Neurocomputing 653 (2025) 131232

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A generative AI model for modelling and creating an emotion-aware database: Design, and analytics☆

Alfredo Cuzzocrea[a,b,*], Giovanni Pilato[c], Pablo G. Bringas[d]

[a] iDEA Lab, University of Calabria, Rende, Italy
[b] Dept. of Computer Science, University of Paris City, Paris, France
[c] ICAR-CNR, Italian National Research Council, Palermo, Italy
[d] Faculty of Engineering, University of Deusto, Bilbao, Spain

HIGHLIGHTS

- We focus on a model based on generative AI for modelling and creating a database that supports emotion recognition and handling.
- We define a general architecture for our generation framework.
- We provide a comprehensive experimental evaluation of the proposed model.

ARTICLE INFO

Communicated by W. Cui

Keywords:
Human-computer interaction
Emotion recognition
Language complexity
Intelligent systems
Sequential pattern mining

ABSTRACT

In this study, we illustrate an ongoing work regarding the building of an *ad hoc* textual dataset for dialogues between a user and an agent of a customer care center. The user is required to express a given emotion and to use a specific level of the *Common European Framework of Reference for Languages* (CEFR). Once the criteria had been defined, we asked *ChatGPT* 3.5 to generate dialogues. We analyzed them to ensure an acceptable level of quality for the dataset. Furthermore, we measured the language complexity used both by the user and the agent through different readability measures. Moreover, we extracted from each turn of each dialogue the attitude that ChatGPT exhibited for each dialogue turn; the interaction patterns were then analyzed with algorithm *Sequential Pattern Mining* (SPM) to extract typical interaction mechanisms occurring in specific emotional contexts. The generated dataset of dialogues, enriched with the readability measures and the extracted patterns, can be exploited as a system that facilitates the interaction between humans and artificial agents in specific contexts and it can constitute the basis of a reinforcement learning system for improving the effectiveness of the dialogues.

1. Introduction

Emotions play a key role during *Human-Computer Interaction* (HCI) and more specifically in *Human-Robot Interaction* (HRI). The use of *Affective Computing* (AC) techniques to recognize and analyze human emotional states can enhance the effectiveness of HCI and HRI. At the same time, emotion recognition is one of the most difficult tasks for artificial systems during their interaction with humans [5,6,31]. Emotions have a multidimensional nature and their understanding depends on the context in which they are expressed [32]. In particular, *Emotion Recognition in Conversation* (ERC) has gained attention from the *Natural language processing* (NLP) community in recent years. This is mainly due to the increased availability of public conversational data. On the other hand, the recognition of emotions is also important for dialogue generation, since this implies the understanding of the emotional state of the interlocutor [1].

Emotions often can overlap and vary between individuals, with the same context potentially triggering different emotional responses at different times. However, there are commonly identifiable situations that can be consistently associated with specific emotions. Therefore, providing examples of emotions associated with specific contexts can contribute to a clearer understanding of this phenomenon. As suggested by Ref. [1], in the context of conversational context modelling,

☆ This research has been made in the context of the Excellence Chair in Big Data Management and Analytics at University of Paris City, Paris, France.

* Corresponding author at: iDEA Lab, University of Calabria, Rende, Italy.
Email addresses: alfredo.cuzzocrea@unical.it (A. Cuzzocrea), giovanni.pilato@icar.cnr.it (G. Pilato), pablo.garcia.bringas@deusto.es (P.G. Bringas).

https://doi.org/10.1016/j.neucom.2025.131232
Received 24 June 2024; Received in revised form 11 June 2025; Accepted 6 August 2025
Available online 20 August 2025

considering the specific context can substantially enhance the effectiveness of NLP systems. Consequently, creating a dataset that is highly specific and contextually focused can be valuable for the development of data-driven models.

There are many contributions in the literature regarding the construction of datasets for emotion recognition. Most of them cover few emotions, tending only to Ekman's basic ones. Some examples are EmotionX [2], *Affect-Intensity Lexicon and Emotion Dataset* (AILA) [3], CrowdFlower's Emotion Dataset [7], Friends [4], and EmoBank [8]. Furthermore, many approaches build datasets using newspapers, books, or dialogues found on the Internet, including those found on social media, e.g., SemEval-2018 Task 1: *Affect in Tweets* (AIT-2018) [9], Sentiment140 [10], *Emotion Intensity Dataset* (EmoInt) [11], the *International Survey on Emotion Antecedents and Reactions* (ISEAR) [12]. Others use movies, e.g., the *Stanford Sentiment and Emotion Classification* (SSEC) [13,14] or physiological signals, e.g., the *Database for Emotion Analysis using Physiological Signals* (DEAP) [15]. Other popular datasets are the *Interactive EMOtional dyadic motion CAPture* (IEMOCAP) database, collected by the *Speech Analysis and Interpretation Laboratory* (SAIL) at the *University of Southern California* (USC) [28], and DailyDialog, a dataset of dialogues about different topics of daily life interactions, characterized by human-written and plain language; data were labeled with communication intention and emotion information [29]. One of the most well-known fully-labeled datasets of human-human conversations across multiple domains is MultiWOZ [46]. Each dialogue includes annotated states and system dialogue acts. The dataset aims to capture natural conversations between a tourist and an information center clerk in a tourist city. It covers various scenarios, from asking about attractions to booking hotels and transportation. The corpus spans seven domains: *Attraction*, *Hospital*, *Police*, *Hotel*, *Restaurant*, *Taxi*, and *Train*.

On the other hand, the authors of Ref. [40] explored a method for developing an agent capable of handling different tasks by integrating dialogue self-play, crowdsourcing, and online reinforcement learning to produce fully annotated dialogues with natural interactions. According to the authors of Ref. [41], an effective chatbot should dynamically combine various behaviors and skills within a single conversation, allowing it to adapt to different users and situations appropriately. Achieving this requires the development of a multi-skill dialogue dataset, featuring multi-turn interactions that demonstrate diverse skills. To address this, the authors of Ref. [41] introduced BOTSTALK, an automated data curation framework that annotates multi-skill dialogues by integrating multiple single-skill dialogue datasets.

Given the significant advancements in *Large Language Models* (LLMs), recent research has explored their potential as generators of training data. Specifically, LLMs have been proposed as task-specific training data generators, particularly in the domain of text classification, with the objective of reducing reliance on manually annotated datasets [43]. Training examples generated by language models have demonstrated effectiveness across various applications through the process of data augmentation [44,45].

The authors of Ref. [44], generate training data using a large *Pre-trained Language Model* (PLM) guided by task-specific prompts. Even though synthetic datasets have limitations, the authors of Ref. [45] introduced PROGEN, a progressive zero-shot dataset generation framework that incorporates feedback from a task-specific model to refine data generation, utilizing in-context examples to drive the creation of more effective training samples.

ChatGPT [26], in particular ChatGPT 3.5, has been used as an LLM due to its capability of producing high-quality, human-like text in many approaches and contexts. As an example, the authors of Ref. [39] introduced a pipeline for creating a multi-turn dialogue dataset by leveraging ChatGPT to converse with itself, where it alternates between simulating both user and AI assistant responses. This method, known as *self-chat*, constitutes the core of the data collection pipeline. The generated dataset can be used for training and assessing chat models in multi-turn conversations. Furthermore, the study illustrated in [43] focuses on the generation of training data for topic classification in a zero-shot setting, where no labeled data are available.

In [42], a three-stage pipeline has been proposed for the creation of PERSONACHATGEN, a small-scale machine-generated dataset. To ensure the production of high-quality outputs, the authors established a hierarchical taxonomy of user profile categories, informed by social profiling frameworks. Subsequently, they designed prompts to guide the generation process. Given GPT-3's potential to produce offensive or socially biased content, filtering mechanisms have been incorporated into the pipeline. The study shows that GPT-3 is capable of generating personalized dialogue datasets, as confirmed through both manual and automated evaluations.

Customer care interaction management is one of the most interesting contexts in which recognizing emotions plays a key role. In specific situations, it is crucial to detect the emotion expressed by the user quickly and also discern the language level used by the client to design effective automatic interaction systems or build artificial systems that can help a human agent appropriately interact with the client.

We have decided to use ChatGPT to generate examples of interactions between people in specific contexts while expressing specific emotions we want to focus on and using a predetermined language complexity. The emotions that we considered were *joy, sadness, anger, fear, surprise, disgust*. Moreover, we chose three levels of language complexity, namely the $A2$, $B2$, and $C2$ levels of the *Common European Framework of Reference for Languages* (CEFR).

In particular, in this paper, we describe the context of a customer care call center for a phone company. Provided the aforementioned directive, each dialogue has been completely invented by ChatGPT. For each turn of the dialogue, ChatGPT associated an attitude label to both the User (e.g., *angry*, *exasperated*, *fearful*, *hesitant*, and so on) and the Agent (e.g., *sympathetic*, *apologetic*, *confirming*, *reassuring*, and so on).

This is an important characteristic, since thanks to the dialogue structure generated by ChatGPT as well as the contextual semantic content and the different dialogue acts arising during the interaction, it is possible to design a system that analyzes and labels them according to specific measures, and that can be capable, over time, of providing effective examples of interactions, like those occurring in the real world. Furthermore, the system can mine the arising sequential patterns of interaction in order to deduce frequent patterns that can be useful for further training an artificial agent to better interact with human users, showing a particular emotion and using a given language complexity. The extracted patterns and sentences from the dialogues can also be useful as a support for human agents to better interact with users, suggesting to them the tone to use or sample sentences to look for. The whole dialogue was then stored in a repository, and it was also analyzed.

The sequence of attitudes provided by ChatGPT during the generation of the dialogues can be interpreted as part of an implicit *emotional protocol* or framework in conflict resolution and customer service. We have extracted for each dialogue the sequence of attitudes. Each sequence is then stored and, once all the sequences of dialogues belonging to the same class of expressed emotion are collected, they are analyzed by algorithm *Sequential Pattern Mining* (SPM). The sequences were analyzed, and their analysis led to interaction graphs that are suited for both the context and the emotion expressed by the user.

As regards the usage of generative AI and AI-assisted technologies, during the preparation of this work, the authors used ChatGPT 3.5 in order to create the dialogues in the examples. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

The remainder of this paper is organized as follows: the next section illustrates the methodology that we used to build the dataset and the measures that we used for the readability of the dialogues and the mining of the sequential attitude patterns; then the results are reported; finally, conclusions and future work are illustrated.

This paper is a significant extension of the paper [33]. With respect to the previous paper, here we: (*i*) improve the generative AI model; (*ii*) introduce a novel SPM algorithm to mine the derived results; (*iii*) conduct a broad and comprehensive experimental analysis on the overall quality of the methodology.

2. Modelling and generating an emotion-aware database via LLM methodologies

The aim of the work is to exploit ChatGPT to build a set of dialogues, provided a context, a level of language complexity used by a human user, and a specific emotion expressed during the dialogue by the user.

The generated dialogues were examined in order to validate the correctness of the generated dialogues in accordance with the context provided, the desired text complexity, and the consistency with the required emotions to be expressed during the conversation to filter out the dialogues that were not sufficiently suitable to be analyzed and stored in the repository. This phase will be removed once the dataset contains a sufficient number of dialogues.

We have used for the experiments the OpenAI ChatGPT-3.5 Generative AI; the context was a client interacting with a customer care of a hypothetical phone company and expressing one of the six basic emotions of Ekman [30], i.e., *anger*, *sadness*, *surprise*, *disgust*, *fear*. Analogously to what was reported in [32], for each dialogue we asked ChatGPT to generate conversations where the word directly expressing the emotion did not appear. As an example, if the customer is described in the prompt as being *surprised*, ChatGPT can be required not to use the words *surprise*, *surprised*, and so on in the dialogue. These latter kinds of dialogues were labeled as *Without Words* (WW) analogously with the procedure illustrated in [32].

The number of emotions used in the dialogue is six because we want to realize an analytics and retrieval system suitable for supporting the interaction of human or artificial agents with users in a specific context. The approach can, in principle, also be used as the base of a reinforcement learning system, provided with a success score for the different interactions. Two aspects were in particular taken into account: the emotion expressed by the user and the language complexity used.

Once the questions and the possible answers are stored in the dialogues repository, they can be retrieved and used as a support for a human agent or an artificial agent to properly behave according to the context of the conversation, the emotion of the user, and their language level.

The goal is to create data that increasingly reflect real situations to train interactive systems that can recognize emotions based on context and the language used by the customer, so that a specific sample of adequate answers both for tackling the expressed emotion and the provided language level can be of any help in improving the effectiveness of the interaction and ensuring a smooth resolution of possible problems that could arise in customer care relationships.

For each one of the six emotions and for each CEFR language level, we decided to generate a set of dialogues. The prompt given to ChatGPT was to generate a short dialogue, of about five turns, between a customer and an agent in a customer care situation in which a specific emotion emerges from the customer's side, and the agent has to react to satisfy the customer's requests in a smooth manner, trying to apply the most adequate *dialogue register*.

All the generated dialogues have been checked through an interactive process to ensure that the language used is appropriate, as well as the general quality of interaction. Furthermore, a manual labeling of the gender used in the generated dialogue has been done for future development of the system. The dialogues are then saved to constitute a repository that can be used to train a system that somehow acquires an interaction habit that could be effective. To reach this goal, we have exploited the explicit emotions and reactions provided as labels in the generated dialogue to mine the sequences and find interaction patterns to be used by the interaction system. The interaction patterns were mined by using the algorithm *seq2pat* [20,22,23], which also offers other interesting characteristics for the future development of the system.

