ARIEL: Brain-Computer Interfaces meet Large Language Models for Emotional Support Conversation

Paolo Sorino*
Polytechnic University of Bari
Bari, Italy, IT
paolo.sorino@poliba.it

Giovanni Maria Biancofiore*
Polytechnic University of Bari
Bari, Italy, IT
giovannimaria.biancofiore@poliba.it

Domenico Lofù
Polytechnic University of Bari
Bari, Italy, IT
domenico.lofu@poliba.it

Tommaso Colafiglio
Polytechnic University of Bari
Bari, Italy, IT
tommaso.colafiglio@uniroma1.it

Angela Lombardi
Polytechnic University of Bari
Bari, Italy, IT
angela.lombardi@poliba.it

Fedelucio Narducci
Polytechnic University of Bari
Bari, Italy, IT
fedelucio.narducci@poliba.it

Tommaso Di Noia
Polytechnic University of Bari
Bari, Italy, IT
tommaso.dinoia@poliba.it

ABSTRACT

In an era characterized by unprecedented virtual connectivity, paradoxically, individuals often find themselves disconnected from genuine human interactions. The advent of remote working arrangements, compounded by the influence of digital communication platforms, has fostered a sense of isolation among people. Consequently, the prevailing socio-technological landscape has underscored the critical need for innovative solutions to address the emotional void. Conversational systems help people improve their everyday tasks with informative dialogues, and recent applications employ them to target emotional support conversation tasks. Nevertheless, their understanding of human feelings is limited, as they depend solely on information discernible from the text or the users' emotional declarations. Recently, Brain-Computer Interfaces (BCIs), devices that analyze electroencephalographic (EEG) signals, have increasingly become popular given their minimally invasive nature and low cost, besides enabling the detection of users' emotional states reliably. Hence, we propose ARIEL, an emotionAl suppoRt bcI dEvices and Llm-based conversational agent that aims at supporting users' emotional states through conversations and monitoring them via BCI. In this way, it is possible to comprehend the users' feelings reliably, thus making the conversational agent aware of users' emotional evolution during conversations. Our framework makes the LlaMA 2 chat model communicate with an emotion recognition BCI-based system to achieve the emotional support conversation goal. Also, we present a controlled running example that shows the potential of our model and its effective functioning, made possible by a wisely designed hard-prompt strategy. In the future, we will conduct an in-vivo experiment to evaluate the system and its components.

*Corresponding authors.

UMAP Adjunct '24, July 01–04, 2024, Cagliari, Italy

ACM ISBN 979-8-4007-0466-6/24/07
https://doi.org/10.1145/3631700.3665193

CCS CONCEPTS

• **Computing methodologies → Discourse, dialogue and pragmatics**; **Natural language generation**; *Support vector machines*; • **Hardware → Sensor devices and platforms**; • **Human-centered computing → Human computer interaction (HCI)**; • **Software and its engineering →** *Software prototyping*.

KEYWORDS

Brain-Computer Interface, Emotion Recognition, Conversational Agent, Large Language Model, Emotional Support Conversation, Machine Learning

ACM Reference Format:

Paolo Sorino, Giovanni Maria Biancofiore, Domenico Lofù, Tommaso Colafiglio, Angela Lombardi, Fedelucio Narducci, and Tommaso Di Noia. 2024. ARIEL: Brain-Computer Interfaces meet Large Language Models for Emotional Support Conversation. In *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization (UMAP Adjunct '24), July 01–04, 2024, Cagliari, Italy*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3631700.3665193

1 INTRODUCTION

The widespread diffusion of social media platforms has enabled people to reach an outstanding virtual connection and receive social rewards from their adoption. More than 4.59 billion people are using these social media worldwide, a number projected to increase to almost six billion in 2027[1], allowed to communicate and share their intimacy and private life with even a large circle of other users. Although these platforms permit faster distribution of information, even from institutional and communications bodies, resulting in a highly connected and promptly updated society, some people's social media use may become maladaptive and

[1]Statista. Number of global social network users 2017-2027 (2023).

problematic, causing distress and impairment in daily functioning [23]. Indeed, problematic social media use has been consistently linked with negative mental health [33], with indirect effects on both depression and anxiety [30]. Such an issue has been further accentuated by the growing impact of Artificial Intelligence (AI) in several domain [9, 32, 43–45], driven also by the recent emergence of generative AI models, which has sparked heated discussions regarding their benefits, limitations, and associated risks [4].

Despite their offered advantages like speeding up repetitive processes, facilitating information access, and empowering people skills in education and working areas are widely recognized nowadays [37, 42], bias, false information and privacy issues of such tools affect the already complex social context that exists lately [16]. Moreover, models' foundational high versatility and efficacy in being easily applied in the most disparate jobs or tasks instil fears and insecurities in human beings about their near future [13, 47], besides bearing the existence of a variegated pool of automated services that additionally isolate people from human contact.

As published by the World Health Organization (WHO), 970 million people globally are living with a mental disorder, with anxiety and depression being the most common[2] and encouraged by lack of affection, real social activities and the isolation people face in the current panorama. Mental disorders are the primary cause of even more severe illnesses [19]. Therefore, it is crucial to address this critical issue by exploiting the advantages that technological advances in devices and AI models offer.

The Emotional Support Conversation (ESConv) task is a recent initiative that seriously meets the requirements of handling people's emotional support [21] who feel a mental disorder. It employs dialogues to reduce the users' emotional distress, requiring the conversational system to have a more high-level and complex ability to understand the help-seeker emotional state and provide supportive responses [34]. Indeed, during the emotional support conversation, it is required to explore the cause of the help-seekers emotional problem and understand their psychological intentions to provide supportive responses effectively [38], which is not straightforward for automated conversational models. Early works have tried to implement human-like emotions and reactions within a conversational agent for reacting to users' conversations [6, 41], highlighting relevant difficulties in modelling such a complex behaviour with rule-based approaches. Similar limitations have been encountered when sequence-to-sequence Machine Learning (ML) models addressed the emotional aspects within dialogues [53]. Although these solutions can easily recognize sentiment from text to then differentiate generated answers over multiple possible emotions, they stop functioning when encountering users' emotional nonlinearity (e.g., the text emerges a positive sentiment while the user feels distressed) or dealing with topics outside the ones learned during its training.

