

Customer Emotion-Aware Conversational Language Model using Generative AI and Retrieval-Augmented Generation (RAG)

Sung-Hyun Jee*

*Professor, Dept. of Software&Security Convergence, Baewha Women's University, Seoul, Korea

[Abstract]

Due to the development of artificial intelligence technology, conversational AI systems that understand and respond to human emotions are becoming increasingly important. While current AI assistants are being widely adopted across industries, technology that reflects human emotions and behaviors in conversations is still limited. This paper proposes a real-time conversational language model that considers the emotional state of the person conversing by utilizing RAG (Retrieval-Augmented Generation) technology, which is used to provide customized prompts. This research aims to infer human emotions based on analysis of text and tone that occur during conversations with people, and enable emotion-based customer response. The significance of this research lies in studying and applying AI technology that understands human emotions and behaviors.

▶ **Key words:** Emotion Recognition, Retrieval-Augmented Generation, Conversational AI, Multimodal AI, Human-Computer Interaction, Artificial Intelligence

[요 약]

최근 인공지능 기술과 멀티모달 데이터 처리 기술의 급속한 발전으로, 텍스트, 음성, 영상 등 다양한 입력 데이터를 활용하여 사용자의 감정 상태와 상황에 맞는 응답을 제공하는 대화형 AI 시스템의 중요성이 크게 부각되고 있다. 본 연구는 검색 증강 생성 (RAG) 기술에 감정 인식 기능을 통합한 새로운 챗봇 시스템을 제안하며, 고객 상담 등 실제 서비스 환경에서 단순 텍스트 응답을 사용자의 감정 상태(예: 좌절, 긍정, 혼란)에 따라 검색 유사도 임계값을 동적으로 조정하여, 부정적 감정 시에는 세분화된 맥락 정보를 포함하고 긍정적 감정 시에는 간결한 핵심 정보를 제공한다. 고객 상담 시나리오를 대상으로 한 실험 결과, 감정 인식 정보를 반영한 시스템은 기존의 RAG 기반 시스템 대비 응답의 만족도가 인간 평가자 기준 28.11%, LLM 기준 21.25% 향상된 사용자 만족도를 달성하는 등, 인간의 감정과 행동을 보다 정밀하게 이해하여 공감 능력 있는 응대를 구현할 수 있음을 입증하였다. 이러한 결과를 바탕으로 차세대 AI 기반 고객 서비스 플랫폼 개발 및 인간-AI 상호작용의 새로운 발전 방향을 제시하고자 한다.

▶ **주제어:** 감정 인식, 검색 증강 생성, 대화형 인공지능, 멀티모달 인공지능, 인간-컴퓨터 상호작용, 인공지능

• First Author: Sung-Hyun Jee, Corresponding Author: Sung-Hyun Jee
*Sung-Hyun Jee (sunny6205@baewha.ac.kr), Dept. of Software&Security Convergence, Baewha Women's University
• Received: 2025. 06. 20, Revised: 2025. 07. 20, Accepted: 2025. 08. 22.

I. Introduction

최근 AI 기반 챗봇 시스템은 고객 상담, 금융, 의료 등 다양한 분야에서 확산되고 있으나, 텍스트 또는 음성 단일 모달리티에 의존하는 기존 시스템은 사용자의 감정 상태를 반영하지 못해 심각한 한계를 드러내고 있다. 2024년 Gartner 보고서 [1]에 따르면 챗봇을 통한 고객 상담의 70%가 초기 문제 해결에 실패하며, 이 중 45%는 감정적 공감 부족으로 인한 사용자 이탈로 연결된다. 2023년 Forrester Research [2]는 감정을 고려하지 않은 챗봇 응답이 52%의 고객 불만을 유발하고 평균 상담 시간을 40% 이상 증가시킨다고 지적하였다. 금융 및 e-커머스 분야에서는 이러한 문제가 더욱 두드러지는데, 2024년 한국 소비자원 [3]의 분석에 의하면 감정 기반 응답이 결여된 시스템은 68%의 오답률을 기록하며 전통적 콜센터 대비 3배 높은 실패율을 보인다. 이러한 배경에서 본 연구는 텍스트, 음성, 영상 데이터를 통합한 멀티모달 감정 분석과 RAG(Retrieval-Augmented Generation) 기술을 결합해 사용자 감정을 실시간으로 반영하는 시스템을 제안한다. 제안 시스템은 감정 상태에 따른 동적 지식 검색과 GPT-4 기반 응답 최적화를 통해 혼란, 지루함, 불만 등 다양한 감정적 맥락을 이해하고 적응형 상호작용을 구현함으로써 사용자 경험 혁신을 목표로 한다.

