Communication Papers of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 159–168 DOI: 10.15439/2024F7168 ISSN 2300-5963 ACSIS, Vol. 41

Psychological Needs as Credible Song Signals: Testing Large Language Models to Annotate Lyrics

E. Cho Smith, Yinxuan Wang, and Eric T. Matson
0009-0008-7817-690X
0009-0003-8740-4912
0000-0001-9200-4903
Computer and Information Technology,
Purdue University,
West Lafayette, Indiana 47906-3603, USA
Email: {wilso287, wang3910, ematson}@purdue.edu

***Abstract*—Our preliminary study presents a new perspective in music information retrieval by investigating how contemporary song-making and listening emulate our innate responses, similar to the primal vocalizations of primates, drawing from music's origins as credible signaling. The diversity of musical expressions within a single culture suggests that it arises from group dynamics reflecting individuals' psychological fitness. Derived from the temporal need-threat framework of ostracism—an evolutionarily stable strategy to influence individuals in a group, we argue that individual differences in song-making and listening can be reduced to songs' lyrical expressions in terms of four basic psychological needs: self-esteem, self-control, seeking to belong, and seeking recognition. We propose a four-binary-decision model to classify English song lyrics for hierarchically organizing the variations of musical expressions. Annotating 260 English song lyrics using ChatGPT-4s with human validation and fine-tuning GPT-3.5-turbo to develop an automated classifier have identified some limitations in current large language models.**

***Index Terms*—psychological needs recognition, music information retrieval, large language models**

I. Introduction

AMID a lively school atmosphere, a solitary individual with earbuds sits apart, symbolizing an invisible barrier separating them from others, with only the sound of music accompanying their solitude. This hypothetical person finds refuge in a musical asylum [1] from ostracism, which, as defined by Oxford Languages, means the temporary banishment from a city-state decided by popular vote in ancient Greece. A scenario often portrayed in media mirrors the real-life tendency of young individuals to use music as a coping mechanism in socially exclusive contexts [2], [3], [4]. Unlike explicit rejection, ostracism involves being ignored by groups and dyads[5], [6] through actions such as avoiding eye contact, using the silent treatment, and withholding information [7]. While ostracism does not result in immediate physical harm, its prolonged effects can profoundly impact our psychological well-being, leading to feelings of alienation, depression, helplessness, and unworthiness, which are often referred to as "social death" [5]. Although research on the relationship between ostracism and music listening is sparse, one study exemplifies how music can counteract ostracism's adverse effects [3]. Specifically, among "metal-heads," music promotes a shared sense of identity and cohesion, shielding them from the despair associated with ostracism [3]. Music's psychological function can become more salient during emotionally challenging circumstances like ostracism, one of the most pervasive forms of social exclusion among young individuals in their critical phases of identity formation and social adaptation. The hypersensitivity to ostracism [8], particularly, in this age group makes them susceptible to its negative impacts, necessitating interventions and earlier detection.

Listening to "song lyrics" resonates with listeners, serving multiple functions such as regulating emotions [9], [10], [11], evoking nostalgic memories [12], reflecting one's identity [13], [3], fostering social bonds [14], functioning as cultural markers [15] and many others. Research with 834 participants identifies 129 unique music functions, projecting to three main dimensions: arousal regulation, self-awareness, and social connections [16]. Yet, internal functions are far more relevant than external, challenging music's origin as social bonding [16]. Another study emphasizes music's psychological functions among adolescents in fulfilling needs including self-regulation, sense of agency, and belonging regardless of an individual's conscious recognition of music's influence [13].

To date, despite the noteworthy role of songs in expressing psychological needs when coping with social exclusion [2], [3], [4], current Music Information Retrieval (MIR) research has not fully appreciated the lyrical content of songs that listeners immerse themselves in and sing along to. This oversight limits the understanding of how contemporary songs function as a medium for the vicarious vocalization of listeners' psychological needs. Switching the vantage point from recognizing musical emotions, our study proposes examining psychological needs closely related to musical emotions as contextual and semantic motivations. The proposed shift provides a more objective representation of listeners' states, beyond subjective emotional experiences.

To develop a proof of concept for our approach to automating Psychological Needs Recognition (PNR) expressed in songs, our preliminary study explores English song lyrics to annotate the binary states of four basic psychological needs: self-esteem, self-control, seeking to belong, and seeking recognition. These needs are derived from the temporal need-threat framework of ostracism [17], a quintessential group dynamic prevalent in all modern societies [18]. Following the framework, we suggest that threats to these fundamental needs can activate music's primeval functions as non-random signals, motivating individuals to create or listen to songs (vicarious vocalization) that express the critical states of their psychological needs. We further elaborate on theoretical and empirical grounds for our conceptual model in the Related Work section to elucidate our proposed new perspective in MIR, which can organize popular songs hierarchically, from the root node to fundamental psychological needs, taking precedence over musical emotions and styles.

II. Related Work

Our study builds on the premise that "listening to songs," differs from hearing music as ambient sounds, is an intentional communication, deeper and more intrinsic than the tangible outcomes we may recall afterward. Following Mehr et al., we speculate music's evolutionary origins in primeval functions such as "territorial advertisements and contact calls, two types of vocal signals that are widespread in primates and other species" [19, p. 140]. The credible signaling hypothesis explains that natural selection refines human abilities to create and appreciate music's rhythms for inter-group alliance/intra-group cohesion and melodies for caregiving and seeking [20]. The notion supports our premise that music's proximate functions stem from its evolutionary origins to communicate "credible signals" of individuals' fitness about group dynamics. We suggest our song-making and listening in today's music consumption signals the resonance of individuals seeking fundamental psychological needs in flux.

A. Rethinking Musical Emotions as Signaling and Signals

According to signaling theory in evolutionary biology, signaling involves one party (the sender) transmitting information to another party (the receiver) through specific, non-random actions or symbols [21]. These signals act as credible and honest indicators aligning with the mutual interests of both senders and receivers [21]. In song-making and listening, musical emotions, constructed as the power of music features that evoke intended emotion among listeners [22], act as both signaling and signals. Musicians communicate their internal states through compositions and performances, while the audience perceives these musical emotions as signals, evoking specific emotional responses and influencing their internal states. Effective musical communication occurs when the internal states of both parties align, often described as "emotional resonance." Although perfect alignment is rare due to unintentional noise or missing contexts in the signals, we argue, musical emotions can intentionally be aligned by storytelling to create a shared emotional resonance between creators and listeners of song lyrics.

Emotions are not ends but means in internal processes, serving as internal signals (not verifiable) for individuals to change or maintain their states at any given time. Labeling musical emotions risks us falling into the complexity of emotions [23], which has muddled the representation of musical emotions for consensus. Musical emotions are modeled as discrete [24], [25], two-dimensional [26], [27], or three-dimensional [28] constructs. Additionally, a recent study suggests that emotions are higher dimensional, semantic spaces that are neither discrete nor simplified to two or three dimensions [29]. From the 1970s into the 21st century, theorists like Paul Ekman and Carroll Izard assert the existence of universal, primary, distinct emotions fulfilling adaptive roles as basic emotions [24]. Izard's differential emotions theory extends the basic emotions concept, advocating for a collection of biologically embedded emotions, such as fear, anger, joy, sadness, disgust, and surprise [25] for they are evolutionarily honed for our survival [24]. While disputing the recognition of musical emotions in MIR, we have overlooked the central psychological facet of emotions, "feelings" and warrants attention for connecting cognition and emotion [30].