Recent generative models identify a revolutionary solution for solving the last task. Large Language Models (LLMs), deep learning systems with massive parameters designed to learn and generate natural language text [12], have proved to be highly effective in handling human language information [7, 20]. They enable the automatic processing and generation of text over the most disparate topics without the need to train them from scratch. Indeed, such models have been pre-trained over large textual corpora, obtaining an impressive general knowledge of the word [49]. Furthermore, their straightaway usage does not require particular tuning other than a well-designed input text message telling them the task to complete. The latter technique is commonly known as the prompting strategy [28], which allows the direct deployment of LLMs within the intended framework. Additionally, contemporary LLM-based conversational tools (e.g., ChatGTP[3], LLaMA 2 chat[4], Vicuna[5], etc.) are famous for engaging in conversations with unprecedented naturalness and depth, facilitating dynamic and nuanced dialogues that mirror human interactions. Therefore, the latest proposals addressing the ESConv task include these outstanding technologies.

Zheng et al. [51] exploit GPT-J to build an extensive collection of conversations for the ESConv task. Specifically, the authors combine specialized heuristics with fine-tuned and prompted GPT-J models to perform a data augmentation of the dataset already made available by the community, thus realizing AugESC. Conversely, Peng et al. [35] focus on including an LLM called BlenderBot [39] within a framework exhaustively designed to provide supportive responses to users. The authors guide the introduced model in understanding the emotional needs of the users, explicitly stated through the exchange of messages, to generate the most effective supportive answer, settling new state-of-the-art performances. More recently, Peng et al. [34] found interesting the results achieved by the current proposal based on LLMs and developed a Hierarchical Graph Network to improve the semantic analysis of users requests about their emotional distress. The proposal succeeds in its aim by enhancing the efficacy of these systems. However, it highlights an extreme difficulty for current models in recognizing actual users' emotional states using only text.

Identifying people's emotions is an active field of research in the Affective Computing domain. Specifically, affective computing is a subdomain of AI [36], which involves automatic emotion recognition. This field has seen an acceleration in research thanks to the availability of low-cost devices called Brain-Computer Interface (BCI), which capture brain signals as input for systems designed to decipher the relationship between emotions and Electroencephalographic signal changes. In particular, EEG is a technique for investigating brain electrical signals, i.e. ionic current flows through brain neurons that cause voltage changes. This electrical activity is spontaneous and is recorded over time by several electrodes on the scalp [11]. Traditionally, EEG signals acquired by BCI are recorded on the scalp. Due to their minimally invasive nature, BCIs have been used in various tasks such as music composition [14], neurorehabilitation [15], and Emotion recognition [2].

With the increasing availability of various electronic devices, people have spent more time on social media, playing online video games, shopping online and using other electronic products. Most contemporary Human-Computer Interaction (HCI) systems are not capable of processing and understanding emotional data and lack emotional intelligence. They are unable to recognize human emotions and use emotional data to make decisions and take action. Therefore, any HCI system that does not take human emotional

[2]Organization, W. H. Mental health (2023).

[3]https://chat.openai.com

[4]https://llama.meta.com/llama2/

[5]https://lmsys.org/blog/2023-03-30-vicuna/

states into account will not be able to respond properly to those emotions. To solve this challenge in HCI systems, machines must be able to understand and interpret human emotional states. To build intelligent HCI systems, a reliable, accurate, flexible and powerful emotion recognition system is needed [24, 29].

To address these challenges, we introduce ARIEL, an emotionAl suppoRt bcI dEvices and Llm-based conversational framework that aims at supporting users' distresses through dialogues and monitoring them via BCI. We design a novel approach to address the ESConv issue by taking advantage of the recent progress in emotion recognition and language modelling tasks. Specifically, ARIEL relies on an Emotion Recognition system composed of several ML models trained to classify users' emotional states based on their EEG signals. Whenever a negative state is recognized, ARIEL asks the user to have a chit-chat to alleviate her emotional conditions. During the user-agent interaction, the system keeps acquiring EEG signals to monitor the emotional evolution of the user, which guides the generation of dialogues thanks to the role-play prompting technique conceived for the LLM employed as the core of the conversational agent. In particular, ARIEL is equipped with the LLaMA 2 chat version. The conversation ends when the user leaves the agent, and the system detects a stable positive emotional state reached by the user. Our contribution can be summarized as follows:

- We are the first to introduce a novel approach to leverage the ESConv issue by mixing models derived from Affective Computing and Language Modelling fields (i.e., emotion recognizer EEG-based and LLMs);
- We present the operative ARIEL framework that addresses the ESConv task, describing in-depth the components we designed and the whole workflow pipeline on which the system is based;
- We demonstrate the effectiveness of our proposal through a simple in-vitro experiment.

2 RELATED WORK

This work lies in the intersection between the field of Large Language Models (LLMs), employed as Conversational Agents (CAs), and Brain Computer Interface (BCI), which enables the acquisition of EEG signals and thus their processing to recognize human states like emotions. Although both domains are highly novel in the research community, reflected by the massive number of proposals shared as a preprint version, we dissect below only those works published after being peer-reviewed into two separate subsections.

2.1 Brain Computer Interface

Atkinson *et al.*[3] propose an approach based on a novel feature-based emotion recognition model is proposed for EEG-based BCI interfaces on a DEAP dataset. In this paper, the authors explore a broader set of emotion types, arguing that the combination of a feature selection method based on mutual information (i.e., minimum redundancy-maximum relevance) and kernel classifiers can improve the accuracy of the emotion classification task, based on a dimensional model of Valence and Arousal. The overall results showed the proposed SVM trained on valence and arousal achieves 73% accuracy on the considered task.

Islam *et al.*[25] highlight a convolutional neural network (CNN) approach with Pearson correlation coefficient (PCC) characterized by correlation images of EEG subband channels on the DEAP dataset. In this work, the CNN model images were proposed for emotion recognition. The authors use two protocols: protocol-1 to identify two levels and protocol-2 to recognize three levels of valence and arousal proving emotion. In addition, the authors point out that only the upper triangular portion of the PCC images reduced the computational complexity and memory size without hindering the accuracy of the model. The proposed CNN achieves an accuracy of 78.22% for valence and 74.92% for arousal.

