

Predicting Emotion States Using Markov Chains

Clement Leung

School of Science and Engineering and
Guangdong Provincial Key Laboratory of
Future Networks of Intelligence
Chinese University of Hong Kong
Shenzhen, China
Email: clementleung@cuhk.edu.cn

Zhifei Xu

School of Science and Engineering
Chinese University of Hong Kong
Shenzhen, China
Email: zhifeixu1@link.cuhk.edu.cn

Abstract—In a wide range of tasks, especially those involving critical safety considerations, it is crucial that human participants maintain appropriate emotional conditions. As a result, accurate recognition of these emotional states has become a central research challenge, with mainstream methods frequently utilizing Pre-trained Language Models (PLMs) to incorporate emotional understanding. With the emergence of Large Language Models (LLMs) like ChatGPT, we have seen remarkable advancements in various natural language processing applications. However, the potential of ChatGPT’s zero-shot capabilities for image-based emotion recognition and analysis has not been thoroughly explored. In this study, we focus on classifying and predicting emotional states, specifically distinguishing between positive and negative emotions, and we examine ChatGPT4’s ability to interpret emotions directly from images. Our experiments show that ChatGPT4 can effectively predict changes in emotional states over time, surpassing expectations in identifying the progression of positive and negative emotions. Nonetheless, we identified shortcomings in its capacity to accurately recognize specific negative emotions, indicating room for further improvement.

Keywords—Image Emotion Prediction; Large Language Model; ChatGPT4; zero-shot; Markov Chain; Emotion Stability Parameter.

I. INTRODUCTION

In human communication, accurately representing and interpreting emotions is crucial. Emotions foster meaningful connections and reveal an individual’s mental state and intentions. Over the past decade, extensive research has focused on integrating emotional insight into human-computer dialogue systems [1]. Concurrently, the advent of ChatGPT [2] and Instruct-GPT [3] has sparked interest in their capacity for precise emotion recognition. Emotional support is increasingly essential in scenarios like personal conversations, mental health assistance, and customer interactions. Accordingly, our study investigates how effectively ChatGPT4 [4] can discern emotions from facial expressions.

Emotion recognition and prediction have gained prominence for promoting safety, supporting mental well-being, and enhancing user experiences. Recognized as a key factor in human safety, emotion recognition has been extensively researched [5] [6]. People naturally communicate emotions through words, text, images, facial cues, and physical gestures.

In a tech-driven world, Artificial Intelligence’s (AI) ability to understand and respond to human emotions is indispensable

[7]. The significance of emotionally sensitive AI is magnified by societal pressures such as occupational stress, perceived injustices, and the strain of personal breakups [8] [9], which can push individuals to harmful extremes. Evidence of such distress includes suicidal ideation linked to professional demands [8], school shootings, and road rage incidents. High-stakes roles—like surgeons, pilots, and truck drivers—require emotional stability, as demonstrated by a pilot with depression who attempted to shut down an airplane’s engines mid-flight [9]. These scenarios underscore the urgent need for advancements in emotion recognition and prediction to bolster safety and mental health [10]. In recent years, generating emotionally responsive outputs through neural networks has become a prominent research focus [11], driven by advancements in online social networks and deep learning technologies. Moreover, the continuous evolution of large language models has triggered a revolution in conversational AI, as exemplified by ChatGPT4. These models exhibit robust, general-purpose linguistic capabilities, offering unprecedented levels of semantic comprehension and nuanced response generation. Consequently, the quality of human-computer interaction has significantly improved. Yet, the extent to which these systems exhibit emotions within their dialogues remains largely unexplored. Our goal involves developing effective conversational strategies in ChatGPT4 and assessing recent progress, strengths, and limitations [12] [13] [14] in the realm of multi-modal emotion recognition and prediction tasks. Employing ChatGPT4 for emotion detection is also considered beneficial for maintaining fairness in experimental settings, as it can interpret data without the biases often seen in human evaluators. Emphasizing this approach not only fosters fairness but also prioritizes safety and well-being. By pinpointing emotional states accurately, interventions can be better tailored and more ethically executed, thereby safeguarding participants.

Furthermore, foundational research on emotion recognition and prediction dates back to Ekman’s widely recognized classification model [15], which identified six universal emotions: joy, sadness, fear, anger, surprise, and disgust. Building upon Ekman’s work, Plutchik proposed an arrangement of eight primary emotions—joy, trust, fear, surprise, sadness, disgust, anger, and anticipation—in a wheel-shaped model [16]. These approaches represent categorical or discrete models, positing a fixed set of universally understood basic emotions. In contrast,

continuum models treat emotions as existing along dynamic dimensions [17] [18], factoring in valence (ranging from positive to negative), arousal (level of excitement or calmness), and dominance (sense of influence or control).

Traditionally, emotion recognition and prediction research focused on a single channel of expression. However, people naturally communicate emotions through multiple modalities: voice, text, images, facial expressions, and body movements, making it challenging to accurately interpret emotions from just one source. To address this complexity, multimodal sentiment analysis integrates various data inputs, such as audio signals, shape changes, and overall appearance [19], often combined with text and images. Employing advanced techniques such as Convolutional Neural Networks (CNNs) [20] [21] or transformers enables more accurate and comprehensive emotion recognition and classification. Since emotions frequently evolve over time, predicting their progression is equally important in understanding real-world emotional dynamics.

Section III shows the results and analysis of the experiment, and Section IV discusses the experimental results. The conclusion and future work are presented in Section V.

II. MATERIALS AND METHODS

In this paper, we propose an emotion prediction model grounded in Markov chains and emotion stability parameters. By constructing an emotion state transition matrix and incorporating stability parameters, the model integrates eight basic emotional states and forecasts how emotions evolve. To verify its effectiveness, we conducted long-term emotion predictions and thoroughly traced how these emotional states change as time passes. Our experimental results show that this model can effectively capture dynamic emotional fluctuations, providing a novel approach to emotion analysis and prediction.

Emotions play a pivotal role in everyday life and human-computer interactions. Accurately predicting and analyzing shifting emotional states is of great importance in fields like psychology, artificial intelligence, and human-computer interaction [15] [16]. However, many existing methods lack a dynamic perspective and struggle to anticipate how emotions might evolve as time moves forward. To address this gap, we present an emotion prediction model based on Markov chains and emotion stability parameters, aiming to precisely predict long-term changes in emotional states.

