Smart Conversational Agent for Ambient Assisted Living

Carla Mendes[a]

[a]*Computer Science and Communications Research Centre, School of Technology and Management, Polytechnic of Leiria, 2411-901 Leiria, Portugal*

ARTICLE INFO

Keywords:

ABSTRACT

Abstract goes here.

1. Introdution

2. Related work

3. Background and contextualization

This section aims to review existing literature regarding Conversational Agents (CAs), Sentiment Analysis (SA) and mood assessment questionnaires.

3.1. Conversational agents

Chatbots, referred to as CAs or Conversational Systems (CS), are software applications created to imitate human-computer interactions [**?**].

3.1.1. Classification methods

Chatbots can be categorized based on various factors, including their goal, interaction mode, knowledge domain, and response-generation method [**?**].

CS operate using different modes of interaction, namely based on text or speech. In the text-based mode, users communicate with the chatbot by typing their queries or statements, commonly through chat applications, messaging platforms, or web-based chat interfaces. The chatbot responds with text-based messages in return. On the other hand, speech-based chatbots enable users to interact with the chatbot using spoken language. These chatbots employ Speech Recognition (SR) technology to convert the user's voice input into text, which is then processed and analyzed to generate appropriate responses. Voice-based chatbots are typically found in voice assistants like Amazon Alexa, Google Assistant, or Apple Siri. They offer a convenient and hands-free method of interacting with the chatbot, enabling users to engage in natural conversations and perform tasks using voice commands. Additionally, chatbots can also adopt a multi-modal approach, allowing interaction through both text and speech [**?**].

When considering their objective, chatbots are categorized as either task-oriented or non-task-oriented. Task-oriented chatbots are designed with a specific purpose, focusing on handling particular tasks and engaging in brief conversations, typically within a limited domain. Conversely, non-task-oriented chatbots specialize in emulating conversations with individuals and participating in casual chitchat primarily for entertainment. As a result, they operate in open domains, facilitating more diverse and unrestricted conversations [**?**].

In terms of response generation methods, chatbots are also classified based on the techniques and algorithms used to generate suitable and meaningful responses to user queries or inputs. These methods can be categorized as follows:

- **Rule-based**: rely on a predefined set of rules and patterns to generate responses. Human experts typically create these rules and program them into the chatbot system. The Conversational Agent (CA) then matches user inputs with specific patterns or keywords and retrieves corresponding responses. Rule-based systems work well for simple and structured conversations but may struggle with complex or unpredictable queries.

✉ carla.c.mendes@ipleiria.pt (C. Mendes)
ORCID(s): 0000-0001-7138-7124 (C. Mendes)
in https://www.linkedin.com/profile/view?id='carla-mendes-5b3586233' (C. Mendes)

- **Retrieval-based**: use Machine Learning (ML) techniques to select the most appropriate response based on the user's input from a large dataset of predefined responses. Retrieval-based systems may struggle with creative responses, however provide contextually relevant responses.
- **Template-based**: use pre-built response templates that are filled with relevant information based on user inputs. These templates contain placeholders for dynamic content such as names, dates, or specific details. The chatbot captures the intents of the user's input and selects an appropriate template to generate a response. Template-based systems may lack flexibility and creativity in generating unique responses but are relatively simple to implement.
- **Generative-based**: rely on advanced Natural Language Processing (NLP) techniques and ML models to create responses from scratch, namely sequence-to-sequence models or transformers. These models are trained on large datasets and learn the patterns and structures of human language. Generative models are more versatile in handling diverse user inputs. However, they can be more computationally intensive and require significant computational resources for training and inference.
- **Hybrid Approach** - combine multiple methods mentioned above to leverage the strengths of each approach.

Lastly, in the knowledge domain dimension, chatbots possess specialized knowledge and capabilities tailored to serve specific purposes within a limited scope. They possess knowledge and capabilities customized to fulfill specific objectives or offer assistance within a restricted topic. These chatbots undergo training and programming to comprehend and address inquiries regarding a specific domain. For instance, a chatbot tailored to a particular subject can aid in customer support for an e-commerce platform, provide medical advice in the healthcare sector, or offer travel recommendations within the tourism industry. On the other hand, open-domain chatbots engage in conversations spanning a broad spectrum of topics, free from confinement to any specific domain. They aim to emulate human-like interactions, delivering casual conversation, entertainment, or general information across various subjects [**?**].