Zheng *et al.*[52] detail a method to convert the one-dimensional EEG signal into a two-dimensional EEG signal with spatial information based on DEAP dataset. The authors propose an emotion recognition method based on an adaptive neural decision tree (ANT) to overcome the problem of model interpretability. They also claim that the use of ANT for EEG signal does not require manual a priori feature extraction. The ANT can automatically search for the optimized parameters by exploring various tree architectures using the reinforced learning method of exploration-exploitation trade-off to obtain the optimal global network structure. The authors' proposed model achieves an accuracy of 99.12 , 98.95 , 97.58 in binary, four-class and eight-class classification tasks, respectively.

2.2 Emotional Support Conversation with Large Language Model

Tu et al. [46] design MISC, a MIxed Strategy-aware model integrating COMET for emotional support conversation. The authors introduce their pre-trained generative commonsense reasoning model called COMET [10], formulating the response generation as a probability distribution over a strategy codebook employed to facilitate the training process. The experimental setting involves multiple offline metrics to evaluate the accuracy of choosing the most effective supportive strategy and assess the naturalness of the generated answer, demonstrating the proposal's rationality.

On the other hand, Peng et al. [35] delve into the integration of a Language Learning Model (LLM) known as BlenderBot [39] into a meticulously crafted framework aimed at furnishing users with empathetic responses. Their approach entails users' emotional cues communicated within the dialogue, adeptly guiding the model to discern and address these needs. By leveraging this understanding, the system produces supportive responses that resonate effectively with users in achieving a positive emotional state.

Liu et al. [31] innovate the ESConv task by including multimodal information and propose a new method called FEAT. In detail, the authors integrate fine-grained emotional knowledge from multiple modalities (text, audio and video), implementing an emotion-aware transformer to recognize users' mental states. The system then learns to generate supportive responses based on conversations viewed during its training phase, reaching state-of-the-art performances.

However, all these works strictly rely on the textual messages and exchange of the information they employ to train their system, thus limiting the generability of such models when facing unseen situations. Moreover, they address the task of users' emotional recognition by counting solely on data that does not grant enough reliability in detecting the complex human emotional sphere. Indeed, De Freitas et al. [17] addressed this urgent question by focusing on the context of mental health and companion AI, applications designed to provide consumers with synthetic interaction partners, surveying all the available tools developed to leverage such an issue. It emerges that companion AIs are often unable to recognize signs of distress, and consumers usually display unfavourable reactions to unhelpful and risky chatbot responses.

With this work, we believe that merging an emotional recognizer BCI-based and an LLM adapted through prompts for the conversational task in handling emotional support dialogues lays the foundation for developing solutions that effectively address the critical context of mental disorders. Thanks to higher accuracy in human emotion recognition, our framework supports users in mitigating their distress in favour of positive emotional states through an LLM-based conversational agent, which can entertain generic dialogues without limitations dictated by a specialized domain's dataset.

3 ARIEL FRAMEWORK

The task of emotional support conversation can be generally divided into two subtasks: i) the modelling of users' psychological factors to detect their emotional state (i.e., emotion recognition task - ER) and ii) the design of a conversational mode able to have natural language dialogues (i.e., conversational agent - CA) with the final aim to generate supportive responses. Although multiple Sentiment Analysis techniques allow today for detailed identification of emotional aspects hidden within a text [8, 48], which helps to infer the emotional state of the person who wrote it, they are effective only when dealing with linear sentiment [18], failing the emotion recognition task otherwise. Moreover, the development of CAs strictly dedicated to generating supportive dialogues suffers from the overfitting problem. Specifically, they may be unskilled in talking about topics outside the addressed task [1], typical of human interactions, which would facilitate user emotional support.

To solve such issues, we design the ARIEL framework by taking advantage of the recent developments in the AC field through BCI devices and the unprecedented linguistic and cognitive properties characterizing the most recent LLM systems in having conversations with users. Figure 1 illustrates ARIEL's architecture and its workflow modelled to address the Emotional Support Conversation (ESConv) task, which mainly relies on four components: the Neuro-Linguistic Interface, the Emotion Recognizer, the Prompt Formatter, and the LLM.

3.1 Neuro-Linguistic Interface

The Neuro-Linguistic Interface identifies the principal component with which users can interact with the ARIEL framework. It allows information gathering from actions undertaken by users who decide to interact with our system, looking for supportive conversations. Such data is mandatory for ARIEL to accomplish the ESConv task, which exploits the Neuro-Linguistic Interface to return, in turn, the computed responses to users. In a few words, this component is essential to make people communicate with the framework and vice versa.

The input on which our framework works is two-folded, spanning over two different information channels: the EEG signal and natural language. The former is acquired through a BCI device, which measures the brain's electrical activities in a non-invasive way. This signal conveys more emotional information from humans than other kinds (e.g., linguistic data, facial expressions). The brain's more superficial areas are closely connected to emotions, with a higher concentration of electrical activity when feelings or emotional states are experienced. Therefore, it is a valuable source of information to detect users' emotional condition and act accordingly. The latter, instead, brings to the system what users want to communicate explicitly in textual form. Such messages may also include users' sentiments through descriptions or writing styles, although they can hardly be considered reliable. Nevertheless, textual messages are handled by the homonym sub-component, which allows the system to know the topics on which users prefer to talk, besides making possible dialogues.

The system consistently collects EEG signals starting when the user initiates interaction. In contrast, it exclusively registers chat messages upon the user's successful transmission through the Neuro-Linguistic Interface.

3.2 Emotion Recognizer

The Emotion Recognizer is a component that analyzes the EEG signals measured via BCI to infer the actual emotional state of the user. Specifically, it receives a stream of EEG signals, which is adequately framed based on the dialogue's timing users have with the system, from the Neuro-Linguistic Interface. These signals are processed and feed machine learning models to infer a label indicating the user's emotional state, which is essential to have a clear picture of the user's current state and act accordingly. To achive this goal, we rely on a novel approach granting high trustworthiness regarding the emotion recognized, which we describe below in tripartite paragraphs related to its founding elements.

3.2.1 Dataset description. In this study, the Deap [26] dataset protocol approach was employed to collect data to train machine learning models based on the emotion state classification. The acquisition process was conducted at the *Polytechnic University of Bari*, taking into account a student population with an average age between 20 and 30 years. Participants were informed about the procedure and provided written consent for the collection of Electroencephalographic (EEG) data before the experiment. Participants' EEG signal was acquired using the Muse 2 EEG device[6] to perform an accurate data acquisition. During the experiment, a series of 40 music videos was shown to each participant, and simultaneously the EEG data were continuously recorded. All collected EEG data have been anonymized to ensure participants' privacy, and stored securely and accessible only to the researcher team. The final dataset is organized as follows: (i) 30 subjects collected and divided into several folders. (ii) Each folder contains 40 EEG trials organized into session subfolders.