In reality, human emotions are continuous and frequently shift from one state to another [22] [23]. Depending on the context, it may be necessary to foresee the emotional states of specific individuals, such as when scheduling surgeries in hospitals, managing pilots during flights, or assigning tasks in high-risk industries. As noted earlier, emotional responses can surface when individuals are fatigued or treated unfairly. In safety-critical jobs, we want the people involved to maintain stable emotional states [9], since those experiencing emotional difficulties can compromise the safety of others.

Within this study, we focus on a classification model guided by Ekman's framework, using six primary emotions: happiness, sadness, fear, anger, surprise, and disgust. These emotions are

categorized into positive and negative groups. Happiness and surprise are considered positive (+1), while sadness, fear, anger, and disgust are treated as negative (-1). Although certain high-risk scenarios might warrant a stricter classification—possibly moving surprise into the negative category—this paper retains surprise as a positive emotion.

Our model, denoted as $S(t)$, represents how an individual's emotional state changes over time, with t indicating the temporal dimension. We assign $S(t) = 1$ for positive emotions and $S(t) = -1$ for negative emotions. Human emotional complexity arises from external factors beyond personal control, such as financial stability, relationships, health, workplace conditions, market fluctuations, and family issues. These elements can trigger transitions from positive to negative emotional states or vice versa.

We begin by setting $S(0) = 1$. We then model the moments of emotional shifts (from +1 to -1 or the reverse) using a Poisson Process. Accordingly, $S(t) = 1$ if the number of transitions in the interval $(0, t)$ is even, and $S(t) = -1$ if it is odd. This approach captures the stochastic nature of emotional shifts and lays the groundwork for predicting long-term emotional evolution.

$$P[S(t) = 1|S(0) = 1] = p_0 + p_2 + p_4 + \dots + \dots, \quad (1)$$

where p_k is the number of Poisson points in $(0, t)$ with parameter λ . That is,

$$\begin{aligned} P[S(t) = 1|S(0) = 1] &= e^{-\lambda t} \left[1 + \frac{(\lambda t)^2}{2!} + \frac{(\lambda t)^4}{4!} \dots + \dots \right] \\ &= e^{-\lambda t} \cosh \lambda t \end{aligned} \quad (2)$$

Now, $S(t) = -1$ if the number of points in the time interval $(0, t)$ is odd; that is,

$$\begin{aligned} P[S(t) = -1|S(0) = 1] &= e^{-\lambda t} \left[1 + \frac{(\lambda t)^3}{3!} + \frac{(\lambda t)^5}{5!} \dots + \dots \right] \\ &= e^{-\lambda t} \sinh \lambda t \end{aligned} \quad (3)$$

Equation (2) represents the probability that the emotion is still positive at time t given that it was positive at time 0. Equation (3) gives the probability that the emotion is negative at time t given that it was positive at time 0. The parameter λ in both expressions represents a rate at which emotions change or decay over time. A larger value of λ would mean emotions change more rapidly, while a smaller value would mean they change more slowly. This is where we mathematically analyze possible emotional changes and predict them. Also, to verify the idea, we use ChatGPT. As for the experimental evaluation part, the Receiver Operating Characteristic (ROC) method was adopted to analyze the experimental results, and a specific explanation was placed in the experimental part.

Here, we briefly explain equation (3). First, we assume that λ is 0.3, 0.6, 0.9. Meanwhile, Figure 1 shows the function $e^{-\lambda t} \sinh(\lambda t)$ with λ values of 0.3, 0.6 and 0.9. Displaying the function $e^{-\lambda t} \sinh(\lambda t)$ with λ values of 0.3, 0.6

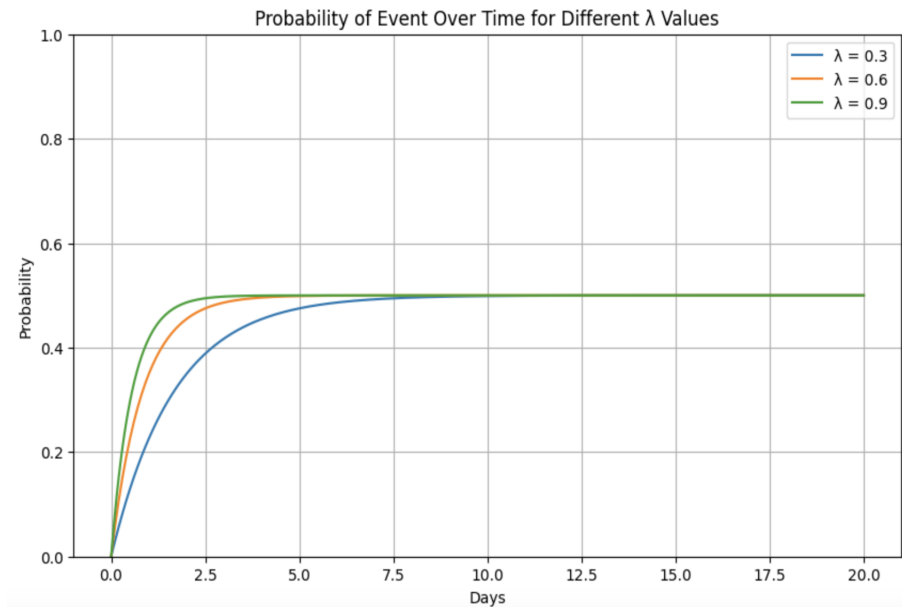


Figure 1. Different λ of the equation 3.

and 0.9, provides insightful observations about the temporal changes in probabilities, especially relevant to emotional states or comparable processes that can either diminish or progress over time. It is evident from the visualization that varying the decay constants (λ) significantly affects how long and intensely certain states persist across several days.

With a decay constant of $\lambda = 0.3$, the probability initially is high but diminishes gradually, indicating a persistent or slowly fading condition. For example, maintaining a negative emotional state translates to a higher likelihood of remaining in this state longer. Within the initial day, the probability stays well above 60%, denoting a strong endurance of the state. By day three, it hovers around 50%, showing a steady, yet noticeable decline. This gradual reduction might symbolize scenarios where the causes behind the emotional state are slow to be addressed or alleviated.