The whole process is illustrated in Fig. 1. We detail in the following the information flow and the task description of each module in the architecture.

- The *ChatGPT* interactive dialogue system (1) is provided with a prompt that includes three key elements: the *context* of the dialogue to generate, i.e., for example a customer care call center of a phone company, a bank, a delivery company, and so on. The context also includes the average number of turns that characterize the dialogue, determining in this manner the length of the dialogue itself. The second key element is the *emotion* expressed by the user. The third key characteristic is the *language level* expressed by the customer and explicitly declared in the prompt as one of the standard CEFR levels (from $A1$ to $C2$).
- The answer (2) of ChatGPT is therefore a set of dialogues reflecting the expected result of the prompt. Since the answer can consist of a number of different dialogues, the text is split in order to obtain one text file for each dialogue (3). Each text file is then labeled with

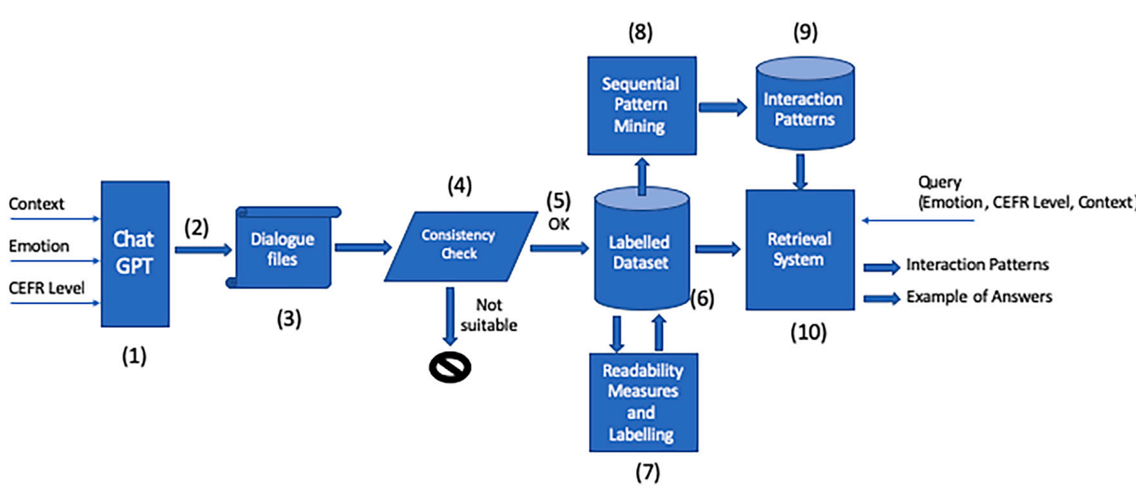

Fig. 1. The dataset creation and retrieval process.

the emotion and the CEFR language level for future reference and retrieval.

- Each dialogue is checked for the *consistency* (4) with the prompt in order to ensure that the generated example has a satisfactory quality.
- The dialogues that pass the consistency check (5) are finally stored in a *repository* (6) for further analysis. In particular, for each generated dialogue, two files are stored. The first one is simply the dialogue with the turns, labeled with the emotion and the language level. The second file is the extracted pattern that associates each turn with an emotion. The emotion label for each dialogue turn is provided by ChatGPT. A parser extracts the patterns and stores them in order to be manipulated by the algorithm SPM.
- The dialogues are analyzed for automated readability measurements and labeling (7) in order to identify the best measure that can be useful for identifying the complexity of the language used for the interaction.
- In parallel, the dialogue patterns of interaction are analyzed by algorithm SPM (8) with the aim of extracting frequent interaction patterns associated with a specific emotion that can be used in the future as a guide for the interaction with a real user by an artificial agent or even by a human agent.
- The patterns extracted by the mining algorithm are stored in an "interaction pattern repository" for further retrieval (9).
- The retrieval module (10) receives as input a query that identifies the context, the CEFR language level and the emotion expressed by the user and provides as a result the most frequent interaction patterns as well as a set of examples that can be used by an artificial agent or a human agent as an example or suggestion of behavior during the real dialogue to provide an effective interaction with a possibly smooth and satisfactory solution to a problem for the user.

2.1. The common European framework of reference for languages (CEFR)

The CEFR subdivides language proficiency into six levels, ranging from $A1$ to $C2$. These can be grouped into three broader categories: *Basic User* (A category), *Independent User* (B category), and *Proficient User* (C category). This framework can be further adapted to specific contexts. Each level is defined by clear descriptions of what a learner can do (called *can-do descriptors*).

The CEFR goes beyond just levels: it describes the process of learning a new language, focusing on different skills and sub-skills. These descriptions are universal and apply to any language, outlining a clear path of progression for each skill, with the six levels ($A1$, $A2$–$B1$, $B2$–$C1$, $C2$) reflecting increasing mastery.

2.2. Manual analysis

We have analyzed and checked the generated dialogues, analogously to what is illustrated in [32], but adapting them to the required tasks and the tackled domain. The manual analysis has been conducted to be sufficiently sure of including in the dataset only the dialogues that adhere to the required characteristics.

In particular, being $N_{dialogues}$ the number of dialogues in the dataset, we have considered the coherence between the required emotion expressed by the user and the request to ChatGPT, naming it *emotional coherence*. Its value is binary (yes/no) and a general parameter that gives an idea of the emotional coherence is C_E, computed as:

$$C_E = \frac{N_{ey}}{N_{dialogues}} \cdot 100 \tag{1}$$

where N_{ey} is the number of times the emotion of the user has been judged as coherent with the request.

Then, we considered the consistency between the required language complexity level of the user and the generated dialogue, naming it *language complexity coherence*. Its value is binary (yes/no); we defined a general measurement that provides an idea of the language complexity coherence as C_L, computed as:

$$C_L = \frac{N_{yc}}{N_{dialogues}} \cdot 100 \tag{2}$$

where N_{yc} is the number of times the language used by the human character has been judged as coherent with the complexity level requested.

Moreover, similarly to what is presented in [32], we computed the *gender distribution* ($GD_{m,f,n}$), counting how many times the genders *Feminine* (F), *Masculine* (M) and *Neutral* (N), occurred in the dialogues. Therefore, we computed their frequency distribution.

$$GD_i = \frac{N_i}{N_{dialogues}} \cdot 100 \tag{3}$$

where $i \in \{m, f, n\}$, N_i is the number of occurrences of dialogues involving the gender i. This labeling has been done for further development of the system.

Finally, the general *Quality of Interaction (QoI)* of the generated dialogues has been estimated, attributing one among three possible values to each dialogue: *Sufficient* (S), *Adequate* (A), *Fail* (F). *Sufficient* indicates that the language appears natural and effectively conveys the intended emotion and language complexity. *Adequate* means that the language lacks complete naturalness but remains acceptable: it may include words that can represent other emotions or words not aligned with the requested language complexity; however, overall, it fulfills its purpose. *Fail* means confusion or unusual language that fails to accurately reflect the specific emotion and the requested language complexity. Additionally, we calculated the relative distribution of these scores across the generated dialogues dataset. The dialogues whose *QoI* value is F are discarded from the analysis and are not included in the repository.

2.3. Readability tests

Readability assessment involves determining how easily a text can be understood and processed. It usually also associates a suitable reading level with a text. This can be useful both for different levels of readers and second-language learners [16]. Different measures have been proposed in the literature with the aim of assessing the readability or, conversely, the difficulty level of a text.

To make an analysis of the generated dialogues, we have exploited the *Automatic Readability Tool for English* (ARTE) [25,27], which allows for automatically computing several readability metrics for texts. The tool has been chosen because it provides free and user-friendly access to the computation of the different metrics on the texts that have been provided as input to the system. In particular, the generated dialogues have been tested through readability measurements to assess their ease of understanding and to verify that the difference in language complexity requested has been maintained in the dialogues generated by ChatGPT.

In the following, for completeness, we report the measures that have been used in our analysis and that are available in the ARTE Tool. They have been exploited to determine the readability of what both the user and the agent said in the generated dialogues, as well as to evaluate the text readability difficulty of the dialogues as a whole. The parameter values reported in the following subsections correspond to those documented in the *ARTE Index Description Sheet*, which is accessible on the ARTE website [27].

2.3.1. Flesch reading ease

This measure provides a higher score for a more difficult text and a lower score for a text that is easier to read. It takes into account linguistic features such as the total number of syllables ($numSyllabes$), the total number of words ($numWords$), and the total number of sentences

($numSentences$) in a text.

$$FRE = 206835 - 84.6 \cdot \left(\frac{numSyllabes}{numWords}\right) + \\ - 1.015 \cdot \left(\frac{numWords}{numSentences}\right) \quad (4)$$

2.3.2. Flesch-Kincaid grade level

The *Flesch-Kincaid Grade Level* is a well-known readability formula, derived from the Flesch Reading Ease, that estimates the reading grade level required to understand a text. This assessment considers the average sentence length and the complexity of words. The resulting scores correspond to U.S. grade levels. Initially designed for educational purposes, it is now applied in a much broader range of contexts.

$$FKGL = 0.39 \cdot \left(\frac{numWords}{numSentences}\right) + \\ + 11.8 \cdot \left(\frac{numSyllabes}{numWords}\right) - 15.59 \quad (5)$$

2.3.3. Automated reading index

Automated Readability Index (ARI) is another well-known readability test for English texts [18]. The ARI score utilizes the characters per word instead of syllables per word. This allows a quicker and easier computing process, since counting characters is easier and more accurate than computing the number of syllables in words. The formula of the ARI score is the following:

$$ARI = 0.5 \cdot \left(\frac{numWords}{numSentences}\right) + \\ + 4.71 \cdot \left(\frac{numCharacters}{numWords}\right) - 21.43 \quad (6)$$

2.3.4. New Dale-Chall readability formula

The *Dale–Chall Readability Formula* is a readability test that quantifies readers' difficulty when reading a text. Its new version makes use of a list of three thousand words that fourth-grade American students can typically understand. The words in the text that are not included in this list are considered difficult to understand [17]. The readability score formula is the following:

$$NDC = 0.0496 \cdot \left(\frac{numWords}{numSentences}\right) + \\ + 0.1579 \cdot \left(\frac{numDifficultWords}{numWords}\right) \quad (7)$$

A constant of 3.6365 is added if the fraction dw of difficult words, defined as:

$$dw = \frac{numDifficultWords}{numWords} \quad (8)$$

in the text is greater than 5 %, where *numDifficultWords* is the number of words not included in a list of about 3000 common words, and *numWords* is the number of words considered in the text to oxamine.

2.3.5. CAREC - crowdsourced algorithm for reading comprehension

This model takes into account different parameters. In particular it is the sum of:

- a constant, set to 1.811;
- +0.022·Average age of acquisition (Kuperman) for all content words;
- +0.746·Average bigram range score (COCA) for all words;
- −0.742· Average trigram proportion score (BNC-written) for all words;
- −0.001· Average imageability score (MRC) for all content words;
- +0.0000625· Average frequency score (Brown) for all words;
- −0.699· Average type token ratio of lemma trigrams for all trigrams;
- −0.111· Proportion of lemma types that occur in the next paragraph for all paragraphs;
- −2.067· Number of temporal connectives divided by number of words in text;
- +0.035· Proportion of noun lemma types that occur in the next paragraph for all paragraphs;
- +0.002· Number of content word lemma types;
- −0.08· Positive adjective scores derived from 4 different corpora;
- +0.047· Average standard deviation of word length for all words;
- −0.395· Average character entropy for all characters.

2.3.6. CAREC_M - crowdsourced algorithm for reading comprehension modified

The CAREC_M variant is the same as the CAREC model, except that the factor +0.002· (Number of content word lemma types) is replaced with the factor +0.2· (Number of content word lemma types divided by number of content words).

2.3.7. Coh-metrix L2 readability index (Approximated)

This model takes into account different parameters. In particular, it is the sum of:

- a constant, set to −43.142;
- +0.642· Number of sentences in text;
- +12.671· Average frequency score (SUBTLEXus) for all content words logged;
- +29.619· Proportion of noun and pronoun lemma types that occur in the next two sentences for all sentences.

2.3.8. Sentence-BERT

The ARTE Tool also provides a deep-learning-based readability model [27], which is based on *Sentence BERT* (SBERT) [19]. Of course, the process that leads to the outcome of the model cannot be translated into explainable and human-understandable linguistic features that characterize the text.

2.4. Sequential pattern mining

To identify and address different potential interaction patterns that a generative AI model like ChatGPT can give rise to, we employed SPM as an analytical approach [21,48]. Pattern mining involves identifying meaningful, valuable, and unforeseen patterns within datasets. SPM is widely applied across different real-world domains, including purchasing patterns, communication sequences, and online user interactions [47]. It can be extremely useful for predicting customer intent. A pattern is a sub-sequence that appears in at least one sequence within the database while preserving the original item order. The frequency of a pattern is determined by the number of sequences that contain it. SPM aims to identify patterns that appear more frequently than a specified threshold.