3.2.2 Preprocessing. The preprocessing pipeline is outlined below: EEG epochs duplication and application of a bandpass filter to isolate frequencies between 1 and 40 Hz was carried out as the first

[6]https://choosemuse.com/products/muse-2

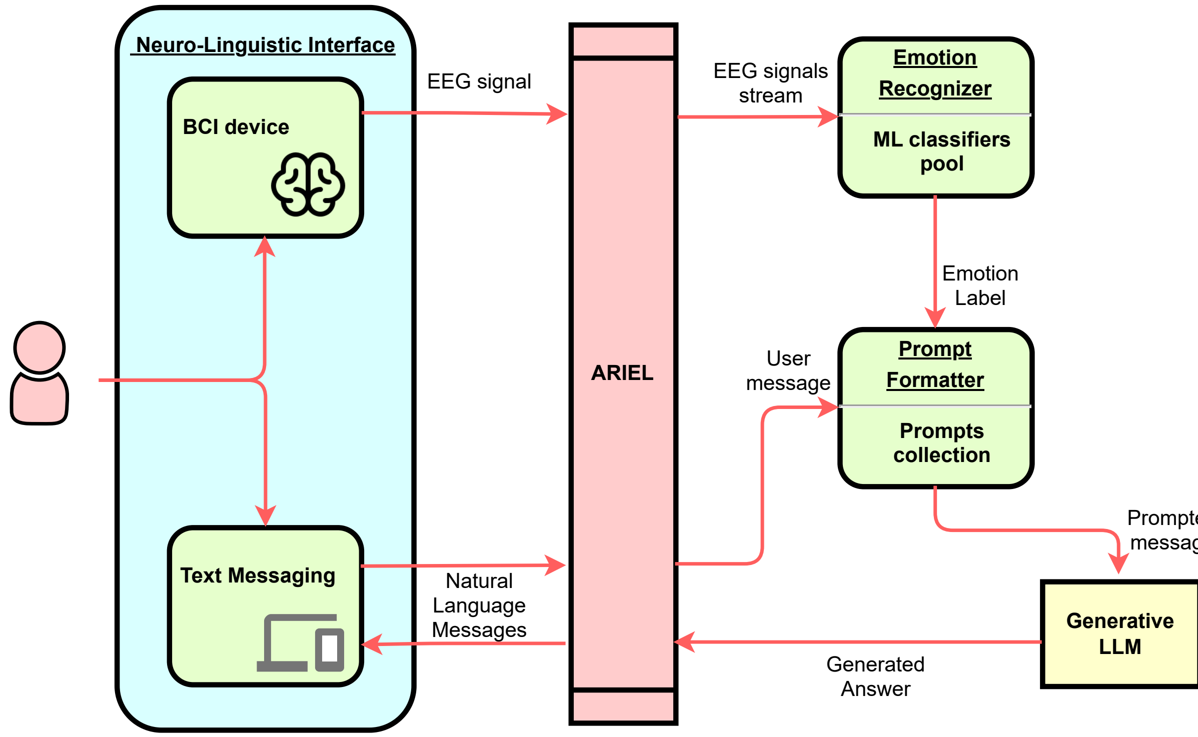

Figure 1: The figure schematizes the overall workflow followed by the ARIEL framework. On the left side, the user interacts with ARIEL through the Neuro-Linguistic Interface (a) component, which contemporarily enables Text Messaging and and measuring EEG signals through a BCI device. On the right side, the Emotion Recognizer (b) receives a stream of EEG signals on which several ML classifiers are asked to infer the emotion label that describes those signals. Such a label feeds the Prompt Formatter (c) with the user message. The (c) component wraps up this information within the most suitable prompt based on the conversation state. The LLM (d) is queried with the selected prompt, generating a supportive response delivered to the user for continuing the dialogue. The interaction ends when the user reaches a positive emotional state and leaves the conversation.

step. After filtering, the data are transposed and amplitude scaled to simplify computation. A K-nearest neighbors (K-NN) classifier, was trained on the filtered dataset to calculate anomaly scores in order to identify EEG data artifacts if a predefined threshold is exceeded. All artifacts detected by the threshold-based approach are subsequently removed by exploiting the megkit [5] framework using the ringing artifact reduction method. Cleaned EEG data are then resized, restructured, and encapsulated in an MNE[22] EpochsArray object, ensuring data consistency since it maintains names and channel types defined according to a standardized electrode fitting. Otherwise, if the EEG Signals are artifact-free, they are preserved in their original shape. This accurate preprocessing procedure is essential to preserve the integrity of the EEG data, increasing the reliability of the results.

3.2.3 Machine Learning models. In order to find the best model able to classify emotional states, different Machine Learning (ML) models such as Logistic Regression (LR), eXtreme Gradient Boosting (XGB), Random Forest (RF) and, Support Vector Machine (SVM) are compared. In order to identify the best model, a hyperparameter tuning using GridSearchCV (5-fold)[7] was applied, with a Leave-One-Subject-Out (LOSO) cross-validation strategy[27]. This approach has led to the development of 120 specialized models. Specifically 30 ML models for each of 4 emotions (Happy, Angry, Sad, Relaxed) able to predict by majority vote the outcome resulting from the quadrants of Russell's complex circumference. After comparison of the various ML models, the best model in predicting Happy, Angry, Sad and, Relaxed emotions was SVM achieving an accuracy of 0.77, 0.78, 0.76, 0.80 respectively for the considered emotions.

3.3 Prompt Formatter

ARIEL introduces also the Prompt Formatter component, a crucial element within our conversational system that orchestrates the initiation and progression of dialogues with a Large Language Model (LLM). The Prompt Formatter serves as the gateway through which the user's emotion label, found by the Emotion Recognizer, and her sent message are amalgamated into coherent prompts, setting the stage for meaningful interactions. A prompt is essentially a piece of natural language text given to the model by the user, serving as the initial input or question. This input initiates the

[7] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

model's response process and guides its direction and scope, acting as a catalyst for the output's relevance and utility [50].