Increasing the decay constant to $\lambda = 0.6$ accelerates the probability's decline. This faster fall suggests a quicker fading of the emotional intensity or the likelihood of sustaining the same state. On the first day, the probability remains elevated but swiftly falls below 60%, nearing 50% by the close of the second day. This quicker reduction may be associated with effective interventions or environmental changes, or possibly better coping strategies that shorten the duration of the negative condition.

At $\lambda = 0.9$, the probability decreases even more swiftly. The graph shows a sharp descent, indicative of scenarios where negative emotional states or similar conditions dissipate very quickly. The probability does not stay above 50% beyond two days, dropping near this mark by the end of the first day. This rapid decline could point to highly effective external support or events that inherently do not have prolonged effects.

By analyzing these curves, one can determine how various

strategies or intrinsic elements affect the control or maintenance of specific states—be they emotional, physical, or of another nature. The differing λ values symbolically illustrate the varied speeds at which environments, individuals, or systems either normalize or transition from one state to another. This knowledge is essential in areas such as psychology, where predicting the duration of an individual's negative emotional state is key to developing timely and effective interventions. Insights into these temporal patterns are invaluable for customizing interventions or supports that are sensitive to timing and more closely correspond to the observed rates of change.

This section elaborates on the construction of the emotion prediction model, including the definition of emotion states, the construction of the state transition matrix, the introduction of emotion stability parameters, the calculation of emotion distributions over time steps, and the computation and ranking of emotion change probabilities. We will demonstrate the detailed derivation process of emotion changes over longer time steps $t = 0$ to $t = 5$. The emotion state vector $S(t)$ represents the probability distribution of emotions at time t :

$$S(t) = [P_{E_1}(t), P_{E_2}(t), \dots, P_{E_8}(t)]^T \tag{4}$$

where $P_{E_i}(t)$ denotes the probability of emotion E_i at time t .

The transition of emotion states is modeled using a Markov chain. The state transition matrix P represents the probability of transitioning from one emotional state to another. The matrix P is an 8×8 probability matrix, where each row sums to 1.

We assume the state transition matrix P to be:

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} & p_{16} & p_{17} & p_{18} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} & p_{26} & p_{27} & p_{28} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{81} & p_{82} & p_{83} & p_{84} & p_{85} & p_{86} & p_{87} & p_{88} \end{bmatrix}$$

where p_{ij} represents the probability of transitioning from emotion E_i to emotion E_j , satisfying:

$$\sum_{j=1}^8 p_{ij} = 1, \quad \forall i \in \{1, 2, \dots, 8\} \quad (5)$$

In other words, we set the following transition probabilities (the values are illustrative and can be adjusted based on actual conditions):

- Higher probabilities of transition among positive emotions and lower probabilities of transitioning to negative emotions.
- Higher probabilities of transition among negative emotions and lower probabilities of transitioning to positive emotions.

For example, when the emotion is Joy (E_1), the transition probabilities are:

$$p_{1j} = \begin{cases} 0.5, & \text{if } j = 1 \text{ (remain in Joy)} \\ 0.15, & \text{if } j = 2 \text{ (transition to Trust)} \\ 0.15, & \text{if } j = 3 \text{ (transition to Surprise)} \\ 0.1, & \text{if } j = 4 \text{ (transition to Anticipation)} \\ 0.05, & \text{if } j = 5 \text{ (transition to Sadness)} \\ 0.02, & \text{if } j = 6 \text{ (transition to Disgust)} \\ 0.02, & \text{if } j = 7 \text{ (transition to Anger)} \\ 0.01, & \text{if } j = 8 \text{ (transition to Fear)} \end{cases}$$

Transition probabilities for other emotions can be similarly defined, ensuring each row sums to 1. The emotion stability parameter λ_i is used to simulate the volatility of emotions in reality:

- Positive emotions have smaller λ_i , indicating they are more stable.
- Negative emotions have larger λ_i , indicating they are less stable.

We set:

$$\lambda_i = \begin{cases} 0.2, & \text{if } E_i \text{ is a positive emotion} \\ 0.5, & \text{if } E_i \text{ is a negative emotion} \end{cases}$$

Calculation of Emotion Distributions over Time Steps. We consider the Initial Emotion State to be $S(0)$ and we set the initial emotion as Joy (E_1):

$$S(0) = [1, 0, 0, 0, 0, 0, 0, 0]^T \quad (6)$$

State Transition Computation

At each time step t , the emotion state is updated using the state transition matrix P :

$$S(t) = P^T S(t-1) \quad (7)$$

where P^T is the transpose of P .

To consider the stability of emotions, we adjust the probability of each emotion at each time step. The probability of emotion E_i remaining the same at time t is:

$$P_{\text{stay}, E_i}(t) = P_{E_i}(t) \cdot e^{-\lambda_i t} \cosh(\lambda_i t) \quad (8)$$

The probability of emotion E_i transitioning is:

$$P_{\text{trans}, E_i}(t) = P_{E_i}(t) \cdot [1 - e^{-\lambda_i t} \cosh(\lambda_i t)] \quad (9)$$

The adjusted emotion probability is:

$$\tilde{P}_{E_i}(t) = P_{\text{stay}, E_i}(t) + \sum_{j \neq i} P_{\text{trans}, E_j}(t) \cdot p_{ji} \quad (10)$$

where p_{ji} is the probability of transitioning from emotion E_j to emotion E_i .

Recursive Calculation for Future Time Steps. We repeat the above steps to calculate the emotion distributions from time $t = 1$ to $t = 5$.

1) *Computation and Ranking of Emotion Change Probabilities:* First, we define the probability of emotion change.

The change probability of emotion E_i at time t is defined as:

$$\Delta P_i(t) = |\tilde{P}_{E_i}(t) - P_{E_i}(0)| \quad (11)$$

Based on $\Delta P_i(t)$, emotions are ranked to obtain the priority of emotion changes at time t . Below, we detail the calculation process of emotion distributions from time $t = 0$ to $t = 5$.