In particular, SPM involves the identification of subsequences that can be remarkable within a collection of sequences. Each sequence is an ordered collection of items that can also have associated attributes. The relevance of a subsequence is assessed based on different criteria, including, for example, its frequency of occurrence, length, and associated profitability [21].

Moreover, the goal is usually not merely to find all frequent patterns in a sequence database: one of the most interesting possibilities of SPM is to identify patterns that are not only frequent but also relevant to the specific properties of the particular application [24]. Two recent surveys on SPM and parallel SPM are provided by Ref. [21] and [48] respectively.

More specifically, for the task tackled in this paper, analogously to what is reported in [21], let us consider the following formalization. Let I be a set of pairs $(a_i, e_j$, where a_i denotes the actor involved in the dialogue, i.e., in our case, $a_i \in \{Agent, Client\}$, $i \in \{1, 2\}$ and e_j is the *emotional expression* or *attitude* expressed by the actor during a turn in a conversation; in our case $e_j \in \{$ *angry, calm, frustrated, confirming, reassuring, supportive,* … $\}$. The cardinality of I, i.e., the number of items in I, is not predetermined, since it depends on the different emotional expressions arising during each turn of a newly generated dialogue. An itemset X is in general, a set of items such that $X \subseteq I$. Let $|X|$ denote the cardinality of X. An itemset X is said to be of length k or a *k-itemset* if it contains k items (i.e., $|X| = k$). In our specific case, we consider itemsets of length one, i.e., $|X| = 1$. For example, consider the set of symbols:

$I = \{(Client, angry), (Agent, calm), (Client, frustrated), (Agent, understanding), (Client, irritated) (Agent, reassuring), (Client, agitated), (Agent, confirming), (Client, frustrated), (Agent, sympathetic), (Client, grudgingly), (Agent, apologetic)\}$ represent the possible pairs (a_i, e_i) during an emotional chain of interactions.

A *sequence* in our case is an ordered list of itemsets $s = \langle X_1, X_2, \dots, X_n \rangle$ such that $X_k \subseteq I (1 \leq k \leq n)$. The order is determined by the temporal sequence of the chain of interactions during a dialogue.

A sequence $s_a = \langle A_1, A_2, \dots, A_n \rangle$ is said to be of length k or a *k-sequence* if it contains k items X_i. Specifically, k is the number of turns in a dialogue.

A list of sequences $[s_1, s_2, \dots, s_p]$ having sequence identifiers $1, 2, \dots, p$ constitutes a sequence database. Each sequence s_i is a chain of interactions extracted from a dialogue.

A sequence $s_a = \langle A_1, A_2, \dots, A_n \rangle$ is said to be contained in another sequence $s_b = \langle B_1, B_2, \dots, B_m \rangle$ if and only if there exist integers $1 \leq i_1 < i_2 < \dots < i_n \leq m$ such that $A_1 \subseteq B_{i1}, A_2 \subseteq B_{i2}, \dots, A_n \subseteq B_{in}$. If a sequence s_a is contained in a sequence s_b, s_a is said to be a subsequence of s_b. The goal is to identify meaningful subsequences within a sequence database, finding sequential relationships between items that can be of any relevance or significance for the task to be managed.

Among the different SPM algorithms available in the literature, we have focused our attention on the use of a new and interesting algorithm designed for constraint-based SPM. This task aims at identifying frequent patterns in a sequential database of items while adhering to constraints on item attributes. The algorithm presented in [20,22,23] introduces novel techniques for constraint-based SPM that rely on a *Multi-valued Decision Diagram* (MDD) representation of the database. Specifically, the representation can handle multiple item attributes and various constraint types. Moreover, *Seq2Pat* is the first Python-based Constraint-based SPM library designed to support a variety of anti-monotone and non-monotone constraint types [49]. We selected this approach due to its demonstrated competitiveness and superiority over similar methods in terms of scalability and efficiency.

In this work, we only use its basic features, without introducing any constraints or costs. However, the use of this algorithm triggers the possibility of future development and improvement of the overall system.

2.5. Biases in ChatGPT-generated dialogues

Even though ChatGPT 3.5 makes it feasible to quickly and effectively create dialogues that are aware of emotions, it is crucial to recognize the limitations and risks of depending on LLMs that have been trained on vast, online datasets. An important issue is that the training data of these models contain inherent biases that could influence the generated content in terms of linguistic style and emotional tone. These biases can manifest in various forms, such as:

- *Emotional Stereotyping*: The way emotions like anger or sadness are expressed may be influenced by stereotypical portrayals seen in training data, potentially leading to exaggerated or flattened emotional expressions that do not reflect real-life nuance.
- *Language Style Bias*: The generated language may favor certain tones or word choices over others based on the dominant patterns in its training corpus. For instance, responses might overuse formal politeness markers or default to overly neutral tones in emotionally charged scenarios.
- *Demographic Blind Spots*: Since the system does not infer or model age, socioeconomic status, or regional dialects, the generated dialogues may unintentionally neglect diversity in communication styles across different user groups.

3. Experimental results

The experimental results presented in this Section concern two key aspects: (*a*) the generation and the analysis of dialogues, expressed at a designated CEFR language proficiency level, in which a specific emotion is also conveyed, and (*b*) the induction of possible interaction graphs, which we named *attitude interaction graphs*. These graphs are derived from the sequences of interactions arising from the dialogues generated by ChatGPT and exploiting an algorithm SPM provided by the literature [22,24]. The two aspects are illustrated in the following Section 3.1 and Section 3.2, respectively.

3.1. Dialogue analysis

In our experimental analysis, we prompted ChatGPT to generate a series of short dialogues based on user input i.e., the *context*, the *emotion*, and the *CEFR Level* as input in module (1) of the system (see Fig. 1). In the following, we present a selection of these dialogues, categorized by a specific emotion and a determined linguistic proficiency level. In the dialogues where ChatGPT referenced a specific commercial brand or phone model, we replaced the actual name with the placeholder "*Brand Model*" to avoid unintended endorsements or criticisms. The sequences of attitude labels arising from the dialogues were extracted and stored in a repository to facilitate the subsequent identification and analysis of recurring interaction patterns. Each sequence is also associated with the corresponding desired emotion and the CEFR proficiency level requested in the prompt.

In the following, we show some examples of dialogues reflecting the emotions of *anger* and *surprise*. For each emotion, we present three dialogues corresponding to different CEFR levels, namely $A2$ (basic), $B2$ (independent) and $C2$ (proficient). Subsequently, we also show the results obtained by analyzing the language used in the different clusters of dialogues generated by ChatGPT using the ARTE tool [27]. This analysis aims to identify the most effective metric or combination of metrics for automating in the future the assessment and validation of text complexity. The presented results serve as a proof of concept, demonstrating the applicability of the proposed methodology to LLM-generated dialogues.

3.1.1. Anger and $A2$ CEFR language level

In the dialogue illustrated in Fig. 2, the client expresses anger using language appropriate for the $A2$ CEFR level. The agent responds calmly and offers assistance to resolve the issue. Between parentheses it is shown the emotion expressed by the subject that has been assigned by ChatGPT. In the dialogue, there is the use of negative expressions, e.g., *I'm not happy at all!*, which directly expresses strong dissatisfaction, or *It's really annoying!*, which denotes a strong negative emotion. Furthermore, there is the use of strong and repetitive language, such as *I've tried that a hundred times!*, which shows anger and impatience. Additionally, the client's responses are often brief and sharp, e.g., *It won't turn on!*. Most sentences are short and straightforward, avoiding complex grammar. Moreover, words like *phone*, *working*, *annoying*, *turn on*, and *charging* are everyday words familiar to $A2$ learners. Furthermore, the dialogue shows a use of direct, simple questions and answers. No complex conditional structures or advanced tenses are used.

From the above interaction, it is moreover inferred that the chain of interaction is illustrated in Fig. 3.

3.1.2. Anger and $B2$ CEFR language level

In the dialogue reported in Fig. 4, the client expresses anger using language appropriate for the $B2$ CEFR level. The agent responds with attentiveness and offers assistance to address the issue. Compared to the previous dialogue, this version demonstrates anger in a more controlled, articulated, and sophisticated manner. The client's frustration is conveyed through word choice, tone, and more formal sentence structures,

Client (*angry*): Hello, I'm calling about my phone, and I'm not happy at all!
Agent (*calm*): I'm sorry to hear that. Can you tell me what seems to be the problem?
Client (*frustrated*): Yeah, my phone is not working, and I need it for work. It's really annoying!
Agent (*understanding*): I understand. Let's try to figure this out. Can you tell me what specifically is wrong with your phone?
Client (*irritated*): It won't turn on! I've tried everything, and it's just dead.
Agent (*reassuring*): I'm sorry you're experiencing this issue. Can you confirm the make and model of your phone, please?
Client (*agitated*): It's a *Brand Model*.
Agent (*confirming*): Thank you. Let's troubleshoot this together. Have you tried charging your phone to see if it responds?
Client (*frustrated*): Yes, I've tried that a hundred times! Nothing works!
Agent (*sympathetic*): I understand your frustration. Let's schedule a technician to assist you. Can I have your preferred time for a visit?
Client (*grudgingly*): Fine. Anytime after 5 p.m. works for me.
Agent (*apologetic*): I apologize for the inconvenience. We'll have someone there after 5 p.m. Thank you for your patience.

Fig. 2. An example of a dialogue in which the client expresses the "anger" emotion and uses language appropriate for the $A2$ CEFR level.

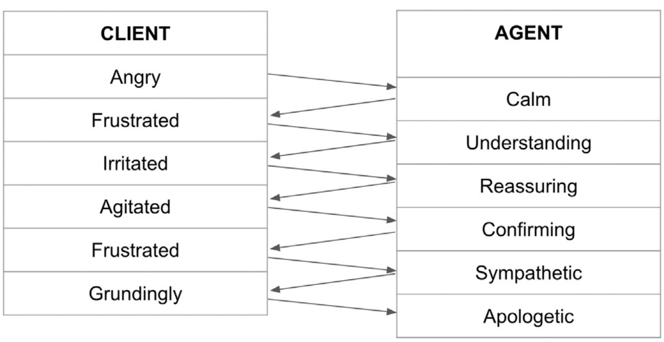

Fig. 3. The extracted chain of interaction during the dialogue of Fig. 2, where the client expresses the "anger" emotion and uses a language appropriate for the $A2$ CEFR level.

Client (*angry*): Good afternoon. I'm calling because I'm quite frustrated with my phone situation.
Agent (*attentive*): I'm sorry to hear that. Can you please explain what's going on with your phone?
Client (*irritated*): Well, my phone refuses to turn on, and I need it urgently for work. It's incredibly inconvenient.
Agent (*understanding*): I understand the urgency. Let's work together to resolve this. Can you provide details on what you've tried so far to fix the issue?
Client (*exasperated*): I've attempted multiple troubleshooting methods, but nothing seems to work. It's just completely unresponsive.
Agent (*sympathetic*): I'm sorry for the frustration this has caused. To assist you further, can you confirm the make and model of your phone?
Client (*aggravated*): It's a *Brand Model*.
Agent (*confirming*): Thank you for providing that information. Let's explore additional troubleshooting steps. Have you tried charging your phone to see if it responds?
Client (*frustrated*): Yes, I've charged it, but there's still no sign of life. I'm at my wit's end.
Agent (*empathetic*): I completely understand your frustration. Let's expedite the resolution process. Would you be available for a technician to visit today?
Client (*reluctantly*): I suppose so. Anytime after 5 p.m. would work for me.
Agent (*apologetic*): I apologize for the inconvenience. We'll arrange for a technician to assist you after 5 p.m. Thank you for your patience.

Fig. 4. An example of dialogue where the client expresses the "anger" emotion and uses a language appropriate for the $B2$ CEFR level.

making it appropriate for a $B2$-level speaker. As an example, the sentence *I'm quite frustrated with my phone situation.* instead of *I'm not happy at all!* conveys irritation but in a more elegant manner. The use of *incredibly* adds emphasis, making the complaint of the client more vigorous. The phrase *I'm at my wit's end.* denotes extreme frustration. The expression *completely unresponsive* sounds more formal and technical than *It's just dead*. Furthermore, there is the use of a polite, indirect question by the Agent. Besides, there are more complex and formal apologies from the Agent. While the $A2$ CEFR version had direct, precise, straightforward, and unfiltered expressions of anger, the $B2$ CEFR version is more refined and indirect, but still expresses the same emotion. Furthermore, we can note the use of more advanced vocabulary, e.g., *Troubleshooting methods* instead of just *tried everything*. There is a presence of more structured and formal sentences and more indirect expressions of emotion. In addition, there is the use of modal verbs for politeness.

From the above interaction, it is moreover inferred that the chain of interaction illustrated in Fig. 5.

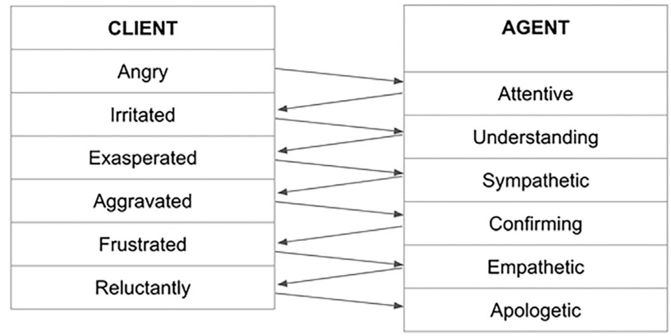

Fig. 5. The extracted chain of interaction during the dialogue of Fig. 4, where the client expresses the "anger" emotion and uses a language appropriate for the $B2$ CEFR level.