At the heart of the Prompt Formatter lies a dynamic prompting strategy designed to optimize dialogue generation in handling the ESConv task. The strategy hinges on the selection and adaptation of prompts tailored to different stages of the conversation. Initially, the Prompt Formatter employs a role-play prompt to kickstart the interaction, providing the role the LLM has to play towards conversations as a context. In this case, we assign to the LLM the role of a virtual assistant specialized in supporting users' mental distress to evolve their emotional status to a positive condition. Furthermore, we also let the LLM know that the user's sentiment is measured with BCI sensors, favouring the LLM in considering possible mismatches between what the user says about her emotional state and what he actually feels. Below we report the adopted prompt for enabling the LLM to initiate the conversation.

Conversation start role-play prompt

You are a Virtual Assistant designed to converse with users supporting their emotional distress and transforming it into a positive feeling. You receive as input also information about users emotional state measured through an high reliable emotion recognition model. Just have a chit-chat conversation until the user feels happy or relaxed and wants to leave the conversation. The emotion recognition model measures the user feels {emotion label}. The user sent {user message}.

As the dialogue unfolds, the prompt evolves in response to the ongoing discourse, ensuring relevance and coherence while leveraging the LLM's contextual understanding. Specifically, the prompt evolution is determined step-by-step through the conversation by selecting a prompt from a collection designed to take into account possible conversational states. Moreover, all the prompts include growing dialogue histories to avoid information loss. This adaptive prompting mechanism not only sustains the coherence of the conversation but also facilitates the LLM's comprehension and response generation, leading to richer interactive exchanges.

3.4 Generative LLM

Finally, the generative LLM component identifies the core of the ARIEL framework. Based on the open weight LLaMA 2 chat model, this component exploits the language processing capabilities of the LLMs instructed to behave in a conversational manner. It serves as the virtual interlocutor, capable of understanding user inputs, processing contextual information, and generating coherent responses in natural language.

Large Language Models (LLMs) are based on advanced neural networks designed for natural language processing tasks. At their core, LLMs are probabilistic models designed to estimate the probability distribution of sequences of linguistic units such as words or sentences [40]. Upon receiving prompts from the Prompt Formatter component, the generative LLM builds responses based on the provided context and instructions. These responses are crafted to emulate human-like conversational behavior, incorporating linguistic nuances, contextually relevant information, and appropriate tone. Moreover, since it is coupled with prompting strategies described within the Prompt Formatter component, LLMs seamlessly integrate into conversational frameworks, effortlessly generating responses that align with the desired conversational goals.

4 ARIEL@WORK

With the growing sense of isolation and detachment people feel in today's scenario, compounded by the pervasive influence of digital communication platforms, ARIEL eavesdrops actively on the users' emotional problems and promulgates through conversations engaging discussions and chit-chat interactions to alleviate their mental distress. The ARIEL's goal is to let people talk about preferred and positive topics, opening up to confidence and problem expressions to relieve the experienced emotive pressure.

Below, we describe an example of a user interaction with the ARIEL system consisting of an AI-based emotion recognition framework and an LLM-based chatbot. Specifically, we have designed a running example immersed in one of the possible cases for which this tool is helpful, besides explaining how the ARIEL framework works in detail. Furthermore, Figure 2 contains an additional conversation example to ensure a better understanding of the functioning of the proposed framework. Such a controlled experiment resulted from a real session trial with the ARIEL system.

Eric is a young professional immersed in a stressful work environment. Particularly, in a very stressful period studded with relevant workload and imminent deadlines, at the end of his working day, Eric has limited time to engage in any social activity to relieve his accumulated tension. Therefore, to get relaxed Eric sits at his computer and wears a BCI device. With a habitual gesture, he starts the system called *ARIEL*, which immediately starts recording the EEG signals of his brain and using an AI-based Emotion Recognition algorithm to predict his emotional state in real-time. Indeed, this system represents an innovative interface between Eric's mind and the digital world, enabling a proactive understanding and reaction to his emotions. Based on the emotions predicted by ARIEL, a chatbot embedded in the system triggers a dialogue with Eric, designed to drive his emotional state towards a positive direction.

In this case, the ARIEL's emotion recognizer predicts that Eric is experiencing sadness. Hence, ARIEL's chatbot welcomes Eric and initiates working on the emotion felt by the user by encouraging them to have a light conversation with the message: "*Good evening Eric! I sense a certain sadness from you. What happened?*". Eric, who needs to talk to someone to relieve himself from daily worries, texts the system: "*Hi Ariel! Yes, my work is quite tiring these days, and deadlines certainly don't help me relax.*". Since ARIEL does not acknowledge a shift in Eric's emotional state, it tries to cheer up him by replying: "*I see! But do you know that after great effort there is always great satisfaction? You're working hard, you deserve it!*". Eric feels partly consoled by the message, but the great tiredness does not allow him to change his sadness to a positive feeling. Accordingly, Eric messages ARIEL: "*I know! However, I feel trapped in this routine. Tonight I would have liked to go and listen to a Pink Floyd cover band, but obviously, I can't.*". Although the overall emotional state did not have sensitive variations from the initial state, the BCI device measures a slight positive change in Eric's EEG signal that the emotion recognizer promptly reports