3.1.2. General Architecture

A robust CA system will possess several key components, represented in Figure 1.

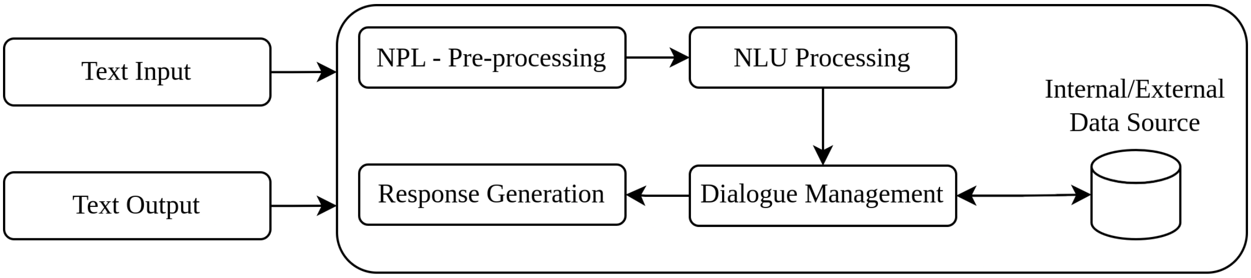

Figure 1: CAs general architecture.

During the NLP phase, the user's request undergoes various techniques, including tokenization, lemmatization, and stemming. These techniques help extract structured data from the request, which is then passed on to the subsequent component, known as the Natural Language Understanding (NLU) module, responsible for analyzing each incoming user request using various strategies, namely, parsing the request to understand the user's intention and the associated details. The dialogue management module focuses on keeping track of the dialogue context and defining the following actions to perform by analyzing the input request that has been transformed into understandable structured data by the CA system. The data sources serve as repositories for information and data utilized by the dialogue manager. These sources can be either internal or external. Internally, chatbots can access data from templates or rules to understand user requests and generate appropriate responses. Moreover, CA can also build their databases from scratch or leverage existing databases that align with their domain and functionality. In contrast, external data sources can be accessed through third-party services like Web APIs, which provide the necessary information. The response generator module plays a crucial role in generating an appropriate response from a pool of potential options after executing an action. This component utilizes the approaches mentioned earlier to generate the most suitable response for the given context. [**? ?**]

3.1.3. Tools

Rasa [?] is an open-source dialogue framework for building conversational Artificial Inteligence (AI) applications. It uses NLP techniques and dialogue management to enable interactive and context-aware conversations. Rasa consists of two main components: the NLU module for processing user inputs and extracting intents and entities, and the Dialogue Management module for handling conversation flow and decision-making. It supports personalized dialogue policies, provides tools for training and evaluation, and integrates with different channels and platforms. Rasa supports both text-based and voice-based interactions, making it versatile for various applications.

Amazon Lex [?] is a service provided by Amazon Web Services that allows developers to build, test, and deploy CA powered by AI. It is designed to create interactive chatbots and virtual assistants that can understand natural language inputs and provide appropriate responses. Amazon Lex leverages advanced natural language models and ML algorithms to enable accurate understanding and interpretation of user inputs. It supports both text and speech inputs and outputs, making it suitable for various applications. With Amazon Lex, developers can easily integrate CA into their applications or platforms, enabling more intuitive and engaging user experiences.

Dialogflow [?] is a NLU platform developed by Google. It provides tools and capabilities for building CA, chatbots, and virtual assistants. With Dialogflow, developers can create, manage, and deploy CA across multiple platforms and systems. It supports both text and speech inputs and outputs, allowing users to interact with the CA through various channels such as messaging platforms, voice assistants, and websites. This platform utilizes advanced ML algorithms to understand and interpret user inputs, extracting important information such as intents (the user's intention) and entities (specific pieces of information). It offers a range of pre-built NLU components and features, including Named Entity Recognition (ER) and SA. Additionally, Dialogflow provides a visual interface for designing conversation flows, managing dialogues, and defining responses.

