



Project Proposal

Title

Developing an Intelligent Customer Support Chatbot Using Large Language Models with Intent and Emotion Analysis and Dynamic RAG-Based Knowledge Retrieval

Prepared By

Name: **Shahad AlHarbi**

ID No: **252000854**

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

At

College of Informatics

Midocean University

UAE

Supervisor Name: **Dr. Khaled Eskaf**

College of Informatics, Midocean University, Comoros

1. Project Overview

The requirement for customer support is growing rapidly across many organisations, primarily due to the increasing number of customers using digital channels for information. As a result, customer support has transitioned to a digital platform that uses chatbots as a method of delivering immediate and scalable support. Unfortunately, most current chatbots do not deliver effective support as they cannot offer accurate support in terms of understanding user intent, identifying the emotional context of a user's inquiry, or providing the user with reliable solutions.

With the advancement of Large Language Models (LLMs), the fluency of a chatbot's conversation has increased; however, these chatbots are still susceptible to hallucinations, and are unable to recognise users' emotional states. Most Retrieval-Augmented Generation (RAG) models use a static source of knowledge when providing support and therefore fail to modify the information retrieved to match the user's intent or emotional state; as a result, RAGs have limited applicability in actual customer support scenarios.

This proposal outlines the development of an intelligent customer support chatbot that utilises LLMs, intent recognition, emotional context, and dynamic RAGs. The goal of this intelligent customer support chatbot is to provide customers with accurate, contextual, and emotional responses to help enhance the quality of customer support provided to customers.

2. Project Goal

The objective of the project is to develop a customer service bot capable of understanding user intent, inferring emotional states and responding in real time with a knowledge base that is retrieved dynamically. The project's objectives are as follows:

1. To develop a conversational interface powered by artificial intelligence, allowing for an intuitive and free-form communication style.
2. To design an Intent recognition module that decodes what the user is trying to accomplish within a customer care context.
3. To integrate an emotion recognition feature to allow the bot to modify its language based on how the user feels.
4. To implement a Dynamic Retrieval-Augmented Generation (RAG) functionality to minimize hallucinations and improve the accuracy of information delivered.
5. To evaluate how well the prototype performs compared to conventional rule-based systems and LLM chatbots that don't include retrieval augmentation or emotion recognition.

3. Approach

For this project, the proposed system is taking a systematic approach that blends natural language processing with conversational AI techniques.

3.1 Data Collection

For this study, research will collect its data using publicly available collections i.e., publicly available databases, such as GoEmotions and CLINC150 containing datasets for customer support conversations, intent classification, and emotion analysis. the use of publicly available databases allows researchers to gather datasets in an ethically responsible manner, while avoiding potential issues surrounding the invasion of privacy of real-world customers.

3.2 Data Processing

Before training any models, the text data must be pre-processed. The process will consist of cleaning, tokenizing, and normalizing the raw text data. Once the pre-processing has been completed, will check that the emotion and intent labels for all data are consistent with each other. To ensure that the proposed system are accurately assessing our models, will divide the datasets into three different groups at random: training, validation, and testing

3.3 Model and Method Selection

The chatbot will utilize a pre-trained large language model to produce the chatbot's output. The models used for intent recognition and emotion analysis will utilize transformer architecture to classify intent and emotions based on the input given to it by the user. To improve the chatbot's responses, the proposed system will also implement Retrieval-Augmented Generation (RAG), which will allow the chatbot to pull related documents stored in a vector database and include those in the generated response.

3.4 Testing and Evaluation

To evaluate the performance, the proposed system will be compared to a baseline chatbot that is built using a Large Language Model and has no intent recognition, emotion analysis, or retrieval-augmented generation capabilities. Quantitative evaluations will be based on the classification accuracy, precision, recall, and F1-score for intent and emotion detection. Qualitative evaluations will analyze the relevance of responses to the context of the user's queries, the contextual appropriateness of responses, and the emotional alignment of responses to user queries. Both methods of evaluation will use simulated customer support scenarios, while qualitative evaluation will additionally involve a panel of experts, due to time constraints.

4. Project Steps and Timeline

The proposed research is separated into structured phases. These phases will be completed over 10 Weeks. Each phase aligns with the research objectives and is practically included in this timeline according to the available timeframe. The research timeline established within this 10 week timeline maintains an equal balance of data preparation, analytics development, system design, and research documentation to facilitate the development of an intelligent customer support chatbot.