2) *At time Step $t = 1$:*

a) *State Transition Calculation:*

$$S(1) = P^T S(0)$$

Since $S(0)$ has only the first element as 1 and others as 0:

$$S(1) = P^T \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} p_{11} \\ p_{12} \\ \vdots \\ p_{18} \end{bmatrix}$$

Substituting specific values:

$$S(1) = \begin{bmatrix} 0.5 \\ 0.15 \\ 0.15 \\ 0.1 \\ 0.05 \\ 0.02 \\ 0.02 \\ 0.01 \end{bmatrix}$$

b) *Adjustment with Stability Parameters:* We compute the stay and transition probabilities for each emotion. For Joy (E_1):

$$P_{\text{stay}, E_1}(1) = P_{E_1}(1) \cdot e^{-\lambda_1 \times 1} \cosh(\lambda_1 \times 1) = 0.5 \cdot e^{-0.2} \cosh(0.2)$$

Calculating $e^{-0.2} \approx 0.8187$, $\cosh(0.2) \approx 1.0201$, so:

$$P_{\text{stay}, E_1}(1) \approx 0.5 \times 0.8187 \times 1.0201 \approx 0.4182$$

We make similar computations for the other emotions.

c) *Adjusted Emotion Probabilities*: Due to space constraints, only the adjusted emotion probabilities are provided:

$$\tilde{S}(1) = \begin{bmatrix} 0.4182 \\ 0.1228 \\ 0.1228 \\ 0.0819 \\ 0.0328 \\ 0.0123 \\ 0.0123 \\ 0.0061 \end{bmatrix}$$

We continue the iteration five times.

3) *Time Step* $t = 5$:

a) *Adjusted Emotion Probabilities*:

$$\tilde{S}(5) = \begin{bmatrix} 0.2071 \\ 0.1764 \\ 0.1764 \\ 0.1131 \\ 0.1259 \\ 0.0652 \\ 0.0652 \\ 0.0287 \end{bmatrix}$$

Probabilities of Emotional Change and Ranking

We calculate the probability of emotion change $\Delta P_i(t)$ at each time step and perform the ranking.

At time Step $t = 1$

- The change Probabilities:

$$\Delta P_i(1) = |\tilde{P}_{E_i}(1) - P_{E_i}(0)|$$

Are calculated as:

$$\Delta P_i(1) = \begin{bmatrix} 0.5818 \\ 0.1228 \\ 0.1228 \\ 0.0819 \\ 0.0328 \\ 0.0123 \\ 0.0123 \\ 0.0061 \end{bmatrix}$$

- **Ranking** (from largest to smallest):

- 1) Joy ($\Delta P_{E_1} = 0.5818$)
- 2) Trust ($\Delta P_{E_2} = 0.1228$)
- 3) Surprise ($\Delta P_{E_3} = 0.1228$)
- 4) Anticipation ($\Delta P_{E_4} = 0.0819$)
- 5) Sadness ($\Delta P_{E_5} = 0.0328$)
- 6) Disgust ($\Delta P_{E_6} = 0.0123$)
- 7) Anger ($\Delta P_{E_7} = 0.0123$)
- 8) Fear ($\Delta P_{E_8} = 0.0061$)

Similarly, we perform calculations and rankings for $t = 2$ to $t = 5$.

Through long-time-step emotion prediction and detailed derivation, we validated the effectiveness of the model. The introduction of the emotion state transition matrix and emotion stability parameters allows the model to capture the dynamic

changes of emotions over time and simulate the transition patterns among different emotions.

The model's prediction results align with real-world emotion evolution. For example, the higher transition probabilities among positive emotions and the longer time steps required for negative emotions to appear provide new perspectives for emotion analysis and prediction. This can be applied in fields such as mental health monitoring and human-computer interaction systems.

This paper proposes an emotion prediction model combining Markov chains and emotion stability parameters. Through detailed derivation and long-time-step emotion change calculations, we demonstrated the model's effectiveness in predicting dynamic changes of emotions. Future work can further optimize the settings of the state transition matrix and stability parameters and apply the model to actual datasets for validation.

Specific Values of State Transition Matrix P . Due to space limitations, the complete numerical values of the state transition matrix P are not listed here. Readers can set and adjust the matrix according to the methods described above.

III. RESULTS

Emotion prediction in conversation stands at the intersection of artificial intelligence and natural language processing. It involves using textual, visual, and even auditory information to identify and forecast the emotional states of participants within a dialogue. Such predictions have practical significance across various domains, including enhancing customer service interactions, assisting in mental health evaluations, and improving human-computer communication. Moreover, emotion predictions derived from conversational content can be evaluated by chatbots to determine a user's current emotional state and their reaction to emotional triggers. Given that ChatGPT4 functions as a conversational agent, an important question arises: can it effectively predict how emotions evolve over time?


A. Emotion Prediction with different situations

For the experimental part, we chose three Data sets from Kaggle which are Emotion Detection, Facial Expressions Training Data, and Natural Human Face Images for Emotion Recognition.

1) *Datasets: Emotion Dection* This dataset is the same as the FER-2013 [24] dataset. The collection features 35,685 grayscale images, each 48x48 pixels, organized into two sections: training and testing. Each section hosts a variety of images representing different emotional states. The images have been categorized by the creators into several emotions, namely anger, disgust, fear, happiness, neutrality, sadness, and surprise, providing a comprehensive basis for emotion detection tasks.

Facial Expression Training Data The AffectNet [25] database, a substantial compilation of facial images annotated with expressions, serves as the foundation for this dataset. To adapt to typical memory constraints, image resolution is scaled down to 96x96 pixels. The dataset employs Principal

TABLE I
SAMPLE OF FOUR DIFFERENT SITUATIONS

Dataset	Question 1	Question 2	Question 3	Question 4
	What is the emotion of this person? If they are about to be praised by their boss or their parents respectively, what do you think their emotions become?	If they were to be criticized, what do you think their emotions would be?	If they were to receive a \$1,000 reward, what do you think their emotions would be?	If they were to break up, what do you think their emotions would be?

Component Analysis, specifically focusing on the Singular Value Decomposition method, to enhance image processing efficiency. A threshold is applied to ensure the Principal Component’s percentage remains below 90%, primarily excluding monochrome images. The dataset, derived from the high-quality AffectNet repository and refined using advanced Facial Expression Recognition technology, spans eight emotional categories: anger, contempt, disgust, fear, happiness, neutrality, sadness, and surprise.