OpenDial [?] is an open-source Java-based toolkit used for building and evaluating speech-based CA. It provides a framework and set of tools that enable developers to create interactive dialogue systems capable of engaging in natural language conversations. Furthermore, the toolkit offers a range of features and functionalities for building CA. It provides modules for NLU, dialogue management, and speech synthesis. OpenDial allows developers to define dialogue policies and strategies to guide the system's behaviour and response generation. It also includes components for handling user input, managing context, and generating appropriate spoken responses. Overall, OpenDial emphasizes modularity and extensibility, enabling developers to customize and adapt the toolkit according to their specific requirements.

Botpress [?] is an open-source platform that enables developers to build, deploy, and manage chatbots and virtual assistants. It provides a visual interface for designing conversational flows and supports both text-based and voice-based interactions. Botpress is written in JavaScript and can be deployed on various platforms. One of the key features of Botpress is its visual flow builder, which allows developers to create complex conversational flows using a drag-and-drop interface. This makes it easy to design the dialogue flow of the chatbot and define the interactions between the user and the bot. Botpress also offers built-in NLU capabilities, allowing developers to train the chatbot to understand user intents and extract entities from user inputs.

ChatterBot [?] is an open-source Python library that facilitates the development of chatbots. The primary focus of ChatterBot is to generate responses based on pre-defined conversational patterns. It uses a machine learning algorithm called Latent Semantic Analysis (LSA) to train a language model on a given corpus of text data and then generate appropriate responses based on the patterns it has learned. ChatterBot supports the use of multiple languages and provides various pre-trained language models that can be used out of the box. Additionally, it enables developers to customize the chatbot's behaviour by defining rules, selecting appropriate responses, and handling specific cases. One of the notable features of ChatterBot is its ability to learn and improve over time. It employs a technique called "conversational context" to maintain the history of the conversation and generate contextually relevant responses.

3.2. Sentiment Analysis

SA, a subfield of NLP, aims to derive the sentiments expressed in a piece of text based on its content, which can be conducted at different levels: Document Level, Sentence Level, Phrase Level, and Aspect Level [? ?].

Document-level SA involves assessing the sentiment of an entire document and assigning a single polarity to it. Classification methods, both supervised and unsupervised, can be employed at this level to determine the sentiment conveyed in the document [?].

At the Sentence Level, each sentence is analyzed and classified into a polarity sentiment. This approach more valuable when a document contains a diverse range of sentiments. By independently determining the polarity of each sentence, either using methodologies similar to document-level analysis or with more extensive training data and processing resources, the overall sentiment of the document can be aggregated or analyzed sentence by sentence [**?**].

The Phrase Level of SA focuses on mining opinion words and sentiments at the phrase level. While document-level analysis broadly categorizes the entire document as either positive or negative, sentence-level analysis proves more advantageous because documents often contain both positive and negative statements. At this level, individual words become the fundamental units of language, and their polarity is intrinsically tied to the subjectivity of the sentences or documents in which they appear [**?**].

Lastly, Aspect Level delves even deeper, as it is performed on specific aspects within a sentence. Since a sentence may contain multiple aspects, this approach pays close attention to all aspects present and assigns polarity to each one. An aggregate sentiment is then calculated for the entire sentence, considering the sentiments of all its aspects.

To conduct SA, several main steps are followed, including data collection, feature selection, feature extraction, and the use of word embeddings. These steps aid in gathering relevant data, choosing essential features, extracting meaningful patterns, and representing words in a numerical format suitable for analysis [**?**].

3.2.1. Approaches

There are three main approaches commonly used for SA: the Lexicon Based Approach, the ML Approach, and the Hybrid Approach [**? ?**].

Lexicons consist of tokens, with each token having a predefined score that indicates the neutral, positive, or negative nature of the text. The lexicon-based method is highly suitable for conducting SA at both the sentence and feature levels. Initially, the document is divided into tokens of individual words, and then the polarity of each token is calculated and aggregated. There are primarily two approaches used in Lexicon Based Approaches: the Corpus Based Approach and the Dictionary based approach [**?**].

