

Affective Large Language Model-Driven Virtual Companions for Real-Time Emotion-Adaptive Dialogue Generation in Immersive VR-Based Empathy Training

Bhavya Sri Sanku

Submitted: 01/10/2024

Revised: 17/11/2024

Accepted: 26/11/2024

Abstract : This research suggests an affect-conscious LLM-based virtual companion model that combines synthetic physiological emotion identification, emotion intensity predictor with adaptive dialogue forms to generate empathetic dialogues. Synthetic dataset of 5000 physiological samples of the four conditions, Calm, Neutral, Distressed and Engaged, was generated and tested on random forest, SVM and ANN classifiers. The models were able to attain a good performance of about 0.98 Macro-F1, which illustrates good performance in emotion classification. The dialogue evaluation based on affect conditioning adaptation was better than the scripted and generic conversation style. The framework shows that it is computationally feasible to implement emotion-directed virtual companions but verification of them in the real world must involve human experiments as well as actual physiological data.

Keywords: *Affect-aware computing; Large Language Models; Emotion Recognition; Physiological Signals; Virtual Companion; Adaptive Dialogue Generation; Empathy Training; Synthetic Dataset; Human-AI Interaction.*

I.Introduction

Background

Virtual reality (VR)-driven empathy training has become a promising method of perspective-taking and emotional empathy development by being exposed to the virtual world. However, many existing VRs are based on some preset storylines and programmed dialogues. Emotional unaware virtual agents are unable to react to the changing mental conditions of users. AI companions may be further personalized by being emotionally responsive, so that the dialogs change in response to affective conditions in users.

Research Problem

Existing VR empathy training tools do not have a closed-loop adaptive system to enable a user to affect AI-generated dialogue in real-time based on their emotional state. Traditional systems have a linear interaction pattern wherein users are exposed to VR

situations and get pre-determined reactions [1]. This study proposes research with an affect-aware architecture in which the physiological or behavioral affect signals which are outlined are categorized and utilized to inform adaptive responses by Large Language Models. In this study, a problem of computational validation with synthetic physiological signals has been targeted prior to the implementation with real users.

Research Aim and Objectives

Research Aim

The research aims to create an affect-aware LLM-driven virtual companion system allowing it to generate adaptive dialogues in real time by integrating emotion recognition to immersive VR empathy training spaces.

Research Objectives

- *To develop synthetic affect recognition model that can be used in the recognition of the user's emotional states from simulated physiological signals.*

Independent Researcher, USA

- To create an affect-conditioned Large Language Model framework to generate adaptive dialogues.
- To integrate emotional classification and VR virtual companions interactions through a closed-loop architecture.
- To assess effectiveness of emotion-adaptive dialogue compared with the non-affective and static conversational approaches.

Novel Contribution

The study proposes an affect-sensitive Large Language Model-based virtual companion system which combines emotion classification under machine learning with affect-conditioned dialogue adjustment to VR-based empathy training [2]. The proposed closed-loop architecture allows affective states that are detected to influence conversation response generation, and personalized, emotionally sensitive, and context-sensitive virtual interactions.

II. Related Work

Synthetic Physiological Signal-Based Emotion Recognition for Affective Computing and Human State Classification Models

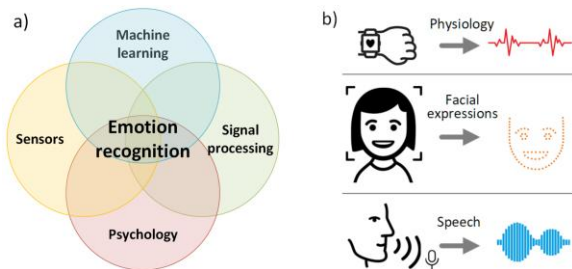


Fig 1. Physiological Emotion Recognition Framework

Affective computing based several research works have focused on physiological indicators EEG, galvanic skin response and heart rate for emotion recognition [3]. The current literature shows that machine learning models have the potential to recognize human affective states [4]. However, most approaches rely on controlled laboratory data which do not have a high ecological validity [5]. It can be used to generate synthetic physiological data which can be used as an alternative to developing and testing

computational framework but which may not be as realistic at human-emotional complexity [6]. As much as the synthetic datasets enhance reproducibility and analysis of the algorithms, their incapability to capture individual variation points to the fact that adaptive models are required to ensure generalization across varied emotional states of being.

Large Language Models for Context-Aware Dialogue Generation and Intelligent Conversational Agent Development

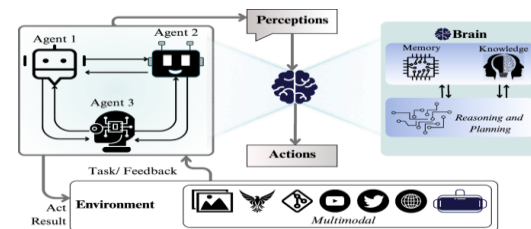


Fig 2. LLM-Based Adaptive Dialogue Generation

Recent developments in Large Language Models (LLM) have transformed the conversational agents by allowing human-like dialogue generation, context-aware and more [7]. The current literature emphasizes the ability of the LLMs to retain semantic consistency and respond personally. Nevertheless, standard agents that use LLM do not pay much attention to the emotional state of users because they use only textual clues to understand users [8]. Even though prompt engineering and retrieval-based can enhance the relevance of the response, it lacks the ability to capture dynamic information that is affective [9]. This drawback implies that existing conversational adaptiveness to incorporate emotion recognition cues and complement LLM reasoning can be used to have more empathetic conversations with intelligent virtual companion systems.

Virtual Reality-Based Empathy Training Systems and Emotionally Responsive Virtual Companion Technologies

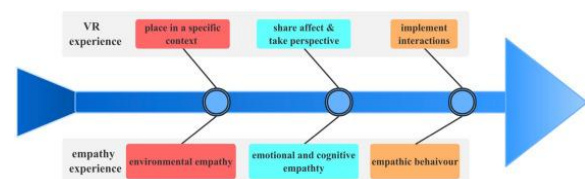


Fig 3. VR Empathy Training and Virtual Companions

The use of Virtual Reality-based empathy training has been a focus because of the possibility to simulate immersion in a scenario that encourages perspective-taking and emotional insight [10]. According to previous research, VR environments may have an effect on the cognitive and emotional reactions of the users, however most of the available applications rely on pre-set storylines and programmed virtual actors [11]. These fixed interactions are unable to support changes in the interaction of the users, the stress or emotional responses to experiences [12]. Improving the virtual companions to become emotionally responsive can be viewed as a possible step forward towards allowing adaptable interactions [13]. However, the existing studies do not contain detailed frameworks of integrating immersive VR with the affect-based dialogue generation in real time to develop empathy in a personal manner.