When we rethink musical emotions as signaling and signals, it makes sense to move beyond conventional labels since emotional experiences are subjective and internal, making them difficult and costly to verify in affective computing [31]. Instead of labeling musical emotions, we should focus on internal signals that hold meaning for both the creators and listeners of popular songs in their contexts, addressing the daily psychological needs of living humans. Our proposed perspective emphasizes personal and unique connections within individuals and variations within a single society, arising when people create and listen to popular songs. This approach acknowledges the potentially shareable but individualized nature of musical experiences. By proposing psychological needs as credible song content, we move away from labeling and toward understanding musical emotions that value internal meaning over naming. Moreover, the limited contextual understanding of musical emotions as signals by listeners has led to more costly approaches in affective computing, such as analyzing listeners' physiological signals [32] and ambient noises in their environments, raising concerns about data privacy and protection [33].

B. Compiling Credible Signals, Mirror Neurons and Empathy

Music's evolutionary roots can be traced back to basic functions such as "territorial advertisements and contact calls," which are common vocal signals among primates and other species [19]. The credible signaling hypothesis suggests that natural selection has refined human abilities to produce and appreciate musical rhythms, which help form inter-group alliances and enhance intra-group cohesion and melodies, which are important for infant caregiving and seeking [19]. As the universality of proto-human signals has diversified over time, the value of credible signaling persists. Contemporary humans continue to create music to exchange musical signals for communication. We assert that the modern functions of popular songs, particularly those using human voices, reflect their ancestral functions. Human songs serve as a means to communicate "credible, honest signals" about individuals' fitness within group dynamics, the universal norms of any human civilization. While there exists a degree of intentionality that separates music from language [34], research suggests that music, as a form of intentional communication within group dynamics, can assist listeners in achieving self-awareness [35], self-regulation [13], a sense of agency [13], and belonging [13], [36], thereby contributing to psychological well-being [37].

Mirror neurons enable individuals to grasp the meaning and intentions behind actions by internally mirroring those actions in their own brains [38]. Found in specific regions of the monkey brain, these neurons fire both when an action is performed and when it is observed, aiding in the prediction and understanding of others' intentions [39]. In humans, the mirror neuron system is activated by visual and auditory signals, for comprehending others' actions and emotions [38]. This neural mechanism is thought to support higher cognitive functions like empathy [40] and the shared understanding necessary for speech perception [41]. For example, individuals with autism are impaired in emotional expressions and social communication. Research suggests their impaired functions are linked to the human mirror neuron system and engaging their impaired regions through music-making activities enables many children with autism to enhance their understanding of others' facial expressions of emotions to improve their communication and social skills [42], [43].

In summary, we support that the origin of music lies in its universal function as credible, non-random signals that engage our ancestral brains, activating our mirror neurons to share intentions. These intentions are linked to fundamental psychological needs, especially in today's material affluence. Thus, contemporary song-making and listening should systematically aim to decipher psychological needs signals as musical intentions. We address the need for computational models that can automate the detection of psychological needs expressed in song signals. By doing so, we can help individuals cope with the increasingly automated (no human in the loop) music platforms of today.

C. Reasoning Basic Psychological Needs as Song Signals

Feelings are visceral sensations tied to emotions, linking to motivation and aiding in psychological processes such as individualization [30]. Feelings organize cognitive processes and guide adaptive actions [30]. We argue that the link between feelings and adaptive actions forms individuals' strategies to meet psychological needs. However, experiencing emotions does not guarantee their conscious recognition or expression, which depends on their intensity and our language skills [30]. We argue, therefore, impacting our capability to have musical expressions to convey our emotional states, as we often struggle to articulate emotions while feeling in them. Aside from the disagreement in representations of emotions' dimensions, Plutchik's psychoevolutionary theory of emotions asserts that certain emotions are primary and fundamental, as primary colors, with more complex emotions evolving from the basic primary emotions such as fear, anger, and joy [44]. The primary emotions are evident across different evolutionary stages and form the core of adaptive survival responses [44].

Although the direct applications of the adaptive paths are not yet fully validated, the role of musical emotions as adaptive signals provides a theoretical framework for modeling musical expressions as individualized, adaptive responses that vary within a society, the fundamental unit of our evolution as social beings. A book chapter on self-esteem speculates its construction as "a means of interpreting mood, which encourages and inhibits conduct in various situations. Mood is a response to positive and negative experiences; self-esteem is a construction of mood fitted to a culture and its themes" [45, p. 310]. From the earliest recordings of Greek history to today's digital era, ostracism, the deliberate or indiscriminate disregard of individuals by others in a group activates our threat responses elicited by painful feelings to fortify basic psychological needs[5].

Williams' temporal need-threat framework [17] outlines ostracism experiences into immediate-reflexive reactions, intermediate-reflective copings, and prolonged-chronic stages. The immediate-reflexive stage involves the abrupt, universal experience of painful threats to four basic psychological needs: self-esteem, self-control, seeking to belong, and seeking recognition [17]. The painful sensations trigger a range of negative emotions, including pain, anger, and sadness [18]. The intermediate-reflective stage is characterized by contextually influenced, individualized strategies to restore the threatened needs, such as seeking belongingness, others' recognition [17], or solitude [46]. Adaptive responses in the intermediate stage include tend-and-befriend [47], fight, freeze, or flight reactions [5], which parallel essential neurophysiological states described as core affect (positivity, negativity, energy, and fatigue) that influence our reflexes, perception, and cognition [48]. Prolonged-chronic ostracism gradually depletes the resources of those affected, leading to feelings of depression, alienation, and unworthiness [17]. At the immediate-reflexive stage of experiencing ostracism, we highlight the paths of painful feelings as visceral responses before individuals are emotionally experienced by labeling the sensations as anger or sadness in self-reports. Moving into the intermediate-reflective coping stage, emotional responses such as sadness, anger, or hurt feelings, along with cognitive appraisals of ostracism, work to mitigate its adverse effects within individuals [49], which is where, we argue, surface individuals' variations of musical expressions chosen to make or listen to songs.

The temporal need-threat framework is invaluable for establishing psychological needs as credible indicators in songs, reflecting adaptive responses of individuals within normative group dynamics. Studies have shown that music can help individuals cope with social exclusion [3], [4] and regulate negative emotions during the COVID-19 lockdown [50], [51] when most industrialized countries required social distancing measures. The framework explains group dynamics and supports the concept of adaptive individualization by addressing the self-regulation of psychological needs as motivations. While many emotion theories implicitly suggest that emotions act as motivators, Williams' framework explicitly details how ostracism activates inherent motivations. These psychological needs drive individuals not only to alleviate negative emotions but also to fortify their impaired psychological needs, which can be replenished by interacting with a group they feel belong to. By understanding these motivations, today's music platforms can identify musical emotions related to psychological needs as credible signals to assist listeners in discovering therapeutic songs.

D. Transfer Learning by Large Language Models

Transfer learning is a technique where a computational model trained for a particular task is reused as the starting point for another task, applying the knowledge (e.g., pre-trained weights and biases of deep neural networks) gained from tasks to the following tasks that are deemed similar. Presented in "Attention is All You Need" [52], the Transformer architecture, the foundational structure of Large Language Models (LLMs), marks a departure from sequence-based deep learning architectures. Unlike its predecessors, Transformers utilize attention mechanisms that simultaneously compute the entire input sequence [52]. The attention mechanism enables the model to focus on various parts of the input sequence when predicting an output, greatly enhancing its contextual understanding [52]. The introduction of the Transformer architecture has led to the development of GPT (Generative Pre-trained Transformer), BERT (Bidirectional Encoder Representations from Transformers), and other models. These models have set new standards in language translation, question-answering, and text generation [53]. By this "somewhat generalized intelligence," transformers are used in emotion detection and analysis from texts, with their superior understanding of contexts[54], [55], [56]. In short, Transformers, the basis of popular LLMs such as GPT and BERT, pave the way for transfer learning in machines' natural language processing (NLP).