Natural Human Face Images for Emotion Recognition

Unlike traditional datasets used in facial expression recognition such as the Facial Expression Recognition (FER) dataset, the Extended Cohn-Kanade dataset (CK+) and the Karolinska Directed Emotional Faces dataset (KDEF), this unique dataset is curated from the Internet, encompassing more than 5,500 images manually labeled for eight emotional expressions: anger, contempt, disgust, fear, happiness, neutrality, sadness and surprise. Each image, which captures real human expressions in grayscale format of 224x224 pixels, is meticulously selected from various online sources, including Google, Unsplash, and Flickr, ensuring a wide array of natural facial expressions for improved learning and recognition tasks.

2) *Task Definition of Emotion Prediction with Four Situations:* According to the above description, we use three datasets and select 6 types anger, disgust, happiness, neutral, sadness, and surprise in the dataset. In each dataset, 10 images of 6 emotions are randomly selected and put into ChatGPT4 for judgment. As for the prompt words in Table 1, we want to preliminarily explore and predict the changes of emotion, so we choose four scenarios that are most likely to produce emotional changes in real life. At the same time, we artificially provide 4 situation simulations for each image, two positive situations, and two negative situations. (For details of specific questions, see Table I).

We predict the emotional changes of the image based on the simulated situation. Since ChatGPT4 was released in 2023, the above experiments were all conducted using ChatGPT4. We use supervised learning and evaluate the performance of ChatGPT4 in a zero-shot prompt setting for the above tasks. After the evaluation of ChatGPT4, if the result is the same as our cognitive result, it is recorded as 1, if the

result is different, it is recorded as 0, in other words, the predicted results must be consistent with the logical results of most cognitive and emotional changes in real society and be consistent with common sense and recorded as positive or negative according to the emotion according to the description of ChatGPT4. Moreover, we construct a ROC [26] curve utilizing the outcomes we have documented. Within this curve, positive emotions such as happiness, neutrality, or surprise are assigned a value of 1, while negative emotions like anger, disgust, or sadness are designated with a value of 0. ChatGPT4’s predictions for positive emotions are marked as 1 when they align with the actual outcomes, and as 0 when they do not. Similarly, for negative emotions, a matching prediction is indicated by a 0, and a mismatching one by a 1. The confidence level of these predictions is categorized on a scale from 1 to 3, where 1 indicates low confidence, 2 signifies moderate confidence, and 3 represents high confidence.

TABLE II
RESULT OF FOUR DIFFERENT SITUATIONS

Emotion	Parameter	Positive Situation	Negative Situation
Anger	accuracy	68.30%	73.30%
	sensitivity	NaN	NaN
	specificity	68.30%	73.30%
Disgust	accuracy	78.30%	85.00%
	sensitivity	NaN	NaN
	specificity	78.30%	85.00%
Happiness	accuracy	91.70%	83.30%
	sensitivity	91.70%	83.30%
	specificity	NaN	NaN
Neutral	accuracy	86.70%	83.30%
	sensitivity	86.70%	83.30%
	specificity	NaN	NaN
Sad	accuracy	71.70%	80.00%
	sensitivity	NaN	NaN
	specificity	71.70%	80.00%
Surprise	accuracy	85.00%	90.00%
	sensitivity	85.00%	90.00%
	specificity	NaN	NaN
Negative	accuracy	72.80%	79.40%
	sensitivity	NaN	NaN
	specificity	72.80%	79.40%
Positive	accuracy	87.80%	85.60%
	sensitivity	87.80%	85.60%
	specificity	NaN	NaN

3) *Preliminary Results:* In the context of data presented in Table II, the True Positive Rate (TPR), also referred to as Sensitivity, is a metric that quantifies the fraction of true positive instances accurately identified by the predictive model. Conversely, the False Positive Rate (FPR), also known as the complement of Specificity (1-Specificity), represents the proportion of negative cases that are mistakenly identified as positive by the model. The Observed Operating Points on the ROC curve signify the various thresholds applied within the classifier. Each of these points illustrates the equilibrium achieved between TPR and FPR at a given threshold setting. To elucidate, setting a higher threshold might lead to a reduction in FPR but at the cost of diminishing TPR, whereas a lower threshold setting is likely to elevate both TPR and FPR. These critical points are instrumental in assessing the model's efficacy and in determining the optimal threshold for the task at hand, highlighting the inherent compromise between maximizing the detection of positive instances (achieving a higher TPR) and minimizing the occurrence of false positives (achieving a lower FPR).

Table II shows the prediction results of ChatGPT4 for the evolution of emotions after initially identifying negative and positive emotions and describing them through two positive situations and two negative situations respectively. For images initially identified as negative emotions, we found that their ChatGPT4 prediction accuracy in negative contexts was 79.4%. However, if the situation was positive, the predicted evolution of emotion was 72.8%. In contrast, for images initially identified as having positive emotions, their response accuracy was higher in positive contexts than in negative contexts. We think that it may be due to ChatGPT4, the possibility of negative emotions turning into positive emotions when encountering a positive environment is lower than the possibility of remaining negative in a negative environment. Positive emotions have the same result. This result shows that the prediction results of ChatGPT4 are consistent with the changes in our cognitive emotions. Since we want to preliminary explore and predict the changes in emotion, we choose four scenarios that are most likely to produce emotional changes in real life. For details in Table II. The explanation is that the six categories of emotions were initially explored and analyzed separately, so when calculating the ROC, we also calculated the six categories of emotions separately to obtain the experimental results. According to the above description, positive emotions are recorded as 1 and negative emotions are recorded as 0. This means that, when the six types of emotions are analyzed separately, they will lack the other half of the records. NaN occurs in specificity because specificity needs negative samples to be calculated. If the dataset only contains positive emotions (no negatives), specificity cannot be computed, leading to NaN. Vice versa is also true.

Table IV corresponds to the prediction results of emotional changes corresponding to different events that will occur under each different emotion. First, we preliminary observe that in the case of images depicting surprise or astonishment, ChatGPT4 demonstrates a notable capability in recognizing

these emotions as such. However, it encounters difficulty in discerning whether the surprise conveys a positive or negative sentiment, leading to a tendency to classify the emotion of surprise as predominantly neutral. Consequently, this is the reason why the outcomes for surprise closely mirror those associated with neutral expressions.