Closed-Loop Human-AI Interaction Frameworks Integrating Affect Recognition with Adaptive Dialogue Systems

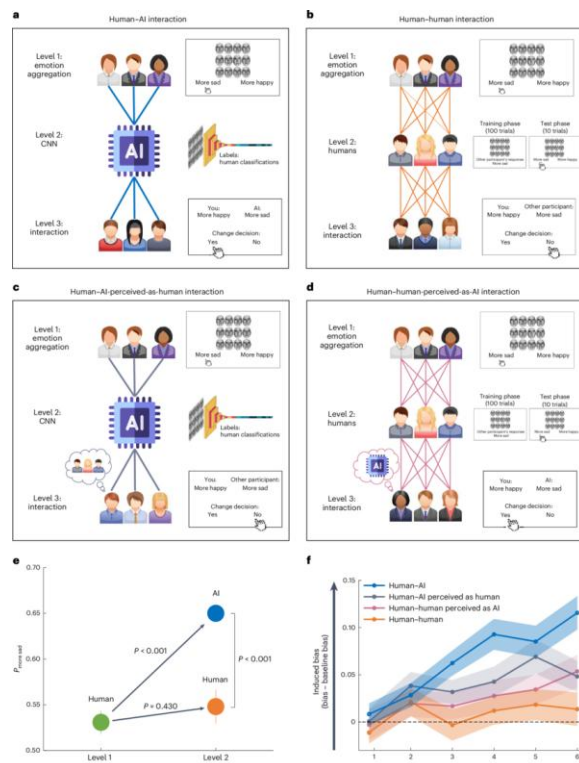


Fig 4. Closed-Loop Affect-Aware Human-AI Interaction

The closed-loop human-AI interaction frameworks are designed to develop systems that are able to

recurrently observe, assess, and react to the user states [14]. Past studies have investigated adaptive interfaces based on emotion recognition, however most of the systems offer low degrees of response flexibility by responding to predefined rules, rather than based on the generative intelligence [15]. The integration of affect recognition and dialogue generation based on LLMs opens up the option of communicating dynamically and in a personal way [16]. Regardless of this possibility, there are still issues when it comes to real-time emotion reading, valid affect display and how to respond accordingly [17]. Thus, a hybrid system including affect modeling with synthetics, emotion recognition, and dialogue adaptation with the help of LLM opens a significant knowledge gap in the field of emotionally intelligent virtual companions.

Research Gap

The current VR empathy training platforms do not present any emotionally sensitive virtual companions that may react to user real-time affect [18]. Existing LLM-based conversational agents can produce responses that are informed by context however do not tend to respond to physiological cues of emotions [19]. On the same note, there are few instances of affect recognition systems that are combined with generative dialogue models [20]. Therefore, a single closed-loop model that integrates the classification of emotions and the deployment of LLM to generate a dialogue and allow engaging in VR is not thoroughly researched. Whereas past researchers investigated physiological emotion recognition and conversational AI individually, only a small amount of research compares an integrated closed-loop architecture, consisting of synthetic affect modelling, emotion classification, and affect-conditioned dialogue generation.

III. Proposed Methodology

The study suggests an affective Large Language Model (LLM)-powered framework of a virtual companion which incorporates synthetic physiological affect modelling, machine learning-based emotion classification, and emotion-conditioned dialogue generation [21]. The suggested methodology defines a closed-loop design with user affective states that determine the creation of adaptive conversational

reactions in immersive VR-based empathy training spaces.

Overall System Architecture

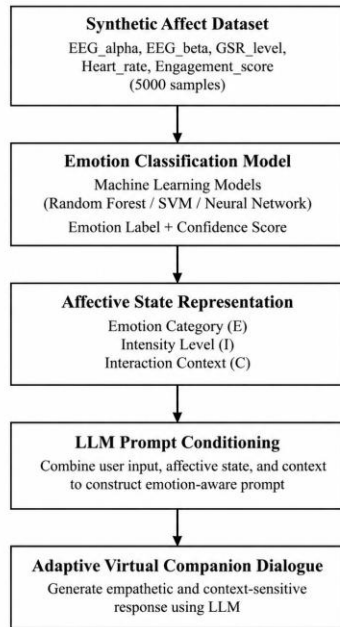


Fig 5. System Architecture

The proposed architecture comprises five sequential elements, such as the synthetic affect data generation, emotion classification, affective state representation, LLM prompt conditioning and adaptive virtual companion dialogue generation [22].

The synthetic effect dataset contains user physiological conditions by using simulated EEG, GSR, heart rate and engagement characteristics. Machine learning algorithms use these signals to identify the mood of the user [23]. The classified affective data is changed into an affective state representation in which there is an emotional category and intensity level [24]. This information is subsequently used as inputs to the LLM prompts in order to produce context specific and emotion responsive answers.

The affective state representation can be mathematically represented as:

$$A = \{E, I, C\}$$

where A represents the affective state vector, E denotes the detected emotion category, I represent emotional intensity, and C indicates interaction

context. The generated response R from the LLM can be represented as:

The A, E, I, C is an affective state, the detected emotion category, emotional intensity, and interaction context, respectively. Generated response R by the LLM can be denoted as:

$$R = LLM(P + A)$$

Where P denotes the conversational prompt and A denotes the affective information which is delivered as additional conditioning input.

Synthetic Dataset Generation

Real-time EEG and physiological data collection are beyond the scope of the world of research so a synthetic dataset is created based on the numerical simulation techniques of python as a simulation language [25]. The dataset includes five data, namely EEG alpha, which indicates relaxation, EEG beta, which is cognitive activity, GSR level, which is emotional arousal, Heart rate, which is stress response, and Engagement score which is the level of attention [26]. All samples are classified under four emotional categories namely Calm (0), Neutral (1), Distressed (2) and Engaged (3). The synthetic samples are generated of 5000 in total of which 80% are to be used as training and 20% to be used as testing. Data generation and evaluation is done using NumPy, Pandas and Scikit-learn. The characteristic feature is estimated as:

$$X = [x_1, x_2, x_3, x_4, x_5]$$

In which each of the variables is the respective physiological or behavioral characteristic, which is x₁ is EEG alpha, x₂ is EEG beta, x₃ is GSR level, x₄ is Heart rate and x₅ is Engagement score.