Week	Phase	Description
1	Data Collection & Initial Setup	Collect public datasets that contain information about customer support calls and texts, determining the type of

		information contained within each dataset as well as which libraries and tools are needed within the environment you've set up for development purposes.
2	Data Cleaning & Preprocessing	Perform text preprocessing tasks such as data cleaning, normalization, tokenization, and handling missing or noisy entries. Prepare datasets for use in intent recognition, emotion analysis, and chatbot interaction scenarios.
3	Exploratory Data Analysis & Literature Alignment	Use the data collected to perform a detailed examination of the types of information that would be included in a customer support call or message (i.e., the frequency distribution) by looking for common characteristics of customers' intent, emotional response patterns, and typical formats of conversations. You will use your findings in conjunction with relevant academic literature to develop a theoretical background and the related work sections.
4	Intent Recognition Framework Design	Utilize your knowledge of current academic research to help determine what types of NLP techniques and model architectures would be best suited for the design of the intent recognition component of your system and record the methodology decisions made regarding types of customer support-related intent categories.
5	Emotion Analysis Framework Design	Create a conceptual framework for emotion analysis, identifying potential classification approaches. Examine the possible ways that emotional expression may affect the responses of chatbots and the quality of a user's interaction with a chatbot.
6	LLM Integration for Response Generation	Determine what the role of the Long-Term Language Model (LLM) is as the primary engine behind the conversational experience. Record the way user input, the detected intent, and emotional context are combined to produce a chatbot's response.
7	RAG-Based Knowledge Retrieval Design	Determine how to design the Retrieval-Augmented Generation (RAG) mechanism by defining how document embeddings will be created, storing them in vector form, and the method by which document embeddings will be retrieved from the storage medium. Explore how the use of dynamic knowledge retrieval can enhance the relevance of responses and decrease the occurrence of hallucinatory responses.
8	System Integration & Research Analysis	Combine all of the components developed for intent recognition, emotion analysis, LLM-based generation, and RAG into a cohesive research framework and analyze how the use of the research framework will affect the overall quality of the integrated method according to experiential and literature findings.
9	Evaluation Design & Documentation	Define evaluation criteria such as response relevance, consistency, and qualitative user satisfaction indicators.

		Document the evaluation methodology without extensive benchmarking or large-scale implementation.
10	Final Documentation & Presentation	Finalize the project report, including methodology, system design, discussion, and expected contributions. Prepare materials for final submission and presentation or viva.

5. Expected Results

This project is expected to achieve the following objectives:

- An intelligent customer service chatbot will be designed and developed that has the ability to recognise user intentions, analyse user emotional states, and respond to users based on augmented information retrieval.
- The responses generated by an intelligent customer service chatbot will have greater relevance and accuracy than those produced by traditional chatbot systems, which rely solely on generative or static knowledge.
- The number of incorrect and unverified answers provided to customers will decrease, as current content will be used as the basis for retrieving information rather than relying on past experience to provide answers.
- The responses from an intelligent customer service chatbot will be more contextually appropriate and emotionally congruent with the user, thus improving users' impression of their interaction with the intelligent customer service chatbot because emotion analysis is integrated into the answer generation process.
- Understanding how customers interact with intelligent customer service chatbots and determining the trade-offs that exist between employing large language models instead of requiring customers to learn from the responses generated by these chatbots to detect customer intent, understand customer emotion, and retrieve factual information.
- The research findings will provide insight into limitations and practical challenges encountered in relation to data quality, chatbot behaviour and evaluations, as well as future research related to intelligent customer service chatbots.

6. Resources Needed

To successfully complete this project, the following resources will be essential across different stages of development, training, and evaluation:

Tools and Software

- Python programming language
- Jupyter Notebook or similar development environment
- NLP and machine learning libraries such as Hugging Face Transformers, Scikit-learn, and PyTorch
- Vector database tools such as FAISS for information retrieval

Datasets

- Public customer support conversation datasets
- Intent classification datasets
- Emotion analysis datasets

Computing Resources

- Personal computer for development and testing
- Optional cloud-based platforms such as Google Colab for model training

Limitations and Risks

Publicly available datasets do not typically cover every possible scenario in customer support (especially rare or unusual instances) and therefore often lack sufficient information. The classification labels that the proposed system utilize to evaluate intent and emotion, for example, vary quite considerably in terms of subjectivity; therefore, our output classifications may not always be identical. The preliminary analysis shows that while these pre-trained language models perform reasonably well when responding to language inputs, they do not have complete knowledge of the underlying context within which they are operating. Due to the limited time available to conduct this research project, the proposed system will not be able to run as many experiments and expand our model to a more complex level.

Risk Mitigation Strategies

To mitigate the risks identified, the proposed system will rely on existing and previously successful customer service-related datasets instead. During the data preprocessing phase, we will concentrate on ensuring that the text is clean and addressing any inconsistencies in labelling the text into categories (based on the expected response) prior to collecting the data; we will also be using Retrieval-Augmented Generation implementations for the generation of responses in order to increase accuracy and minimize error rates. Rather than implementing a complex solution, the proposed system will focus on design and conceptual development during the remaining timeframe to ensure the maximization of the time usage.