II. Related works

1. AI Assistant bot survey

최근 AI 기반 챗봇 시스템은 고객 상담, 금융, 의료 등 다양한 분야에서 빠르게 확산되고 있다. 금융 분야에서는 은행 및 보험사가 챗봇을 활용해 계좌 개설, 대출 상담, 사기 탐지 등의 서비스를 제공하며, 2023년 JP모간 체이스 보고서 [4]에 따르면 AI 챗봇 도입으로 운영 비용이 30% 감소한 것으로 나타났다. 의료 분야에서는 환자 문의 응대, 진료 예약, 증상 초기 평가에 챗봇이 활용되며, 메이요 클리닉의 연구 [5]는 챗봇을 통한 환자 관리 효율성이 45% 향상되었다고 밝혔다. 고객 상담 분야에서는 이커머스 기업들이 24/7 챗봇 서비스를 도입해 주문 추적, 반품 처리, 맞춤형 추천을 제공함으로써 고객 만족도를 50% 이상 개선했다는 데이터가 알리바바 그룹의 보고서 [6]에서 확인된다. 그러나 감정 기반 맥락 이해 부재, 다중 모달리티 데이터 통합의 기술적 한계 등으로 인해 완전한 인간적 상호작용 구현에는 아직 과제가 남아있다.

2. Problems with AI Assistant bot

챗봇이 감정적인 맥락을 이해하지 못하고 적응형 상호작용을 구현하지 않을 경우, 다음과 같은 심각한 한계점이 발생한다. 첫째, 사용자의 감정 상태를 고려하지 않은 응답은 맥락 이해 부족으로 이어져, 사용자가 혼란스러워하거나 불만을 느낄 가능성이 크다. 예를 들어, 사용자가 설명을 이해하지 못하고 혼란스러워할 때, 챗봇이 추가적인 설명을 제공하지 않으면 문제 해결에 실패할 뿐만 아니라 사용자의 불만을 가중시킨다. 둘째, 감정적 공감이 결여된 응답은 사용자 이탈률을 증가시킨다. 감정을 반영하지 않은 기계적인 응답은 사용자에게 무관심한 느낌을 주어, 상담을 중단하거나 다른 채널로 전환하는 결과를 초래한다. 셋째, 감정 기반 적응형 상호작용이 없을 경우, 상담 효율성이 저하된다. 사용자의 감정 상태에 맞춰 응답을 조절하지 못하면 상담 시간이 불필요하게 길어지고, 반복적인 질문과 설명이 필요해져 전체적인 서비스 품질이 떨어진다. 마지막으로, 감정을 고려하지 않은 챗봇은 고객 신뢰도 하락을 유발한다. 사용자는 감정적 공감이 없는 응답에 대해 불만을 느끼고, 서비스 제공자에 대한 신뢰를 잃을 가능성이 높아진다. 이러한 한계점들은 챗봇이 단순 정보 전달 도구로 전락하게 하며, 인간적 상호작용을 모방하는 AI의 본질적 목표를 달성하지 못하게 만든다. 이러한 메커니즘을 통해 RAG 모듈은 정적 답변 생성의 한계를 극복한다. 감정 상태에 따른 검색 범위의 유동적 조정은 단순한 정보 전달을 넘어 사용자의 심리적 맥락에 적응하는 인간 중심(Human-Centric) 상호작용을 가능하게 합니다. 결과적으로 시스템은 사용자의 정서적 필요와 지적 호기심을 동시에 충족시키며, 교육, 고객 지원, 개인 비서 등 다양한 분야에서 확장 가능한 솔루션으로 작동합니다.

3. Multimodal Emotion Recognition

멀티모달 감정 인식(Multimodal Emotion Recognition, MER)은 텍스트, 음성, 영상 등 다양한 입력 채널을 융합하여 사용자의 감정 상태를 정밀하게 파악하고자 하는 기술로, MERLOT Reserve [7], MERT [8], EmoLLM [9]과 같은 최신 연구들이 표현 통합 기법을 고도화하고 있다. 그러나 실제 사용자와의 상호작용에서 높은 만족도를 제공하는 실시간 감정 적응형 멀티모달 챗봇 시스템은 아직 실현되지 않고 있으며, 이는 다음과 같은 구조적 한계에서 기인한다.