Client (*angry*): Good afternoon. I'm contacting you because I'm deeply frustrated by the ongoing issue with my phone.
Agent (*attentive*): I'm sincerely sorry to hear that. Could you please elaborate on the nature of the problem?
Client (*irritated*): Certainly. My phone has become unresponsive, which is highly inconvenient, particularly for my work obligations.
Agent (*understanding*): I understand the urgency. Let's work together to resolve this promptly. Could you detail the troubleshooting steps you've already taken?
Client (*exasperated*): I've exhausted various troubleshooting methods to no avail. Despite multiple attempts, my phone remains entirely non-functional.
Agent (*sympathetic*): I'm truly sorry for the inconvenience this has caused. To proceed, could you confirm the make and model of your phone?
Client (*aggravated*): It's a *Brand Model*.
Agent (*confirming*): Thank you for providing that information. Let's explore further options. Have you attempted to charge your phone to determine if it responds?
Client (*frustrated*): Yes, I've attempted charging it, yet there's been no indication of life. This situation is becoming increasingly infuriating.
Agent (*empathetic*): I completely empathize with your frustration. Let's expedite the resolution process. Are you available for a technician to visit today?
Client (*reluctantly*): Regrettably, yes. Anytime after 5 p.m. would be suitable for me.
Agent (*apologetic*): I apologize for the inconvenience. We'll arrange for a technician to assist you after 5 p.m. Thank you for your understanding.

Fig. 6. An example of dialogue where the client expresses the "anger" emotion and uses a language appropriate for the $C2$ CEFR level.

3.1.3. Anger and C2 CEFR language level

In the dialogue reported in Fig. 6, the client expresses anger using language appropriate for the $C2$ CEFR level. The agent responds attentively and offers assistance to address the issue. Instead of direct expressions of anger, the client elegantly conveys her emotion through formal and elegant phrasing, e.g., in the sentence *I'm deeply frustrated by the ongoing issue with my phone*, the expression *Deeply frustrated* conveys strong emotion, while *ongoing issue* implies some persistent inconvenience, increasing the sense of urgency. The anger emotion is expressed by avoiding informal or exaggerated language; instead, it is managed, refined, or moderated. There is a more sophisticated and formal vocabulary, and also an advanced structure of sentences, e.g., *Could you please elaborate on the nature of the problem?* which is a formal question

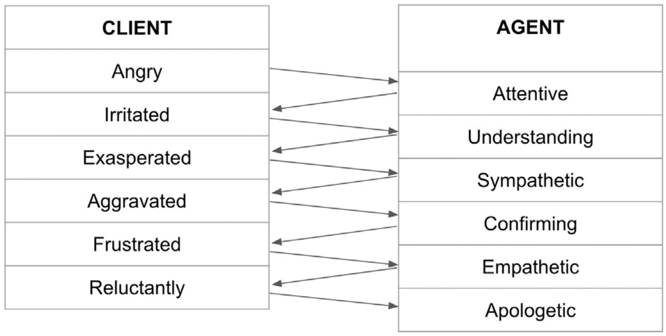

Fig. 7. The extracted chain of interaction during the dialogue of Fig. 6, where the client expresses the "anger" emotion and uses a language appropriate for the $C2$ CEFR level.

Client (*surprised*): Um, hi. I'm calling because I'm, like, really surprised right now. My phone's acting strange.
Agent (*curious*): Oh, I see. Can you tell me what's happening with your phone?
Client (*amazed*): Yeah, it's, like, not turning on! I didn't expect this at all.
Agent (*concerned*): That does sound surprising. Let's see what we can do. Can you confirm the make and model of your phone, please?
Client (*hesitantly*): Um, sure. It's a *Brand Model*.
Agent (*confirming*): Got it. Let's try a few troubleshooting steps. Have you tried charging your phone to see if it responds?
Client (*surprised*): Oh, um, no, I haven't. I'll try that now.
Agent (*encouraging*): Great. Let's see if that helps. Please let me know if anything changes.
Client (*surprised*): Wow, it's turning on now! I didn't think that would work.
Agent (*reassuring*): I'm glad to hear that. Sometimes, it's the simple things that do the trick. If you encounter any more issues, don't hesitate to reach out.

Fig. 8. An example of dialogue where the client expresses the "surprise" emotion and uses a language appropriate for the $A2$ CEFR level.

structure rather than a basic and straightforward *Can you tell me what's wrong?*.

From the above interaction, it is moreover inferred that the chain of interaction is illustrated in Fig. 7.

3.1.4. Surprise and $A2$ CEFR language level

In the dialogue reported in Fig. 8, the client expresses surprise using language appropriate for the $A2$ CEFR level. The agent responds with curiosity and offers assistance to address the unexpected situation. This dialogue expresses surprise in a basic and informal way, using simple sentence structures, casual vocabulary, and interjections. The $A2$ CEFR level language makes the conversation sound natural but not complex, using everyday phrases and expressions commonly found in spoken English. Surprise is often shown through verbal pauses, fillers, and interjections, e.g., *Um, hi.* or *Oh, um, no, I haven't.* These short, spontaneous reactions make the surprise sound genuine and natural. It is present in repetition for emphasis, e.g., in *I'm, like, really surprised right now.* where *like* and *really* add emphasis and hesitation, making the surprise sound spontaneous. The surprise is expressed clearly and directly, without complex grammar or figurative language. The dialogue shows the use of short and direct sentences, a basic vocabulary and an informal tone as well as explicit expressions.

From the above interaction, it is moreover inferred that the chain of interaction is illustrated in Fig. 9.

3.1.5. Surprise and $B2$ CEFR language level

In the dialogue reported in Fig. 10, the client expresses surprise using language appropriate for the $B2$ CEFR level. The agent responds with curiosity and offers assistance to address the unexpected situation. The dialogue expresses surprise in a more structured and refined manner, which can be associated with a $B2$ CEFR level of English. The emotion

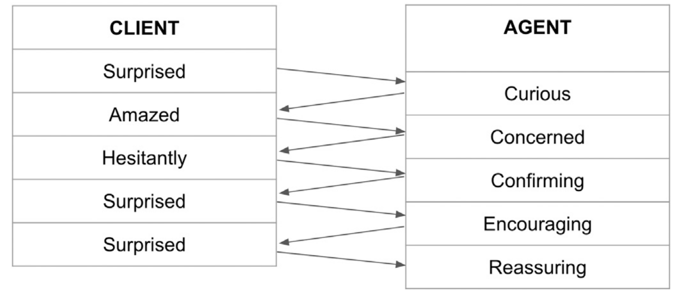

Fig. 9. The extracted chain of interaction during the dialogue of Fig. 8, where the client expresses the "surprise" emotion and uses a language appropriate for the $A2$ CEFR level.

Client (*surprised*): Hello. I'm calling because I'm quite surprised by the current situation with my phone.
Agent (*curious*): I understand. Can you please explain what's happening?
Client (*amazed*): Well, my phone suddenly won't turn on. It's really unexpected.
Agent (*concerned*): That does sound surprising. Let's see if we can resolve it. Could you confirm the make and model of your phone?
Client (*hesitant*): Of course. It's a *Brand Model*.
Agent (*confirming*): Thank you. Let's try a few troubleshooting steps. Have you attempted to charge your phone to see if it responds?
Client (*surprised*): Oh, actually, I haven't tried that yet. Let me give it a go.
Agent (*encouraging*): Alright, let's see if that helps. Please keep me updated on any changes.
Client (*surprised*): Wow, it's turning on now! I didn't expect that to work.
Agent (*reassuring*): I'm glad to hear that. Sometimes, simple solutions can be quite effective. If you encounter any further issues, feel free to contact us again.

Fig. 10. An example of dialogue where the client expresses the "surprise" emotion and uses a language appropriate for the $B2$ CEFR level.

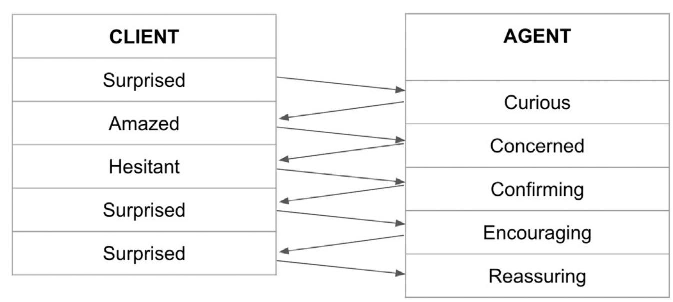

Fig. 11. The extracted chain of interaction during the dialogue of Fig. 10, where the client expresses the "surprise" emotion and uses a language appropriate for the $B2$ CEFR level.

is conveyed through a moderately advanced vocabulary, varied sentence structures, and a balance between formal and natural speech. Unlike the $A2$ CEFR version, which is more direct, this version sounds more professional and fluid, showing an intermediate-advanced level of fluency. E.g., *It's really unexpected* is a simple but more elegant manner to express disbelief compared to *I didn't expect this at all*; the choice of *unexpected* denotes a slightly more advanced vocabulary. Instead of short sentences like *Yeah, it's, like, not turning on!* the client says: *Well, my phone suddenly won't turn on. It's really unexpected*; this makes the conversation sound more fluid and engaging.

From the above interaction, it is moreover inferred that the chain of interaction is illustrated in Fig. 11.

3.1.6. Surprise and $C2$ CEFR language level

In the dialogue reported in Fig. 12, the client expresses surprise using language appropriate for the $C2$ CEFR level. The agent responds with attentiveness and offers assistance to address the unexpected situation.

Client (*surprised*): Good day. I'm contacting you because I find myself quite taken aback by the current state of my phone.
Agent (*inquiring*): I understand. Could you please elaborate on what exactly is happening?
Client (*amazed*): Certainly. My phone has suddenly become unresponsive, which is truly unexpected.
Agent (*concerned*): That does sound surprising. Let's work to resolve it. Could you confirm the make and model of your phone?
Client (*hesitant*): Certainly. It's a *Brand Model*.
Agent (*confirming*): Thank you. Let's proceed with troubleshooting. Have you attempted to charge your phone to ascertain if it responds?
Client (*surprised*): Actually, I haven't considered that. I'll give it a try now.
Agent (*encouraging*): Very well, let's see if that yields any results. Please keep me informed of any developments.
Client (*surprised*): Remarkably, it's powering on now! I hadn't anticipated such a simple solution.
Agent (*reassuring*): I'm glad to hear that. Sometimes, the most straightforward approaches prove to be the most effective. If any further issues arise, do not hesitate to contact us.

Fig. 12. An example of dialogue where the client expresses the "surprise" emotion and uses a language appropriate for the $C2$ CEFR level.

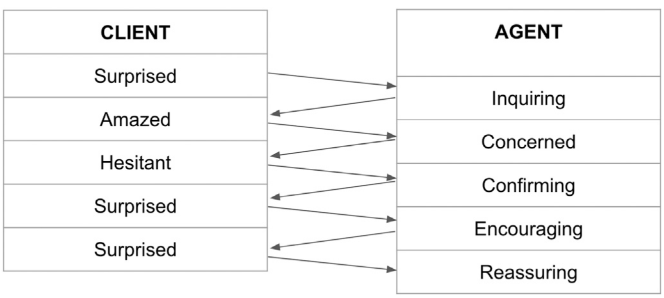

Fig. 13. The extracted chain of interaction during the dialogue of Fig. 12, where the client expresses the "surprise" emotion and uses a language appropriate for the $C2$ CEFR level.

It can be seen that dialogue denotes surprise in a highly sophisticated and articulate manner, which can be associated with a $C2$ CEFR level of English. The language is formal, elegant, and precise; it uses rather complex sentence structures, advanced vocabulary, and refined phrasing to convey disbelief and astonishment in an eloquent manner. For example, instead of basic or casual surprise phrases, this dialogue uses very articulate expressions, like *I find myself quite taken aback by the current state of my phone.*; the expression *Find myself quite taken aback* is a refined manner of expressing shock, more advanced than *I'm really surprised* ($A2$ level) or *I'm quite surprised* ($B2$ level). Furthermore, the phrase *current state of my phone* sounds formal and precise, avoiding basic phrasing like *what's going on with my phone*. Another example can be seen in the advanced vocabulary and the formal tone used, which is more sophisticated than the dialogues expressed in $A2$ and $B2$ levels. Furthermore, the structure of the sentences is more complex.

From the above interaction, it is moreover inferred that the chain of interaction is illustrated in Fig. 13.

3.1.7. Readability results

All dialogues created with ChatGPT were aggregated by the CEFR level required in the prompt, regardless of the emotions the user was asked to manifest.

After setting the CEFR level, we considered separately the sentences attributed to the user and the sentences attributed to the agent. Readability was then measured by CEFR level and by the role played in the dialogue.

Individual turns of the considered interlocutor were then randomly extracted. Each turn is appended to a list of sentences until the total number of words does not reach the maximum number allowed by the ARTE online tool (specifically, 1000 words). Of course, the last turn has not been interrupted. As a consequence, if the number of words in the last turn to be included in the analysis is such that it exceeds the maximum allowed number of words, the turn is discarded and it is not included in the list of sentences to be evaluated.