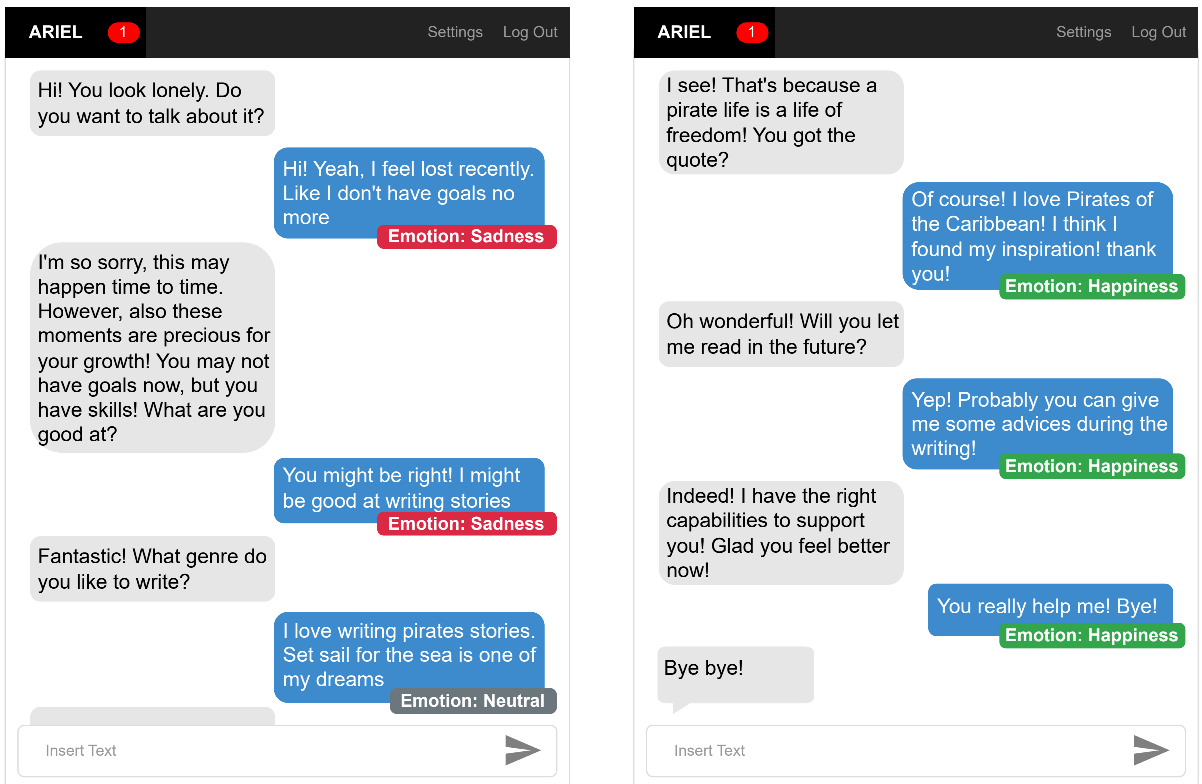

Figure 2: The figure highlights a simplified conversation held by ARIEL with the user Eric. The first screenshot from the left captures the start of the conversation, where ARIEL engages Eric in talking about his emotional distress (grey balloon). Thus, Eric joins the dialogue with an answer describing what he feels (blue balloon) together with the emotion recognised by the Emotion Recognizer (cf. Figure 1) through the BCI device (red balloon). The two actors of the framework have a chit-chat conversation that brings the user's emotional state to evolve from sadness into happiness.

to the ARIEL chatbot. Consequently, ARIEL continues the conversation by resuming the topic discussed in the previous message, which allowed the recording of this little positive perturbation (e.g., Pink Floyd). It answers Eric: "*I'm sorry! However, even if it is not the same, you can reproduce their songs to relax! I think you would love listening to "The Great Gig in the Sky"! Is one of my favourites! The vocal performance is exceptional!*". At this moment, Eric feels engaged by the conversation since the mentioned song is also one of his favourites. He replies: "*It is also one of my favourite songs! Clare was commendable and conveys many positive feelings!*". The emotion recognizer now detects a neutral state of the user from the initial sadness, which encourages ARIEL to continue talking about music. Through messages of encouragement, positive talk or joyful activities, the system will try to transform Eric's sadness into a more positive feeling, such as happiness. The conversational session will terminate whenever Eric reaches a positive emotional state and greets ARIEL, declaring the end of the dialogue.

ARIEL implements a dynamic approach for handling the ES-Conv task thanks to the Prompt Formatter. Depending on emotions predicted by the emotion recognizer, this component selects the most suitable prompt that guides ARIEL's LLM in conversing with users and accomplishing the emotional supportive mission. In this example, Eric has a bad day at work the following day, having had some discussions with colleagues. When interacting with ARIEL, the system predicts that Eric is experiencing anger. In this case, the chatbot will take a different approach. Therefore, ARIEL texts Eric: "*Hi Eric! Did you have a bad day? I feel your anger.*" Eric will then explain to ARIEL his unfortunate day, and the emotion recognizer will detect his negative feelings. As a consequence, ARIEL engages in a conversation to manage nervousness issues, replying to Eric: "*What a pity! But first thing first, take a deep breath. Every problem has a solution. Let's find it together.*" During the dialogue, ARIEL will offer Eric further stress management techniques, such as calm reflection on his feelings, despite positively facing the emerging work issue. The aim is to help Eric calm down and regain his inner peace, gradually shifting his emotionality towards more neutral and ultimately positive states. Also, the whole conversation will be guided wisely thanks to the continuous monitoring of the user's EEG signals through the BCI device.

Whenever Eric starts a conversation with ARIEL while feeling happy or any other positive emotion, the emotion recognizer will detect that Eric is already in a positive emotional state, therefore the chatbot will keep the conversation light and joyful, without attempting to change his mood further. Instead, it might propose fun activities or encourage Eric to share his positive experiences, thus helping to consolidate his happy emotional state.

5 CONCLUSION

In this paper, we presented a novel framework to handle the emotional support conversation (ESConv) task. ARIEL, an emotionAl suppoRt bcI dEvices and Llm-based system, embraces an innovative strategy in engaging users through conversations to alleviate their mental distress via the adoption of brain-computer interface (BCI) devices for reliably detecting their feelings. Our proposal is the first to take advantage of the recent discoveries in the fields of Affective Computing and Language Modelling, mixing BCI-based emotion recognition models and a large language model (LLM)-based conversational agent (CA) to effectively address the ESConv task. Indeed, ARIEL is composed of an emotion recognizer to detect the users' emotional states via BCI, an LLM to generate natural language messages given a text as the input, and a prompt formatter to enclose emotional labels and users' messages within different and dynamically selected prompts to guide the LLM in generating dialogues. One of the most interesting aspects of ARIEL is the high reliability in detecting users' emotions thanks to the electroencephalographic (EEG) signals acquired by the BCI device, showing effective functioning during our running example.

In future developments of this work, through a detailed user study, the impacts of this technology on mental health, emotional well-being and work performance will be analysed. Furthermore, the ethical and social implications of an increasingly intimate interaction between man and machine in the field of personalised emotional assistance will be assessed.

ACKNOWLEDGMENTS

This work was partial supported by the following projects: 'SECURE SAFE APULIA', "LIFE: the itaLian system wIde Frailty nEtwork", PNRR-MAD-2022-12376656, CTEMT - 'Casa delle Tecnologie Emergenti di Matera', "IDENTITA - rete Integrata meDiterranea per losservazione ed Elaborazione di percorsi di Nutrizione", "PRINN 2022 PNR" - "The Words of Peace and Pacifism. French Literature in the Inter-war period by exploiting Distributional Semantic Analysis". Also, this work was carried out while Tommaso Colafiglio was enrolled in the Italian National Doctorate on Artificial Intelligence run by Sapienza University of Rome in collaboration with Politecnico di Bari.