기존 멀티모달 챗봇 시스템은 텍스트, 음성, 시각 정보 간의 cross-modal fusion 문제로 인해 정서 추론의 일관성과 신뢰도가 저하되는 한계를 가진다. 예컨대, 텍스트는

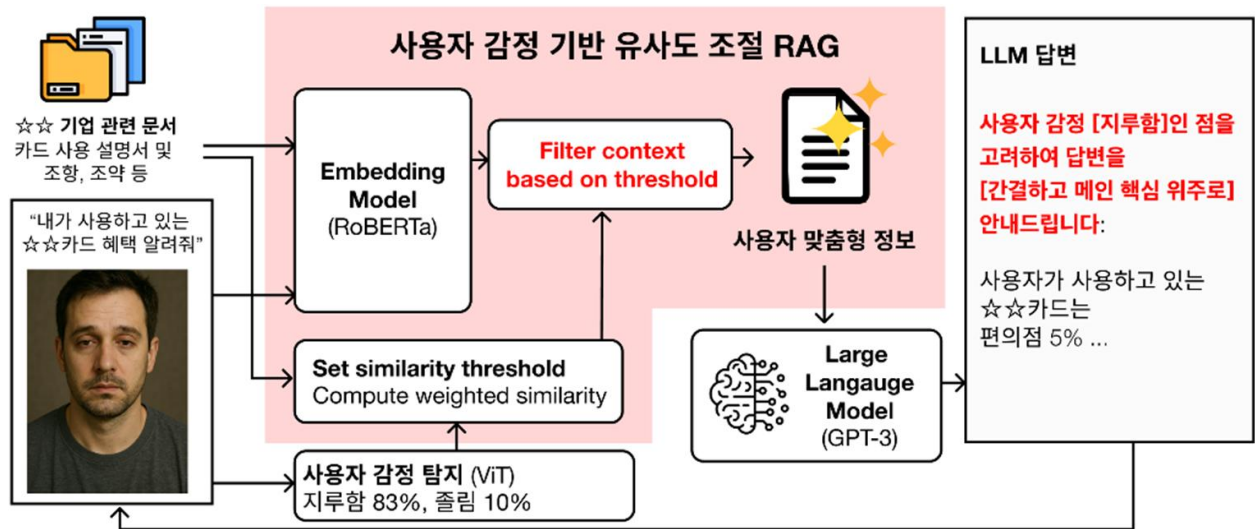


Fig. 1. Overall system of our framework

긍정적이지만 음성 역양은 부정적인 경우처럼 모달리터 간 감정 불일치가 발생할 때, 기존 시스템은 이를 종합적으로 해석하지 못하거나 하나의 모달에 과도하게 의존해 오인식할 위험이 있다. 본 시스템은 이러한 융합 오류를 다루기 위해, 감정 기반 유사도 조절 RAG 모듈을 통해 검색 임계값을 실시간으로 조정하는 방식을 도입하였다. 이는 cross-modal fusion 과정에서 발생하는 감정 인식 불확실성을 검색 단계에서 보정하는 전략으로 작동한다. 또한, 고해상도 멀티모달 입력의 실시간 병렬 처리로 인한 latency 문제는 감정과 응답 간 비동기적 경험을 초래하여 공감 상호작용을 저해한다. 이에 따라 본 시스템은 모듈 간 병렬성보다 순차적 처리 최적화에 초점을 둔 경량 파이프라인 구조를 채택하여 한계를 보완하고자 한다.

4. 검색 증강 생성 (RAG) 기술 동향과 과제

검색 증강 생성은 외부 지식 소스와 생성형 언어 모델을 결합해 정확성과 맥락 민감도를 동시에 달성하고자 하는 프레임워크이다. 기존의 단순 RAG 구조는 고정된 검색-생성 파이프라인을 기반으로 작동했으나, 최근에는 목적 기반 검색 전략과 역할 기반 모듈화를 통해 Advanced RAG 혹은 Modular RAG로 진화하고 있다. 예컨대, RePlug [10], HyDE [11], Fusion-in-Decoder [12] 등은 검색과 생성을 동적으로 통합하거나, 문서 내 reasoning을 반영한 검색어 생성 등으로 정답률과 일관성을 향상시키고 있다. 나아가, LangChain [13], LlamaIndex [14] 등 프레임워크는 도메인별 retriever, reranker, memory 모듈을 독립적으로 구성하는 Modular RAG 구조를 구현해, 상황별 사용자 요구에 맞춘 조합형 시스템 구축을 가능케 한다.

특히 최근에는 단순한 질의응답을 넘어 사용자의 목표를 파악하고 여러 스텝의 추론과 검색을 통해 능동적으로 행동하는 Agentic RAG 패러다임이 부상하고 있다. 대표적으로 LlamaIndex의 AgentRAG [15]와 OpenAI의 Function-calling [16] 기반 루프 구조가 이에 해당하며, 멀티턴 상호작용에서 추론-검색-행동을 반복 수행함으로써 인간적인 문제 해결 흐름을 일부 구현하고 있다. 그러나 현재 대부분의 연구는 정보 정확성과 도메인 적용에 집중되어 있으며, 사용자 감정 상태나 상호작용 흐름의 맥락 변화를 실시간으로 반영하는 구조는 부족한 실정이다.