III. Methodology

A. Psychological Needs as Credible Song Signals Model

As shown in Fig. 1, we formalize the variations in terms of four basic psychological needs using binary decision points for simplicity. There are sixteen classes of lyrical expressions, based on whether the subjects of the lyrics (narrators or protagonists) express the states of High Self-Esteem (HSE), Low Self-Esteem (LSE), High Self-Control (HSC), Low Self-Control (LSC), Positive Seeking to Belong (PSB), Negative Seeking to Belong (NSB), Positive Seeking Recognition (PSR), and Negative Seeking Recognition (NSR). The classes are enumerated as 0:[HSE, HSC, PSB, NSR], 1:[HSE, HSC, PSB, PSR], 2:[HSE, HSC, NSB, PSR], 3:[HSE, HSC, NSB, NSR], 4:[HSE, LSC, PSB, NSR], 5:[HSE, LSC, PSB, PSR], 6:[HSE, LSC, NSB, PSR], 7: [HSE, LSC, NSB, NSR], 8:[LSE, LSC, PSB, NSR], 9:[LSE, LSC, PSB, PSR], 10:[LSE, LSC, NSB, PSR], 11:[LSE, LSC, NSB, NSR], 12:[LSE, HSC, PSB, NSR], 13:[LSE, HSC, PSB, PSR], 14:[LSE, HSC, NSB, PSR], and 15:[LSE, HSC, NSB, NSR].

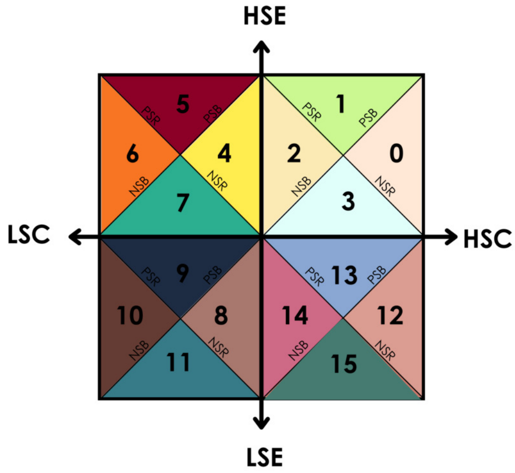

Fig. 1. The sixteen classes are based on four binary states of psychological needs: High Self-Esteem (HSE), Low Self-Esteem (LSE), High Self-Control (HSC), Low Self-Control (LSC), Positive Seeking to Belong (PSB), Negative Seeking to Belong (NSB), Positive Seeking Recognition (PSR), and Negative Seeking Recognition (NSR). The classes are enumerated using the first letter of each sub-dimension as 0:HHPN, 1:HHPP, 2:HHNP, 3: HHNN, 4:HLPN, 5:HLPP, 6:HLNP, 7: HLNN, 8:LLPN, 9:LLPP, 10:LLNP, 11:LLNN, 12:LHPN, 13:LHPP, 14:LHNP, and 15:LHNN.

B. Leveraging Model Distillation to Annotate English Lyrics

We utilize GPT-4 and GPT-4o, the fourth-generation GPT models by OpenAI. The latter, known as "GPT-4-turbo", is an optimized version of GPT-4 designed for greater efficiency and performance. To produce labeled data quickly, our methodology involves model distillation, where larger, pre-trained models such as GPT-4 and GPT-4o are used to generate labels for psychological needs expressed in English songs. Traditionally, labels in Music Emotion Recognition (MER) come from manual annotation processes, where experts or hired individuals assign labels based on predefined emotional categories. These labels require extensive human intelligence, involving subjective interpretations that can be inconsistent and prone to individual biases and lapses in attention. Due to its annotation challenges [57], [58], social tags online are often used to increase efficiency. However, in our case, there are no existing social tags for psychological needs expressed in song lyrics to utilize through crawling techniques. By leveraging generative LLMs in NLP, we streamline the rapid labeling process while ensuring the accuracy of machines by human-in-the-loop validation for testing our proposed computational model. Based on the demonstrated feasibility of using ChatGPT for annotating English text data in various topics [59], [60], [61] and increase the human and computer interaction more naturally, we employ ChatGPT 4.0 and 4o to automate the process of annotating psychological needs expressed in English song lyrics. We test whether the complexity of English lyrics can be generalized to the binary states of four psychological needs: self-esteem, self-control, seeking to belong, and seeking recognition.

C. Prompt Engineering and Fine-Tuning of LLMs

To expedite the automation of annotation, we turn to prompt engineering. This involves crafting precise instructions along with English lyrics as inputs for ChatGPT-4 and 4o, aiming to generate the most accurate labels that categorize English lyrics into sixteen classes as shown in Fig. 1. The quality of outputs produced by generative LLMs is highly dependent on the quality of inputs, including data with instructions and labels. Given that existing strategies are quite general, we address the novel challenge of annotating psychological needs in English song lyrics using ChatGPT-4 and 4o, testing our instructions for generating labels based on the proposed model Fig. 1.

A new paradigm called "prompt-based learning" modifies prompts—inputs given to generative AI to guide their responses [62]. Prompts can be questions, statements, or commands provided by users to generate the AI's outputs, utilizing zero-shot and few-shot learning [62]. Zero-shot learning enables an AI model to classify objects or concepts it has not seen or learned before and it has been shown to produce sentiment time series closely matching those from fine-tuned models, though they slightly underestimate negative examples [63]. The need for LLMs to quickly adapt to various semantic categories with minimal training has led to the development of n-shot learning, including few-shot learning and one-shot learning [64]. Few-shot learning frequently utilizes transfer learning and meta-learning strategies to train pre-trained models to identify new classes using only a few labeled data, or just one labeled example in one-shot learning [64].

According to the OpenAI API documentation, the recommended approach is to begin with prompt engineering, then proceed to fine-tuning if necessary, and refine the fine-tuning process to minimize training and validation losses. Hence, we start by employing prompt engineering, using zero-shot (no-example) prompts with ChatGPT-4 and 4o to leverage their state-of-the-art performance in labeling 16 psychological needs classes in English lyrics. The high performance of ChatGPT-4 and 4o in annotating texts is demonstrated and compared to human annotators for accuracy in a study [61]. Next, we use few-shot (some examples) prompts, providing more detailed instructions and annotated lyrics. Finally, we fine-tune GPT-3.5 as the last step in our model distillation process.

IV. Experiments and Results

A. Curated Dataset of 260 English Song Lyrics

We built our lyrics dataset from the perspective of a hypothetical user. Imagine a user experiencing negative emotions due to ostracism and listening to music to regulate their mood and restore their sense of psychological well-being while cognitively appraising their situation. Some of the songs in our curated dataset were collected from Reddit conversations where users shared song titles and artists while experiencing episodes of ostracism at school, work, or within their families. Based on this scenario, we defined ostracism proxy experiences, such as the loss of close relationships and significant life events that trigger innate threat responses, including breakups, bereavement, and social distancing during the COVID-19 lockdown. We collected lyrics based on keyword searches, such as songs listened to after breakups, during the pandemic, or while grieving. Additionally, about 20% of songs in the dataset were chosen to reflect experiences from the authors' volatile youth periods, identified by asking for songs listened to during our younger years without explicitly priming for ostracism experiences. To maintain the variance of musical expressions in the dataset, 100 English songs were randomly selected from the Music4All dataset [65], ensuring similarity to the within-class base rate.

