 기간	2025.04 – 2025.10
소속	SK텔레콤
역할	에이전트 설계 및 개발

1. 배경 및 목표

- 에이닷의 일정과 알림을 대화로 처리하는 에이전트를 만들었습니다.
- "내일 3시 치과 넣고, 금요일 회의 30분 미뤄줘"처럼 여러 의도가 섞인 복합 발화를 빠르고 정확하게 처리하는 것이 목표였습니다.

2. 핵심 과제

- 1) 복합 발화 처리와 응답 속도
- 2) 외부 일정(Outlook·Google Calendar)과 구독 일정 연동
- 3) 반복 일정·알림 등 시간 정보의 정확한 처리

3. 기여 내용

A. "Plan-and-Execute" 패턴과 "Refine" 단계를 적용한 에이전트 아키텍처 (과제 1)

- a. Plan: 사용자의 요청을 하위 작업(Sub-Task)으로 나누고 작업 간 순서를 부여
- b. Execute(순차): 순서가 있는 하위 작업은 차례대로 끝까지 실행
- c. Execute(병렬): 서로 영향을 주지 않는 하위 작업은 병렬로 실행해 응답 속도 향상
- d. Refine: 실행 결과를 바탕으로 재계획(Re-planning)해 작업 정확도 향상

B. 외부 일정·구독 일정 연동 설계 (과제 2)

- a. 외부 일정(Outlook, Google Calendar)과 구독 캘린더(હે택·이벤트 일정)를 LLM이 구분해 연동 여부와 조회·변경·삭제 가능 여부를 판단하도록 설계

C. 일정 추출 RAG 설계 (과제 1)

- a. 필터링: 일정 관리 주체, 구독 여부, 조회 일자 등으로 후보 일정 도출
- b. 병렬 검색: 후보 일정을 N개로 나누고 LLM을 병렬로 호출해 제한 응답 시간 내에 요청한 일정 추출
- c. 검증: 추출한 일정이 요청에 해당하는지 판단하고 근거를 함께 추출해 정확도 향상

D. 시간 정보 처리를 위한 프롬프트 엔지니어링 (과제 3)

- a. 반복 일정과 알림을 정확히 다루도록, 일정 국제 표준(iCalendar)과 ISO 8601 기간 표기를 쓰게 프롬프트 설계

4. 기술 스택

프레임워크 · 플랫폼
방법론

LangGraph, Gradio, FastAPI
Prompt Engineering, RAG, Plan-and-Execute

5. 성과

- 일정 등록의 67%가 에이전트 경유 (에이전트 6.6만 건 / 전체 9.8만 건)
- 서비스 전체 일정 기능 MAU 2.5배 성장 (2025.01 20.3만 → 2025.09 51.1만)

6. 참고 자료



재현 화면 · 합성 데이터



Plan-and-Execute: 독립 작업은 병렬 실행, 결과로 재계획

- 에이닷 에이전틱 워크플로우 도입 (SKT 뉴스룸)
news.sktelecom.com
- 에이닷 캘린더 반복 일정 구현 (DEVOCEAN)
devocean.sk.com
- Plan-and-Execute Agents (LangChain)
blog.langchain.com
- SKT-TimeTree 파트너십 (SKT 뉴스룸)
news.sktelecom.com
- 에이닷 4.0 에이전틱 워크플로우 (SKT 뉴스룸)
news.sktelecom.com

에이닷 메인 에이전트

진행 기간	2024.04 – 2024.12
소속	SK텔레콤
역할	에이전트 설계, 프롬프트 엔지니어링, 학습 데이터 정의, 모델 파인튜닝

1. 배경 및 목표

- 에이닷의 대표 LLM 에이전트로, 일상 대화와 환율·날씨·시간·길 찾기·뉴스·지하철 혼잡도 등 17가지 기능을 제공합니다.

2. 핵심 과제

- 함수 호출 정확도와 할루시네이션
- API 비용과 응답 속도
- 멀티턴 대화 품질

3. 기여 내용

A. 답변 정확도와 UI/UX 연결을 고려한 에이전트 설계 (과제 1)

- 유사한 함수를 통폐합해 함수 선택을 단순화
- 대화 외 앱 정보(재생 중인 목록, UI 화면)를 대화에서 활용하도록 설계
- 인자 선택 정확도가 낮은 함수는 작업을 나누어 순차적으로 호출하도록 설계

B. 비용과 답변 정확도를 위한 프롬프트 엔지니어링 (과제 1·2)

- API 비용을 줄이기 위해 프롬프트를 최소화하고 파인튜닝으로 답변 경향성 주입
- 할루시네이션이 잦은 시간 관련 함수는 인자 추출을 NER 방식으로 설계