Documentation Tools:

- Microsoft Word for proposal and final report writing in accordance with university formatting guidelines.
- Zotero or Mendeley for organizing references and managing citations.
- Jupyter Notebook for documenting data preprocessing steps, experiments, and methodological decisions.
- Git for version control and structured management of project files (optional).

7. Literature Review

The studies cited below assist in the understanding and definition of this project. The following mentioned references will provide the theoretical framework and research methodology the proposed system will be using to construct the Intelligent Customer Support Chatbots based on the usage of Large Language Models, together with the Intent Recognition and Emotion Recognition techniques, and Retrieval-Augmented Generation methods, which will employ to accomplish this project.

#	Reference	Title	Name of Researchers	Summary	Methodologies	Accuracy	Comments
1	Lewis et al. (2020)	Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks	Patrick Lewis, Ethan Perez, et al.	Proposed a Retrieval-Augmented Generation framework combining neural retrieval with text generation to improve factual accuracy in NLP tasks.	Dense Passage Retrieval, Transformer-based Generation	Reported improvement in factual correctness over generative-only models	Not evaluated specifically on customer support dialogue systems
2	Shuster et al. (2022)	Language Models that Learn from Retrieval	Kurt Shuster et al.	Investigated integrating retrieval mechanisms into large language models to enhance dialogue quality and factual grounding.	Transformer LLMs, Retrieval Modules	Demonstrated reduced hallucination and more accurate responses	Limited focus on intent or emotion-aware interaction
3	Gao et al. (2023)	A Survey on Retrieval-Augmented Generation	Yunfan Gao et al.	Provided a comprehensive survey of RAG architectures, applications, and challenges across NLP tasks.	Systematic Literature Review	Not applicable (survey-based study)	Lacked experimental validation in real customer service scenarios
4	Zhang et al. (2023)	Intent Detection Using Transformer Models	Zhang et al.	Applied transformer-based models for intent classification in conversational systems.	BERT, RoBERTa	Achieved high intent classification accuracy across benchmarks	Emotion analysis not considered
5	Qin et al. (2024)	Emotion-Aware	Qin et al.	Studied the role of emotion recognition in	Deep Learning	Improved emotion classification	No integration with

		Conversational Agents		dialogue systems and its impact on user satisfaction.	Emotion Classifiers	on performance	external knowledge sources
6	Liu et al. (2024)	Joint Intent and Emotion Detection in Dialogue Systems	Liu et al.	Proposed joint modeling of intent and emotion using transformer-based architectures.	Transformer-based Joint Learning	Outperformed separate intent and emotion models	Computationally expensive; limited real-time evaluation
7	Wu et al. (2024)	Customer Support Chatbots with Retrieval-Augmented Generation	Wu et al.	Implemented a RAG-based chatbot framework for customer support applications.	Vector Databases, LLMs, RAG	Reported higher response relevance and accuracy	Emotion awareness not included
8	Ahmed et al. (2023)	Emotion-Aware Customer Support Chatbots	Ahmed et al.	Developed an emotion-aware chatbot to adapt responses based on user mood.	Emotion Analysis, Rule-based Response Logic	Improved Emotion classification accuracy	Relied on a static knowledge base
9	Chen et al. (2025)	Dynamic Knowledge Retrieval for Conversational AI	Chen et al.	Explored dynamic retrieval mechanisms for updating chatbot knowledge in real time.	Vector Search, LLM Integration	Demonstrated improved response accuracy	Did not consider intent or emotion modeling
10	Rashkin et al. (2019)	Towards Empathetic Open-Domain Dialogue Systems	Hannah Rashkin et al.	Introduced emotion-aware dialogue generation to improve empathy in conversations.	Emotion Labeling, Seq2Seq Models	Improved emotion recognition accuracy	Not designed for customer support; older architecture

7.1 Overview of Existing Research

The development of conversational AI today is driven by the advancement of large language models (LLMs). While these LLMs have demonstrated that they can identify and create language similar to human written language, they do have their limitations when asked to produce an answer accurately, particularly for very specific fields and with little or no outside knowledge to assist them in answering questions correctly. Additionally, many traditional generative methods can generate language that makes sense, but the generated information is not backed by any factual basis, this is often referred to as "hallucination."

