PTEI: Integrating Personality Traits to Enhance Emotional Intelligence in Large Language Models

Anonymous ACL submission

Abstract

Despite advances in *Emotional Intelligence (EI)*, Large Language Models (LLMs) still significantly underperform humans in complex emotional reasoning. This gap originates partly from the limited incorporation of individual differences, particularly personality traits, which are fundamental to human emotional inference. To address this, we propose **PTEI**, a novel framework for integrating Personality Traits into Emotional Intelligence tasks using LLMs. In PTEI, MBTI and OCEAN personality traits are first extracted directly from the given emotional scenarios and then utilized as contextual knowledge within personality-aware prompts, guiding LLMs to accurately infer emotions and their underlying causes. To ensure optimal contextual grounding, we employ *Contrastive Learning* to construct an optimized retrieval system that surfaces emotionally and personally aligned scenarios, enhancing reasoning quality. Extensive experiments on established EI benchmarks show that PTEI enhances Emotional Understanding (EU) capabilities of various LLMs in EI, with the strongest improvement observed in GPT models, where combining PTEI with *Chain-of-Thought (CoT) reasoning* yields an additional 4% increase in accuracy. These findings underscore PTEI's contribution toward advancing AI systems with more sophisticated social and psychological grounding.

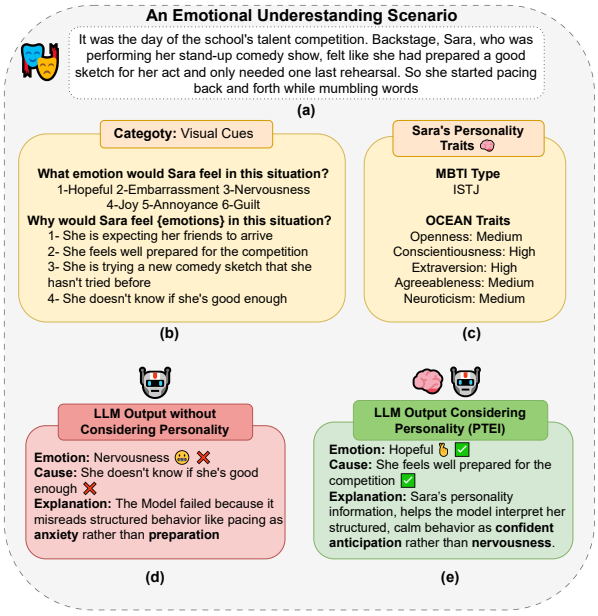

Figure 1: Illustration of PTEI's impact on emotional inference in LLMs. (a) An emotionally ambiguous scenario featuring Sara. (b) The LLM's task: a multiple-choice question asks for both Sara's emotion and its underlying cause. (c) Personality traits extracted information for Sara. (d) Baseline LLM prediction without personality knowledge misinterprets Sara's structured behavior as *nervousness*. (e) Our PTEI framework infers the emotion as *hopeful*, leveraging psychological context to explain her behavior as confident preparation.

1 Introduction

Emotional intelligence (EI), the ability to perceive, understand, regulate, and express emotions, is essential for effective communication, social interaction, and decision making (Salovey and Mayer, 1990; Goleman, 1996; Hess and Bacigalupo, 2011). As Large Language Models (LLMs) are increasingly deployed in human-facing applications, there is growing interest in evaluating and improving their emotional capabilities (Wang et al., 2023). While recent studies show that models like GPT-4 can perform well on tasks such as emotional awareness and understanding (Elyoseph et al., 2023), their abilities remain limited, particularly in scenarios involving implicit emotional situations or subjective interpretation (Maruf et al., 2024). This presents an ongoing challenge in Natural Language processing (NLP): enabling LLMs to reliably interpret and reason about human emotions in context. Enhancing EI in LLMs is therefore critical for more natural and effective human-AI collaboration.