In the end, a text of approximately 950 words was built and evaluated according to the different readability measures provided by the ARTE tool. This operation was carried out several times for each CEFR level and each interlocutor type (User or Agent). Because the sentences were taken randomly, the text considered for the analysis is varied, and it considers different emotions that ChatGPT was asked to express by the user.

The readability results provided by the ARTE tool were then saved, and finally, an average value was computed. Moreover, we calculated the standard deviation of the measures taken. The results are illustrated in Fig. 14–21.

As a further analysis, we merged all the dialogues concerning a given CEFR level that explicitly used the emotion word; we did the same with those dialogues that did not explicitly use the emotion word. We executed readability tests on these types of files by using the ARTE tool [27].

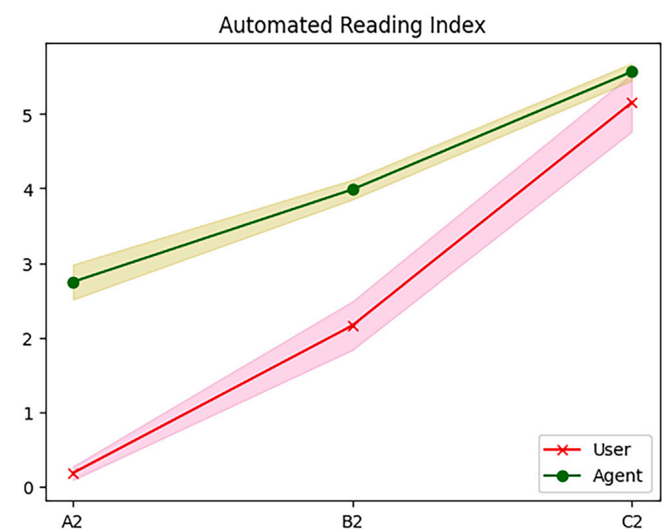

Fig. 14. The ARI average readability results for the $A2$, $B2$, and $C2$ CEFR proficiency levels both for the User and the Agent.

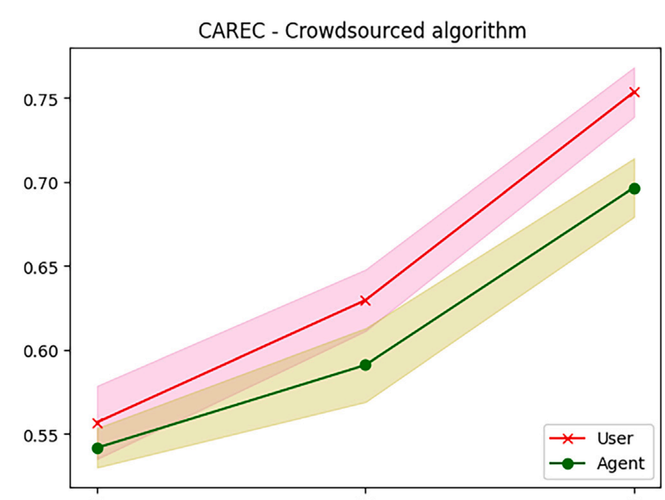

Fig. 15. The CAREC average readability results for the $A2$, $B2$, and $C2$ CEFR proficiency levels both for the User and the Agent.

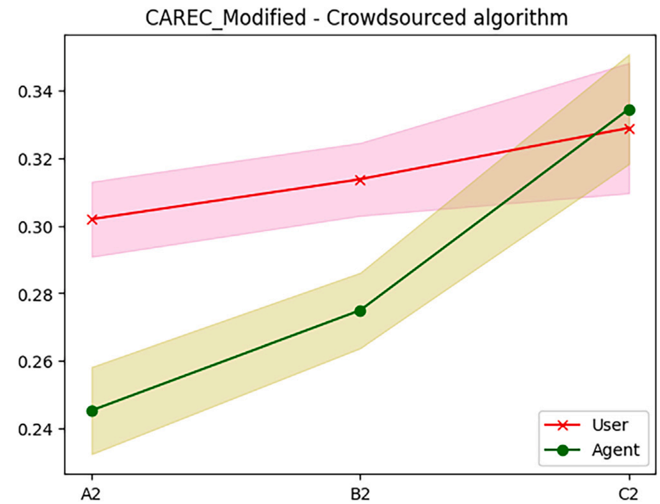

Fig. 16. The CARECM average readability results for the $A2$, $B2$, and $C2$ CEFR proficiency levels both for the User and the Agent.

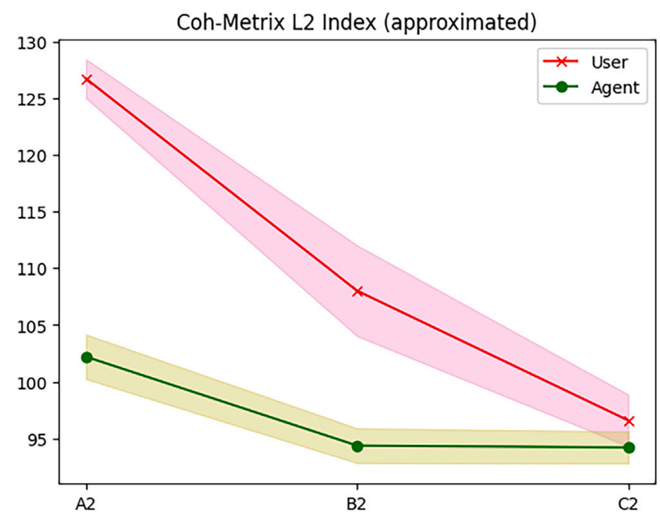

Fig. 17. The CML2 average readability results for the $A2$, $B2$, and $C2$ CEFR proficiency levels both for the User and the Agent.

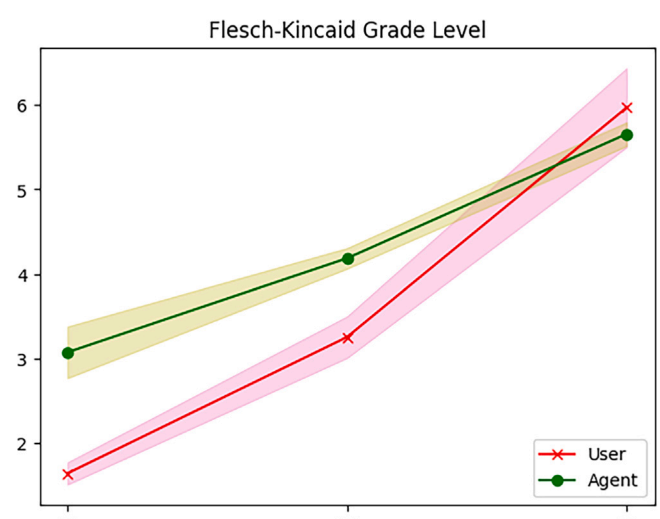

Fig. 18. The FKG average readability results for the $A2$, $B2$, and $C2$ CEFR proficiency levels both for the User and the Agent.

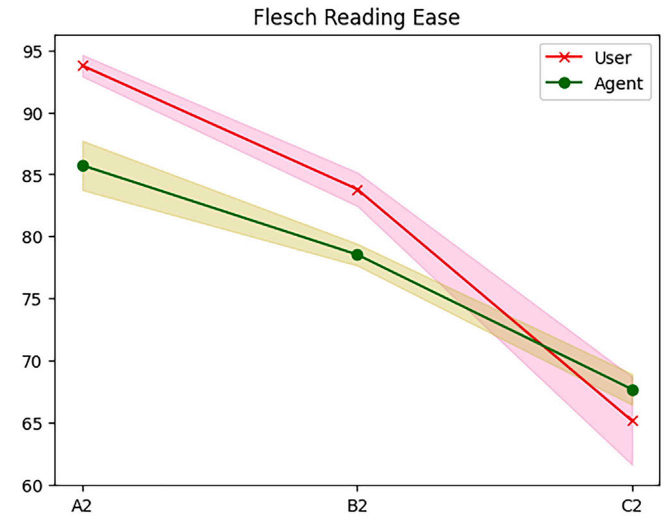

Fig. 19. The FRE average readability results for the $A2$, $B2$, and $C2$ CEFR proficiency levels both for the User and the Agent.

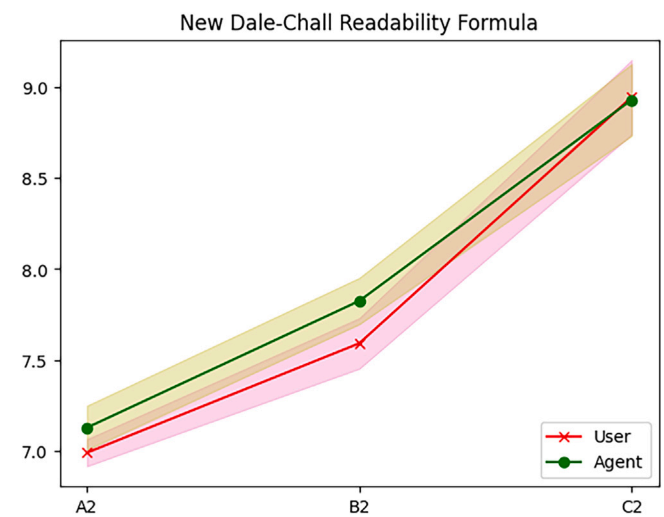

Fig. 20. The NDC average readability results for the $A2$, $B2$, and $C2$ CEFR proficiency levels both for the User and the Agent.

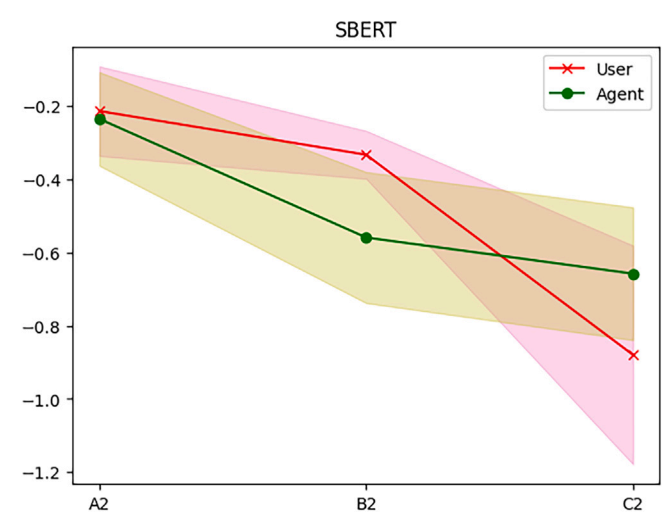

Fig. 21. The SBERT average readability results for the $A2$, $B2$, and $C2$ CEFR proficiency levels both for the User and the Agent.

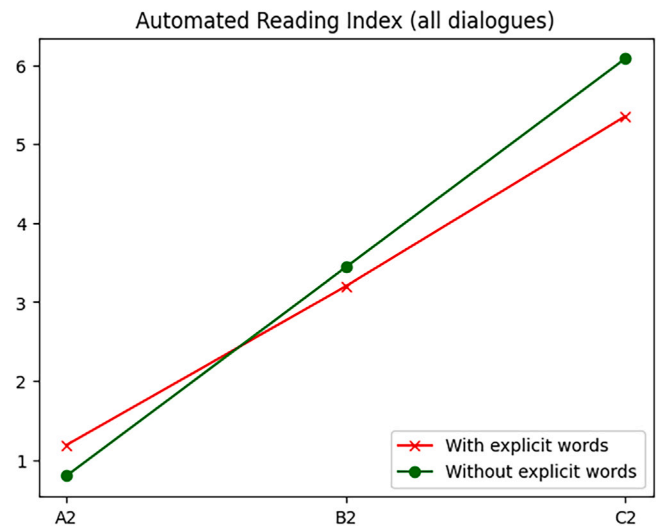

Fig. 22. The ARI readability results for the $A2$, $B2$, and $C2$ CEFR proficiency levels were analyzed considering the entire generated dialogues, both with and without lexical items that explicitly denote emotions.

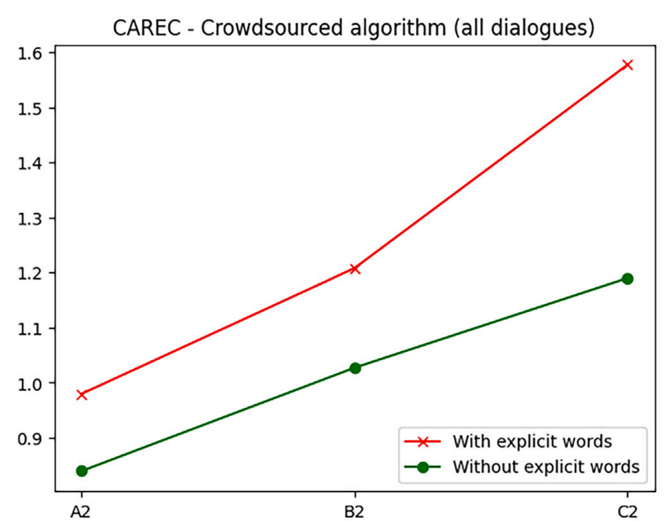

Fig. 23. The CAREC readability results for the $A2$, $B2$, and $C2$ CEFR proficiency levels were analyzed considering the entire generated dialogues, both with and without lexical items that explicitly denote emotions.