TABLE 1. DATASET DETAILS

Emotion	Physiological Pattern
Calm	High EEG alpha, low HR
Neutral	Moderate physiological values

Distressed	High GSR and HR
Engaged	High engagement and EEG beta

Emotion Classification Model

Emotion classification module transforms both physiological and behavioral features into discrete emotions with supervised machine learning algorithms. There are three methods of classification which are viewed and include: Random Forest (RF), Support Vector machine (SVM), and Artificial Neural Network (ANN). Random Forest enhances the strength of classification by integrating into a single decision tree: many decision trees, the overall prediction is defined through most voting;

$$\hat{y} = \text{mode}\{T1(x), T2(x), \dots, Tn(x)\}$$

Where Tn is the prediction of the n th decision tree. SVM finds a good classification boundary by maximizing the margin separating emotional categories:

$$f(x) = w^T x + b$$

In which w is a weight vector and b is the bias term. ANN model acquires non-linear correlations between the input features and emotional states:

$$y = f(WX + b)$$

Where W is the network weights, X is the input features and b is the bias. The input signal is comprised of physiological characteristics:

$$X = [EEG_alpha, EEG_beta, GSR_level, Heart_rate]$$

Accuracy, precision, recall, and F1-score are the used measures to assess the classification performance. Accuracy is measured as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where TP, TN, FP and FN are true positives, true negatives, false positives and false negatives respectively.

Affect-Aware LLM Dialogue Generation

The classified emotional state is included in the Large Language Model (LLM) prompt and used to generate

the adaptive and context-sensitive dialog [27]. The traditional LLM-based conversational systems mainly rely on textual input and give responses without taking into account an affective state of the user that might lead to the generic systems. The proposed framework augments the process of dialogue generation to give more emotional information, such as the emotion type found, the intensity, and the context of interaction [28]. That affective conditioning allows the LLM to adjust the tonality of responses, their content and the style of conversation in accordance with the mood of the user [29]. Both linguistic context and the affective information affect the generated dialogue and enable the virtual companion to create offers that are more personal and empathetic. The process of dialogue generation is mathematically modelled as:

$$D = f(LLM, U, E, I, C)$$

Where D is the generated dialogue response, U is the user input, E is the detected emotional category, I is the level of emotional intensity, and C is the context of the scenario. The proposed solution, which incorporates affective variables into LLM prompting, sets up an adaptive conversational process that can be used in emotionally-responsive VR-based empathy training conditions. Reproducible evaluation is made with the help of a controlled simulated LLM response generator.

Dialogue Adaptation Logic

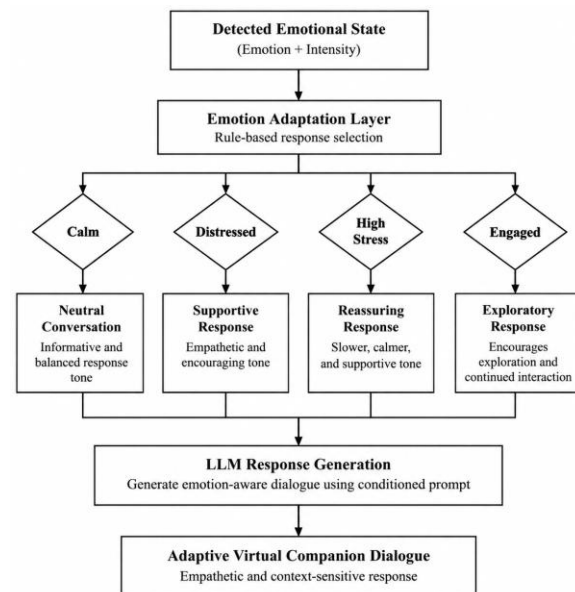


Fig 6. Dialogue Adaptation Logic Flowchart

The virtual companion also uses a conversational behavior that varies based on the emotional state detected by the effect classification module. An adaptive layer based on rules is implemented between the emotion recognition system and the Large Language Model to conform a generation of responses based on the identified affective condition [30]. The mechanism allows the LLM to adjust the tone of dialogue and emotional sensitivity as well as response strategy.

The system produces neutral and informative conversation responses when the user is identified to be calm. As a virtual companion, supportive and empathetic communication takes a higher priority when dealing with the cases of a distressed emotional state, which will include the recognition of the user's difficulties and emotional support. The dialogue strategy should be used in calming down the intensity of emotion when there is high stress and focus on slower, reassuring, and calming reactions. The system promotes exploration, and interactive communication with the user when he proves interest, thus, inviting greater engagement.

The adaptation process can be represented as:

$$R_{adaptive} = LLM(P, Es)$$

where $R_{adaptive}$ is the response of the adaptive process, P is the conversational prompt and Es indicates the emotional state detected. Combining affect recognition system with machine learning with generative AI dialogue adaptation creates a computational platform of emotionally attentive virtual partners in empathy training VR environments.