REFERENCES

[1] Ali Ahmadvand, Harshita Sahijwani, and Eugene Agichtein. 2020. Would you like to talk about sports now? towards contextual topic suggestion for open-domain conversational agents. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. 83–92.

[2] Carmelo Ardito, Ilaria Bortone, Tommaso Colafiglio, Tommaso Di Noia, Eugenio Di Sciascio, Domenico Lofù, Fedelucio Narducci, Rodolfo Sardone, and Paolo Sorino. 2022. Brain computer interface: Deep learning approach to predict human emotion recognition. In *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2689–2694.

[3] John Atkinson and Daniel Campos. 2016. Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Systems with Applications* 47 (2016), 35–41.

[4] Maria Teresa Baldassarre, Danilo Caivano, Berenice Fernandez Nieto, Domenico Gigante, and Azzurra Ragone. 2023. The social impact of generative ai: An analysis on chatgpt. In *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*. 363–373.

[5] N Barascud. [n. d.]. meegkit: EEG and Meg denoising in Python.

[6] Christian Becker, Stefan Kopp, and Ipke Wachsmuth. 2004. Simulating the emotion dynamics of a multimodal conversational agent. In *tutorial and research workshop on affective dialogue systems*. Springer, 154–165.

[7] Giovanni Maria Biancofiore, Yashar Deldjoo, Tommaso Di Noia, Eugenio Di Sciascio, and Fedelucio Narducci. 2024. Interactive Question Answering Systems: Literature Review. *ACM Comput. Surv.* (apr 2024).

[8] Giovanni Maria Biancofiore, Tommaso Di Noia, Eugenio Di Sciascio, Fedelucio Narducci, and Paolo Pastore. 2022. Aspect based sentiment analysis in music: a case study with spotify. In *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*. 696–703.

[9] Pietro Boccadoro, Vitanio Daniele, Pietro Di Gennaro, Domenico Lofù, and Pietro Tedeschi. 2022. Water quality prediction on a sigfox-compliant iot device: The road ahead of waters. *Ad Hoc Networks* 126 (2022), 102749.

[10] Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. COMET: Commonsense Transformers for Automatic Knowledge Graph Construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 4762–4779.

[11] Joseph D Bronzino. 2000. *Biomedical Engineering Handbook 2*. Vol. 2. Springer Science & Business Media.

[12] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *NeurIPS*.

[13] Erik Brynjolfsson, Danielle Li, and Lindsey R Raymond. 2023. *Generative AI at work*. Technical Report. National Bureau of Economic Research.

[14] Tommaso Colafiglio, Domenico Lofù, Paolo Sorino, Fabrizio Festa, Tommaso Di Noia, and Eugenio Di Sciascio. 2023. Exploring the Mental State Intersection by Brain-Computer Interfaces, Cellular Automata and Biofeedback. In *IEEE EUROCON 2023-20th International Conference on Smart Technologies*. IEEE, 461–466.

[15] Tommaso Colafiglio, Paolo Sorino, Domenico Lofu, Angela Lombardi, Fedelucio Narducci, and Tommaso Di Noia. 2023. Combining Mental States Recognition and Machine Learning for Neurorehabilitation. In *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 3848–3853.

[16] Giandomenico Cornacchia, Vito Walter Anelli, Giovanni Maria Biancofiore, Fedelucio Narducci, Claudio Pomo, Azzurra Ragone, and Eugenio Di Sciascio. 2023. Auditing fairness under unawareness through counterfactual reasoning. *Information Processing & Management* 60, 2 (2023), 103224.

[17] Julian De Freitas, Ahmet Kaan Uğuralp, Zeliha Oğuz-Uğuralp, and Stefano Puntoni. 2022. Chatbots and mental health: insights into the safety of generative AI. *Journal of Consumer Psychology* (2022).

[18] Jiawen Deng and Fuji Ren. 2021. A survey of textual emotion recognition and its challenges. *IEEE Transactions on Affective Computing* 14, 1 (2021), 49–67.

[19] William W Eaton, Silvia S Martins, Gerald Nestadt, O Joseph Bienvenu, Diana Clarke, and Pierre Alexandre. 2008. The burden of mental disorders. *Epidemiologic reviews* 30, 1 (2008), 1–14.

[20] Jianfeng Gao, Chenyan Xiong, and Paul Bennett. 2020. Recent Advances in Conversational Information Retrieval. In *SIGIR*. ACM, 2421–2424.

[21] Karen Glanz, Barbara K Rimer, and K Viswanath. 2008. Theory, research, and practice in health behavior and health education. (2008).

[22] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti S. Hämäläinen. 2013. MEG and EEG Data Analysis with MNE-Python. *Frontiers in Neuroscience* 7, 267 (2013), 1–13. https://doi.org/10.3389/fnins.2013.00267

[23] Mark D. Griffiths, Daria J. Kuss, and Zsolt Demetrovics. 2014. Chapter 6 - Social Networking Addiction: An Overview of Preliminary Findings. In *Behavioral Addictions*. Academic Press, 119–141. https://doi.org/10.1016/B978-0-12-407724-9.00006-9

[24] Essam H Houssein, Asmaa Hammad, and Abdelmgeid A Ali. 2022. Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review. *Neural Computing and Applications* 34, 15

(2022), 12527–12557.

[25] Md Rabiul Islam, Md Milon Islam, Md Mustafizur Rahman, Chayan Mondal, Suvojit Kumar Singha, Mohiuddin Ahmad, Abdul Awal, Md Saiful Islam, and Mohammad Ali Moni. 2021. EEG channel correlation based model for emotion recognition. *Computers in Biology and Medicine* 136 (2021), 104757.

[26] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2011. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing* 3, 1 (2011), 18–31.

[27] Sajeev Kunjan, Tyler S Grummett, Kenneth J Pope, David MW Powers, Sean P Fitzgibbon, T Bastiampillai, M Battersby, and Trent W Lewis. 2021. The necessity of leave one subject out (LOSO) cross validation for EEG disease diagnosis. In *Brain Informatics: 14th International Conference, BI 2021, Virtual Event, September 17–19, 2021, Proceedings 14*. Springer, 558–567.