B. ChatGPT-4 Zero-shot Annotation with Human in the Loop

For approximately six months, we have annotated lyrics using ChatGPT-4 and its recently optimized version, 4o. Due to human attention issues, we tested only a handful of lyrics per day to train ChatGPT-4s to perform as lyrics annotators using the following instructions:

> As a lyrics annotator, your task is to categorize the subjects' (the narrators or protagonists) states expressed or closely assumed in lyrics into one of the sixteen combinations based on four binary sub-dimensions and they are:
>
> 1) The subjects can either have High Self-Esteem (HSE) or Low Self-Esteem (LSE).
> 2) The subjects can either have High Self-Control (HSC) or Low Self-Control (LSC).
> 3) The subjects can either have Seeking to Belong (PSB) or Not Seeking to Belong (NSB).
> 4) The subjects can either have Seeking Recognition (PSR) or Not Seeking Recognition (NSR).
>
> Therefore, when you classify each lyrics, you must determine the either-or sub-dimensions in four dimensions: self-esteem, self-control, seeking to belong, and seeking recognition.

A song titled "Creep" by Radiohead on Genius, narrates perhaps one of the lowest levels of self-esteem, describing oneself as a creep. The following excerpt is how ChatGPT-4 annotates the song in four dimensions. In the Self-Esteem dimension: The singer repeatedly labels themselves as a "creep" and a "weirdo" indicating feelings of self-deprecation classified as Low Self-Esteem (LSE). In the Self-Control dimension: There's a longing for control and a perfect self that suggests current dissatisfaction and lack of control over one's life and self-perception indicating Low Self-Control (LSC). In Seeking to Belong dimension: Although the singer expresses feeling out of place "I don't belong here," there isn't an active search to belong; rather, there's a resignation to their perceived outsider status supporting Not Seeking to Belong (NSB). In the Seeking Recognition dimension: The focus is not on seeking validation or attention from others but rather on an internal struggle with their own identity and worth suggesting Not Seeking Recognition (NSR). ChatGPT-4 determines the classification as LSE, LSC, NSB, and NSR, which corresponds to Class 11.

Another song titled, “So What” by P!nk on Genius, displays high self-esteem describing oneself as a rock star. The following excerpt is the analysis of ChatGPT-4 in four dimensions. In the Self-Esteem dimension: The refrain “I am a rock star” suggests High Self-Esteem (HSE), implying confidence and self-worth. In the Self-Control dimension: The lines “I wanna start a fight” and “I’m gonna drink my money” could be seen as impulsive, pointing to Low Self-Control (LSC). In the Seeking to Belong dimension: The repeated statement “I don’t need you” and the celebratory tone about being single indicate Not Seeking to Belong (NSB). In the Seeking Recognition dimension: The strong assertions of identity, a rock star, and existence in the face of adversities imply a demand for recognition of their independence and resistance suggesting seeking recognition (PSR). Therefore, ChatGPT-4 determines the classification as HSE, LSC, NSB, and PSR, which corresponds to Class 6.

TABLE I
DISTRIBUTION OF SIXTEEN PSYCHOLOGICAL NEEDS CLASSES

Annotated by ChatGPT 4.0 & 4o with Human Feedback																
class	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
count	24	7	7	36	7	18	12	14	28	33	18	18	9	9	8	12

As a result, the distribution of 260 lyrics across the 16 classes, before splitting into training and testing datasets, is shown in Table 1. To increase the accuracy of annotations, we implement three rounds of refining prompts and annotating each set of lyrics with ChatGPT-4 and 4o using human validation and feedback (human in the loop). This process enables us to settle on the most feasible classification. Class 3 and Class 9 are the two most common classes, with 36 and 33 instances, respectively. When internal resources (such as self-esteem and self-control) are higher in Class 3 [HSE, HSC, NSB, NSR], the subjects of the lyrics do not express a need to belong or to be validated by others. Conversely, when internal resources are lower in Class 9 [LSE, LSC, PSB, PSR], the subjects of the lyrics signal a need to belong and seek validation. Considering 160 songs listened to during ostracism proxy experiences (such as breakups, grief, and social distancing during the pandemic) or when feeling ostracized (songs discussed in Reddit conversations), it is feasible that individuals feeling threatened by these experiences might listen to songs that either express their internal psychological needs or their external psychological needs. These songs signal the subjects’ psychological fitness (expressed in lyrics) and resonate with listeners by figuratively allowing themselves to step into the signalers’ shoes, thereby activating their mirror neurons at full throttle [38].

C. GPT-4o Assistant and Gemini 1.5 Few-shot Annotation

Using human feedback in the loop, ChatGPT-4 and 4o have shown a tendency to forget previous instructions over prolonged use, requiring us to start over with prompting. Additionally, we have encountered inconsistent answer formats and irrelevant analyses. To improve the effectiveness of generating large datasets with generative AI, we enter the next phase of model distillation to train more efficient models to annotate lyrics. We use two different instructions and datasets:

- **Trial 1:** Instructions given to ChatGPT-4s with 80 lyrics.
- **Trial 2:** Detailed Instructions with 80 lyrics that are sliced into [verse] + [chorus] for data augmentation.
- **Trial 3:** Detailed Instructions with 80 lyrics.

We utilize Gemini 1.5 by Google and GPT-4o Assistant by OpenAI, setting the randomness of outputs (temperature) to 0.2 and the probability mass (top p) to 0.1 to produce more deterministic results. These hyperparameters control the randomness of the model’s outputs, with lower values (0.2 and 0.1 respectively) focusing token selection on the most probable options, limiting it to the top 10%.

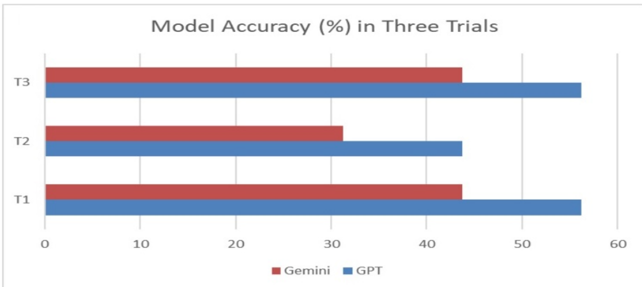

Fig. 2. Performance Comparison: GPT-4o and Gemini 1.5

As shown in Fig. 2, we use the simple average percentage for accuracy to evaluate the performance of two few-shot learning models. GPT-4o’s accuracy was higher than Gemini 1.5’s in all trials, with scores of 56.25%, 43.75%, and 56.75% compared to 43.75%, 31.25%, and 43.76%. Both Gemini 1.5 and GPT-4o performed worse in Trial 2 when we used detailed instructions and data augmentation techniques with split lyrics. It makes sense for lyrics to be categorized as one class with four dimensions, requiring all the lyrics (not splits). The highest score for GPT-4o in Trials 1 and 3 was 56.75%. Given that we have 16 classes, the accuracy indicates that while the model performs much better than random guessing (which would have an accuracy of 6.25% for 16 classes), there is certainly substantial room for improvement.