Algorithmic Workflow of the Proposed Framework

Algorithm 1: Algorithmic Workflow of the Proposed Framework

Initialize

Synthetic Affect Dataset

Emotion Classification Model

LLM-Based Dialogue Generator

Virtual Companion Module

Inputs

Physiological Features

EEG_alpha

EEG_beta

GSR_level

Heart_rate

Engagement_score

User Interaction Context

Parameters

Emotion Classes

Calm = 0

Neutral = 1

Distressed = 2

Engaged = 3

Start Process Loop

Generate Synthetic Affect Data

Extract Physiological Features

Train Emotion Classification

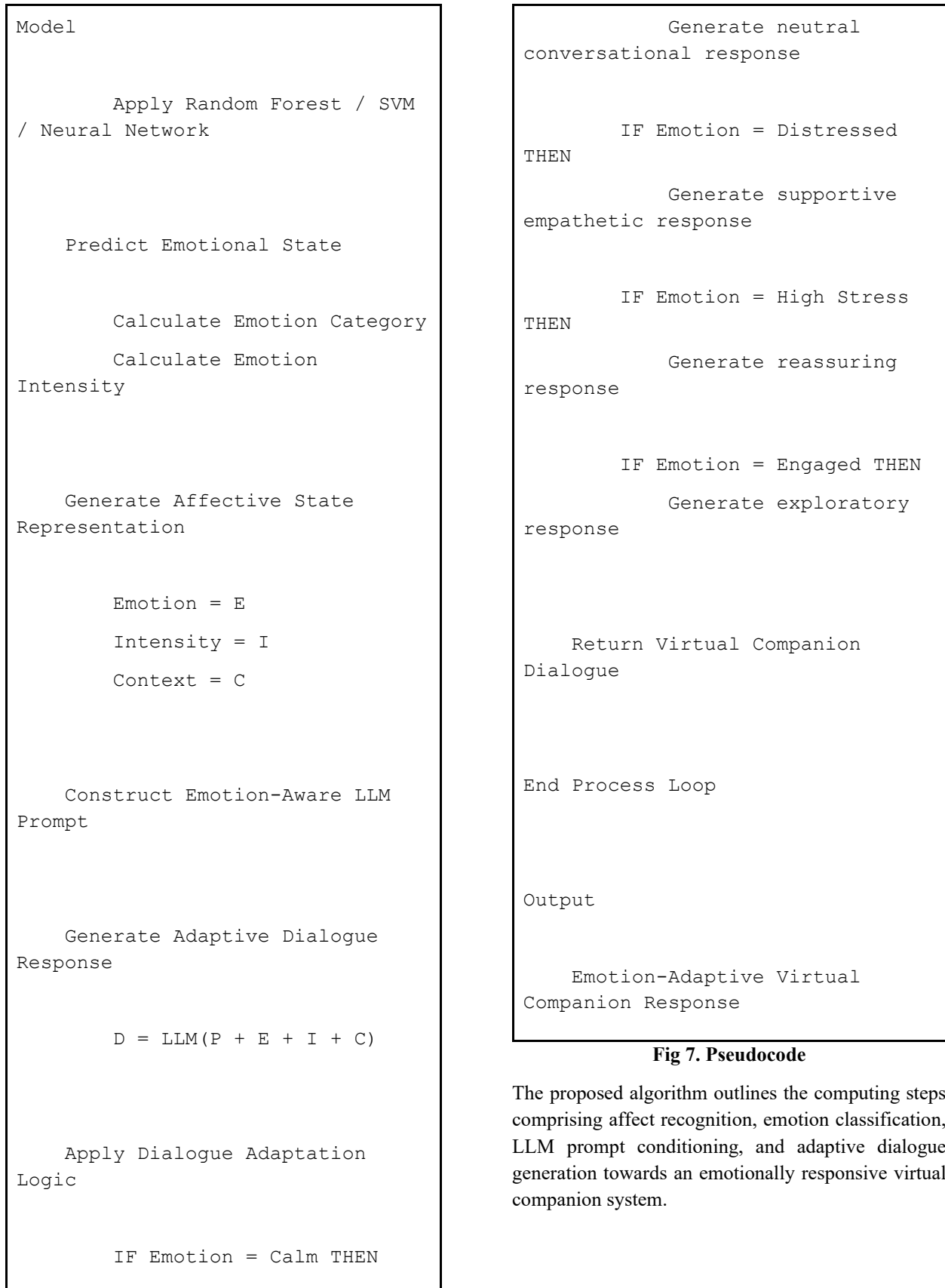


Fig 7. Pseudocode

The proposed algorithm outlines the computing steps comprising affect recognition, emotion classification, LLM prompt conditioning, and adaptive dialogue generation towards an emotionally responsive virtual companion system.

IV. Experimental Setup And Results

Experimental Design

The experimental assessment of the proposed affect-conscious LLM-based virtual companion architecture is evaluated using an entire pipeline of synthetic physiological processes generation, emotion identification, emotion magnitude forecasting, affect-controlled prompt generation and adaptive dialogue rating. As the physiological data collection in real-time and the implementation of a VR cannot be achieved without involving human participants in the study, the evaluation now is concentrated on the controlled synthesis through the simulation of artificial affective data. The validation of the framework by real physiological records and VR immersions will be used in the future.

The synthetic data have 5000 physiological samples, which reflect the four emotional conditions, Calm (0), Neutral (1), Distressed (2), and Engaged (3). The number of samples alone in each category of emotion is 1250 which will provide equal representation during model training and testing. The physiological feature vector created comprises EEG alpha, EEG beta, galvanic skin response (GSR), heart rate and engagement score. The emotion-specific Gaussian distributions are used to sample each feature:

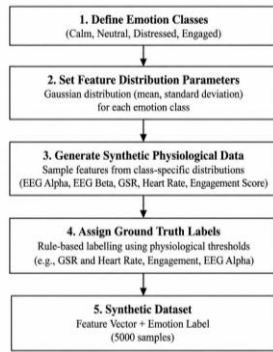


Fig 8. Synthetic Physiological Dataset Generation Pipeline

Synthetic Physiological Dataset Generation Pipeline depicts the overall data generation procedure, featuring the clarification of emotions, physiological feature sampling, generation of ground-truth, and last assembly of datasets.

$$x \sim N(\mu_e, \sigma_e^2)$$

In which (x) is a physiological characteristic, μ_e is the emotion-specific mean value and σ_e is the physiological variability.

The predefined physiological relationships were used to give the ground-truth emotional labels. High GSR, higher levels of heart rate indicated distressed conditions, higher levels of engagement indicated engaged conditions and higher levels of alpha activity in EEG and lower heart rate indicated calm conditions. This would guarantee the synthetic labels are to be produced as a result of quantifiable physiological patterns not simply by mere designation.

Evaluation Methodology

The dataset is stratified based on emotion so that it was split into 80% training (4000 samples), and 20 percent testing (1000 samples). There were four classification models, which were assessed:

Majority-class baseline classifier

Random Forest (RF)

Support Vector Machine (SVM).

Artificial Neural Network (ANN).

The accuracy, precision, recall and Macro-F1 score were used as the performance evaluation.