이에 본 연구는 감정 기반 유사도 임계값 조정 메커니즘을 도입하여, 사용자의 정서 상태에 따라 검색 범위, 문서 선택, 응답 톤 등을 동적으로 조정하는 감정 적응형 RAG 구조를 제안한다. 이는 단순 정보 검색에서 나아가, 사용자의 심리적 요구와 정서적 컨텍스트를 함께 반영하는 인간 중심 상호작용(Human-Centric Interaction)을 가능케 하며, 교육, 상담, 퍼스널 어시스턴트 등 다양한 분야에서 공감 기반 고품질 AI 시스템 구축의 실마리를 제공한다.

III. System architecture

본 연구에서 제안하는 시스템은 사용자와의 상호작용 전 과정을 실시간으로 처리하는 통합 대화 시스템으로, 입력 데이터 수집부터 감정 분석, 응답 생성, 최종 출력에 이르는 전 단계가 유기적으로 연결되었음을 그림1에서 확인할 수 있다. 전체 시스템은 입력 처리 모듈, 감정 분석 모듈, 감정 기반 유사도 조절 RAG 모듈, 응답 생성 모듈로 구성되며, 각 모듈은 다음과 같은 기능을 수행한다.

1. Input Processing Module

입력 처리 모듈은 사용자와의 상호작용에서 발생하는 다중 모달 데이터 (즉, 텍스트, 이미지)를 수집 및 전처리한다. 사용자가 키보드나 채팅 인터페이스를 통해 입력한 텍스트 데이터는 별도의 변환 없이 그대로 수집되며, 음성 데이터의 경우 Google Speech-to-Text API를 통해 실시간으로 텍스트로 변환된다. 웹캠이나 스마트폰 카메라를 통해 수집된 얼굴 이미지는 다음 감정 분석 모듈에서 자세한 감정 분석이 이루어진다.

2. Emotion Analysis Module

감정 분석 모듈은 이미지를 기반으로 얼굴 표정 데이터를 얻게 된다. OpenCV를 사용해 입력된 프레임에서 얼굴을 감지한 후, Kaggle 및 Hugging Face에 공개된 facial_emotions_image_detection 데이터셋을 기반으로 Vision Transformer(ViT) [17]구조 기반으로 구현되어 있으며, Ekman의 7가지 기본 감정(happy, sad, angry, surprise, fear, disgust, neutral)을 91% 정확도로 분류할 수 있다. 얼굴 입력은 RGB 또는 grayscale 이미지로 수집되며, 전처리 과정에서 정규화(normalization), 크기 조정(resize), 차원 확장(expand_dims) 등이 적용된다. 최종적으로 softmax 출력을 기반으로 감정 레이블을 결정하고, 이 정보는 다른 모달리티의 감정 추론 결과와 함께 융합되어 최종 감정 판단에 사용된다. 얼굴 표정 감정 분석 결과 예시는 그림 2와 같다.

3. Emotion-based Similarity Adjustment RAG

감정 기반 유사도 조정 RAG 모듈은 사용자의 질의 내용과 실시간 감정 상태를 통합적으로 분석하여 개인화된 정보 검색과 응답 생성을 지원하는 차세대 대화 시스템의 핵심 구성요소이다. 기존 시스템이 질의의 의미적 맥락만 처리하는 데 그쳤다면, RAG 모듈은 감정 분석 모듈과 연동하여 사용자의 정서적 상태를 반영함으로써 개인화된 상호작용을 구현한다.

문서 검색 단계에서는 sentence-transformers 라이브러리에서 제공하는 사전학습된 RoBERTa [19]에 한국어 데이터셋인 KLUE [20]를 학습한 임베딩 모델이 활용되었다. 이 모델은 사용자 질의와 문서의 의미 정보를 벡터화하며, 벡터 간 유사도 평가는 코사인 유사도로 수행된다. 검색 결과로 도출된 후보 문서들은 감정 결과인 e 기반 임계값 함수 $T(e)$ 에 따라 필터링된다.



Fig. 2. Emotion prediction results

임계값 함수는 각 감정 e_i 에 대해 모델이 예측한 확률(prediction)에 감정별 중요도를 나타내는 가중치를 곱하여 개별 감정 점수를 구한다. 여기서 가중치(score)는 감정에 대해 사전 할당된 값으로 변경 가능하다: happy: 0.8, sad:0.4, angry:0.3, surprise:0.5, fear:0.5, disgust: 0.3, neutral: 0.6. 예컨대, 예측확률이 0.7이고 가중치가 0.8인 “happy”의 경우 개별 감정 점수는 0.56이 되며, 이와 같은 방식으로 모든 감정에 대해 최종 $T(e)$ 값을 구하게 된다.

$$T(e) = \sum_{e_i} (\text{prediction}(e_i)) \times \text{score}(e_i)$$

이와 같은 규칙 기반 임계값 조정은 사용자의 감정 상태가 부정적일수록 더 많은 문서 후보를 수용하도록 하고, 반대로 긍정적 상태에서는 필터링 강도를 높여 핵심 정보 위주의 응답을 유도하여 감정에 따라 유연한 임계값 산출이 가능하다. 본 모듈은 또한 향후 사용자 피드백 데이터에 기반하여 이 임계값 함수의 파라미터를 학습적으로 최적화하는 방향으로 확장이 가능하다.