To address the shortcomings of traditional generative approaches, researchers have developed hybrid models that integrate retrieval methods and generative models. In a hybrid system, data is retrieved and brought into the generative process as a way to influence the output of the conversation model. When researchers first began to examine these systems, they identified how building on actual data helps produce more accurate and reliable responses. Based on the initial findings, subsequent studies have grouped together many different types of architectural models and demonstrated better accuracy and

performance across a number of different NLU (Natural Language Understanding) tasks. However, much of what has been done to date has focused on enhancing the factual content of a model via the retrieval and integration of factual content sources into its generative output. There is less focus on improving the conversational aspects of a model, such as a model's ability to demonstrate empathy for others or understand the user's needs accurately.

Several studies conducted by researchers have focused on the ability of machine learning algorithms to detect user intent and emotion through the analysis of text data. Transformer based models have demonstrated exceptional performance on the task of intent classification. In addition, emotion-aware systems create interactions that users consider more empathetic, resulting in increased satisfaction. Despite the advancement of each area of research, the categorization of each area of capability has been done as an individual study without addressing how multiple areas may coexist in a unified conversational interface.

As previously mentioned, the literature describes a significant advancement in each of the components associated with the development of intelligent chatbots, however, it also highlights the need for an integrated approach to incorporate these different types of capabilities into actual customer service settings.

7.2 Key Research Findings

The literature review identified four primary findings:

- Retrieval-Augmented Generation (RAG) dramatically improves factual accuracy in Language Model (LM) responses by adding external sources of knowledge to the model context (reducing hallucinations).
- Transformer-based models (e.g. BERT and RoBERTa) consistently provide good intent classification performance and are therefore capable of identifying user goals within conversation systems.
- Emotion-aware chatbots produce more empathetic-sounding and emotionally engaged responses from users, which leads to increased user satisfaction and engagement.
- Most existing studies do not attempt to simply enhance the individual factors as previously discussed, nor focus on using all three factors together within the same conversation systems.

7.3 Critical Research Gap: Regional and Cultural Context

Despite notable progress in conversational AI research, several gaps remain:

- **Lack of integration:** Existing studies rarely combine retrieval-based generation, intent recognition, and emotion analysis into a unified chatbot framework.
- **Limited focus on customer support applications:** Many models are evaluated using general dialogue datasets rather than domain-specific customer service scenarios.
- **Cultural and linguistic bias:** Most datasets and assessments in AI are developed with the cultural and linguistic biases of Western cultures; therefore, they don't represent the perspectives of those from different regions.

- **Evaluation limitations:** Previous research mainly evaluates AI-generated content using metrics and does not adequately assess the degree to which responses were appropriate or relevant to the context and how satisfied users were with their interactions.

Consequently, these gaps indicate the need for additional investigations into the development of AI that will be effective through combining both technical integration with the multiple contexts in which users interact with the AI software.

7.4 Research Justification and Expected Contribution

This research is justified by the need to address the identified gaps through a more holistic approach to customer support chatbots. The expected contributions of this study include:

- Proposing an integrated framework that combines Large Language Models, intent recognition, emotion analysis, and Retrieval-Augmented Generation within a single conversational system.
- Demonstrating how dynamic knowledge retrieval can enhance response accuracy while maintaining sensitivity to user intent and emotional state.
- Providing a structured evaluation approach that considers not only classification accuracy but also response relevance, consistency, and user experience.
- Offering insights that can inform future research and development of intelligent customer support systems, particularly in contexts that require both factual reliability and empathetic interaction.

8. References

- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W. T., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems (NeurIPS)*.
- Shuster, K., Poff, S., Chen, M., Kiela, D., & Weston, J. (2022). Language Models that Learn from Retrieval. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Gao, Y., Li, Y., Zhang, J., Lin, C., & Liu, Z. (2023). A Survey on Retrieval-Augmented Generation. *ACM Computing Surveys*.
- Zhang, H., Li, X., & Zhao, Y. (2023). Intent Detection Using Transformer Models for Conversational Systems. *Journal of Artificial Intelligence Research*.
- Qin, L., Wang, Y., & Liu, X. (2024). Emotion-Aware Conversational Agents: Models and Applications. *IEEE Transactions on Affective Computing*.
- Liu, J., Chen, R., & Sun, M. (2024). Joint Intent and Emotion Detection in Dialogue Systems Using Transformer-Based Models. *Knowledge-Based Systems*.
- Wu, T., Zhang, Y., & Huang, S. (2024). Customer Support Chatbots with Retrieval-Augmented Generation. *Information Processing & Management*.
- Ahmed, S., Khan, R., & Ali, N. (2023). Sentiment-Aware Customer Support Chatbots. *Expert Systems with Applications*.
- Chen, Y., Zhou, Q., & Li, H. (2025). Dynamic Knowledge Retrieval for Conversational AI. *Artificial Intelligence Review*.
- Rashkin, H., Smith, E. M., Li, M., & Boureau, Y. L. (2019). Towards Empathetic Open-Domain Dialogue Systems. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*.