

LLM Agent 플랫폼 기반 금융분야 민원상담 챗봇 설계에 관한 연구

김경엽, 윤수연*

국민대학교, *국민대학교

kkyub13@kookmin.ac.kr, *1104py@kookmin.ac.kr

A Study on Designing a Financial Complaint Consultation Chatbot Using LLM Agent Platforms

Gyeong Yeob Kim, Soo Yeon Yoon*

Kookmin Univ., *Kookmin Univ.

요약

본 연구는 Retell AI 플랫폼을 기반으로 다양한 LLM 모델(GPT, Gemini 등)과 고성능 AI TTS 모델(ElevenLabs 등)을 연동하여 음성기반 AI Agent를 설계하고 구현하는 방법을 제안한다. Retell의 블록 기반 콜 플로우 구성 도구를 통해 비 개발자도 손쉽게 IVR(Interactive Voice Response) 로직을 구성할 수 있으며, 다양한 언어, 억양, 나이의 보이스 모델 선택이 가능하다. 또한 API 및 SDK를 통해 외부 시스템과의 연계성도 뛰어나며, 본 연구에서는 이 플랫폼의 구조, 주요 기능, 적용 예시를 기술하고 실제 서비스 구현 결과를 공유한다.

I. 서론

최근 인건비 및 운영비 상승으로 인해 고객센터 운영에 대한 기업의 부담이 증가하고 있다. 기존 고객센터는 인간 상담사가 내부 데이터베이스를 수동으로 탐색하여 응답하는 방식이 주를 이루었으며, 이로 인한 비효율성과 고객 불편이 지속되어 왔다. 이러한 문제를 해결하기 위한 대안으로 대규모 언어모델을 활용한 AI 상담 에이전트가 주목받고 있다. 특히 최신 LLM은 향상된 응답 정확도, 환각(hallucination) 현상 감소, 반복 작업 처리 효율성, 맥락 기반 의사결정 능력 등에서 기존 시스템을 능가하는 성능을 보인다. 국내외 기업들은 상담 자동화를 위해 AI 에이전트를 도입하고 있으며, AI Contact Center(AICC) 구축 사례가 빠르게 확산되고 있다. 이와 같은 기술적 전환은 고객 응대의 품질 향상과 기업 운영 효율성 제고라는 측면에서 중요한 의미를 가진다. 본 연구는 Retell AI 플랫폼을 기반으로 음성형 LLM 에이전트를 구성하고, 이를 통해 실제 상담 업무 자동화와 AICC의 실용성을 실증적으로 검토하고자 한다.

II. LLM Agent 기술 및 구조 분석

2.1 LLM Agent 기술 개요

2.1.1 LLM Agent의 개념

LLM 에이전트는 대형 언어 모델에 기반한 AI 도구라고 볼 수 있다. 기존 AI 모델이 질문에 맞춰 간단하게 답변을 하는 수준이었다면, LLM 에이전트는 정보를 탐색 후 능동적으로 답변을 구성하여, 타 시스템과 상호적으로 작동할 수 있는 보다 능동적인 기능을 갖춘다.

2.1.2 LLM Agent 플랫폼

대표적인 LLM 기반 음성 에이전트 플랫폼으로는 Moshi, AssemblyAI, Rasa Pro, Voiceflow, Retell AI 등이 있다. 이들 플랫폼은 각각의 목적과

환경에 따라 다양한 기능을 제공하지만, 본 연구에서는 실험 환경에 적합한 조건을 갖춘 Retell AI를 도구로 선정하였다. Retell AI는 한국어 음성 인식(STT)과 음성 합성(TTS) 성능이 안정적이며, 대규모 언어 모델과의 연동 및 사용자 정의 설정이 가능하다. 또한 SDK 기반 구조를 통해 빠르게 에이전트를 구성하고 실험을 수행할 수 있다는 점에서 연구 목적에 부합했다. 다른 플랫폼들도 일정 수준의 음성 처리 및 에이전트 구축 기능을 제공하지만, 본 연구에서 요구하는 언어 지원 및 구성 유연성 측면에서는 Retell AI가 상대적으로 적합하다고 판단하였다.

2.1.3 LLM Agent 주요 구성 요소

대화형 AI 에이전트는 대화 엔진, 음성처리, 프롬프트 설정, 사전지식 연결, 대화 흐름 설계 등의 요소로 구성된다. 첫째, GPT-4o, Claude 등 대규모 언어모델은 사용자 발화를 이해하고 자연스러운 대화를 생성한다. 둘째, 음성 인식(STT) 및 음성 합성(TTS)을 통해 실시간 다국어 음성 상호작용을 지원한다. 셋째, 프롬프트 내 페르소나 설정은 말투, 감정, 응대 방식 등을 정의하여 일관된 에이전트 성격을 형성한다. 넷째, 외부 문서 및 FAQ 기반의 사전지식 연동은 도메인 특화 응답의 정확성과 일관성을 높인다. 다섯째, 시각적 대화 흐름 설계 도구를 활용해 조건 기반 분기와 행동을 구조화할 수 있다.