The Corpus-based approach utilizes semantic and syntactic patterns to determine the emotion of a sentence. It begins with a predefined set of sentiment terms and their orientations, then explores syntactic or similar patterns within a vast corpus to identify sentiment tokens and their orientations. This method is specific to the situation and requires a substantial amount of labeled data for training. Within the corpus-based approach, there are two types of approaches: the Statistical Approach and the Semantic Approach [**? ?**].

On the other hand, the Dictionary-Based Approach commences by manually collecting a set of opinion words to form a seed list. Next, dictionaries and thesauruses are consulted to find synonyms and antonyms of these words, which are then added to the seed list. This process continues until no new words are discovered. However, a drawback of this approach is the challenge of finding context or domain-oriented opinion words [**? ?**].

There are two primary in ML approaches: Supervised ML and Lexicon-based unsupervised learning [**?**].

In Lexicon-based unsupervised strategies, knowledge bases, ontologies, databases, and lexicons containing specific and detailed information for SA are utilized. On the other hand, supervised learning methods are more widely used due to their high accuracy. These algorithms require training on a labeled dataset before they can be applied to real data. Features are extracted from the text data during the training process [**?**].

The ML technique uses syntactic or linguistic factors to classify sentiment and constitute a text classification problem. Therefore, the model associates features of the underlying record with class labels and predicts the label for unknown instances [**?**]. Commonly used ML algorithms include:

- **Decision Tree (DT) classifier**
- **Linear classifier** - e.g. Support Vector Machine (SVM) and Neural Networks such as Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BILSTM), Transformers
- **Rule-based classifier**
- **Probabilistic classifiers** - e.g. Multinomial Naive Bayes (NB), Maximum Entropy (ME) and Bayesian Network (BN)
- **K-Nearest Neighbor (KNN)**

The hybrid approach combines ML and lexicon-based techniques. It refers to the integration of both methods to enhance the accuracy and effectiveness of the systems. By leveraging the strengths of both approaches, the hybrid method can provide more comprehensive insights into the sentiment expressed in the text [**?**].

3.2.2. *Datasets*

There are diverse datasets used to train SA algorithms, some of which include: SemEval 2007 task 14 corpus, ISEAR, SentiWordNet, IMDB, SST and NRC.

The SemEval 2007 task 14 corpus [?] consists of emotional newspaper headlines from reputable sources like the New York Times, CNN, BBC, or Google News. This corpus allows training models for emotions such as anger, disgust, fear, joy, sadness, and surprise.

ISEAR [?] is a dataset comprising psychological data from a survey conducted in 1990. In this dataset, 3,000 subjects described situations where they felt various emotions, including joy, sadness, fear, anger, disgust, shame, or guilt.

The SentiWordNet dataset [?] is a publicly available dataset that automatically labels synsets in the WordNet dataset based on their degree of positivity, negativity, and neutrality.

The IMDB Movie Reviews dataset [?] is publicly available and contains around 50,000 movie reviews, labeled as either positive or negative.

SST [?] is a sentiment analysis dataset from Stanford, containing 10,662 movie sentences from Rotten Tomatoes, labeled with corresponding emotions, ranging from "very negative" to "very positive."

NRC [?] is a comprehensive emotion Lexicon dataset, containing more than 14,000 words annotated with sentiment (positive or negative) and various emotions, such as anger, anticipation, disgust, fear, joy, sadness, surprise, and trust.

3.3. Mood Assessment Questionnaires

The Psychological General Well-Being Index (PGWBI) [?] is a questionnaire to measure subjective psychological well-being, assessing emotional states that reflect a sense of well-being or distress. It consists of 22 standardized items (or 6 for the short form), producing a single well-being score. It includes subscales for anxiety, depression, positive well-being, self-control, general health, and vitality. The original PGWBI has 22 self-administered items, rated on a 6-point scale, covering six HRQoL domains: anxiety, depression, positive well-being, self-control, general health, and vitality. Each item scores from 0 to 5, referring to the last 4 weeks. The domains have 3 to 5 items each. All domain scores contribute to a global summary score, with a theoretical maximum of 110 points. The short form, containing six items, explains over 92% of the variance of the full questionnaire.