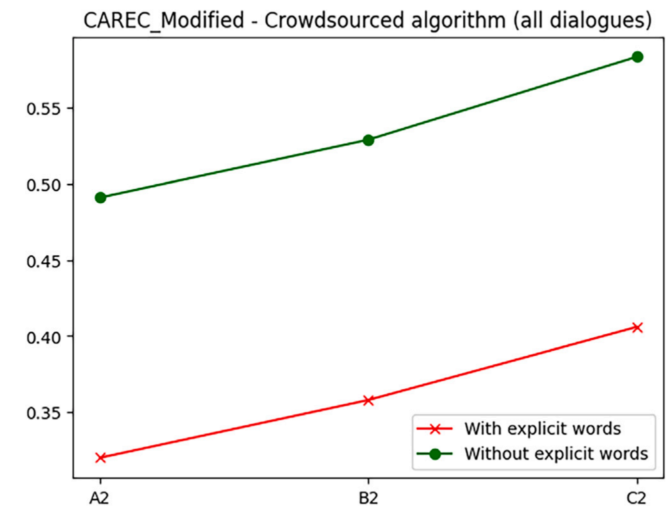

Fig. 24. The CARECM readability results for the $A2$, $B2$, and $C2$ CEFR proficiency levels were analyzed considering the entire generated dialogues, both with and without lexical items that explicitly denote emotions.

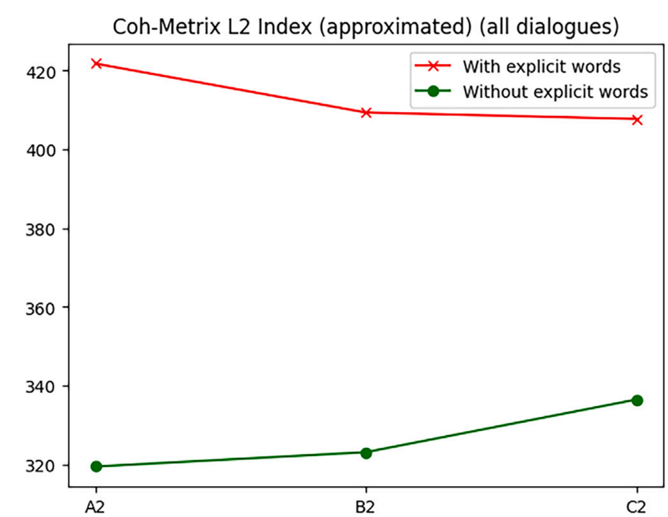

Fig. 25. The CML2 readability results for the $A2$, $B2$, and $C2$ CEFR proficiency levels were analyzed considering the entire generated dialogues, both with and without lexical items that explicitly denote emotions.

The readability results for the $A2$, $B2$, and $C2$ CEFR levels considering the entire generated dialogues with or without words that explicitly refer to an emotion are reported in Fig. 22–29.

3.2. SPM results

Starting from the *chain of interactions* associated with the dialogues generated by ChatGPT, for each one of the six basic emotions, we created a graph, which we named *attitude interaction graph*. The graph has been induced by exploiting the algorithm SPM that we have selected for extracting the most frequent sequential emotional patterns occurring during the generated interaction. In the following, we illustrate the construction process of the attitude interaction graph: provided a given emotion, all the patterns that have been found by the *seq2pat* algorithm [22,24] are collected together independently of the CEFR level used in the dialogue. A pattern is therefore analyzed in order to find the *building blocks* of the attitude interaction graph. Each pattern has an occurrence score associated, then pairs are extracted together with their occurrence score. Once all the pairs have been extracted, the attitude interaction graph is then built. If more identical pairs are present, they are considered once and their occurrence scores are summed.

After the graph structure has been constructed, to each out-link i of each node k is associated a weight w_i so that each score so_j^k of each out-link is summed and the fraction w_j^k is subsequently computed:

$$w_i^k = \frac{so_i^k}{\sum_j so_j^k} \tag{9}$$

Examples of graphs are provided in Figs. 30 and 31.

We used the *seq2pat* algorithm by considering a maximum span of three in the sequences and also a threshold of three as the minimum frequency of the patterns to include. The value of these parameters has been experimentally determined and they can change according to the extent of the dataset, i.e., the number of dialogues and their length.

For each emotion we have also computed a similarity matrix computing the Tanimoto Similarity (Eq. 10) between the sets of patterns

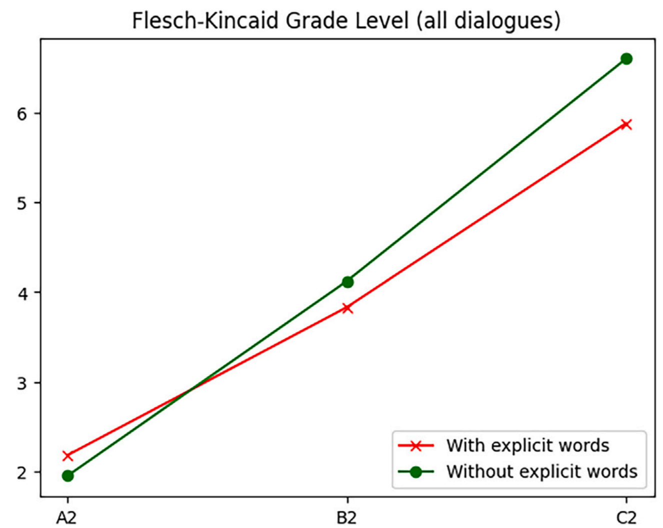

Fig. 26. The Flesch-Kincaid Grade (FKG) readability scores for the $A2$, $B2$, and $C2$ CEFR proficiency levels were analyzed considering the entire generated dialogues, both with and without lexical items that explicitly denote emotions.

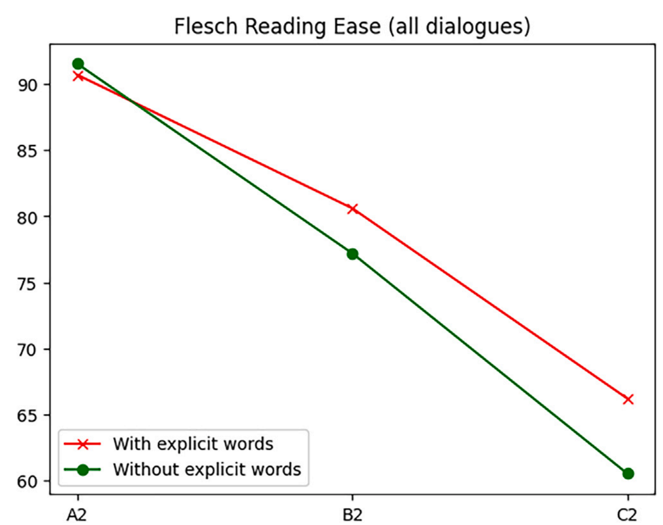

Fig. 27. The FRE readability results for the $A2$, $B2$, and $C2$ CEFR proficiency levels were analyzed considering the entire generated dialogues, both with and without lexical items that explicitly denote emotions.

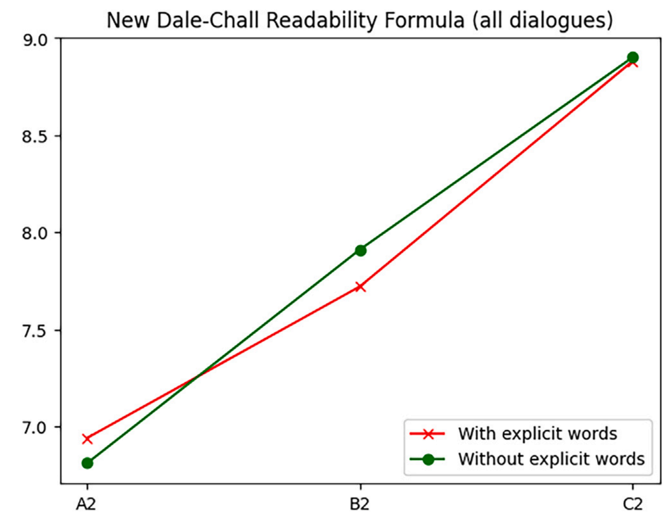

Fig. 28. The NDC readability results for the A2, B2, and $C2$ CEFR proficiency levels were analyzed considering the entire generated dialogues, both with and without lexical items that explicitly denote emotions.

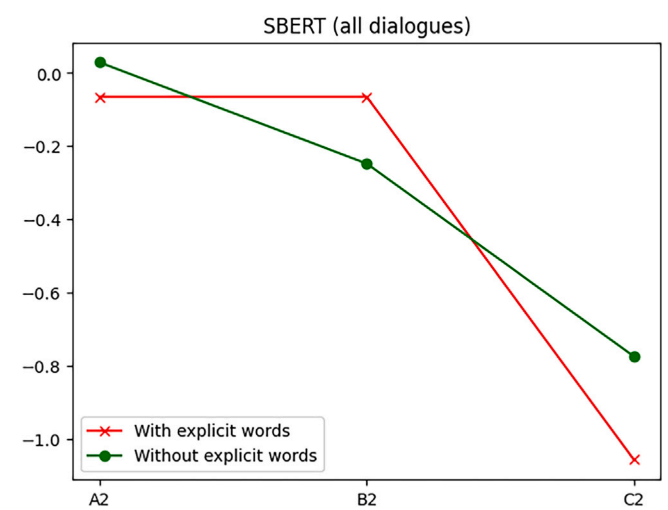

Fig. 29. The SBERT readability results for the $A2$, $B2$, and $C2$ CEFR proficiency levels were analyzed considering the entire generated dialogues, both with and without lexical items that explicitly denote emotions.

set_{L_1}, set_{L_2} found by the *seq2pat* algorithm for each class of language complexity requested from the generative AI, where $L_1, L_2 \in \{A1, A2, B1, B2, C1, C2\}$ i.e., are associated with the considered CEFR language complexity identifier. The obtained matrix, of course, is symmetric.

$$T(set_{L_1}, set_{L_2}) = \frac{\|set_{L_1} \cap set_{L_2}\|}{\|set_{L_1} \cup set_{L1_2}\|} \tag{10}$$

We denote with $A2$ the sets of patterns mined for the dialogues characterized by an $A2$ CEFR language level; with $B2$ the sets of patterns mined for the dialogues characterized by a $B2$ CEFR language level; with $C2$ the sets of patterns mined for the dialogues characterized by a $C2$ CEFR language level. The obtained results are illustrated in Table 1.

As can be seen from the results, for the *joy*, *surprise*, *sadness*, and *disgust* emotions, the extracted sequential patterns are the same for the $B2$ and $C2$ levels. Moreover, for the surprise emotion, the extracted patterns are very similar for all three CEFR levels. Different results have been obtained for the *fear* emotion. On the other hand, for *angry* and *fear* emotions, the sets of patterns associated with the $A2$ and $B2$ CEFR levels have little in common.

In the Table 2–7, we report some results regarding the most frequent patterns extracted from eleven dialogues generated by considering a specific emotion and a given CEFR language level using the *seq2pat* algorithm, identified with a maximum span of three and a minimum occurrence of three.

4. Limitations and research directions

This study presents a *methodology-focused contribution* rather than a *model-specific tool*. As a consequence, we do not directly compare our results to other generative models or emotion-aware datasets. The framework has been designed to be flexible and agnostic to the underlying LLM, allowing it to evolve alongside advances in generative AI. While ChatGPT-3.5 is used in this study due to accessibility constraints at the time of writing, the proposed framework is not constrained to a specific generative model. Because of its modular and flexible