4. Generation Module

응답 생성 모듈은 검색된 문서와 감정 분석 모듈에서 전달된 감정 정보를 바탕으로, OpenAI의 text-davinci-003 모델(GPT-3) [21]을 활용하여 감정 상태에 최적화된 자연어 응답을 생성한다. 본 모듈은 프롬프트 엔지니어링(prompt engineering) 기법을 적용하여, 감정 상태에 따른 응답의 구조, 어조, 설명 밀도를 유연하게 조절할 수 있도록 설계되었다.

예를 들어, 사용자가 불안하거나 혼란스러운 감정을 보이는 경우, 시스템은 보다 단계적이고 명확한 설명을 포함한 응답을 생성하며, 반대로 만족스럽거나 중립적인 감정 상태에서는 불필요한 정보를 줄이고 간결한 요약형 응답을 제공한다. 예컨대, 사용자가 “fear” 상태일 경우, 프롬프트는 단계적이고 예제 중심의 설명을 유도하도록 구성되며, “happy” 상태일 경우에는 간결하고 요약적인 응답 생성을 유도한다.

생성된 응답은 텍스트 형식으로 UI에 출력될 뿐 아니라, Google Text-to-Speech API를 활용하여 음성화되어 멀티모달 사용자 피드백을 제공한다. 감정 레이블에 기반한 억양 및 말 속도 조정도 가능하며, 대화 전반에서 정서적 공감 경험을 강화한다. 본 모듈은 피드백 루프를 통해 이후 응답 생성의 스타일을 점진적으로 조정한다. 이 피드백 기반 적응 구조는 사용자의 장기적인 상호작용 패턴에 기반한 응답 개인화를 가능하게 하며, 반복 상호작용에서도 지속적으로 응답 품질을 개선하는 데 기여한다.

본 모듈은 Streamlit 기반 웹 상에서 구현되었으며, 사용자로부터의 텍스트 및 음성 입력을 처리하고, 실시간 감정 분석 결과 및 검색 응답 결과를 직관적으로 시각화한다. 그림 3에서 구현 예시를 확인할 수 있다.

IV. Experiments

본 연구는 감정 기반 유사도 조절이 적용된 검색 증강 생성시스템의 효과를 다각적으로 검증하기 위해, 여러가지 시나리오를 활용하여 완성되었다. 평가용 데이터셋은 감정 맥락(좌절, 공감, 기쁨, 혼란, 호기심 등)과 질의 복잡성이 상호 교차되도록 설계한 다섯 개 시나리오(각 시나리오당 세 문항, 총 15문항)로 구성하였다.

감정 기반 RAG의 유사도 임계값 조절 효과를 정량화하기 위해 검색된 문서 수와 응답 생성 시간을 기록하여 시스템 효율성을 평가하였다. 이를 통해 본 연구는 (1) RAG가 LLM 단독 사용 대비 정보 신뢰도와 맥락 이해도를 개선하는지, (2) 감정 기반 유사도 조절이 기존 RAG 대비 사용자 감정에 적응적인 응답 생성을 가능하게 하는지를 입증하고자 한다. 각 사용 사례에 대해 세 가지 접근법을 사용하여 응답을 생성하였다 (그림 3). (1) 기본 LLM(GPT-3.5 Turbo)을 활용하여 RAG 없이 직접 질의한 경우 (i.e., No RAG), (2) 기존 RAG를 적용하되 감정 기반 유사도 조절 없이 고정 임계값을 사용한 경우 (i.e., plain

OO은행 고객센터 보이는 ARS

고객님, 무엇을 도와드릴까요?

언제나 고객님의 문익에 최선을 다하겠습니다.

유저 정보

현재 고객님의 상태: 40대 남성, 지침 56%, 화남 23%, 무표정 21%
 고객님 정보:
 - 상품명: KB Star 정기예금
 - 계좌번호: 123-456-7890
 - 가입일: 2023.01.15
 - 만기일: 2024.01.15 (지난 7월)
 - 현재 잔액: 10,000,000원 (원금+이자)



업로드된 유저 사일 화면

유저 요청 사항

정기 예금이 이번 달에 만기가 되었는데, 만기일을 놓쳐서 이자가 제대로 지급될지 걱정이 됩니다. 어떻게 해야 할까요?