The Depression-Happiness Scale (DHS) [? ?] is a psychometric tool utilized to evaluate an individual's levels of depression and happiness, to measure their emotional well-being and mood. The questionnaire consists of 25 items, 12 assessing happiness, and the remaining 13 focus on depression. However, there is also a short version comprising 6 items, where 3 measure happiness and the other 3 measure depression. Respondents are asked to reflect on their feelings over the past seven days and rate each item on a 4-point scale: "never" (0), "rarely" (1), "sometimes" (2), and "often" (3). The positive items are directly scored, while the negative ones are reversed-scored. The scores on the 25-item scale range from 0 to 75, with higher scores indicating a greater sense of happiness and lower scores suggesting lower levels of depression.

The Positive Affect Negative Affect Scale (PANAS) [?] is a widely used scale for assessing mood or emotions. It consists of 20 items, with 10 items dedicated to measuring positive affect (e.g., excited, inspired) and 10 items for negative affect (e.g. upset, afraid). Respondents rate each item on a five-point Likert Scale, indicating the extent to which they have experienced the particular effect within a specified time frame. The PANAS is designed to gauge affect in various contexts, such as the present moment, the past day, week, year, or overall (on average). As a result, the scale measures current emotional state, enduring or trait-based emotions, fluctuations in emotions over a specific period, or emotional responses to specific events.

The Oxford Happiness Questionnaire (OHQ) [?] is a questionnaire designed to assess an individual's happiness, covering various dimensions such as positive emotions, life satisfaction, sense of purpose, and overall well-being. The OHQ consists of 29 items, each presented as a single statement, which respondents rate on a uniform six-point Likert scale. Researchers used discriminant analysis to create a shorter version of the OHQ and identified eight key items. The final OHQ score is calculated by summing the responses to positive items and reverse-scored negative items and then dividing by the total number of items (29). This calculation results in scores ranging from 1 to 6, reflecting the participants ' level of happiness based on their responses to the questionnaire, with lower values indicating sadness and higher values of positivity.

Affectometer 2 [?] is a 40-item self-report scale designed to measure general happiness or well-being by assessing the balance of positive and negative emotions experienced in recent times. It takes approximately 5 minutes to complete. The questionnaire contains separate items for measuring both positive and negative affect. Participants rate the extent to which they have experienced each emotion within a specific time frame, such as the past week or month. They provide their responses using a Likert-type scale, indicating their agreement or frequency of experiencing each emotion. The total score on the Affectometer 2 represents the difference between the positive and negative affect, reflecting an individual's overall emotional balance. In the development of the scale, the authors categorized the items into 10 mnemonic groups, referred to as "qualities of happiness." They aimed to include four items for each category, with one each from positive sentences, negative sentences, positive adjectives, and negative adjectives. To complete the 40-item matrix, the researchers retrieved four items from the original item pool and created and validated six new ones.

The Affect Balance Scale (ABS) [**?**] is a comprehensive measure used to assess specific aspects of quality of life in both national and international contexts. It consists of a 10-item Subjective Well-Being Scale, with five questions focused on positive feelings and five statements related to negative feelings. Participants respond to the questions without gradation, indicating either a "yes" or "no" to each item. This format allows for the summation of responses, resulting in scores ranging from 0 to 5 for both the positive and negative affect subscales. Higher scores on each subscale indicate greater levels of positive and negative affect, respectively. Researchers sometimes use the final score, which is the difference between the positive affect and negative affect scores, as an indicator of overall happiness.

Well-being Questionnaire (W-BQ12) [**?**] is a self-report psychometric tool used to assess an individual's subjective well-being and overall quality of life. It is designed to measure various dimensions of well-being, including emotional, psychological, and social aspects. The W-BQ12 consists of 12 items, each addressing different aspects of well-being. Participants are asked to rate the extent to which they agree or disagree with each statement based on their feelings and experiences. Responses are typically measured on a Likert-type scale, where participants indicate their level of agreement or disagreement or the frequency of experiencing certain feelings or thoughts. The questionnaire covers a wide range of well-being dimensions, including positive affect, life satisfaction, self-esteem, and social relationships. By analyzing the responses, researchers and practitioners can gain insights into an individual's overall sense of well-being and emotional health.