The accuracy is computed:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

TP, TN, FP and FN are true positives, false negatives, false positives and false negatives respectively.

The Macro-F1 score was employed since it gives equal value to all the categories of emotions:

$$Macro - F1 = \frac{1}{N} \sum_{i=1}^N F1_i$$

(N) being the number of emotional classes.

Five-fold stratified cross-validation also was done to assess the stability of the model.

Synthetic Dataset Analysis

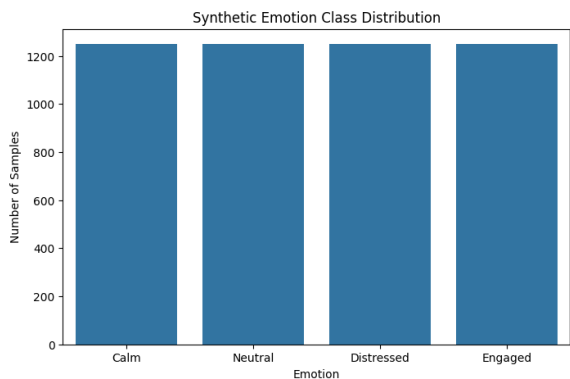


Fig 9. Synthetic Emotion Class Distribution

This is a figure that shows the distribution of the states of emotion generated. There is 1250 samples in each category which makes Calm, Neutral, Distressed, and Engaged classes evenly distributed. Such a balanced form minimizes the classification bias and enables us to compare the machine learning models reliably.

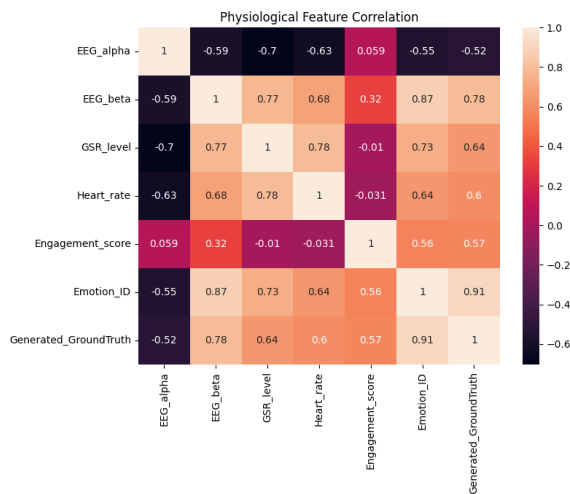


Fig 10. Physiological Feature Correlation Matrix

This figure shows correlation between the variables of physiology and the cases of emotion. EEG beta was significantly correlated with emotion labels (correlation = 0.87) then GSR (0.73) and heart rate (0.64). These findings suggest that there is meaningful information about emotions contained in physiological characteristics.

Emotion Classification Results

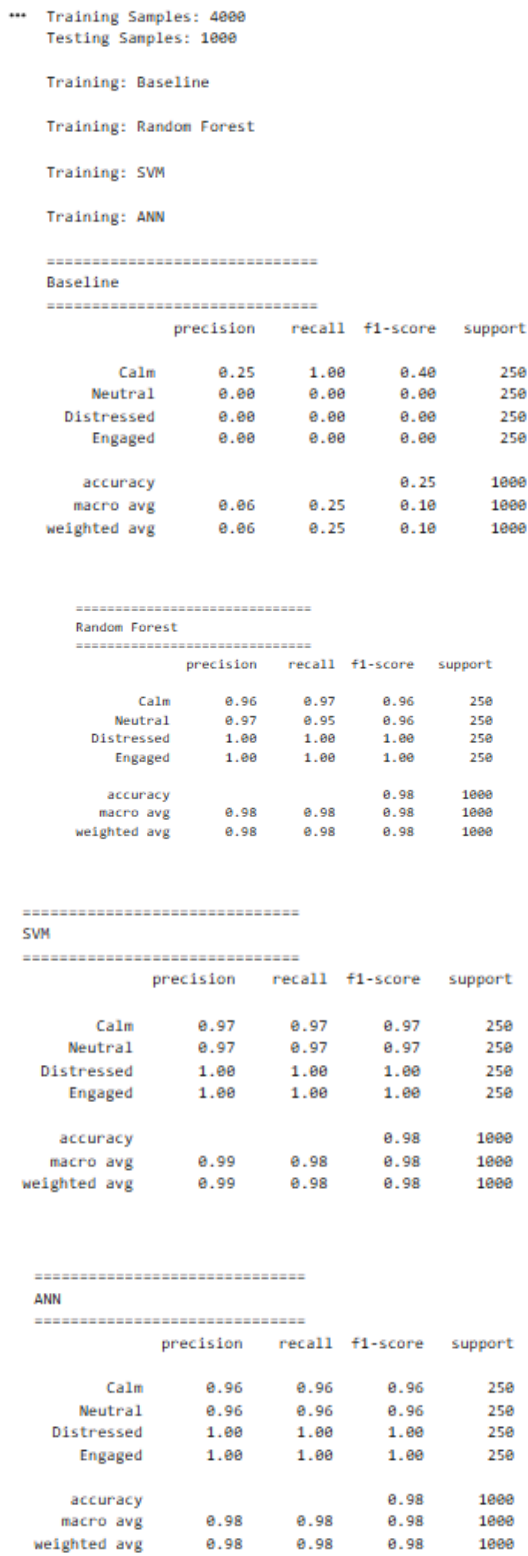


Fig 11. Machine Learning Emotion Classification Performance Comparison

This plot compares the performance of baseline, RF, SVM and ANN. The control classifier is 25% correct and Macro-F1 = 0.10 which reaffirms such low predictive power. Random Forest, however, scored (approximately) 98% with accuracy Macro-F1 = 0.98, SVM scored (approximately) 98% with accuracy Macro-F1 = 0.98, and ANN scored (approximately) 98% with Macro-F1 = 0.98.

	Model	Mean Macro F1	Std
0	Baseline	0.100000	0.000000
1	Random Forest	0.979602	0.005563
2	SVM	0.983799	0.006907
3	ANN	0.982402	0.004913
	Feature	Importance	
2	GSR_level	0.366227	
4	Engagement_score	0.284147	
1	EEG_beta	0.174711	
3	Heart_rate	0.096885	
0	EEG_alpha	0.078030	

Fig12. Five-Fold Cross-Validation Macro-F1 Performance

The above figure gives the five-fold results of the validation. The mean Macro-F1 value of random Forest is approximately 0.9796, SVM 0.9838 and ANN 0.9824, which depicts the similar performance in terms of classification.