고객님, 답변드립니다.

고객님의 정기 예금이 만기일을 지났을 경우, 다음과 같은 상황을 고려해볼 수 있습니다:

1. 자동 재예치 여부 확인

자동 재예치 설정 여부: 고객님의 예금 가입 시 자동 재예치(원금 또는 원금+이자)를 선택하셨다면, 만기일이 지나도 자동으로 해당기간 만기 이자가 계속 누적될 수 있습니다. 만기일이 지난 시점은 추가적인 조치 없이도 이자가 자동으로 지급되는 방식입니다.

Fig. 3. Streamlit implementation results

RAG), (3) 감정 기반 RAG를 적용하여 감정 분류 모델이 유사도 임계값을 동적으로 조정한 경우 (i.e., User emotion-based RAG, 감정 기반 RAG)이다. 시나리오 예시 및 모델 별 응답 예시는 표 1에서 확인할 수 있다.

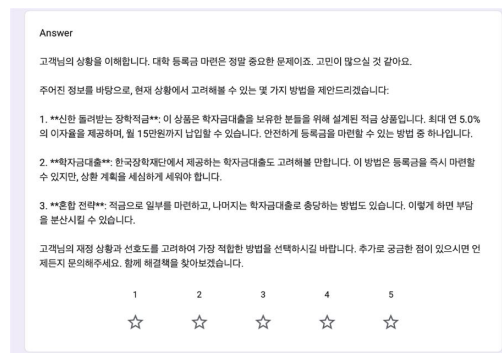


Fig. 4. Response evaluation form

평가는 인간 평가와 다중 LLM 기반 자동 평가를 결합한 실험 설계를 채택하였다. 인간 평가는 성별·연령 균형을 맞춘 30명의 참가자(남 15, 여 15; 20대 10명, 30대 10명, 40대 이상 10명)를 대상으로 실시하였다. 참가자는 제공된 시나리오와 응답을 구글 설문지를 통해 제공받아 1부터 5까지 만족도를 평가하는 방식으로 진행되었다. 평가자에게 방법론 정보가 공개되지 않도록 블라인드 처리되었고, 방법론 순서는 랜덤 처리되었다. 그림 4에서 설문지 예시를 확인할 수 있다.

Table 1. Model inference results under scenario (Figure 3) and the LLM evaluation score.

	Model results	Zero-shot LLM Evaluation results
No RAG	고객이 정기 예금 만기일을 놓쳤다면, 우선 해당 은행의 고객센터(1599-XXXX)나 홈페이지를 통해 문의하여 자동 연장 여부와 이자 지급 조건을 확인하는 것이 좋습니다. 만약 해지가 필요하다면, 은행 지점 방문이 필요할 수 있으므로, 이를 고려하여 조치를 취하는 것이 바람직합니다.	- 고객 감정 고려: 고객의 걱정이거나 번거로움에 대한 공감 전혀 없음. - 명확성: 고객센터나 홈페이지 확인 제안하나 구체적 설명 부족으로 모호. - 실용성: 이동/온라인 접근 어려움을 고려하지 않은 추상적 조언. 별점: 별점: ★★☆☆☆ (2/5)
Plain RAG	고객님의 정기 예금이 만기일을 지났을 경우, 다음과 같은 상황을 고려해볼 수 있습니다: 1. 자동 재예치 여부 확인 - 자동 재예치 설정 여부: 고객님의 예금 가입 시 자동 재예치(원금 또는 원금+이자)를 선택하셨다면, 만기일이 지나도 자동으로 재예치가 되어 이자가 계속 누적될 수 있습니다. 추가적인 조치 없이도 예금이 유지되며, 이자는 재예치된 금액에 따라 계속 계산됩니다. - 자동 해지 설정 여부**: (이하 생략) 2. **만기 후 이자 지급** - **만기 후 이자율****: (이하 생략) 3. **은행 방문 필요성** - **해지 또는 재예치 변경****: (이하 생략) - **상담****: (이하 생략) 결론 - 자동 재예치 여부 확인: 먼저, 예금 가입 시 자동 재예치를 설정하셨는지 확인해보세요. 설정하셨다면 추가 조치 없이 이자가 계속 누적됩니다. - 만기 후 이자율 확인: (이하 생략) - 은행 방문: (이하 생략) 추가적인 문의가 필요하시면, 00고객센터로 연락하시거나 은행 지점을 방문하시면 도움을 받으실 수 있습니다.	- 고객 감정 고려: 감정 언급 약간 있으나 깊은 공감 부족. - 명확성: 자동 재예치, 이자율, 방문 필요성 상세히 설명해 명확. - 실용성: 정보는 풍부하나 이동 불편함 대안 부족. 별점: ★★★★★☆ (4/5)
Ours	고객님의 걱정을 이해합니다. 만기일을 놓치셨다면, 정기 예금은 자동으로 연장될 가능성이 높습니다. 하지만 정확한 상황을 확인하고 이자를 받으시려면 몇 가지 조치를 취하셔야 합니다. 1. 자동 재예치 확인: 0000 정기예금의 경우, 만기일 이후 자동으로 재예치될 수 있습니다. 만기일 이후에도 이자가 계속 발생할 수 있으니, 먼저 인터넷뱅킹이나 KB스타뱅킹 앱을 통해 계좌 상태를 확인해보세요. 2. 이자 지급 확인: (이하 생략) 3. 은행 방문: (이하 생략) 4. 온라인 해지: (이하 생략) 고객님의 상황에 맞는 최선의 방법을 찾으시길 바랍니다. 추가로 궁금한 점이 있으시면 언제든지 문의해주세요. 함께 해결책을 찾아보겠습니다.	- 고객 감정 고려: 걱정과 번거로움을 공감하며 위로 잘 반영. - 명확성: 온라인/전화 확인, 방문 최소화 등 단계적 조언으로 명확. - 실용성: 이동성 고려해 전화 상담 우선 제안, 실질적 해결책 제공. 별점: ★★★★★ (5/5)