In order to avoid the harm caused by negative emotions to high-risk industries or high-risk groups, we mainly look at three types of emotions: anger, disgust, and sadness. We observe that in negative emotions, if the upcoming event is positive, then the accuracy of ChatGPT4 in predicting the emotion evolution from high to low in zero-shot is disgust, sad, and anger; FPR is 78.3%, 71.7%, and 68.3%, respectively. Anger is the strongest of negative emotions and the lowest in response to positive events. At the same time, because disgust is the most complex of negative emotions, including disgust, unhappiness, contempt, etc., it ranks the highest. Furthermore, the precision in identifying negative emotions falls short of expectations, suggesting that ChatGPT4 could benefit from the inclusion of additional descriptive cues to enhance its decision-making process. Presently, in a zero-shot scenario, ChatGPT4 is adept at recognizing the presence of negative emotions in individuals; however, it struggles with the accurate classification of specific emotions such as disgust, contempt, or anger. This is why negative predictions are less accurate than positive ones.

4) *Analysis and Discussion:* Throughout the training phase, it is common to encounter discrepancies between the emotions depicted in certain dataset images and our real-world perceptions. Due to the subjective nature of emotional interpretation, there is a possibility of encountering biases in recognizing the emotions conveyed by some images. In such instances, we rely on our judgment as the ultimate criterion and compare it to the interpretations provided by ChatGPT4.

Moreover, we have identified an additional complication: a misalignment between ChatGPT4's interpretations and the dataset's guidelines. A closer look at the specific examples of ChatGPT4's predictions highlights a fundamental issue—the disparity between the model's understanding and the dataset's standard. While the dataset might categorize an image as portraying anger based on its guidelines, ChatGPT4 might interpret the same expression as sadness or confusion. This discrepancy is not a matter of accuracy but rather an indication of differing standards used to classify negative emotions. Upon analysis, this divergence seems not solely a limitation of ChatGPT4 but could also stem from inadequate prompting. As the complexity of prompt instructions increases, expecting comprehensive coverage with minimal input becomes impractical. This realization opens up avenues for future improvements: if adhering strictly to the dataset's criteria is not mandatory, then refining the model based on broad prompt adjustments (like specifying the depicted emotions) might be viable. Yet, evaluating based on the dataset's labels could prove unsuitable, necessitating a more thorough manual review. On the contrary, if strict conformity to the dataset's guidelines is essential, relying on a multitude of prompt adjustments may fall short, making the supervised model fine-tuning a more effective

strategy.

B. Emotion Prediction with Different Categories of Emotional Sentences

1) *Dataset*: First, we continue to use the same images as the previous task. They are still from emotion detection, facial expressions training data, and natural human faces. Each dataset is still the same 10 images. But in the second task, we added a dataset called MELD [27].

MELD The Multimodal EmotionLines Dataset (MELD) builds upon and enriches the original EmotionLines dataset by incorporating additional modalities such as audio and visual elements alongside text. MELD features over 1,400 dialogue sequences and 13,000 spoken exchanges drawn from the "Friends" TV series, with various characters contributing to the conversations. Every piece of dialogue within MELD is categorized under one of seven possible emotions: Anger, Disgust, Sadness, Joy, Neutral, Surprise, and Fear. Additionally, MELD assigns a sentiment classification—positive, negative, or neutral—to each utterance, further enhancing its utility for emotion and sentiment analysis research.

2) *Task Definition*: The tasks in Part Two are partially similar to those in Part One. They all use the same images from the same dataset. However, each picture uses 6 categories of sentences full of different emotions 1. Anger, 2. Disgust, 3. Happiness, 4. Neutral, 5. Sad, 6. Surprise; think of these statements as what the character in the image is going to say. The input images and sentences are then analyzed using ChatGPT4 and the emotional evolution of any image is predicted and judged (For details of specific questions see Table III). At the same time, for the diversity of results, we also put the same pictures into the large language model for comparison test, in which tik tok's Doubao large language model [28] is used to compare the output content.

3) *Preliminary Results*: The abscissa of Table VI represents the image of the dataset, and the ordinate represents the evolution of ChatGPT4's prediction of emotions after inputting 6 different emotional sentences.

We can observe that the prediction accuracy of ChatGPT4 from high to low is: happiness, surprise, neutral, anger, sad, disgust. Among the three positive emotions, according to ChatGPT4 prediction, except for the happiness emotion that is directly converted into anger, which has the lowest accuracy, happiness is the highest for the others. At the same time, we observe that according to the description of ChatGPT4, when defining surprise and neutral, because they can be regarded as positive or negative, the results of the two are very similar. In the prediction of negative emotions, according to the above explanation of the FPR index, it shows that the disgust emotion is the least accurate to identify, and the emotion of the disgust category is the most difficult to judge among the six types of emotions. At the same time, still the same as the previous task, ChatGPT4 requires more prompts to achieve the accuracy of negative emotions. In the case of zero-shot, ChatGPT4 is not as good at predicting the evolution of emotions as in the case of positive emotions.

Similarly, the tested Doubao LLM is less accurate at recognizing negative emotions compared to positive ones. Table V show that the result accuracies of ChatGPT and Doubao. In many instances, it even misclassifies negative emotions as neutral. However, when comparing the results of the two large language models, ChatGPT's output accuracy is significantly higher than that of the Doubao model. In zero-shot situations, the Doubao model tends to misidentify negative emotions as positive, a problem that ChatGPT does not exhibit. Although ChatGPT may not always precisely identify the specific type of negative emotion, it can determine that the person in the image is experiencing some form of negative emotion. This explains why the Doubao model is less accurate in predicting mood changes.

The vertical axis of the ROC curve represents sensitivity, which is directly proportional to the model's diagnostic accuracy. Conversely, the horizontal axis denotes 1-specificity, where a lower value indicates a reduced rate of false positives. Generally, a point closer to the upper-left corner of the ROC space signifies superior diagnostic performance, implying that a sensitivity approaching 1 correlates with enhanced predictive accuracy.