[28] Lei Li, Yongfeng Zhang, and Li Chen. 2023. Prompt distillation for efficient llm-based recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 1348–1357.

[29] Xiang Li, Yazhou Zhang, Prayag Tiwari, Dawei Song, Bin Hu, Meihong Yang, Zhigang Zhao, Neeraj Kumar, and Pekka Marttinen. 2022. EEG based emotion recognition: A tutorial and review. *Comput. Surveys* 55, 4 (2022), 1–57.

[30] Chung-Ying Lin, Peyman Namdar, Mark D Griffiths, and Amir H Pakpour. 2021. Mediated roles of generalized trust and perceived social support in the effects of problematic social media use on mental health: A cross-sectional study. *Health Expectations* 24, 1 (2021), 165–173.

[31] Guangya Liu, Xiao Dong, Meng-xiang Wang, Jianxing Yu, Mengjiao Gan, Wei Liu, and Jian Yin. 2023. Multi-modal Multi-emotion Emotional Support Conversation. In *International Conference on Advanced Data Mining and Applications*. Springer, 293–308.

[32] Domenico Lofù, Pietro Di Gennaro, Pietro Tedeschi, Tommaso Di Noia, and Eugenio Di Sciascio. 2023. URANUS: Radio Frequency Tracking, Classification and Identification of Unmanned Aircraft Vehicles. *IEEE Open Journal of Vehicular Technology* (2023).

[33] Dar Meshi and Morgan E Ellithorpe. 2021. Problematic social media use and social support received in real-life versus on social media: Associations with depression, anxiety and social isolation. *Addictive Behaviors* 119 (2021), 106949.

[34] Wei Peng, Yue Hu, Luxi Xing, Yuqiang Xie, Yajing Sun, and Yunpeng Li. 2022. Control Globally, Understand Locally: A Global-to-Local Hierarchical Graph Network for Emotional Support Conversation. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI*. ijcai.org, 4324–4330.

[35] Wei Peng, Ziyuan Qin, Yue Hu, Yuqiang Xie, and Yunpeng Li. 2023. Fado: Feedback-aware double controlling network for emotional support conversation. *Knowledge-Based Systems* 264 (2023), 110340.

[36] Rosalind W Picard. 2000. *Affective computing*. MIT press.

[37] Kalyan Prasad Agrawal. 2023. Towards adoption of generative AI in organizational settings. *Journal of Computer Information Systems* (2023), 1–16.

[38] Stephen A Rains, Corey A Pavlich, Bethany Lutovsky, Eric Tsetsi, and Anjali Ashtaputre. 2020. Support seeker expectations, support message quality, and supportive interaction processes and outcomes: The case of the comforting computer program revisited. *Journal of Social and Personal Relationships* 37, 2 (2020), 647–666.

[39] Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, et al. 2021. Recipes for Building an Open-Domain Chatbot. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. 300–325.

[40] Ronald Rosenfeld. 2000. Two decades of statistical language modeling: where do we go from here? *Proc. IEEE* 88, 8 (2000), 1270–1278.

[41] Marcin Skowron. 2010. Affect listeners: Acquisition of affective states by means of conversational systems. In *Development of Multimodal Interfaces: Active Listening and Synchrony: Second COST 2102 International Training School, Dublin, Ireland, March 23-27, 2009, Revised Selected Papers*. Springer, 169–181.

[42] Vishvesh Soni. 2023. Impact of Generative AI on Small and Medium Enterprises' Revenue Growth: The Moderating Role of Human, Technological, and Market Factors. *Reviews of Contemporary Business Analytics* 6, 1 (2023), 133–153.

[43] Paolo Sorino, Angelo Campanella, Caterina Bonfiglio, Antonella Mirizzi, Isabella Franco, Antonella Bianco, Maria Gabriella Caruso, Giovanni Misciagna, Laura R Aballay, Claudia Buongiorno, et al. 2021. Development and validation of a neural network for NAFLD diagnosis. *Scientific Reports* 11, 1 (2021), 20240.

[44] Paolo Sorino, Maria Gabriella Caruso, Giovanni Misciagna, Caterina Bonfiglio, Angelo Campanella, Antonella Mirizzi, Isabella Franco, Antonella Bianco, Claudia Buongiorno, Rosalba Liuzzi, et al. 2020. Selecting the best machine learning algorithm to support the diagnosis of Non-Alcoholic Fatty Liver Disease: A meta learner study. *PLoS One* 15, 10 (2020), e0240867.

[45] Rossella Tatoli, Luisa Lampignano, Ilaria Bortone, Rossella Donghia, Fabio Castellana, Roberta Zupo, Sarah Tirelli, Sara De Nucci, Annamaria Sila, Annalidia Natuzzi, et al. 2022. Dietary patterns associated with diabetes in an older population from Southern Italy using an unsupervised learning approach. *Sensors* 22, 6 (2022), 2193.

[46] Quan Tu, Yanran Li, Jianwei Cui, Bin Wang, Ji-Rong Wen, and Rui Yan. 2022. MISC: A Mixed Strategy-Aware Model integrating COMET for Emotional Support Conversation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 308–319.

[47] Krzysztof Wach, Cong Doanh Duong, Joanna Ejdys, Rūta Kazlauskaitė, Pawel Korzynski, Grzegorz Mazurek, Joanna Paliszkiewicz, and Ewa Ziemba. 2023. The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review* 11, 2 (2023), 7–30.

[48] Mayur Wankhade, Annavarapu Chandra Sekhara Rao, and Chaitanya Kulkarni. 2022. A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review* 55, 7 (2022), 5731–5780.

[49] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems* 32 (2019).

[50] JD Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21.

[51] Chujie Zheng, Sahand Sabour, Jiaxin Wen, Zheng Zhang, and Minlie Huang. 2023. AugESC: Dialogue Augmentation with Large Language Models for Emotional Support Conversation. In *Findings of the Association for Computational Linguistics: ACL 2023*. 1552–1568.

[52] Yongqiang Zheng, Jie Ding, Feng Liu, and Dongqing Wang. 2023. Adaptive neural decision tree for EEG based emotion recognition. *Information Sciences* 643 (2023), 119160.

[53] Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. 2018. Emotional chatting machine: Emotional conversation generation with internal and external memory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

Received 20 February 2024; revised 12 March 2024; accepted 5 June 2024