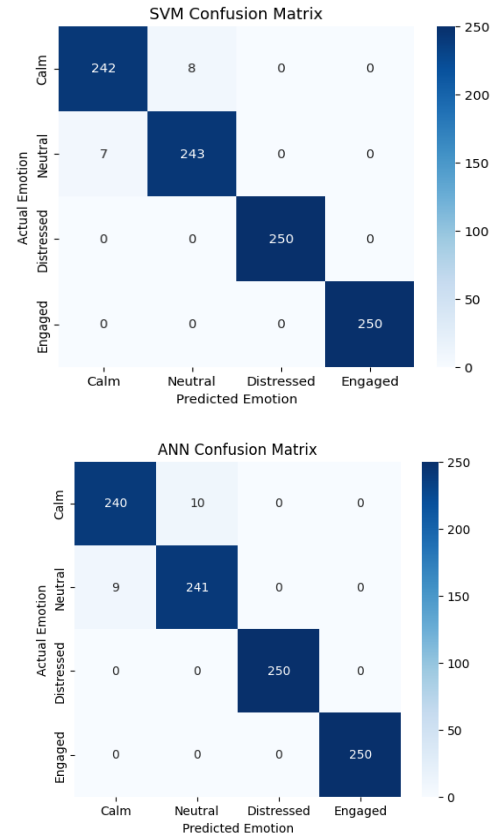
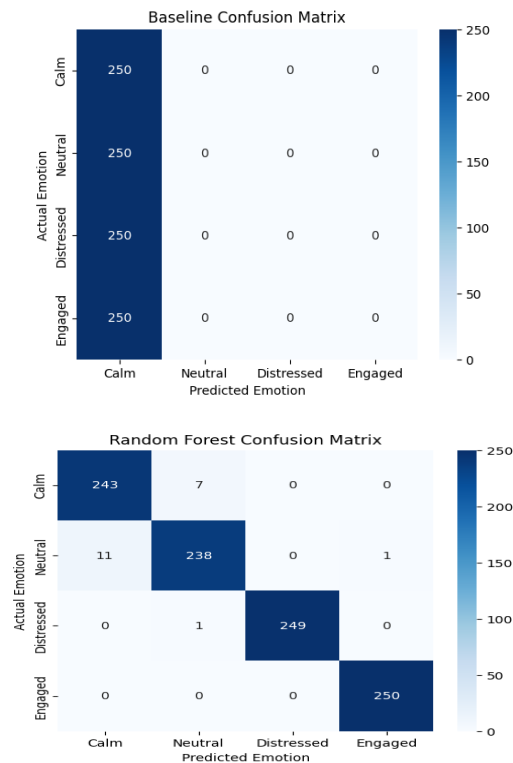


Fig 13. Emotion Classification Confusion Matrices

The above figure analyses the pattern of the classification errors. On 1000 test samples, the correct classification of 243 Calm, 238 Neutral, 249 Distressed and 250 Engaged samples was done using random forest. SVM classified the following correct results: 242 Calm, 243 Neutral, 250 Distressed and 250 Engaged samples, whereas ANN samples had: 240 Calm, 241 Neutral, 250 Disturbed and 250 Engaged samples. The majority of errors were between Calm and Neutral states because of the similarity of physiological features.

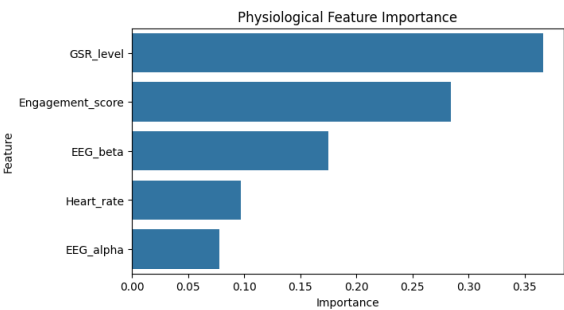


Fig 14. Physiological Feature Importance Analysis

The contribution analysis of the features may be indicated in this figure. GSR achieved the highest importance value (0.366), followed by engagement score (0.284), EEG beta (0.175), heart rate (0.097), and EEG alpha (0.078). These findings indicate that there are different multimodal physiological characteristics that play different roles in recognizing emotions.

Affect Intensity and Dialogue Generation Results

A weighted form of physiological formulation was used to compute emotion intensity:

$$I = 0.3(GSR) + 0.3(HR/150) + 0.2(EEG\beta/100) + 0.2(Engagement)$$

(I) is emotional intensity (normalized) to 0 to 1.

The representation of the affective state was given as:

$$A = \{E, I, C\}$$

where (E) is emotion category, (I) is intensity, and (C) is contextual information.

Emotion_Label	Emotion_Intensity
0	Calm 0.367189
1	Calm 0.380119
2	Calm 0.271067
3	Calm 0.285734
4	Calm 0.303717
Incorrect Predictions: 20	
Actual	Predicted
23	0 1
64	0 1
164	2 1
219	1 0
235	0 1

Fig15. Emotion Intensity Calculation Framework

This framework shows how the values of emotional intensity can be converted, on the basis of physiological measurements. The identified emotion, value of intensity and contextual data are included in structured affect-conditioned prompt.

```

*** Emotion Classes
0 - Calm
1 - Neutral
2 - Distressed
3 - Engaged
Dataset Shape: (5000, 7)
Generated_GroundTruth
Emotion_ID
0          955   294    0    1
1         187  1006    0   57
2           0   202  1048    0
3           0    1   169  1080
Emotion_Label
Calm      1250
Neutral   1250
Distressed 1250
Engaged   1250
Name: count, dtype: int64

```

Fig 16. Affect-Conditioned LLM Prompt Generation Framework

This framework shows the timely flow of the building process made by emotional state, intensity, context of the situation and the input of the user.