TABLE II
ACCURACY IN FOUR PSYCHOLOGICAL NEEDS DIMENSIONS

	GPT-4o	**Gemini 1.5**
Self-Esteem (HSE/LSE)	93.75%	93.75%
Self-Control (HSC/LSC)	87.50%	81.75%
Seeking to Belong (PSB/NSB)	81.25%	81.25%
Seeking Recognition (PSR/NSR)	56.25%	50.00%
Overall Accuracy (%)	79.69%	76.56%

As shown in Table 2, based on the analysis of four-dimensional accuracy in Trial 3, both GPT-4o and Gemini 1.5 achieve an impressive 93.75% accuracy in the Self-Esteem

(HSE/LSE) dimension. However, GPT-4o surpasses Gemini 1.5 in the Self-Control (HSC/LSC) dimension, with accuracy of 87.50% and 81.75%, respectively. The models perform equally moderately in the Seeking to Belong (PSB/NSB) dimension, each scoring 81.25%. However, in the Seeking Recognition (PSR/NSR) dimension, GPT-4o demonstrates a higher accuracy of 56.25% compared to Gemini 1.5's 50.00%, marking closer and exactly to random guessing. Overall, GPT-4o has a marginally higher overall accuracy at 79.69% compared to Gemini 1.5's 76.56%. The declining pattern in accuracy suggests that both models may be overfitting, as learning appears to occur in the order of Self-Esteem, Self-Control, Seeking to Belong, and Seeking Recognition dimension.

The declining accuracy pattern in both models can have potential reasons. Firstly, the dimensions might be ordered by increasing complexity, with Self-Esteem (HSE/LSE) being the simplest to classify and Seeking Recognition (PSR/NSR) the most challenging. Secondly, the models might have limitations in understanding the nuances required for the later dimensions, such as the subtle cues needed to recognize Seeking Recognition (PSR/NSR) due to working memory limitations. Thirdly, because the models were trained on a limited dataset, they might overfit familiar patterns like Self-Esteem but struggle with less common patterns or highly varying patterns in Seeking Recognition. Lastly, because the models were generated and learned reasons sequentially, earlier tasks might receive more optimization, causing progressively lower performance on later tasks.

To improve accuracy, after Trial 1, we developed conceptually more detailed instructions for Trials 2 and 3. We defined the four dimensions (self-esteem, self-control, seeking to belong, and seeking recognition) and provided behavioral inferences to anchor these definitions. For example, we added 33 detailed behavioral anchors to the High Self-Esteem sub-dimension, with similar numbers of anchors for all other sub-dimensions. Here are abbreviated versions of the High Self-Esteem (HSE) sub-dimension behavioral anchors included in Trial 2 and 3 prompts (some of them are shortened for brevity):

- Understand one's self-worth and the value of one's life.
- Express a confident, self-assured positive outlook.
- Express one's contentment and a sense of fulfillment.
- Show one's confidence in vocalizing their feelings.
- Recognize their own flaws but not let the world bring them down.
- Express a strong sense of self and a willingness to support another person.
- Recognize their own needs and desires, and the refusal to settle for less.
- Express an inherent belief in love and its transformative power.

Surprisingly, these detailed anchors did not affect the performance of both GPT-4o and Gemini 1.5. Therefore, we decided to abandon the detailed instructions and focus on increasing the example size per class. Due to the imbalanced dataset of 260 lyrics, we tested few-shot learning with 80 lyrics, balanced with 5 lyrics in each class. At the time of trials, increasing the number of examples was limited due to token limits. With this predicament, we venture the data augmentation technique to increase the number of examples with splitting lyrics. Not surprisingly, as mentioned earlier, splitting lyrics into one verse and a chorus (repeated in many splits depending on the number of verses in the original) did not improve model performance. This intuitively makes sense because the repeated annotations determined by four dimensions might not be relevant among splits, potentially increasing noise.

D. GPT-3.5-Turbo-0125 (Recommended) Fine-tuning

With the results of three few-shot trials, we move to the next step: fine-tuning a smaller version of GPT to finalize our model distillation. Although the model sizes are not publicly available, we start with GPT-4 and 4o, the updated version of GPT-3 consisting of over 175 billion parameters. We test the performance of these larger GPT-4 models in generating labels for our sixteen psychological needs classes. After the authors validate the labels, we harvest 260 lyrics with 16 class labels. Out of the total of 260 imbalanced datasets (refer to the distribution in Table 1 in subsection B), we use sets of 80 lyrics (a perfectly balanced set) and 109 lyrics (an approximately balanced set with 7 lyrics per class, except for classes 1, 2, and 3, which have 6 lyrics each). The following descriptions are based on reports from the GPT-3.5-turbo-0125 (recommended) version after fine-tuning it in two trials:

- **Trial 4 (80 lyrics):** 3,251,361 tokens, Epochs=3, Batch size=1, LR multiplier=2, Train loss=0.4324, Full validation loss=0.9579.
- **Trial 5 (109 lyrics):** 4,968,234 tokens, Epochs=3, Batch size=1, LR multiplier=2, Training loss=0.7395, Full validation loss=0.6827.

In Trial 4, we fine-tune the GPT-3.5-turbo-0125 model using a dataset consisting of 80 lyrics, processing a total of 3,251,361 tokens. The training process spans 3 epochs with a batch size of 1, allowing for individual processing and updating of each lyric. We employ a learning rate multiplier of 2 to enhance the base learning rate and expedite convergence. The training loss achieved is 0.4324, reflecting a relatively low error on the training data. However, the validation loss is 0.9579, indicating that although the model performs well on the training set, it shows signs of overfitting, as evidenced by the higher error on the validation set. This discrepancy underscores the need for further tuning to improve generalization to new data. For the subsequent trial, we increase the dataset size to 109 lyrics to maintain dataset balance and potentially enhance the model's performance.

In Trial 5, we fine-tune the GPT-3.5-turbo-0125 model using a dataset of 109 lyrics, resulting in the processing of 4,968,234 tokens. As in Trial 4, the training is conducted over 3 epochs with a batch size of 1 and a learning rate multiplier of 2. The training loss is 0.7395, higher than in Trial 4. However, the validation loss is notably lower at 0.6827, signifying improved generalization and performance on unseen data compared to Trial 4. The improvement suggests that the larger dataset in Trial 5 contributes to a more robust model. Despite the improvement, significant overfitting remains, alarming that the fine-tuned model is still far from optimal.

Consequently, we use a test set consisting of 16 hand-picked lyrics, chosen for their clear quality and distinct class differences, across five trials, including three few-shot trials, to test the accuracy of GPT-4o and Gemini 1.5. Somewhat unexpectedly, the accuracy of GPT-3.5-0125 in Trial 4 is 12.5%, nearly equivalent to a random response, even considering the smaller size of GPT-3.5 compared to GPT-4. In Trial 5, the accuracy is 31.25%, noticeably higher than in Trial 4. However, this is the same as the lowest accuracy shown in the second trial by Gemini 1.5, where we used splits of lyrics resulting in increased noise. Intuitively, the increase in the number of examples, from 80 to 109 lyrics, has improved accuracy. We are pleased to see that the difference in accuracy between using 80 lyrics and 109 lyrics is 18.75%, a hopeful increase achieved by adding two more lyrics per class, except for Classes 1, 2, and 3, which each had one additional lyrics. However, having closer to 7 examples per class falls short of the recommended minimum of 10 to 50 examples per class, according to some experimenters' anecdotal estimation commented in the OpenAI Forum (click here). In essence, the minimum number of examples depends on the specific tasks and domains. Until we test it with varying sizes of examples, we cannot estimate this with certainty. What is certain, though, is that at this time, the size of the examples matters in increasing the accuracy of the smaller model, considering the accuracy of 12.5% for the GPT-3.5-turbo-0125 model and 56.75% for the GPT-4o Assistant, both trained with 80 lyrics and slightly different interface.