III. 실험 및 결과

3.1 실험 개요 및 환경 설정

본 연구는 Retell AI SDK를 활용하여 가상의 금융사를 설정하고, 고객 민원 시나리오에 기반한 AI 상담 에이전트 실험을 설계하였다. 실험의 핵심 목적은 사전지식 연동 여부에 따라 GPT 기반 에이전트 응답의 의미적 정확성과 도메인 적합성이 어떻게 달라지는지를 분석하는 데 있다. [표 1]에서 확인할 수 있듯이, 에이전트는 GPT-4.1 nano 모델을 기반으로 동작

하며, 두 가지 조건으로 실험되었다. 조건 A는 단순 프롬프트 기반 에이전트(사전지식 미연동), 조건 B는 사전지식을 system prompt에 연동한 구성이다. 사전지식에는 회사 소개, 금융 상품 안내, 연금 해외 주식·세금 환급 관련 QA 페어, 그리고 업무 가이드라인 문서 등이 포함되었다. 이를 통해 총 25건의 민원 시나리오를 수집 및 정제하여 기준 응답 세트를 구축하였다.

[표 1] 실험 환경 구성

항목	내용
플랫폼	Retell AI (SDK 기반 에이전트 실험)
언어 모델	GPT-4.1 nano
프롬프트 구성 방식	조건 A: 단일 프롬프트 조건 B: 단일 프롬프트 + 사전지식 연동
사전지식	금융 민원 FAQ, 업무 가이드라인 문서
응답 지연 설정	평균 1초 이내 (Responsiveness: 1.0)
Temperature	0.0
평가 지표	BERTScore (Precision, Recall, F1)
평가 기준	실제 상담사 응답 또는 업무 정답 예시와의 의미적 유사도 비교
시나리오 수	25개 금융 민원 유형

3.2 데이터셋 구성 및 평가 방식

QA 데이터셋은 사전지식 내용을 요약가공하여 생성되었으며, 검증 질문은 원본 QA의 의미를 유지하면서 표현만 변형하여 일반화 수준과 응답 일관성을 함께 평가할 수 있도록 설계되었다. 사용자 질문은 QA 배열을 기반으로 STT 음성 입력을 통해 제시되었고, 에이전트 응답도 음성으로 출력되었다. 응답의 정량 평가는 BERTScore(Precision, Recall, F1)를 기준으로 이루어졌으며, 응답 지연은 평균 1초 이내로 제한하여 사용자 경험 품질을 유지하였다.

3.3 실험 결과 분석

[표 2]에서 확인할 수 있듯이, 사전지식이 없는 조건과 있는 조건을 비교한 결과, F1-score 기준으로 평균 0.1833 포인트 상승이 확인되었다. 이는 사전지식 연동이 전반적인 응답 품질 향상에 실질적으로 기여했음을 보여준다. 특히 Recall(재현율)의 향상폭이 Precision보다 근소하게 더 컸는데, 이는 사전지식이 보다 많은 정답 정보를 포괄하도록 도움을 주었기 때문으로 해석된다. 평균적으로 Precision은 약 0.17, Recall은 약 0.19, F1-score는 약 0.18 포인트 향상되었다. 이는 다양한 표현의 질문에 대한 포괄력에서도 긍정적인 효과가 있었음을 시사한다. 또한, 리벨린성 추천, VIP 고객 관련 질의 등 특정 금융 도메인 지식이 요구되는 항목에서 사전 지식의 효과가 더욱 뚜렷하게 나타났다. 이는 단순 언어 모델이 아닌, 신뢰할 수 있는 지식 자원과의 연계를 통해 AI 에이전트가 실질적인 상담 효용성을 확보할 수 있음을 시사한다.

[표 2] 사전 지식 여부에 따른 성능 평가

지표	평균 점수		향상치
	사전지식 없음	사전지식 있음	
Precision	0.7077	0.8810	+0.1733
Recall	0.7134	0.9061	+0.1927
F1-score	0.7103	0.8936	+0.1833

IV. 결론

본 연구는 사전지식의 제공 여부가 AI 에이전트 응답 품질에 미치는 영향을 실험적으로 분석하였다. 실험을 통해 사전지식 연동이 응답의 정확성과 포괄성을 높이는 데 실질적으로 기여함을 확인할 수 있었으며, 이는 향후 고신뢰 응답이 요구되는 분야에서 AI 시스템의 신뢰성 확보를 위한 중요한 단서를 제공한다. 특히 금융 보험과 같이 도메인 지식 기반의 판단이 중요한 환경에서는, 사전지식을 효과적으로 반영한 시스템 설계가 사용자 만족도와 신뢰도 상승에 큰 역할을 할 수 있음을 보여준다. 향후 연구에서는 사전지식의 정교화뿐 아니라, 사용자 맥락(상황, 이력, 감정 등)을 함께 반영하는 다층적 대화 전략이 요구된다. 이를 통해 정확성과 자연스러움을 동시에 갖춘 상담 시스템 구축이 가능할 것이며, 이 연구는 이러한 기반을 다지는 데 기여할 수 있을 것이다.

참 고 문 헌

[1] Naveen Krishnan, "AI Agents: Evolution, Architecture, and Real-World Applications," arXiv preprint arXiv:2503.12687, 2025.
[2] Nature Digital Medicine, "Systematic review and meta-analysis of AI-based conversational agents in mental health," 2023.