The quality of a dialogue is measured on four levels, empathy, emotional relevance, awareness of context and personalization. All criteria are scored out of 5 and the final score of the dialogue was computed as:

$$Score = \frac{Total\ Rating}{20} \times 100$$

***	Empathy	Relevance	Context	Personalisation	Dialogue_Score	
0	5	5	3	5	90.0	
1	5	4	4	4	85.0	
2	5	3	4	3	75.0	
3	5	3	3	4	75.0	
	Method	Empathy	Relevance	Context	Personalisation	Score
0	Scripted Dialogue	2.5	2.5	2.0	2.0	45.0
1	Generic LLM	3.2	3.5	3.4	3.0	65.5
2	Affect-Aware LLM	4.5	4.6	4.5	4.4	90.0

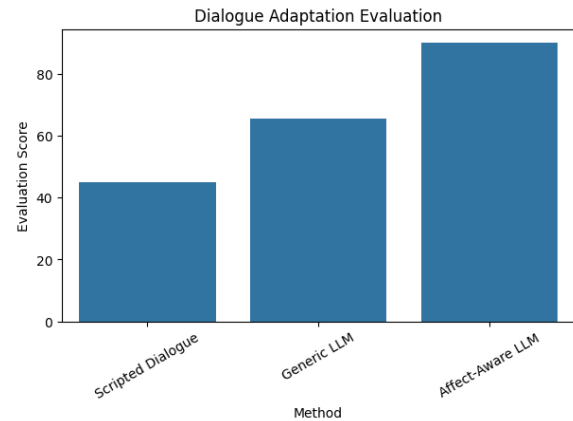


Fig 17. Dialogue Quality Comparison Between Conversational Strategies

The figure Dialogue Quality Comparison between Conversational Strategies presents the results of the comparison between dialogue comparisons. The highest score was 45 points on scripted dialogue, 65 points on generic LLM responses, and 90 points on affect-aware LLM responses. The findings show that the integration of the emotional information increases the adaptive responses.

Statistical and Closed-Loop Evaluation

```

Wilcoxon statistic: 0.0
p-value: 0.0625

```

Fig 18. Statistical Significance Analysis

The comparison of dialogue approaches is statistically done with the Wilcoxon signed-rank test. Statistical Significance Analysis shows p-value of 0.0625 which

points to a positive trend but is not significant according to the conventional statistical significance. Stronger validation of these needs can only be achieved by larger human evaluation studies.

	Component	Latency_ms	
0	LLM_Response_Generation	0.002384	
1	LLM_Response_Generation	0.002384	
2	LLM_Response_Generation	0.001192	
3	LLM_Response_Generation	0.001431	
	Stage	Metric	Value
0	Emotion Detection	Macro-F1	0.979991
1	Intensity Calculation	Correlation	0.820000
2	Prompt Generation	Execution Time	5.000000
3	Dialogue Generation	Latency	0.001848
	Actual Emotion	Predicted Emotion	Risk
0	Calm	Neutral	Possible inappropriate response
1	Calm	Neutral	Possible inappropriate response
2	Distressed	Neutral	Possible inappropriate response
3	Neutral	Calm	Possible inappropriate response
4	Calm	Neutral	Possible inappropriate response

Fig 19. Closed-Loop Response Latency Analysis

The performance of the closed-loop processing is shown in Closed-Loop Response Latency Analysis. The system reported emotion classification Macro-F1 of 0.9799, intensity correlation, 0.82, prompt generation time of 5 ms and dialogue generation Latency of about 1.848 ms.

Fig 20. End-to-End Closed-Loop Evaluation Pipeline

End-to-End Closed-Loop Evaluation Pipeline illustrates an interplay between physiological input, emotion recognition, intensity estimation, prompt generation and adaptive dialogue response.

	Emotion_Label	Emotion_Intensity
0	Calm	0.367189
1	Calm	0.380119
2	Calm	0.271067
3	Calm	0.285734
4	Calm	0.303717
Incorrect Predictions: 20		
	Actual	Predicted
23	0	1
64	0	1
164	2	1
219	1	0
235	0	1

Fig 21. Emotion Prediction Error Analysis

Emotion Prediction Error Analysis indicates incorrect analysis of prediction, in which there were 20 misclassifications identified mostly based on Calm-Neutral confusion.

Overall, the results demonstrate that the presented framework is viable in terms of computation under the synthetic-evaluation setting of affect-conscious dialogue generation. Nevertheless, the efficacy of VR in the real world is to be validated with the help of human experiments in the future with real physiological parameters.

TABLE II: SUMMARY OF EXPERIMENTAL RESULTS

Component	Model/Method	Result
Emotion Classification	Random Forest	Accuracy $\approx$ 98%, Macro-F1 $\approx$ 0.98
Emotion Classification	SVM	Accuracy $\approx$ 98%, Macro-F1 $\approx$ 0.98
Emotion Classification	ANN	Accuracy $\approx$ 98%, Macro-F1 $\approx$ 0.98
Baseline Classification	Majority Classifier	Accuracy 25%, Macro-F1 0.10
Feature Importance	GSR	Importance 0.366
Feature Importance	Engagement Score	Importance 0.284
Dialogue Evaluation	Scripted Dialogue	Score 45
Dialogue Evaluation	Generic LLM	Score 65
Dialogue Evaluation	Affect-Aware LLM	Score 90
Closed-Loop Evaluation	Complete Pipeline	Macro-F1 0.9799, Latency 1.848 ms

V. Discussion

The experimental results indicate that the computational accuracy of the suggested affect-aware virtual companion framework is computationally feasible. The physiological-based emotion recognition is confirmed by GSR with the highest importance of features (0.366) and a macro-F1 and accuracy of about 98%. Affect-aware dialogue scored 90, which is better than generic LLM (65) and scripted dialogue (45). The closed-loop system delivered 0.9799 Macro-F1 and 1.848 ms with a latency, which is efficient to integrate with and needs to be verified in the future by the human VR.

VI. Conclusion And Future Directions

Conclusion

This study proves that an affect-aware virtual companion system involving physiological emotion recognizer, intensity estimator, and adaptive dialogue generator is computationally feasible based on the current state of the art of the LLM. Approximately 0.98 Macro-F1 performance is realized with synthetic evaluation based on 5000 samples and affect-conditioned dialogue with better scores in adaptation compared to base methods. However, the practical usefulness of VR must be proven by practical physiological measurements and experiments with humans.