The budget for all five trials cost us approximately $575.00, which is significantly less than the time and resources we would have spent training the models from scratch. Moreover, quickly obtaining the proof of concept is invaluable. However, because GPT-3.5-turbo-0125 is proprietary, fine-tuning it comes with limitations such as restricted access to closed information and less control over hyperparameters. In this preliminary study, facilitating OpenAI's fine-tuning interface to gain the proof of concept is sensible due to its user-friendliness and quick results. Nevertheless, it is advisable to switch to a more efficient pre-trained transformer model for production-level work, especially in mobile platforms.

V. Discussion

In today's digital era, the internet continues to expand rapidly, reflecting the dynamics of the modern world. This digital realm encompasses various types of unfiltered big data, contributing to the digital representation of human emotional and psychological expression. Just as in the physical world, the digital world requires methods to understand and monitor public emotional and psychological health. Understanding psychological needs is crucial for analyzing the digital footprints of human emotional and psychological conditions.

The development of the proposed model to recognize psychological needs expressed in song lyrics began after I listened to a story about the life-altering moments of a teenage boy who chose to live instead of fading away like dew in the early morning. This story was shared under the title "Power of One" in Prof. Williams' ostracism course at Purdue, inspiring our project to start with the concept of a mobile agent, PO2. After taking a fatal dose of psychotropic medication, the teenage boy saw the sole, smiling face of a kind girl from his school before falling into the abyss. Her smile prompted him to call 911 and save himself. After his recovery, he left a note explaining how her smile had helped him break through his darkest hour and thanking her for smiling at him, everyone's invisible person at school. Our vision with the PO2 project is to help identify and utilize popular songs to assist young people in coping with ongoing offline and online ostracism [66]. Our framework and computational methods are designed to recognize four fundamental psychological needs, marking the first step toward materializing the mission of our PO2 mobile agent. This agent aims to alleviate the harmful effects of ostracism through the therapeutic use of popular song listening. By promoting healthier personal and social adjustment facilitated by the lyrics of recommended songs, the PO2 agent, much like a smiling face, seeks to detect and mitigate the perpetuation of ostracism's detrimental impacts.

In addition to its therapeutic potential, identifying psychological needs as song content can address data privacy and protection issues inherent in current efforts to integrate physical, physiological, and contextual signals of listeners' emotions. For example, Spotify has disclosed a patented speech emotion recognition [67] and ambient noise detection [33]. Despite Spotify's stance on upholding user privacy, this technology has raised concerns among users, artists, and activist groups regarding potential invasions of privacy and discrimination against gender-transitioning individuals [33].

Moreover, in automating music search and retrieval, identifying psychological needs in song lyrics can significantly enhance music recommendation systems by better deciphering listeners' emotional and psychological states. A key player in this domain is Music Emotion Recognition (MER) recommendation systems, which rely on accurately recognizing emotions by extracting the emotional content from music [68] and assessing the emotional states of listeners [69]. MER recommendation systems, however, face a critical drawback due to the disparity between musical emotions and listeners' emotional states. Aside from privacy concerns, these technologies aim to detect and respond to listeners' transient emotional states but fail to address their more enduring psychological needs linked to their emotional states. These latent psychological needs, which are fundamental to listeners' motivation and closely tied to their emotional experiences, align more closely with their goal-setting behaviors and are better suited as metrics for tracking listeners' intentions.

Due to its novelty, our preliminary study presents not-yet-validated speculations that are further challenged by interdisciplinary dialogues and lack substantial corroborating evidence from prior research. Moving forward from the preliminary study, we are optimistic about leveraging generative AI to develop a classifier for psychological needs as credible song signals. Our next step is to continue annotating English song lyrics using an improved and faster version of ChatGPT-4o, incorporating human-in-the-loop processes, to achieve a balanced dataset with 50 to 100 lyrics per class and to expand our final test set to a minimum of 20% of the total dataset. Using model distillation, moving forward, we aim to fine-tune more efficient open-source LLMs with this extensive dataset.

ACKNOWLEDGMENT

Thank you, Dr. Kipling Williams, for your invaluable encouragement in developing my thesis. Your passion and knowledge in ostracism research were transformative to our vision.

REFERENCES

[1] T. DeNora, *Music asylums: Wellbeing through music in everyday life*. Routledge, 2016.

[2] W. M. Randall and N. S. Rickard, "Reasons for personal music listening: A mobile experience sampling study of emotional outcomes," *Psychology of Music*, vol. 45, no. 4, pp. 479–495, 2017. http://dx.doi.org/10.1177/0305735616666939

[3] P. Rowe and B. Guerin, "Contextualizing the mental health of metal youth: A community for social protection, identity, and musical empowerment," *Journal of Community Psychology*, vol. 46, no. 4, pp. 429–441, 2018. http://dx.doi.org/10.1002/jcop.21949

[4] T. ter Bogt, N. Canale, M. Lenzi, A. Vieno, and R. van den Eijnden, "Sad music depresses sad adolescents: A listener's profile," *Psychology of Music*, vol. 49, no. 2, pp. 257–272, 2021. http://dx.doi.org/10.1177/0305735619849622

[5] K. D. Williams, "Ostracism," *Annu. Rev. Psychol.*, vol. 58, pp. 425–452, 2007. http://dx.doi.org/10.1146/annurev.psych.58.110405.085641

[6] ——, *Ostracism: The power of silence*. Guilford Press, 2002.

[7] E. D. Wesselmann, M. R. Grzybowski, D. M. Steakley-Freeman, E. R. DeSouza, J. B. Nezlek, and K. D. Williams, "Social exclusion in everyday life," *Social exclusion: Psychological approaches to understanding and reducing its impact*, pp. 3–23, 2016. http://dx.doi.org/10.1007/978-3-319-33033-4_1

[8] C. M. Büttner, M. Jauch, M. Marinucci, K. D. Williams, R. Greifeneder, P. Riva, and S. C. Rudert, "It will (never) stop hurting: Do repeated or chronic experiences of exclusion lead to hyper-or hyposensitive psychological responses?" *Group Processes & Intergroup Relations*, p. 13684302221140002, 2023. http://dx.doi.org/10.1177/13684302221140002

[9] G. A. Dingle, J. Hodges, and A. Kunde, "Tuned in emotion regulation program using music listening: Effectiveness for adolescents in educational settings," *Frontiers in psychology*, vol. 7, p. 859, 2016. http://dx.doi.org/10.3389/fpsyg.2016.00859

[10] S. Uhlig, E. Jansen, and E. Scherder, ""being a bully isn't very cool…": Rap & sing music therapy for enhanced emotional self-regulation in an adolescent school setting–a randomized controlled trial," *Psychology of music*, vol. 46, no. 4, pp. 568–587, 2018. http://dx.doi.org/10.1177/0305735617719154

[11] S. Garrido, M. Du Toit, and T. Meade, "Music listening and emotion regulation: Young people's perspectives on strategies, outcomes, and intervening factors." *Psychomusicology: Music, Mind, and Brain*, 2022. http://dx.doi.org/10.1037/pmu0000285

[12] E. Michels-Ratliff and M. Ennis, "This is your song: Using participants' music selections to evoke nostalgia and autobiographical memories efficiently." *Psychomusicology: Music, Mind, and Brain*, vol. 26, no. 4, p. 379, 2016. http://dx.doi.org/10.1037/pmu0000167

[13] S. Laiho, "The psychological functions of music in adolescence," *Nordic Journal of music therapy*, vol. 13, no. 1, pp. 47–63, 2004. http://dx.doi.org/10.1080/08098130409478097