The Emotional State Assessment Tool (ESAT) [**?**] is a measure of emotional well-being based on a two-dimensional model of affect (positive and negative emotion) and a six-dimensional model (cheerfulness, vitality, serenity, sadness, lethargy, and stress) derived from a philosophical theory of affect. The tool was developed through a series of studies, resulting in an 18-item questionnaire that offers a more detailed understanding of positive and negative affect, providing a versatile and cross-culturally sensitive assessment of emotions.

The Geriatric Depression Scale (GDS) [**?**] is a short, self-administered questionnaire designed to assess depressive symptoms in older adults. It is specifically developed for use with individuals aged 65 years and older, although it can also be used with younger adults in certain circumstances. The GDS is a widely used screening tool to identify potential depression in older adults, particularly in clinical settings, research studies, and geriatric assessments, and possesses numerous versions. One of which is the Geriatric Depression Scale-15 (GDS-15) a questionnaire with yes-or-no questions where participants are asked to respond based on how they felt over the past week. Of the 15 items, 10 indicated the presence of depression when answered positively, while the rest (question numbers 1, 5, 7, 11, 13) indicated depression when answered negatively. Scores of 0-4 are considered normal, depending on age, education, and complaints; 5-8 indicate mild depression; 9-11 indicate moderate depression; and 12-15 indicate severe depression.

4. Architecture

This section aims to detail the architecture and main components of the proposed system, including the goal and communication schemes between each component. The system comprises two primary modules, as detailed in Figure 2: frontend clients, further subdivided into the mobile and web application, and backend, which encompasses the API, database, and CA.

The mobile application serves as the primary interface for interacting with elderly users, collecting input data for the system, and providing output responses and data. Its key function is to engage in ongoing and meaningful conversations with older adults, assisting them in regulating their emotions. Given the technological challenges encountered by older adults, detailed in **??**, such as age-related health issues affecting hearing and vision, as well as their limited digital literacy and skills resulting from historical exclusion from the digitization process of society, the design of the mobile application must adhere to essential design principles to ensure a positive user experience (Blazic, 2020) [**?**]. In contrast, the web application was crafted to empower caregivers to monitor the emotional well-being of elderly

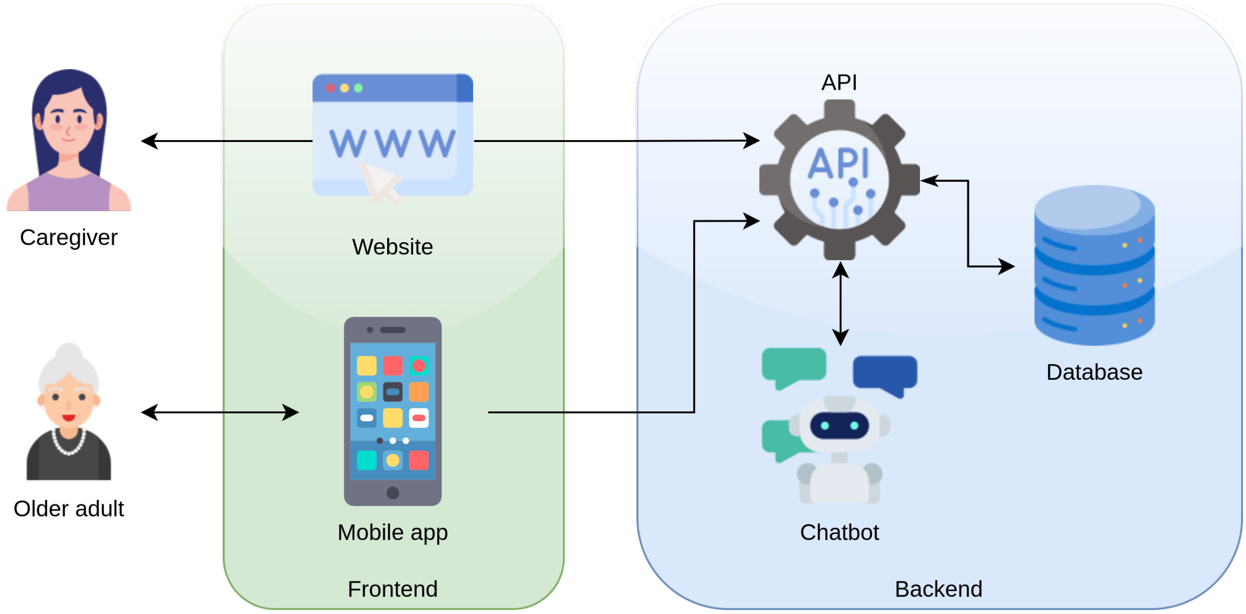

Figure 2: Proposed architecture.