평가자가 산출한 만족도를 평균화한 표1에 따르면, 감정 인식 기반 RAG 시스템 (4.594)이 일반 RAG 시스템 (3.586)과 RAG 미적용 시스템 (2.22)에 비해 전반적으로 높은 선호도를 보였다. 특히, 감정 인식 RAG는 모든 평가 항목에서 일관되게 4.33~4.7의 높은 점수를 기록하며 우수한 사용자 경험을 제공함을 확인할 수 있었다. 일반 RAG 대비 평균 28.11% (3.586 → 4.594), RAG 미적용 대비 106.94% (2.22 → 4.594)의 선호도 향상을 보였는데, 이는 감정 기반 동적 검색 조정이 사용자의 정서적 맥락을 효과적으로 반영했기 때문으로 해석된다. 5번 평가 항목에서 가장 큰 격차(1.33 → 4.7, 253.38% 향상)가 나타난 점은 부정적 감정 시나리오에서 특히 효과적임을 시사한다. 이러한 결과는 감정 인식 기술이 RAG 시스템의 사용자 적응성을 크게 향상시킬 수 있음을 입증한다.

LLM 기반 평가는 ChatGPT(GPT-4) [22], Claude 3.5 Haiku [23], DeepSeek-R1 [24], Grok-3 [25] 등 네 개의 상용 LLM을 활용한 제로샷 방식으로 수행하였다. 각 모델에는 논리성과 실용성을 평가하도록 통일된 시스템-사용자 프롬프트를 제공했고, 모델이 산출한 순위를 5점 선호도 점수로 환산하여 평균화했다. LLM 평가 예시는 표 1에서 확인 가능하며, 해당 예시는 Grok-3 모델 평가 결과이다.

Table 2. Human Evaluator Preference Results

	No RAG	Plain RAG	Ours
1	2.67	3.4	4.67
2	2.2	4.2	4.33
3	2.1	3.33	4.6
4	2.8	3.4	4.67
5	1.33	3.6	4.7
Avg	2.22	3.586	4.594

표 3의 실험 결과에 따르면, 감정 기반 RAG 시스템(평균 4.85)이 일반 RAG 시스템(평균 4.0)과 RAG 미적용 시스템(평균 2.475)에 비해 현저히 높은 선호도를 보였다. 특히 감정 기반 RAG는 모든 평가 항목에서 4.625~5.0의 거의 만점에 가까운 점수를 기록하며 우수한 성능을 입증했다. 일반 RAG 대비 평균 21.25% (4.0 → 4.85), RAG 미적용 대비 95.96% (2.475 → 4.85)의 선호도 향상을 보였으며, 2번과 5번 평가 항목에서는 최대점수인 5점을 달성했다. 이는 감정 인식 기술이 LLM 평가자 기준에서도 시스템의 응답 품질을 크게 향상시킬 수 있음을 보여주며, 특히 복잡한 감정 상황에서 더 효과적으로 작동함을 시사한다. 이러한 결과는 감정 기반 접근 방식이 객관적 평가 기준 하에서도 그 유효성이 검증되었음을 의미한다.