[14] G. Kreutz, "Does singing facilitate social bonding?" *Music and Medicine*, vol. 6, no. 2, pp. 51–60, 2014. http://dx.doi.org/10.47513/mmd.v6i2.180

[15] C. N. DeWall, R. S. Pond Jr, W. K. Campbell, and J. M. Twenge, "Tuning in to psychological change: Linguistic markers of psychological traits and emotions over time in popular us song lyrics." *Psychology of Aesthetics, Creativity, and the Arts*, vol. 5, no. 3, p. 200, 2011. http://dx.doi.org/10.1037/a0023195

[16] T. Schäfer, P. Sedlmeier, C. Städtler, and D. Huron, "The psychological functions of music listening," *Frontiers in psychology*, vol. 4, p. 511, 2013. http://dx.doi.org/10.3389/fpsyg.2013.00511

[17] K. D. Williams, "Ostracism: A temporal need-threat model," *Advances in experimental social psychology*, vol. 41, pp. 275–314, 2009. http://dx.doi.org/10.1016/S0065-2601(08)00406-1

[18] C. H. Hartgerink, I. Van Beest, J. M. Wicherts, and K. D. Williams, "The ordinal effects of ostracism: A meta-analysis of 120 cyberball studies," *PloS one*, vol. 10, no. 5, p. e0127002, 2015. http://dx.doi.org/10.1371/journal.pone.0127002

[19] S. A. Mehr, M. M. Krasnow, G. A. Bryant, and E. H. Hagen, "Toward a productive evolutionary understanding of music," *The Behavioral and brain sciences*, vol. 44, p. e122, 2021. http://dx.doi.org/10.1017/s0140525x21000030

[20] ——, "Origins of music in credible signaling," *Behavioral and Brain Sciences*, vol. 44, p. e60, 2021. http://dx.doi.org/10.1017/S0140525X20000345

[21] W. contributors, "Signalling theory," 2023, accessed: 2024-05-20. [Online]. Available: https://en.wikipedia.org/wiki/Signalling_theory

[22] P. N. Juslin and D. Västfjäll, "Emotional responses to music: The need to consider underlying mechanisms," *Behavioral and brain sciences*, vol. 31, no. 5, pp. 559–575, 2008. http://dx.doi.org/10.1017/s0140525x08005293

[23] T. Eerola and J. K. Vuoskoski, "A review of music and emotion studies: Approaches, emotion models, and stimuli," *Music Perception: An Interdisciplinary Journal*, vol. 30, no. 3, pp. 307–340, 2012. http://dx.doi.org/10.1525/mp.2012.30.3.307

[24] P. Ekman, "An argument for basic emotions," *Cognition & Emotion*, vol. 6, no. 3-4, pp. 169–200, 1992. http://dx.doi.org/10.1080/02699939208411068

[25] C. E. Izard and C. E. Izard, "Differential emotions theory," *Human emotions*, pp. 43–66, 1977. http://dx.doi.org/10.1007/978-1-4899-2209-0_3

[26] R. E. Thayer, J. R. Newman, and T. M. McClain, "Self-regulation of mood: strategies for changing a bad mood, raising energy, and reducing tension." *Journal of personality and social psychology*, vol. 67, no. 5, p. 910, 1994. http://dx.doi.org/10.1037//0022-3514.67.5.910

[27] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980. http://dx.doi.org/10.1037/h0077714

[28] R. Plutchik, *Emotions and life: Perspectives from psychology, biology, and evolution.* American Psychological Association, 2003.

[29] D. Keltner, J. A. Brooks, and A. Cowen, "Semantic space theory: Data-driven insights into basic emotions," *Current Directions in Psychological Science*, p. 09637214221150511, 2023. http://dx.doi.org/10.1177/09637214221150511

[30] C. E. Izard, "Emotion theory and research: Highlights, unanswered questions, and emerging issues," *Annual review of psychology*, vol. 60, pp. 1–25, 2009. http://dx.doi.org/10.1146/annurev.psych.60.110707.163539

[31] Y. Wang, W. Song, W. Tao, A. Liotta, D. Yang, X. Li, S. Gao, Y. Sun, W. Ge, W. Zhang *et al.*, "A systematic review on affective computing: Emotion models, databases, and recent advances," *Information Fusion*, vol. 83, pp. 19–52, 2022. http://dx.doi.org/10.48550/arXiv.2203.06935

[32] V. Chaturvedi, A. B. Kaur, V. Varshney, A. Garg, G. S. Chhabra, and M. Kumar, "Music mood and human emotion recognition based on physiological signals: a systematic review," *Multimedia Systems*, vol. 28, no. 1, pp. 21–44, 2022. http://dx.doi.org/10.1007/s00530-021-00786-6

[33] A. Name, "Spotify patented emotional recognition technology to recommend songs based on users' emotions," 2022, accessed: 2023-10-16. [Online]. Available: https://jolt.richmond.edu/2022/01/11/spotify-patented-emotional-recognition-technology-to-recommend-\protect\penalty\z@songs-based-on-users-emotions/

[34] S. Koelsch, E. Kasper, D. Sammler, K. Schulze, T. Gunter, and A. D. Friederici, "Music, language and meaning: brain signatures of semantic processing," *Nature neuroscience*, vol. 7, no. 3, pp. 302–307, 2004. http://dx.doi.org/https://doi.org/10.1038/nn1197

[35] T. Schäfer, P. Sedlmeier, C. Städtler, and D. Huron, "The psychological functions of music listening," *Frontiers in psychology*, vol. 4, p. 511, 2013. http://dx.doi.org/10.3389/fpsyg.2013.00511

[36] K. Schäfer and T. Eerola, "How listening to music and engagement with other media provide a sense of belonging: An exploratory study of social surrogacy," *Psychology of Music*, vol. 48, no. 2, pp. 232–251, 2020. http://dx.doi.org/10.1177/0305735618795036

[37] R. Granot, D. H. Spitz, B. R. Cherki, P. Loui, R. Timmers, R. S. Schaefer, J. K. Vuoskoski, R.-N. Cárdenas-Soler, J. F. Soares-Quadros Jr, S. Li *et al.*, ""help! i need somebody": music as a global resource for obtaining wellbeing goals in times of crisis," *Frontiers in Psychology*, vol. 12, p. 648013, 2021. http://dx.doi.org/10.3389/fpsyg.2021.648013

[38] I. Molnar-Szakacs and K. Overy, "Music and mirror neurons: from motion to'e'motion," *Social cognitive and affective neuroscience*, vol. 1, no. 3, pp. 235–241, 2006. http://dx.doi.org/10.1093/scan/nsl029

[39] L. Fogassi, P. F. Ferrari, B. Gesierich, S. Rozzi, F. Chersi, and G. Rizzolatti, "Parietal lobe: from action organization to intention understanding," *Science*, vol. 308, no. 5722, pp. 662–667, 2005. http://dx.doi.org/10.1126/science.1106138

[40] L. Carr, M. Iacoboni, M.-C. Dubeau, J. C. Mazziotta, and G. L. Lenzi, "Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas," *Proceedings of the national Academy of Sciences*, vol. 100, no. 9, pp. 5497–5502, 2003. http://dx.doi.org/10.1073/pnas.0935845100

[41] A. M. Liberman and D. H. Whalen, "On the relation of speech to language," *Trends in cognitive sciences*, vol. 4, no. 5, pp. 187–196, 2000. http://dx.doi.org/https://doi.org/10.1016/s1364-6613(00)01471-6

[42] C. Y. Wan, K. Demaine, L. Zipse, A. Norton, and G. Schlaug, "From music making to speaking: Engaging the mirror neuron system in autism," *Brain research bulletin*, vol. 82, no. 3-4, pp. 161–168, 2010. http://dx.doi.org/10.1016/j.brainresbull.2010.04.010

[43] G. L. Wagener, M. Berning, A. P. Costa, G. Steffgen, and A. Melzer, "Effects of emotional music on facial emotion recognition in children with autism spectrum disorder (asd)," *Journal of Autism and Developmental Disorders*, vol. 51, pp. 3256–3265, 2021. http://dx.doi.org/10.1007/s10803-020-04781-0

[44] R. Plutchik, "A psychoevolutionary theory of emotions," 1982.