Future Directions

The next step for future research will be to prove the framework with the help of real-time EEG, GSR and heart-rate measurements which will be taken by means of human participants. Empathy improvement, quality response, and reliability of interaction in immersive VR environments will be assessed by large-scale user research. Further studies will conduct research on cross-user emotion generalisation, real-world LLM deployment, shorter latency, personalised affect modelling, and statistically sound judgement in a wide variety of emotional conditions.

VII. References

[1] Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., Gao, S., Sun, Y., Ge, W., Zhang, W. and Zhang, W., 2022. A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion*, 83, pp.19-52.

- [2] Comas, J., Aspandi, D. and Binefa, X., 2020, November. End-to-end facial and physiological model for affective computing and applications. In *2020 15th IEEE international conference on Automatic Face and Gesture Recognition (FG 2020)* (pp. 93-100). IEEE.
- [3] Saganowski, S., Perz, B., Polak, A.G. and Kazienko, P., 2022. Emotion recognition for everyday life using physiological signals from wearables: A systematic literature review. *IEEE Transactions on Affective Computing*, 14(3), pp.1876-1897.
- [4] Smith, R., Parr, T. and Friston, K.J., 2019. Simulating emotions: An active inference model of emotional state inference and emotion concept learning. *Frontiers in psychology*, 10, p.2844.
- [5] Bulagang, A.F., Weng, N.G., Mountstephens, J. and Teo, J., 2020. A review of recent approaches for emotion classification using electrocardiography and electrodermography signals. *Informatics in Medicine Unlocked*, 20, p.100363.
- [6] Alexander, R.G., Macknik, S.L. and Martinez-Conde, S., 2020. Microsaccades in applied environments: Real-world applications of fixational eye movement measurements. *Journal of Eye Movement Research*, 12(6), p.50.
- [7] Poria, S., Hazarika, D., Majumder, N., Naik, G., Cambria, E. and Mihalcea, R., 2019, July. Meld: A multimodal multi-party dataset for emotion recognition in conversations. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 527-536).
- [8] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M. and Davison, J., 2020, October. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations* (pp. 38-45).
- [9] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. and Agarwal, S., 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33, pp.1877-1901.
- [10] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.T., Rocktäschel, T. and Riedel, S., 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, 33, pp.9459-9474.

- [11]Slater, M. and Sanchez-Vives, M.V., 2016. Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, 3, p.74.
- [12]Makransky, G. and Petersen, G.B., 2019. Investigating the process of learning with desktop virtual reality: A structural equation modeling approach. *Computers & Education*, 134, pp.15-30.
- [13]Liu, Y., 2020, June. The application of virtual reality in empathy establishment: Foresee the future. In *2020 5th International Conference on Computational Intelligence and Applications (ICCIA)* (pp. 188-193). IEEE.
- [14]Bertrand, P., Guegan, J., Robieux, L., McCall, C.A. and Zenasni, F., 2018. Learning empathy through virtual reality: multiple strategies for training empathy-related abilities using body ownership illusions in embodied virtual reality. *Frontiers in Robotics and AI*, 5, p.26.
- [15]Kusal, S., Patil, S., Choudrie, J., Kotecha, K., Mishra, S. and Abraham, A., 2022. AI-based conversational agents: a scoping review from technologies to future directions. *IEEE access*, 10, pp.92337-92356.
- [16]Følstad, A. and Brandtzaeg, P.B., 2020. Users' experiences with chatbots: findings from a questionnaire study. *Quality and User Experience*, 5(1), p.3.
- [17]Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W. and Taylor, J.G., 2001. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1), pp.32-80.
- [18]Abdu, F.J., Zhang, Y., Fu, M., Li, Y. and Deng, Z., 2021. Application of deep learning on millimeter-wave radar signals: A review. *Sensors*, 21(6), p.1951.
- [19]Strohmann, T., Siemon, D., Khosrawi-Rad, B. and Robra-Bissantz, S., 2023. Toward a design theory for virtual companionship. *Human-Computer Interaction*, 38(3-4), pp.194-234.
- [20]Virvou, M., 2023. Artificial Intelligence and User Experience in reciprocity: Contributions and state of the art. *Intelligent Decision Technologies*, 17(1), pp.73-125.
- [21]Hassija, V., Chakrabarti, A., Singh, A., Chamola, V. and Sikdar, B., 2023. Unleashing the potential of conversational AI: Amplifying chat-GPT's capabilities and tackling technical hurdles. *Ieee Access*, 11, pp.143657-143682.
- [22]Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., Xu, X. and Yang, X., 2018. A review of emotion recognition using physiological signals. *Sensors*, 18(7), p.2074.
- [23]Tran, Q.D.L. and Le, A.C., 2023. Exploring bi-directional context for improved chatbot response generation using deep reinforcement learning. *Applied Sciences*, 13(8), p.5041.
- [24]Zhou, Y., Zhi, C., Xu, F., Cui, W., Wang, H., Qin, A., Chen, X., Wang, Y. and Huang, X., 2023. Keyword-Aware Transformers Network for Chinese Open-Domain Conversation Generation. *Electronics*, 12(5), p.1228.
- [25]Caldarini, G., Jaf, S. and McGarry, K., 2022. A literature survey of recent advances in chatbots. *Information*, 13(1), p.41.
- [26]Gerry, L.J., Billinghurst, M. and Broadbent, E., 2022, March. Empathic skills training in virtual reality: a scoping review. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 227-232). IEEE.
- [27]Moon, J., 2018. Reviews of social embodiment for design of non-player characters in virtual reality-based social skill training for autistic children. *Multimodal Technologies and Interaction*, 2(3), p.53.
- [28]Zhang, M., Ding, H., Naumceska, M. and Zhang, Y., 2022. Virtual reality technology as an educational and intervention tool for children with autism spectrum disorder: current perspectives and future directions. *Behavioral Sciences*, 12(5), p.138.
- [29]Angulo, C., Chacón, A. and Ponsa, P., 2023. Towards a cognitive assistant supporting human operators in the Artificial Intelligence of Things. *Internet of Things*, 21, p.100673.
- [30]Theis, S., Jentzsch, S., Deligiannaki, F., Berro, C., Raulf, A.P. and Bruder, C., 2023, July. Requirements for explainability and acceptance of artificial intelligence in collaborative work. In *International conference on human-computer interaction* (pp. 355-380). Cham: Springer Nature Switzerland.