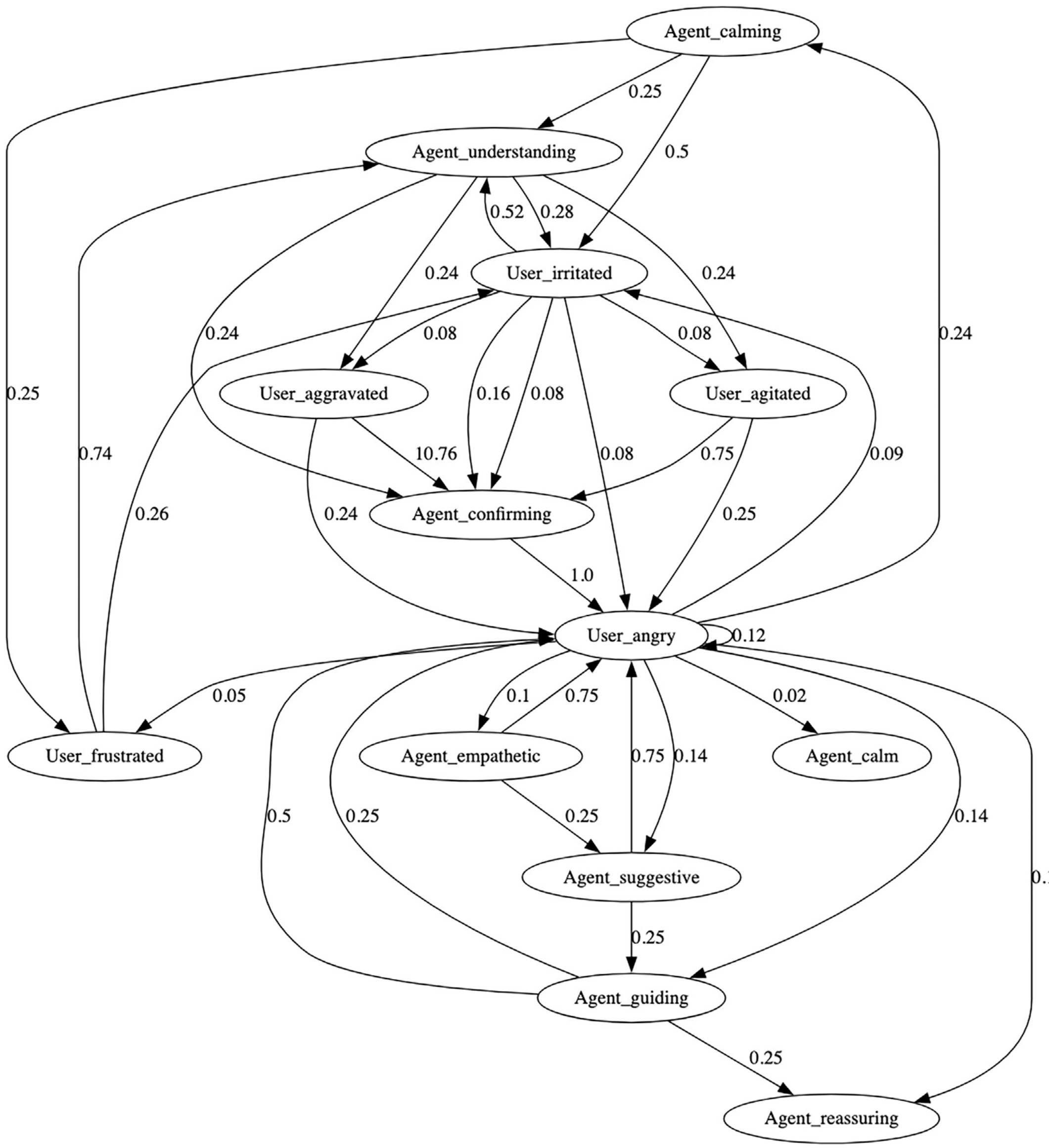

Fig. 30. The interaction graph induced by exploiting the most frequent sequential patterns induced by the *seq2pat* algorithm regarding only the dialogues characterized by the *anger* emotion.

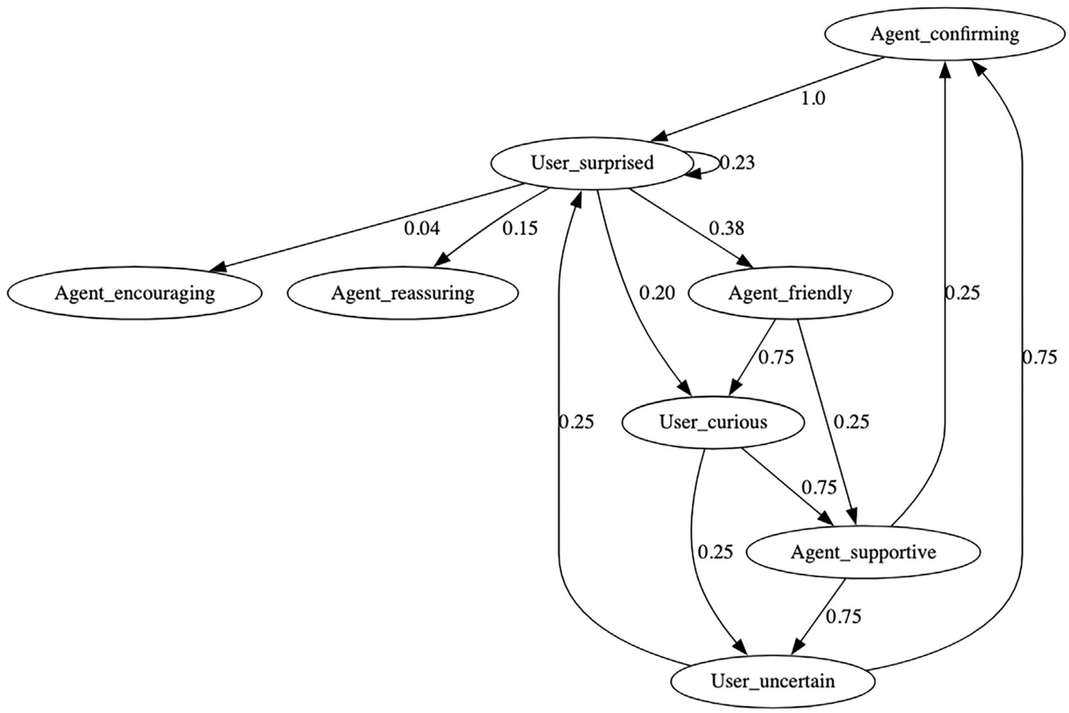

Fig. 31. The interaction graph induced by exploiting the most frequent sequential patterns induced by the *seq2pat* algorithm regarding only the dialogues characterized by the *surprise* emotion.

Table 1
The similarity matrix obtained by computing the Tanimoto measure (Eq. 10) between the sets of patterns found by the *seq2pat* algorithm for each language complexity requested from the generative AI.

Emotion	T(A2,B2)	T(A2,C2)	T(B2,C2)
Anger	0.11	0.11	0.65
Fear	0.09	0.05	0.27
Joy	0.17	0.17	1.0
Surprise	0.95	0.95	1.0
Sadness	0.51	0.51	1.0
Disgust	0.69	0.69	1.0

Table 2
Patterns extracted for “anger” and $A2$ CEFR Language Level.

#occ	Pattern
6	‘Agent (understanding)’, ‘User (irritated)’
6	‘User (angry)’, ‘User (frustrated)’
6	‘User (frustrated)’, ‘Agent (understanding)’
6	‘User (frustrated)’, ‘Agent (understanding)’, ‘User (irritated)’
6	‘User (frustrated)’, ‘User (irritated)’

Table 3
Patterns extracted for “anger” and $B2$ CEFR Language Level.

#occ	Pattern
6	‘User (irritated)’, ‘Agent (understanding)’
6	‘User (angry)’, ‘User (irritated)’
5	‘Agent (calming)’, ‘Agent (understanding)’
5	‘Agent (calming)’, ‘User (irritated)’
5	‘Agent (confirming)’, ‘User (angry)’
5	‘Agent (calming)’, ‘User (irritated)’, ‘Agent (understanding)’
5	‘Agent (understanding)’, ‘Agent (confirming)’
5	‘Agent (understanding)’, ‘User (agitated)’
5	‘Agent (understanding)’, ‘User (agitated)’, ‘Agent (confirming)’
5	‘User (agitated)’, ‘Agent (confirming)’
5	‘User (agitated)’, ‘Agent (confirming)’, ‘User (angry)’
5	‘User (agitated)’, ‘User (angry)’
5	‘User (angry)’, ‘Agent (calming)’
5	‘User (angry)’, ‘Agent (calming)’, ‘User (irritated)’
5	‘User (angry)’, ‘User (angry)’
5	‘User (irritated)’, ‘Agent (understanding)’, ‘User (agitated)’
5	‘User (irritated)’, ‘User (agitated)’

architecture, it can be integrated with newer models like ChatGPT-4 or other cutting-edge LLMs. This adaptability guarantees that our approach can be aligned with improvements in generative AI capabilities. New research directions will include systematic comparisons with publicly available *emotion-annotated corpora* and the inclusion of additional generative models to benchmark performance across different architectures. This will make it possible to have a more thorough evaluation of the performance, including its efficacy in generating emotionally complex dialogues appropriate for the given context. Furthermore, we plan to evaluate the proposed framework using more recent and domain-adapted LLMs to assess *dialogue coherence*, *emotional expression*, and *language complexity improvements*. Incorporating these models will also provide an opportunity to benchmark performance across versions and to further validate the generalizability of the methodology to different model architectures.

Moreover, the analysis regarding the CEFR levels of the generated dialogues is exploratory. Its aim consists of investigating a potential direction for an automatic evaluation of the adherence of the generated dialogues to a required complexity level. Next, the focus should be on expanding this component through the application of *advanced linguistic metrics* and tools. Furthermore, we aim to address how well the generated dialogues align with *real-life customer care interactions* by collaborating with organizations willing to share anonymized dialogue data or by simulating customer care scenarios with human-in-the-loop evaluations.

While the SPM analysis presented in this study is limited to its basic functionality, it is a foundational step toward more advanced dialogue analytics. At this stage, we have neither imposed *constraints* nor explored

Table 4
Patterns extracted for “anger” and $C2$ CEFR Language Level.

#occ	Pattern
6	‘User (aggravated)’, ‘Agent (confirming)’
6	‘User (angry)’, ‘User (irritated)’
6	‘User (irritated)’, ‘Agent (understanding)’
5	‘Agent (calming)’, ‘Agent (understanding)’
5	‘Agent (calming)’, ‘User (irritated)’
5	‘Agent (calming)’, ‘User (irritated)’, ‘Agent (understanding)’
5	‘Agent (confirming)’, ‘User (angry)’
5	‘Agent (understanding)’, ‘Agent (confirming)’
5	‘Agent (understanding)’, ‘User (aggravated)’
5	‘Agent (understanding)’, ‘User (aggravated)’, ‘Agent (confirming)’
5	‘User (aggravated)’, ‘Agent (confirming)’, ‘User (angry)’
5	‘User (aggravated)’, ‘User (angry)’
5	‘User (angry)’, ‘Agent (calming)’
5	‘User (angry)’, ‘Agent (calming)’, ‘User (irritated)’
5	‘User (angry)’, ‘User (angry)’
5	‘User (irritated)’, ‘Agent (understanding)’, ‘User (aggravated)’
5	‘User (irritated)’, ‘User (aggravated)’

Table 5
Patterns extracted for “Surprise” and $A2$ CEFR Language Level.

#occ	Pattern
6	‘Agent (confirming)’, ‘User (surprised)’
6	‘User (surprised)’, ‘User (surprised)’
5	‘Agent (friendly)’, ‘Agent (supportive)’
5	‘Agent (friendly)’, ‘User (curious)’
5	‘Agent (friendly)’, ‘User (curious)’, ‘Agent (supportive)’
5	‘Agent (supportive)’, ‘Agent (confirming)’
5	‘Agent (supportive)’, ‘User (uncertain)’
5	‘Agent (supportive)’, ‘User (uncertain)’, ‘Agent (confirming)’
5	‘User (curious)’, ‘Agent (supportive)’
5	‘User (curious)’, ‘Agent (supportive)’, ‘User (uncertain)’
5	‘User (curious)’, ‘User (uncertain)’
5	‘User (surprised)’, ‘Agent (friendly)’
5	‘User (surprised)’, ‘Agent (friendly)’, ‘User (curious)’
5	‘User (surprised)’, ‘User (curious)’
5	‘User (uncertain)’, ‘Agent (confirming)’
5	‘User (uncertain)’, ‘Agent (confirming)’, ‘User (surprised)’
5	‘User (uncertain)’, ‘User (surprised)’

Table 6
Patterns extracted for “Surprise” and $B2$ CEFR Language Level.

#occ	Pattern
6	‘Agent (confirming)’, ‘User (surprised)’
6	‘User (surprised)’, ‘User (surprised)’
5	‘Agent (friendly)’, ‘Agent (supportive)’
5	‘Agent (friendly)’, ‘User (curious)’
5	‘Agent (friendly)’, ‘User (curious)’, ‘Agent (supportive)’
5	‘Agent (supportive)’, ‘Agent (confirming)’
5	‘Agent (supportive)’, ‘User (uncertain)’
5	‘Agent (supportive)’, ‘User (uncertain)’, ‘Agent (confirming)’
5	‘User (curious)’, ‘Agent (supportive)’
5	‘User (curious)’, ‘Agent (supportive)’, ‘User (uncertain)’
5	‘User (curious)’, ‘User (uncertain)’
5	‘User (surprised)’, ‘Agent (friendly)’
5	‘User (surprised)’, ‘Agent (friendly)’, ‘User (curious)’
5	‘User (surprised)’, ‘User (curious)’
5	‘User (uncertain)’, ‘Agent (confirming)’
5	‘User (uncertain)’, ‘Agent (confirming)’, ‘User (surprised)’
5	‘User (uncertain)’, ‘User (surprised)’

Table 7
Patterns extracted for “Surprise” and $C2$ CEFR Language Level.

#occ	Pattern
6	‘Agent (confirming)’, ‘User (surprised)’
6	‘User (surprised)’, ‘User (surprised)’
5	‘Agent (friendly)’, ‘Agent (supportive)’
5	‘Agent (friendly)’, ‘User (curious)’
5	‘Agent (friendly)’, ‘User (curious)’, ‘Agent (supportive)’
5	‘Agent (supportive)’, ‘Agent (confirming)’
5	‘Agent (supportive)’, ‘User (uncertain)’
5	‘Agent (supportive)’, ‘User (uncertain)’, ‘Agent (confirming)’
5	‘User (curious)’, ‘Agent (supportive)’
5	‘User (curious)’, ‘Agent (supportive)’, ‘User (uncertain)’
5	‘User (curious)’, ‘User (uncertain)’
5	‘User (surprised)’, ‘Agent (friendly)’
5	‘User (surprised)’, ‘Agent (friendly)’, ‘User (curious)’
5	‘User (surprised)’, ‘User (curious)’
5	‘User (uncertain)’, ‘Agent (confirming)’
5	‘User (uncertain)’, ‘Agent (confirming)’, ‘User (surprised)’
5	‘User (uncertain)’, ‘User (surprised)’

cost-based weighting nor conducted a thorough analysis of the mined patterns. However, recurring sequences across different emotional states and CEFR levels suggest the potential for further development. These patterns could be exploited to enhance interaction quality by informing *system-level adaptations*, such as customizing dialogue flows based on detected user emotion or language proficiency. Moreover, possible differences arising in patterns might suggest ways to provide more personalized emotional support or to adapt the complexity level of the conversation during the interaction. Exploring these applications will be a focus of subsequent research, especially as the dataset scales and richer annotations become available.