[45] J. P. Hewitt, "22 the social construction of self-esteem," *The Oxford handbook of positive psychology*, p. 309, 2020. http://dx.doi.org/10.1093/oxfordhb/9780195187243.013.0020

[46] D. Ren, E. Wesselmann, and K. D. Williams, "Evidence for another response to ostracism: Solitude seeking," *Social Psychological and Personality Science*, vol. 7, no. 3, pp. 204–212, 2016. http://dx.doi.org/10.1177/1948550615616169

[47] S. E. Taylor, L. C. Klein, B. P. Lewis, T. L. Gruenewald, R. A. Gurung, and J. A. Updegraff, "Biobehavioral responses to stress in females: tend-and-befriend, not fight-or-flight." *Psychological review*, vol. 107, no. 3, p. 411, 2000. http://dx.doi.org/10.1037/0033-295x.107.3.411

[48] J. A. Russell, "Core affect and the psychological construction of emotion." *Psychological review*, vol. 110, no. 1, p. 145, 2003. http://dx.doi.org/10.1037/0033-295x.110.1.145

[49] E. Yaakobi, "Immediate moderating effects of attribution on ostracism distress," *Australian Psychologist*, vol. 56, no. 3, pp. 193–203, 2021. http://dx.doi.org/10.1080/00050067.2020.1829452

[50] T. Y.-C. Yeung, "Did the covid-19 pandemic trigger nostalgia? evidence of music consumption on spotify," *Evidence of Music Consumption on Spotify (August 21, 2020)*, 2020.

[51] N. Henry, D. Kayser, and H. Egermann, "Music in mood regulation and coping orientations in response to covid-19 lockdown measures within the united kingdom," *Frontiers in psychology*, vol. 12, p. 647879, 2021. http://dx.doi.org/10.3389/fpsyg.2021.647879

[52] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017. http://dx.doi.org/10.48550/arXiv.1706.03762

[53] M. O. Topal, A. Bas, and I. van Heerden, "Exploring transformers in natural language generation: Gpt, bert, and xlnet," *arXiv preprint arXiv:2102.08036*, 2021. http://dx.doi.org/10.48550/arXiv.2102.08036

[54] F. A. Acheampong, H. Nunoo-Mensah, and W. Chen, "Transformer models for text-based emotion detection: a review of bert-based approaches," *Artificial Intelligence Review*, pp. 1–41, 2021. http://dx.doi.org/10.1007/s10462-021-09958-2

[55] J. Huang, J. Tao, B. Liu, Z. Lian, and M. Niu, "Multimodal transformer fusion for continuous emotion recognition," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020. http://dx.doi.org/10.1109/ICASSP40776.2020.9053762 pp. 3507–3511.

[56] D. Cortiz, "Exploring transformers in emotion recognition: a comparison of bert, distillbert, roberta, xlnet and electra," *arXiv preprint arXiv:2104.02041*, 2021. http://dx.doi.org/10.48550/arXiv.2104.02041

[57] J. S. Gómez-Cañón, E. Cano, T. Eerola, P. Herrera, X. Hu, Y.-H. Yang, and E. Gómez, "Music emotion recognition: Toward new, robust standards in personalized and context-sensitive applications," *IEEE Signal Processing Magazine*, vol. 38, no. 6, pp. 106–114, 2021. http://dx.doi.org/10.1109/MSP.2021.3106232

[58] Y. Wang, J. Tian, D. S. Ones, R. N. Landers *et al.*, "Using natural language processing and machine learning to replace human content coders." *Psychological Methods*, 2022. http://dx.doi.org/10.1037/met0000518

[59] M. Alizadeh, M. Kubli, Z. Samei, S. Dehghani, J. D. Bermeo, M. Korobeynikova, and F. Gilardi, "Open-source large language models outperform crowd workers and approach chatgpt in text-annotation tasks," *arXiv preprint arXiv:2307.02179*, 2023. http://dx.doi.org/10.48550/arXiv.2307.02179

[60] M. Li, T. Shi, C. Ziems, M.-Y. Kan, N. F. Chen, Z. Liu, and D. Yang, "Coannotating: Uncertainty-guided work allocation between human and large language models for data annotation," *arXiv preprint arXiv:2310.15638*, 2023. http://dx.doi.org/10.18653/v1/2023.emnlp-main.92

[61] M. Hasanain, F. Ahmed, and F. Alam, "Large language models for propaganda span annotation," *arXiv preprint arXiv:2311.09812*, 2023. http://dx.doi.org/10.48550/arXiv.2311.09812

[62] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–35, 2023. http://dx.doi.org/10.48550/arXiv.2107.13586

[63] S. Mohammadi, M. Mohammadi, S. Sinaei, A. Balador, E. Nowroozi, F. Flammini, and M. Conti, "Balancing privacy and accuracy in federated learning for speech emotion recognition," in *Proceedings of the 18th Conference on Computer Science and Intelligence Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki, and D. Ślęzak, Eds., vol. 35. IEEE, 2023. http://dx.doi.org/http://dx.doi.org/10.15439/2023F444 p. 191–199.

[64] IBM, "Zero-shot learning," 2024, accessed: 2024-07-20. [Online]. Available: https://www.ibm.com/topics/zero-shot-learning

[65] I. A. P. Santana, F. Pinhelli, J. Donini, L. Catharin, R. B. Mangolin, V. D. Feltrim, M. A. Domingues *et al.*, "Music4all: A new music database and its applications," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE, 2020. http://dx.doi.org/10.1109/IWSSIP48289.2020.9145170 pp. 399–404.

[66] K. D. Williams, C. K. Cheung, and W. Choi, "Cyberostracism: effects of being ignored over the internet." *Journal of personality and social psychology*, vol. 79, no. 5, p. 748, 2000. http://dx.doi.org/https://psycnet.apa.org/doi/10.1037/0022-3514.79.5.748

[67] J. Borst, L. Wehrheim, A. Niekler, and M. Burghardt, "An evaluation of a zero-shot approach to aspect-based sentiment classification in historic german stock market reports," in *Communication Papers of the 18th Conference on Computer Science and Intelligence Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki, and D. Ślęzak, Eds., vol. 37. PTI, 2023. http://dx.doi.org/http://dx.doi.org/10.15439/2023F3725 p. 51–60.

[68] D. Han, Y. Kong, J. Han, and G. Wang, "A survey of music emotion recognition," *Frontiers of Computer Science*, vol. 16, no. 6, p. 166335, 2022. http://dx.doi.org/http://dx.doi.org/10.1007/s11704-021-0569-4

[69] M. de Witte, A. d. S. Pinho, G.-J. Stams, X. Moonen, A. E. Bos, and S. van Hooren, "Music therapy for stress reduction: a systematic review and meta-analysis," *Health Psychology Review*, vol. 16, no. 1, pp. 134–159, 2022. http://dx.doi.org/https://doi.org/10.1080/17437199.2020.1846580