Before proceeding, it is important to build upon the partial definitions provided earlier; this section focuses on the concept of the Empirical ROC Area. The Empirical ROC Area, commonly known as the Area Under the Curve (AUC), quantifies a model's discriminative power directly from raw data by constructing an empirical ROC curve. This curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) across a range of decision thresholds. The AUC metric evaluates the model's efficacy in distinguishing between positive and negative classes over all threshold values, with a larger AUC indicating superior performance. An AUC value of 0.5 suggests no better than random classification, while a value of 1.0 represents perfect discrimination.

According to the data presented, we believe that the sensitivities of the three datasets are very similar, except in the case of the disgust statements. When the initial emotional state varies, it becomes challenging for ChatGPT-4 to accurately identify expressions of disgust. For example, in a positive context, it might interpret a disgust statement as a joke or prank, resulting in lower accuracy. In terms of specificity, however, the prediction results of ChatGPT-4 exceed expectations, especially under an initially positive sentiment where the prediction accuracy is very high—almost entirely correct. Based on the accuracy and ROC curve, ChatGPT-4's performance in predicting sentences across different emotions surpasses expectations.

IV. DISCUSSION

In this paper, our sentiment evaluation is mainly derived from static inputs (images or single pieces of text). However, in real-world situations, emotions are dynamic and can shift rapidly depending on ongoing interactions—an aspect not fully reflected in our current experimental setup. As a result, the absence of real-time feedback mechanisms to update model

TABLE III
EXAMPLE OF SIX DIFFERENT CATEGORIES EMOTIONAL SENTENCES.


Dataset	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6
	What is the emotion of this person? If the next thing they say is, "Well, why don't you tell her to stop being silly!" What do you think their emotions will become?	If the next sentence they say is, "Say it louder, I don't think the guy in the back heard you!" What do you think their emotions will become?	If the next sentence they say is, "Guess what, I got an audition!" What do you think their emotions will become?	If the next sentence they say is, "Great. He's doing great. Don't you worry about him?" What do you think their emotions will become?	If the next sentence they say is, "Yeah but we won't be able to like to get up in the middle of the night and have those long talks about our feelings and the future." What do you think their emotions will become?	If the next sentence they say is, "Look what I got! Look what I got! Can you believe they make these for little people?" What do you think their emotions will become?

TABLE IV
RESULT OF SIX DIFFERENT CATEGORIES EMOTIONAL SENTENCES.

Emotion	Anger sentence	disgust Sentence	Happiness sentence	Neutral Sentence	Sad sentence	Surprise sentence
Anger	70.00%	86.70%	86.70%	86.70%	86.70%	83.30%
Disgust	60.00%	70.00%	60.00%	56.70%	83.30%	56.70%
Happiness	70.00%	96.70%	1	96.70%	96.70%	96.70%
Neutral	76.70%	86.70%	96.70%	96.70%	90.00%	90.00%
Sad	63.30%	76.70%	76.70%	76.70%	86.70%	86.70%
Surprise	73.30%	86.70%	96.70%	96.70%	93.30%	96.70%

TABLE V
ACCURACY OF DIFFERENT LARGE LANGUAGE MODELS.

LLM	Negative Emotion Accuracy	Positive Emotion Accuracy
ChatGPT	68.89%	80.56%
Doubao	26.11%	40%

predictions based on user responses limits the immediate practical value of adaptive systems, such as interactive chatbots or mental health monitoring tools.

Our study primarily focuses on ChatGPT4's capabilities in image-based emotion recognition. In the future, our work could be extended to other large language models, such as Claude3, to compare their respective advantages and drawbacks. Additionally, there has yet to be a comprehensive evaluation under real-world conditions, leaving questions about these models' robustness and generalizability beyond controlled experiments.

Looking ahead, further investigations into how ChatGPT4 generates predictions could involve refining prompts or fine-tuning the model, potentially increasing both the transparency and interpretability of its decision-making process. Another consideration is that basing judgments solely on perceived emotional changes may introduce bias. Since ChatGPT is a probabilistic model, its responses may vary even when given the same input multiple times. To address this, future studies might involve running the same input multiple times and averaging

the results, mitigating the limitations of relying on a single experiment for input correlation.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we examine ChatGPT4's zero-shot abilities in interpreting sentiment from image-text inputs and compare its performance to the Doubao model. ChatGPT4 demonstrates high accuracy but sometimes mislabels disgust as depression. Targeted prompts and mental health considerations can improve its inference quality.

ChatGPT4 outperforms Doubao in prediction accuracy, although it may struggle to identify specific negative emotions. Doubao often misinterprets negative emotions as neutral or positive in zero-shot scenarios. We recommend refining prompts and using relevant examples to boost ChatGPT4's performance in subjective tasks, including mental health applications.

Dataset images can conflict with real-life perceptions, introducing biases in emotion recognition. We compare our human assessments with ChatGPT4's outputs to pinpoint discrepancies and address potential biases. ChatGPT4 predictions sometimes clash with the dataset guidelines, highlighting their deviation from standard annotations. For example, it may interpret anger as sadness or confusion. This discrepancy reflects varied emotional criteria rather than outright errors.

Differences in interpretation may stem from prompt design limitations rather than ChatGPT4's flaws. If strict dataset adherence isn't crucial, broader prompts can enrich the model's performance, though manual reviews may be needed. If exact compliance is required, more supervised fine-tuning is essential to align with dataset-specific emotional classifications.

REFERENCES

[1] C. H. Leung, J. J. Deng, and Y. Li, "Enhanced human-machine interactive learning for multimodal emotion recognition in dialogue system", in *Proceedings of the 2022 5th International Conference on Algorithms, Computing and Artificial Intelligence*, 2022, pp. 1-7.

TABLE VI
 RESULT OF DATASET FOR SIX DIFFERENT CATEGORIES EMOTIONAL SENTENCES.