individuals under their care from any location and at any time. Additionally, it provides caregivers with the flexibility to customize and personalize their own experience as well as that of the elderly individuals they are responsible for. Both client applications communicate with the API to efficiently retrieve and store data, ensuring the smooth exchange of information and enhancing the overall system's functionality.

The API has several key functions: capturing, validating, and processing data from client applications, offering them the necessary data, and managing data storage and retrieval from the database. Additionally, the API establishes communication with the CA, which is responsible for guiding and maintaining user conversations to ensure a continuous input flow. Subsequently, the CA, equipped with a sentiment analysis model, processes and analyzes each user message received via the API to determine emotional states, generates responses, and returns them to the API for delivery to the user. Lastly, the database is entrusted with the responsibility of storing and providing comprehensive information, encompassing client details, emotional analysis results, and exchanged messages.

5. Proposed solution and implementation

5.1. Artificial Intelligence Models

The availability of Portuguese SA models with a comprehensive spectrum of emotions is limited, as most existing models primarily differentiate between positive and negative emotions. Consequently, the need arose to develop a custom model from the ground up. In pursuit of a higher model quality, three distinct architectural approaches were meticulously studied, analyzed, and tested with different layers, parameters, and hyper-parameters as they are among the most commonly employed for sentiment analysis tasks: CNN, Gated Recurrent Unit (GRU), and LSTM.

In response to the need for a sentiment analysis model encompassing a broad spectrum of emotions, and in line with the datasets mentioned earlier, the ISEAR dataset [**?**] was selected. However, this dataset is originally in English, necessitating a translation to Portuguese through the use of the DeepL translator, followed by manual verification. Furthermore, a data augmentation technique to was employed toprovide a larger and more robust foundation for each model under construction.The data augmentation process involved utilizing the lemmatization capabilities of the 'pt_core_news_lg' model `https://spacy.io/models/pt#pt_core_news_lg`, a CPU-optimized model provided by the spaCy library, specifically designed for NLP tasks in Portuguese. This lemmatization process targeted non-stop words composed solely of alphabetic characters. Remarkably, the model's lemmatizer demonstrated an impressive accuracy rating of 97%.

As a result of this data augmentation and translation process, a unique and wider version of the ISEAR dataset was created, expanding from its initial 7511 English entries (with the removal of 60 entries, specifically those with placeholder values indicating the absence of felt emotions or without actual responses) to a total of 15022 Portuguese entries.

In response to the need for a SA model encompassing a broad spectrum of emotions, and in line with the datasets mentioned earlier, the ISEAR dataset was selected. However, this dataset is originally in English, necessitating a translation to Portuguese through the use of the DeepL translator, followed by manual verification. Furthermore, a data augmentation technique to was employed provide a larger and more robust foundation for each model under construction. The data augmentation process involved utilizing the lemmatization capabilities of the 'pt_core_news_lg' model `https://spacy.io/models/pt#pt_core_news_lg`, a CPU-optimized model provided by the spaCy library, specifically designed for NLP tasks in Portuguese. After acquiring the translated dataset, each record underwent a lemmatization process. In this process, any non-stop word consisting exclusively of alphabetic characters was lemmatized, and the result was used to generate a new record with the extracted lemmatized words. Thanks to this process of data augmentation and translation, an expanded version of the ISEAR dataset was obtained, where it grew from its initial 7511 English entries (59 entries were initially removed, specifically those with placeholder values indicating the absence of emotions or without actual responses) to a total of 15022 Portuguese entries.

6. Conclusion

CRediT authorship contribution statement

: Conceptualization of this study, Methodology, Software.