C. 모델 파인튜닝과 평가 (과제 3)

- 학습 데이터 포맷을 정의하고 멀티턴 대화 데이터 생성
- Azure OpenAI로 GPT 모델 파인튜닝
- 정량 평가(함수 선택 여부·정확도, 인자 선택·추출 정확도)와 정성 평가(멀티턴 유창성, 함수 결과 안내 정확도, 민감한 질문 답변 적합성, 부적절한 요청 대응)

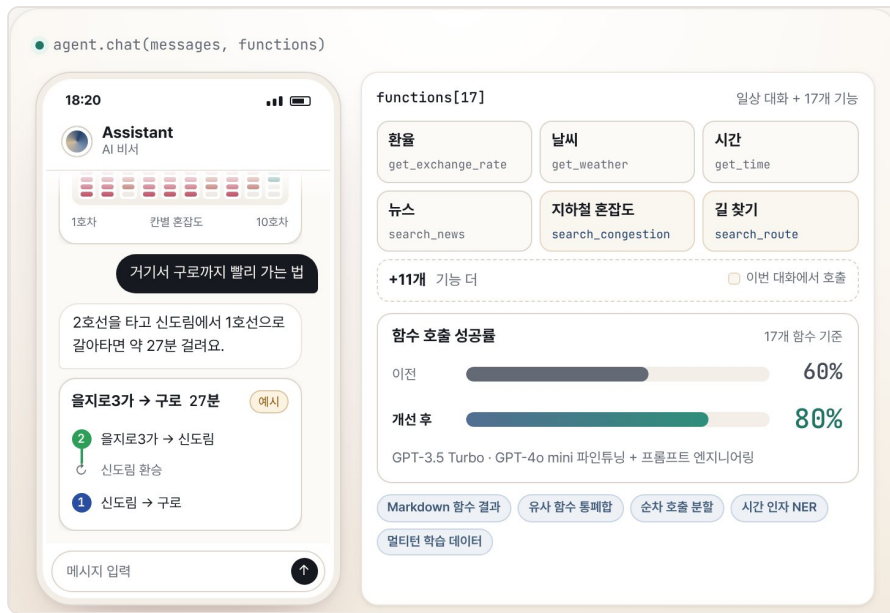
4. 기술 스택

프레임워크 · 플랫폼	Azure OpenAI
방법론	Prompt Engineering, Function Calling, Fine-tuning, NER

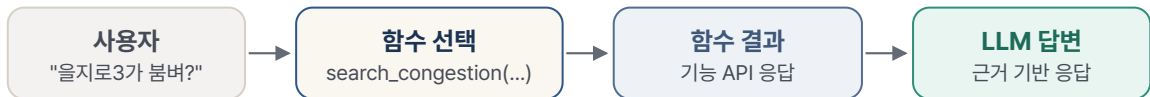
5. 성과

- 파인튜닝과 프롬프트 엔지니어링으로 17개 기능(함수) 호출 성공률 60% → 80%
- 시간 관련 인자를 NER로 추출한 뒤 순차 분석해 할루시네이션 감소·시간 인식 정확도 향상

6. 참고 자료



재현 화면 · 합성 데이터



함수 호출로 답을 만드는 에이전트 · 학습 데이터 설계·파인튜닝·프롬프트 개선으로 17개 함수 호출 성공률 60% → 80%

– 에이닷 LLM 에이전트 체험기 (SKT 뉴스룸)

news.sktelecom.com

영화 예매 에이전트 (T멤버십·에이닷)

진행 기간	2024.01 – 2024.12
소속	SK텔레콤
역할	UX & LLM 통합 Workflow 설계, RAG 개발

1. 배경 및 목표

- T멤버십 영화 예매를 대화로 제공하는 에이전트로, 영화관 직원과 대화하듯 영화를 추천·검색하고 쉽게 예매할 수 있어야 했습니다.
- 대화뿐 아니라 버튼 등 UI도 함께 쓰는 서비스여서 UI/UX와 LLM Workflow가 자연스럽게 이어져야 했습니다.