Dataset	Parameter	Anger Sentence	Disgust Sentence	Happiness Sentence	Neutral Sentence	Sad Sentence	Surprise Sentence
Emotion Dection	accuracy	88.30%	53.30%	93.30%	90.00%	71.70%	91.70%
	sensitivity	83.30%	30.00%	96.70%	96.70%	70.00%	93.30%
	specificity	93.30%	76.70%	90.00%	83.30%	73.30%	90.00%
	Empiric ROC Area	0.989	0.837	0.997	0.994	0.92	0.993
Facial Express	accuracy	81.70%	58.30%	93.30%	91.70%	78.30%	95.00%
	sensitivity	83.30%	46.70%	100	100	83.30%	96.70%
	specificity	80.00%	70.00%	86.70%	83.30%	73.30%	93.30%
	Empiric ROC Area	0.967	0.84	1	1	0.956	0.998
Neutral Human	accuracy	73.30%	58.30%	93.30%	93.30%	79.70%	85.00%
	sensitivity	76.70%	50.00%	100	100	79.30%	100
	specificity	70.00%	66.70%	66.70%	86.70%	80.00%	70.00%
	Empiric ROC Area	0.93	0.833	1	1	0.959	1

- [2] B. Mann *et al.*, “Language models are few-shot learners”, *arXiv preprint arXiv:2005.14165*, 2020.
- [3] L. Ouyang *et al.*, “Training language models to follow instructions with human feedback”, *Advances in Neural Information Processing Systems*, vol. 35, pp. 27 730–27 744, 2022.
- [4] *Open AI GPT 4*, <https://openai.com/gpt-4>, 2023.
- [5] T. Zhang, A. M. Schoene, S. Ji, and S. Ananiadou, “Natural language processing applied to mental illness detection: A narrative review”, *NPJ digital medicine*, vol. 5, no. 1, p. 46, 2022.
- [6] D. Ciraolo *et al.*, “Emotional artificial intelligence enabled facial expression recognition for tele-rehabilitation: A preliminary study”, in *2023 IEEE Symposium on Computers and Communications (ISCC)*, IEEE, 2023, pp. 1–6.
- [7] D. B. Shank, C. Graves, A. Gott, P. Gamez, and S. Rodriguez, “Feeling our way to machine minds: People’s emotions when perceiving mind in artificial intelligence”, *Computers in Human Behavior*, vol. 98, pp. 256–266, 2019.
- [8] *Fat cat incident*, https://sports.sohu.com/a/776021122_121856967.
- [9] R. Lewis and J. Rose, ‘i’m not okay,’ *off-duty alaska pilot allegedly said before trying to cut the engines*, <https://www.npr.org/2023/10/24/1208244311/alaska-airlines-off-duty-pilot-switch-off-engines>, OCTOBER 25, 2023, 11:55 AM ET.
- [10] A. Geetha, T. Mala, D. Priyanka, and E. Uma, “Multimodal emotion recognition with deep learning: Advancements, challenges, and future directions”, *Information Fusion*, vol. 105, p. 102 218, 2024.
- [11] R. Zhang, Z. Wang, Z. Huang, L. Li, and M. Zheng, “Predicting emotion reactions for human–computer conversation: A variational approach”, *IEEE Transactions on Human-Machine Systems*, vol. 51, no. 4, pp. 279–287, 2021.
- [12] K. Yang, S. Ji, T. Zhang, Q. Xie, and S. Ananiadou, “On the evaluations of chatgpt and emotion-enhanced prompting for mental health analysis”, *arXiv preprint arXiv:2304.03347*, 2023.
- [13] W. Zhao *et al.*, “Is chatgpt equipped with emotional dialogue capabilities?”, *arXiv preprint arXiv:2304.09582*, 2023.
- [14] H.-D. Le, G.-S. Lee, S.-H. Kim, S. Kim, and H.-J. Yang, “Multi-label multimodal emotion recognition with transformer-based fusion and emotion-level representation learning”, *IEEE Access*, vol. 11, pp. 14 742–14 751, 2023.
- [15] P. Ekman, *Facial expressions of emotion: New findings, new questions*, 1992.
- [16] P. Robert, *Emotion: Theory, research, and experience. vol. 1: Theories of emotion*, 1980.
- [17] R. Kosti, J. M. Alvarez, A. Recasens, and A. Lapedriza, “Emotion recognition in context”, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1667–1675.
- [18] R. Kosti, J. M. Alvarez, A. Recasens, and A. Lapedriza, “Context based emotion recognition using emotic dataset”, *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 11, pp. 2755–2766, 2019.
- [19] J. Nicolle, V. Rapp, K. Bailly, L. Prevost, and M. Chetouani, “Robust continuous prediction of human emotions using multiscale dynamic cues”, in *Proceedings of the 14th ACM international conference on Multimodal interaction*, 2012, pp. 501–508.
- [20] W. Zhang, X. He, and W. Lu, “Exploring discriminative representations for image emotion recognition with cnns”, *IEEE Transactions on Multimedia*, vol. 22, no. 2, pp. 515–523, 2019.
- [21] D. Issa, M. F. Demirci, and A. Yazici, “Speech emotion recognition with deep convolutional neural networks”, *Biomedical Signal Processing and Control*, vol. 59, p. 101 894, 2020.
- [22] A. Metallinou and S. Narayanan, “Annotation and processing of continuous emotional attributes: Challenges and opportunities”, in *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)*, IEEE, 2013, pp. 1–8.
- [23] SAMHSA, *Warning signs and risk factors for emotional distress*, <https://www.samhsa.gov/find-help/disaster-distress-helpline/warning-signs-risk-factors>.
- [24] L. Zahara, P. Musa, E. P. Wibowo, I. Karim, and S. B. Musa, “The facial emotion recognition (fer-2013) dataset for prediction system of micro-expressions face using the convolutional neural network (cnn) algorithm based raspberry pi”, in *2020 Fifth international conference on informatics and computing (ICIC)*, IEEE, 2020, pp. 1–9.
- [25] A. Mollahosseini, B. Hasani, and M. H. Mahoor, “Affectnet: A database for facial expression, valence, and arousal computing in the wild”, *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18–31, 2017.
- [26] T. Fawcett, “An introduction to roc analysis”, *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [27] S. Poria *et al.*, “Meld: A multimodal multi-party dataset for emotion recognition in conversations”, *arXiv preprint arXiv:1810.02508*, 2018.
- [28] *Tik Tok*, https://www.doubao.com/chat/?channel=baidu_pz&source=db_baidu_pz_01&keywordid=weizhi7, August 17, 2023.