Table 3. LLM Evaluator Preference Results

	No RAG	Plain RAG	Ours
1	3.5	3.875	4.75
2	2.875	4.875	5
3	2	4.25	4.625
4	2	3.25	4.875
5	2	3.75	5
Avg.	2.475	4	4.85

V. Conclusions

본 연구는 감정 인식 기반 동적 RAG 시스템을 제안하여 기존 챗봇 시스템의 한계를 극복하고자 하였다. 실험 결과, 제안 시스템은 인간 평가자 기준 28.11%, LLM 평가자 기준 21.25%의 향상된 만족도를 보였으며, 특히 부정적 감정 상황에서 더욱 뚜렷한 효과를 나타냈다. 이는 감정 상태에 따른 유사도 임계값의 동적 조정이 사용자의 정서적 맥락을 효과적으로 반영했기 때문으로 해석된다. 또한, 시스템은 복잡한 감정 상황에서도 일관되게 높은 성능을 유지하며, 고객 상담 시나리오에서 상담 효율성과 사용자 신뢰도를 크게 향상시켰다. 이러한 결과는 감정 인식 기술이 단순한 정보 전달을 넘어 인간 중심의 상호작용을 구현하는 데 핵심적인 역할을 할 수 있음을 시사한다. 향후 연구에서는 더 다양한 감정 상태와 복잡한 대화 시나리오에 대한 시스템의 확장성을 검증할 계획이다. 본 연구가 제시한 접근법은 AI 기반 고객 서비스 플랫폼의 발전에 기여할 뿐만 아니라, 인간-AI 상호작용 분야의 새로운 방향을 제시하였다는 점에서 의의가 있다.

ACKNOWLEDGEMENT

This research was supported by the 2025 Academic Research Fund of Baewha Women's University.

REFERENCES

- [1] Gartner, "Customer Service Technology Trends 2024," 2024.
- [2] Forrester Research, "Overlook The Impact Of Emotion On Customer Experience At Your Own Peril," 2023.
- [3] 한국소비자원, "2024 AI 챗봇 서비스 품질 평가 보고서," 2024.
- [4] JP Morgan Chase, "AI Chatbots in Banking: Cost Efficiency and Customer Impact," 2023.
- [5] Mayo Clinic, "AI in healthcare: The future of patient care and health management," 2024.
- [6] Cheng et al. "Exploring consumers' response to text-based chatbots in e-commerce: the moderating role of task complexity and chatbot disclosure," 2023.
- [7] Zellers, R., Hessel, J., Cao, J., Park, J. S., Yu, Y., Farhadi, A., & Choi, Y., "MERLOT Reserve: Neural Script Knowledge through Vision and Language and Sound," 2022 arXiv
- [8] Bänziger, T., Grandjean, D., & Scherer, K. R., "Emotion recognition from expressions in face, voice, and body: the Multimodal Emotion Recognition Test (MERT)," 2009 PubMed
- [9] Yang, Q., Ye, M., & Du, B., "EmoLLM: Multimodal Emotional Understanding Meets Large Language Models," 2024
- [10] Shi, W., Wang, J., Li, X., & Goldberg, Y., "REPLUG: Retrieval-Augmented Black-Box Language Models," 2024.
- [11] Gao, L., Ma, X., Lin, J., & Callan, J., "Precise Zero-Shot Dense Retrieval without Relevance Labels (HyDE)," 2023.
- [12] Izcard, G., & Grave, E., "Leveraging Passage Retrieval with Generative Models for Open-Domain Question Answering (Fusion-in-Decoder)," 2021.
- [13] Chase, H., "LangChain: A framework for developing language model applications," 2022.
- [14] Yang, S., "LlamaIndex (formerly GPT-Index): The ultimate LLM framework for indexing and retrieval," 2023.
- [15] Singh, A., "Agentic Retrieval-Augmented Generation: A Survey on Agentic RAG," 2025.
- [16] OpenAI, "Function calling," 2025.
- [17] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., & Houlsby, N., "An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale," 2020.
- [18] Ekman, P., "An Argument for Basic Emotions," 1992.

- [19] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," 2019.
- [20] Park, J., Kang, J., Kim, S., & Cho, S., "KLUE: Korean Language Understanding Evaluation Benchmark," 2021.
- [21] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., ... & Amodei, D., "Language Models Are Few-Shot Learners," 2020.
- [22] OpenAI, "GPT-4 Technical Report," 2023.
- [23] Anthropic, "Claude 3.5 Haiku," 2024.
- [24] DeepSeek Inc., "DeepSeek-R1," 2025.
- [25] XAI, "Grok-3," 2025.

Authors



Sung-Hyun Jee received the B.S., M.S. and Ph.D. degrees in Computer Science from Chungbuk National University, Korea, in 1993, 1995 and 2000, respectively. She also completed post-doctoral course at University

of Missouri in U.S.A. for 2001~2003. She is currently a professor in the Department of Software&Security Convergence, Baewha Women's University, Seoul, Korea. She is interested in IoT Integrational strategies.