2. 핵심 과제

- 1) 대화로 예매 정보를 수집하는 Slot Filling과 개인화
- 2) UI/UX와 LLM Workflow의 연동
- 3) 영화 정보 검색 정확도, 응답 속도, API 비용

3. 기여 내용

A. Slot Filling 에이전트 설계·개발 (과제 1)

- a. 대화로 영화 정보를 안내하면서 예매에 필요한 정보를 모으도록 Function Calling 기반 LLM Workflow 설계
- b. 사용자 기본 정보(위치, 최근 방문 영화관, 선호 극장)를 반영한 추천 정보 제공
- c. 긴 대화에서 수집한 정보(Slot)가 누락되지 않도록 상태 관리 모듈 적용

B. 영화 검색 RAG 설계·개발 (과제 3)

- a. 전처리: 영화명 유의어 사전을 저장한 Trie 구조로 요청에서 영화 이름 태깅
- b. 키워드 추출: LLM으로 요청에서 영화 관련 메타·키워드 추출
- c. 관련 영화 검색: 추출한 메타·키워드로 Vector 검색과 메타 필터링 적용
- d. 최종 추천: LLM으로 요청에 맞는 영화 선정

C. 영화 메타 분석 파이프라인 자동화 (과제 3)

- a. Airflow로 매일 업데이트된 영화 정보에서 LLM과 전처리 모듈로 키워드·요약·메타 추출
- b. 추출한 정보를 Vector로 변환해 Vector DB에 동기화

D. UI/UX와 LLM Workflow 연동 설계 (과제 2)

- a. UI(버튼 등)로 수행한 결과를 LLM Workflow의 프롬프트에 포함해 두 기능이 자연스럽게 이어지도록 설계

E. 로그 기반 데이터 수집과 모델 학습 (과제 3)

- Function Calling 정확도와 멀티턴 유창성을 높이도록, 서비스 로그에서 학습 데이터를 모아 GPT 모델 파인튜닝 (Azure OpenAI)

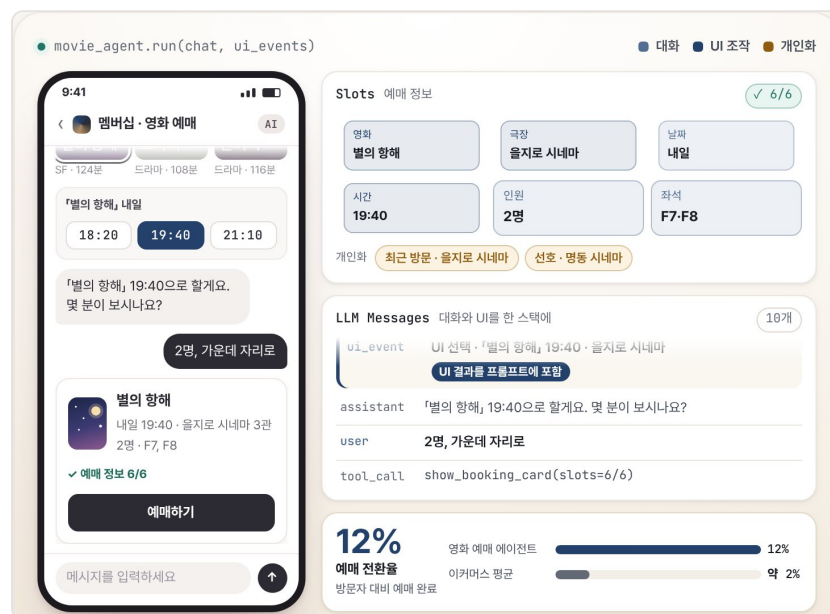
4. 기술 스택

프레임워크 · 플랫폼 Azure OpenAI, Datadog, Airflow
방법론 Prompt Engineering, Function Calling, RAG

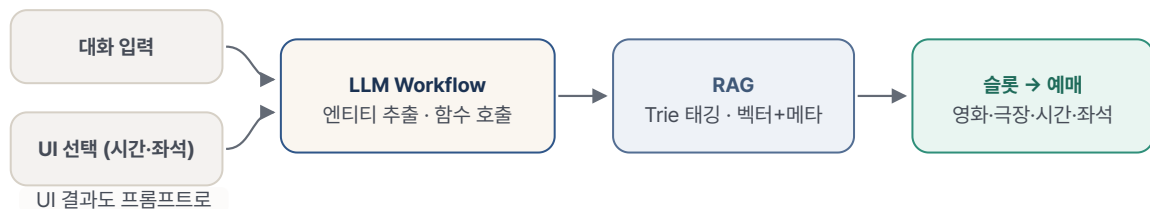
5. 성과

- T멤버십 영화 예매 MAU 8,000명, 영화 예매 에이전트 누적 사용자 11만 명
- 예매 전환율 12% (T멤버십 영화 예매 방문자 대비 예매 완료)

6. 참고 자료



재현 화면 · 합성 데이터



대화과 UI가 하나의 LLM Workflow로 이어지는 예매 흐름

- 에이닷 LLM 에이전트 체험기 (SKT 뉴스룸) news.sktelecom.com
- 영화 예매 에이전트 LLM Workflow (SKT Enterprise) sktenterprise.com

큐피드 유사 질문 답변 추천

진행 기간	2022.05 – 2023.12
소속	SK텔레콤
역할	모델 개발, 추천·평가 파이프라인 구축

1. 배경 및 목표

- 큐피드는 위치를 기반으로 사용자끼리 질문하고 답하는 커뮤니티 서비스입니다.
- 비슷한 질문에 이미 달린 답변을 찾아 주는 AI 기능으로 사용자 편의를 높이하고자 했습니다.