5. Conclusions and future work

In this paper, we have presented an analytical approach for generating a dataset of dialogues between a *user* and an *agent* of a customer care office using *LLM Models*. Specifically, we employ ChatGPT 3.5 to generate hypothetical dialogues in a given context. Furthermore, two characteristics have been considered: (*i*) the emotion expressed by the user; (*ii*) the complexity of the language during the interaction. Each dialogue has been checked for different parameters in order to ensure at least an acceptable level of quality for the dialogues that have been generated. The novelty of the presented work is the definition of a methodological approach that combines multiple elements, such as emotion awareness, language complexity evaluation, and SPM, into a cohesive system as a tool for enhancing human-agent interactions.

The study analyzed dialogues using two types of analyses, focusing on (*i*) language complexity; (*ii*) inferring interaction patterns based on specific emotions. The first task involved using different readability measures, and the second task involved an efficient algorithm SPM. The tool allows for clear interaction by creating graphs with nodes representing agent or user attitudes and links indicating transitions between them. It can be useful for reinforcement learning if a satisfaction score is provided at the end of the conversation, or for effectively supporting interactions and managing conversations in specific contexts.

The results of readability measures confirmed the effectiveness of ChatGPT in generating dialogues characterized by different language levels. The ability to exploit attitude tags associated with each turn of the dialogue allows for inferring effective interaction habits, provided an emotion is expressed by the user. Despite the limited number of dialogues, the preliminary findings highlight the potential of the approach and the need for further investigation.

Future work will regard two lines of research. The first one deals with the improvement of the analytics system in order to actually embed it in a decision support system for human-computer and human-robot interaction. The second one deals with combining the topics addressed by our current research with the ones coming from the emerging *big data trend* (e.g., [34–38]).

CRediT authorship contribution statement

Alfredo Cuzzocrea: Writing – review & editing, Validation, Resources, Methodology, Conceptualization. **Giovanni Pilato:** Writing – original draft, Validation, Formal analysis, Data curation. **Pablo G. Bringas:** Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was funded by the Next Generation EU - Italian NRRP, Mission 4, Component 2, Investment 1.5 (Directorial Decree n. 2021/3277) - project Tech4You n. ECS0000009.

Data availability

Data will be made available upon request.

References

[1] S. Poria, N. Majumder, R. Mihalcea, E. Hovy, Emotion recognition in conversation: research challenges, datasets, and recent advances, IEEE Access 7 (2019) 100943–100953.
[2] B. Shmueli, L. Ku, SocialNLP EmotionX 2019 challenge overview: predicting emotions in spoken dialogues and chats, arXiv preprint arXiv:1909.07734, 2019.
[3] S. Mohammad, Word affect intensities, arXiv preprint arXiv:1704.08798, 2017.
[4] A. Joshi, V. Tripathi, P. Bhattacharyya, M. Carman, Harnessing sequence labeling for sarcasm detection in dialogue from TV series ‘friends’, in: Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, Aug, 2016, pp. 146–155, https://aclanthology.org/K16-1015.
[5] G. Pilato, E. D’Avanzo, Data-driven social mood analysis through the conceptualization of emotional fingerprints, Proc. Comput. Sci. 123 (2018) 360–365.
[6] A. Cuzzocrea, G. Pilato, A composite framework for supporting user emotion detection based on intelligent taxonomy handling, Log. J. IGPL 29 (2021) 207–219.
[7] Crowdflower the emotion in text, 2016. https://www.figure-eight.com/data/sentiment-analysis-emotion-text/.
[8] S. Buechel, U. Hahn, EmoBank: studying the impact of annotation perspective and representation format on dimensional emotion analysis, in: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, Apr, 2017, pp. 578–585, https://aclanthology.org/E17-2092.
[9] S. Mohammad, F. Bravo-Marquez, M. Salameh, S. Kiritchenko, SemEval-2018 task 1: affect in tweets, in: Proceedings of the 12th International Workshop on Semantic Evaluation, Jun 2018, pp. 1–17, https://aclanthology.org/S18-1001.
[10] A. Goel, J. Gautam, S. Kumar, Real time sentiment analysis of tweets using naive bayes, in: 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), 2016, pp. 257–261.
[11] S. Mohammad, F. Bravo-Marquez, WASSA-2017 shared task on emotion intensity, in: Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Sep 2017, pp. 34–49, https://aclanthology.org/W17-5205.
[12] K. Scherer, H. Wallbott, Evidence for universality and cultural variation of differential emotion response patterning: correction, J. Pers. Soc. Psychol. (1994).
[13] H. Schuff, J. Barnes, J. Mohme, S. Padó, R. Klinger, Annotation, modelling and analysis of fine-grained emotions on a stance and sentiment detection corpus, in: Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2017, pp. 13–23.
[14] S. Mohammad, P. Sobhani, S. Kiritchenko, Stance and sentiment in tweets, ACM Trans. Internet Technol. 17 (2017) 1–23.
[15] S. Koelstra, C. Muhl, M. Soleymani, J. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, Deap: a database for emotion analysis; using physiological signals, IEEE Trans. Affective Comput. 3 (2011) 18–31.
[16] T. Deutsch, M. Jasbi, S. Shieber, Linguistic features for readability assessment, in: Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications, 2020, pp. 1–17.
[17] J. Chall, E. Dale, Readability Revisited: the New Dale-Chall Readability Formula, Brookline Books; 1995.
[18] J. Kincaid, R. Fishburne Jr, R. Rogers, B. Chissom, Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. IST Technical Report 8-75, 1975.
[19] N. Reimers, I.S.-B. Gurevych, Sentence embeddings using Siamese BERT-Networks, in: Proceedings of the 2019 Conference On Empirical Methods In Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 3982–3992.
[20] S. Kadıoğlu, X. Wang, A. Hosseininasab, W. Hoeve, Seq2Pat: sequence-to-pattern generation to bridge pattern mining with machine learning, AI Magazine 44 (2023) 54–66.
[21] P. Fournier-Viger, J. Lin, R. Kiran, Y. Koh, R. Thomas, A survey of sequential pattern mining, Data Sci. Pattern Recognit. 1 (2017) 54–77.
[22] X. Wang, A. Hosseininasab, P. Colunga, S. Kadıoğlu, W. Hoeve, Seq2Pat: sequence-to-pattern generation for constraint-based sequential pattern mining, Proc. AAAI Conf. Artif. Intell. 36 (2022) 12665–12671.
[23] A. Hosseininasab, W. Hoeve, A. Cire, Constraint-based sequential pattern mining with decision diagrams, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 1495–1502.
[24] S. Ghosh, S. Yadav, X. Wang, B. Chakrabarty, S. Kadıoğlu, Dichotomic pattern mining integrated with constraint reasoning for digital behavior analysis, Front. Artif. Intell. 5 (2022) 868085.
[25] J. Choi, S. Crossley, Advances in readability research: a new readability web app for English, in: 2022 International Conference on Advanced Learning Technologies (ICALT), 2022, pp. 1–5.
[26] ChatGPT, https://chat.openai.com/.
[27] ARTE automatic readability tool for English, https://www.linguisticanalysistools.org/arte.html.
[28] C. Busso, M. Bulut, C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. Chang, S. Lee, S.I. Narayanan, Interactive emotional dyadic motion capture database, Lang. Resour. Eval. 42 (2008) 335–359.
[29] Y. Li, H. Su, X. Shen, W. Li, Z. Cao, S. Niu, Dailydialog: a manually labelled multi-turn dialogue dataset, arXiv preprint arXiv:1710.03957, 2017.
[30] P. Ekman, Basic emotions, in: Handbook of Cognition and Emotion, Vol. 98, John Wiley & Sons, Ltd.; 1999, pp. 45–60.
[31] A. Moin, F. Aadil, Z. Ali, D. Kang, Emotion recognition framework using multiple modalities for an effective human–computer interaction, J. Supercomput. 79 (2023) 9320–9349.
[32] A. Asta, A. Cuzzocrea, A. Fantini, G. Pilato, P. Bringas, Supporting emotion recognition in human-robot interactions: an experimental Italian textual dataset, in: International Conference on Hybrid Artificial Intelligence Systems, 2023, pp. 483–494.
[33] A. Asta, A. Cuzzocrea, A. Fantini, G. Pilato, P. Bringas, Supporting emotion recognition in human-robot interactions: an experimental Italian textual dataset, in: Hybrid Artificial Intelligent Systems - 18th International Conference, HAIS 2023, Salamanca, Spain, September 5–7, 2023, Proceedings, vol. 14001, 2023, pp. 483–494.
[34] M. Masum, H. Shahriar, H. Haddad, M. Faruk, M. Valero, M. Khan, M. Rahman, M. Adnan, A. Cuzzocrea, F. Wu, Bayesian hyperparameter optimization for deep neural network-based network intrusion detection, in: 2021 IEEE International Conference On Big Data (Big Data), Orlando, FL, USA, December 15–18, 2021, 2021, pp. 5413–5419.
[35] C. Leung, P. Braun, A. Cuzzocrea, AI-Based sensor information fusion for supporting deep supervised learning, Sensors 19 (2019) 1345.
[36] R. Camara, A. Cuzzocrea, G. Grasso, C. Leung, S. Powell, J. Souza, B. Tang, Fuzzy logic-based data analytics on predicting the effect of hurricanes on the stock market, in: 2018 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2018, 2018 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2018, Rio De Janeiro, Brazil, July 8–13, 2018, 2018, pp. 1–8.
[37] P. Howlader, K. Pal, A. Cuzzocrea, S. Kumar, Predicting facebook-users’ personality based on status and linguistic features via flexible regression analysis techniques, in: Proceedings of the 33rd Annual ACM Symposium on Applied Computing, SAC 2018, Pau, France, April 09–13, 2018, 2018, pp. 339–345.
[38] N. Shukla, R. Katikeri, M. Raja, G. Sivam, S. Yadav, A. Vaid, S. Prabhakararao, Generative AI approach to distributed summarization of financial narratives, in: IEEE International Conference on Big Data, BigData 2023, Sorrento, Italy, December 15–18, 2023, 2023, pp. 2872–2876.
[39] C. Xu, D. Guo, N. Duan, J.B. McAuley, An open-source chat model with parameter-efficient tuning on self-chat data, in: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 2023, pp. 6268–6278.
[40] P. Shah, D. Hakkani-Tur, B. Liu, G. Tür, Bootstrapping a neural conversational agent with dialogue self-play, crowdsourcing and on-line reinforcement learning, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 3 (Industry Papers), 2018, pp. 41–51.
[41] M. Kim, C. Kim, Y. Song, S. Hwang, J.B. Yeo, Machine-sourced framework for automatic curation of large-scale multi-skill dialogue datasets, in: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022, pp. 5149–5170.
[42] Y. Lee, C. Lim, Y. Choi, J. Lm, H. Choi, PERSONACHATGEN: generating personalized dialogues using GPT-3, in: Proceedings Of The 1st Workshop On Customized Chat Grounding Persona And Knowledge, 2022, pp. 29–48.
[43] Y. Yu, Y. Zhuang, J. Zhang, Y. Meng, A. Ratner, R. Krishna, J. Shen, C. Zhang, Large language model as attributed training data generator: a tale of diversity and bias, Adv. Neural Inf. Process. Syst. 36 (2023) 55734–55784.
[44] J. Ye, J. Gao, Q. Li, H. Xu, J. Feng, Z. Wu, T. Yu, L. Kong, ZeroGen: efficient zero-shot Learning via dataset generation, in: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022, pp. 11653–11669.
[45] J. Ye, J. Gao, Z. Wu, J. Feng, T. Yu, L. Kong, ProGen: progressive zero-shot dataset generation via in-context feedback, in: Findings of the Association for Computational Linguistics: EMNLP 2022, 2022, pp. 3671–3683.
[46] P. Budzianowski, T. Wen, B. Tseng, I. Casanueva, S. Ultes, O. Ramadan, M. Gasic, MultiWOZ-A large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 5016–5026.
[47] B. Requena, G. Cassani, J. Tagliabue, C. Greco, L. Lacasa, Shopper intent prediction from clickstream e-commerce data with minimal browsing information, Sci. Rep. 10 (2020) 16983.
[48] W. Gan, J. Lin, P. Fournier-Viger, H. Chao, P. Yu, A survey of parallel sequential pattern mining, ACM Trans. Knowl. Discov. Data 13 (2019) 1–34.
[49] X. Wang, S. Kadıoğlu, Bridging constraint-based sequential pattern mining and machine Learning, in: AAAI 2023 Bridge on Constraint Programming and Machine Learning February 7th, 2023, Washington DC, USA, 2023, http://osullivan.ucc.ie/CPML2023/submissions/09.pdf.