

AI 고객상담 Agent 평가 결과 보고서

주요 평가 결과와 개선 과제를 중심으로 정리한 보고 자료

(주)네피리티

2026-07-24

EXECUTIVE SUMMARY

핵심 요약

평가 결과와 주요 개선 사항을 중심으로 핵심 내용을 정리했습니다.

테스트 결과 핵심 정책 문의에 대한 답변 정확도와 응답 구조의 일관성이 전반적으로 개선되었습니다.

다만 복수 조건이 포함된 문의에서는 검색 결과와 최종 답변 간의 연결 기준을 추가로 보완할 필요가 있습니다.

전체 테스트

100건

기존 평가 데이터셋을 동일한 조건으로 적용했습니다.

평가 방식

LLM
Judge

자동 평가 후 개발 담당자가 결과를 검수했습니다.

종합 상태

개선

주요 정책 문의의 응답 안정성이 높아졌습니다.

METHODOLOGY

평가 진행 방식

동일한 테스트 데이터를 기준으로 응답 생성부터 검수까지 진행했습니다.

STEP 01

질문 입력

기존 테스트 데이터셋의 질문을 입력합니다.

STEP 02

응답 생성

구축된 상담 Agent가 최종 답변을 생성합니다.

STEP 03

자동 평가

LLM as a Judge 방식으로 응답을 판정합니다.

STEP 04

담당자 검수

개발 담당자가 자동 판정 결과를 확인합니다.

KEY FINDINGS

주요 평가 결과

평가 과정에서 확인된 개선점과 추가 검토 사항입니다.

개선 확인

답변 구조의 일관성 향상

문의 유형에 따라 필요한 정보를 먼저 확인하고 관련 정책을 안내하는 흐름이 이전보다 안정적으로 적용되었습니다.

개선 확인

정책 근거 활용 강화

정책과 FAQ 검색 결과를 기반으로 답변하는 비율이 높아졌으며 임의 판단 가능성이 감소했습니다.

추가 검토

복합 문의 검색 기준

여러 조건이 동시에 포함된 문의에서는 핵심 검색어가 충분히 생성되지 않는 사례가 확인되었습니다.

추가 검토

예외 상황 응답 기준

일반 정책으로 처리하기 어려운 예외 상황에 대해서는 이관 및 추가 질문 기준을 명확히 할 필요가 있습니다.

DETAIL RESULTS

문제 유형별 개선 방향

구분	확인된 문제	개선 방향	우선순위
검색	복합 조건 검색어 생성 부족	문의 핵심 조건을 분리해 검색어에 반영	높음
응답	일부 답변의 설명이 길어짐	핵심 안내와 추가 설명을 분리	중간
정책	예외 조건의 근거 부족	정책 문서의 예외 처리 기준 보완	높음
운영	이관 조건의 표현 차이	공통 이관 규칙을 모든 응답에 적용	중간

INSIGHTS

주요 발견과 후속 조치

KEY FINDING

답변 생성보다 검색 품질이 중요

최종 답변 모델의 성능뿐 아니라 필요한 정책 정보가 검색 결과에 포함되는지가 전체 정확도에 큰 영향을 주었습니다.

REVIEW POINT

검색 실패 시 임의 답변 위험

적절한 검색 결과가 없는 경우 고정 안내 또는 이관 기준이 안정적으로 적용되어야 합니다.

NEXT ACTION

검색 기준과 예외 처리 규칙을 우선 보완

복합 문의의 검색어 생성 기준을 정교화하고, 검색 결과가 없거나 부적합한 경우의 응답 및 이관 기준을 통합합니다.

결론

현재 상담 Agent는 핵심 정책 문의에 대해 이전보다 안정적인 답변 흐름을 제공하는 것으로 확인되었습니다.

향후에는 복합 문의 검색과 예외 상황 처리 기준을 중심으로 보완해 전체 응답 정확도와 운영 안정성을 높일 필요가 있습니다.