2. 핵심 과제

- 1) 주제·지역별 질문-답변 매칭 정확도
- 2) 주요 모델의 성능 평가 자동화

3. 기여 내용

A. 콘텐츠 기반 추천 파이프라인 구축 (과제 1)

- a. 분류·추출: 질문에서 위치와 카테고리 추출 (BERT 계열 분류·NER 모델)
- b. 필터: 위치, 카테고리, 불용어 기준으로 매칭 후보 필터링
- c. 변환: 장소 정보를 마스킹하고 임베딩 모델로 벡터 변환
- d. 검색: Elasticsearch에서 BM25 + 코사인 유사도로 유사 질문 후보 추출
- e. 재순위화: 질문 간 포함 관계와 질문-답변 정합성을 판단해 순위를 매기고 상위 N개 제공

B. 추천 파이프라인에 쓰인 모델 개발 (과제 1)

- a. 임베딩 모델: 유사한 문장이 가깝게 놓이도록 Siamese Network 방식으로 BERT 계열 모델 학습
- b. 계층형 다중 분류 모델: 계층 구조의 주제 분류 데이터를 만들고 BERT 계열 모델 학습
- c. NER 모델: 질문에서 장소·키워드를 추출하도록 BIO 태깅 방식으로 학습

C. 모델 평가 파이프라인 구축 (과제 2)

- a. 카테고리가 바뀐 데이터와 전체 중 N개 샘플을 외부 업체에 맡겨 정답(Ground Truth) 라벨링
- b. F1·Accuracy 기준으로 평가하고 예측 정확도가 낮은 데이터 추출
- c. 재학습 후 Golden Set 평가를 거쳐 배포 결정

D. 신규 데이터 배치 처리 (과제 1)

- a. 새로 올라온 질문-답변이 추천 후보에 반영되도록 Airflow 배치로 벡터 변환·저장과 Vector DB 재색인 자동화

4. 기술 스택

프레임워크 · 플랫폼

PyTorch, Transformers, Elasticsearch, Airflow

방법론

Siamese Network, Hierarchical Multi-class Classification, BIO NER

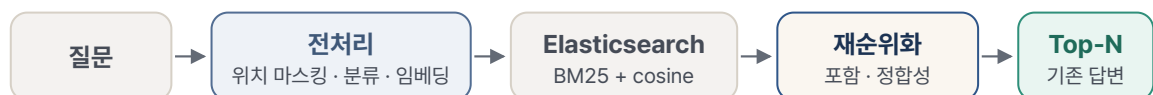
5. 성과

- 중복 질문에 대한 AI 자동 답변율 16% → 31% (2022.08.01 → 2022.08.21)
- AI 답변의 긍정 피드백("도움이 됐어요") 비율 38% → 57%

6. 참고 자료



재현 화면 · 합성 데이터



비슷한 질문의 기존 답변을 찾는 콘텐츠 기반 추천 파이프라인

- 큐피드를 만든 사람들 (SKT 뉴스룸)

news.sktelecom.com

그 밖의 프로젝트

K-Wav2vec 2.0: 자소·음절 Joint Decoding 기반 한국어 음성인식 Interspeech 2022 · 1저자	2022.09
차량·모바일 음성인식 시스템 현대자동차 에어랩	2022.02 – 2022.04
일관성 정규화를 적용한 멀티모달 한국어 감정인식 대한산업공학회지 2021 · 1저자	2021.12
Back-Translated Task Adaptive Pretraining arXiv 2021 · 2저자	2021.07
오픈 도메인 한국어 질의응답 시스템 고려대학교 (대학원 프로젝트)	2020.04 – 2020.07
화학 공정 예지보전 시스템 고려대 DSBA × 한화시스템 ICT부문 산학과제	2020.01 – 2021.09
마켓캐스터 – 뉴스 기반 금융시장 예측 SK C&C	2019.07 – 2020.03
포트폴리오 로보어드바이저 SK C&C	2018.07 – 2019.02
기업 부실징후 조기인식 시스템 SK C&C (고객 KDB산업은행)	2018.01 – 2018.07
KDB산업은행 전자금융 서비스 SK C&C (고객 KDB산업은행)	2016.01 – 2017.03

상세 내용: joungeekim.github.io/portfolio/ko

* 참고 자료의 화면과 도식은 동작 방식을 설명하기 위한 재현 예시이며, 실제 화면·데이터가 아닙니다.