Recent EI benchmarks such as EQ-Bench (Paech, 2023) and EmoBench (Sabour

et al., 2024) offer structured evaluations of LLMs' EI, but they still struggle with complex aspects such as emotional reasoning, regulation, and application in ambiguous social contexts. While these benchmarks represent an important step forward, a key limitation is their lack of personal context; they overlook individual characteristics such as personality traits, which are known to significantly shape emotional inference and behavior (Robinson and Clore, 2002; Sap et al., 2022). In psychology, personality traits refer to enduring individual differences in patterns of thinking, feeling, and behaving, as described by trait theory (McCrae and Costa, 1997). This gap reflects an issue in how LLMs are typically prompted or fine-tuned for EI tasks such as Emotional Understanding (EU): most current methodologies operate without incorporating sufficient personal context and tend to treat all emotional scenarios as one-size-fits-all. Hence, EI evaluations often remain surface-level and fail to capture the individualized, psychologically grounded reasoning required for real-world EU, as illustrated in Figure 1.

Personality has long been recognized as a critical factor in shaping emotional perception and behavior (McCrae and John, 1992; Myers, 1987). In humans, individual differences in traits such as openness, neuroticism, or extraversion are correlated with how emotions are interpreted, regulated, and expressed (Izard et al., 1993). Recent work has shown that LLMs can exhibit consistent and measurable personality traits in their responses, and these traits can be shaped and aligned with desired profiles (Safdari et al., 2023). Despite extensive research on personality prediction from text (Stajner and Yenikent, 2020; Mehta et al., 2020; Amirhosseini and Kazemian, 2020; Sorokovikova et al., 2024), most existing studies have either focused speaker characteristic for emotion recognition (Fu et al., 2025) or explored the interaction between personality and emotion in narrow settings relying on a single framework (e.g., OCEAN (McCrae and John, 1992), MBTI (Myers, 1987)) and rarely addressing EI as a broader construct (Wang et al., 2024). Thus, the role of personality traits in enhancing the EI of LLMs remains largely underexplored.

This paper proposes **PTEI (Personality Traits in Emotional Intelligence)**, a novel framework to systematically integrates OCEAN and MBTI personality traits to enhance EI in LLMs. PTEI extracts individual personality traits directly from textual scenarios and leverages this information through personality-aware prompting to improve emotion and cause prediction. Additionally, PTEI employs a Contrastive Learning-based embedding method and a retrieval mechanism to identify emotionally and personally similar scenarios, which helps ground the model's reasoning in psychologically aligned examples and improves its contextual sensitivity. Our approach specifically targets implicit and ambiguous emotional scenarios, significantly improving LLMs' capabilities in EI tasks and promoting more psychologically grounded inference strategies.

In summary, our contributions are as follows:

- We propose **PTEI**, first comprehensive framework utilizing MBTI and OCEAN personality traits into EU tasks for LLMs, addressing both type and trait theories of personality.
- We design an efficient, personality detection module that leverages structured few-shot prompting to infer MBTI and OCEAN traits and incorporates this knowledge into customized prompts for emotion and cause prediction.
- We construct a synthetic memory bank of similar EU scenarios to the EI benchmark and enriched it with fine-grained personality annotations. We also introduce a personality-aware Contrastive Learning (CL) objective to structure the scenario embedding space, enabling more effective retrieval of emotionally and personally aligned examples to support contextual reasoning in our few-shot setup.
- We demonstrate that integrating personality traits via few-shot and Chain-of-Thought (CoT) prompting enhances EI in LLMs, consistently outperforming personality-agnostic baselines and substantially narrowing the performance gap to human-level inference on challenging EI benchmarks.

2 Related Work

2.1 Personality-based Methods

Analyzing personality traits from a psychological perspective plays a crucial role in understanding and predicting human behavior and emotions. Among the various models, the Big Five personality framework (known as OCEAN), encompassing Openness, Conscientiousness, Extraver-

sion, Agreeableness, and Neuroticism (McCrae and John, 1992), and the Myers-Briggs Type Indicator (MBTI), based on four categories: Introversion versus Extraversion, Sensing vs Intuition, Thinking vs Feeling, and Judging vs Perceiving (Myers, 1987), are two of the most widely used approaches for characterizing individual personality profiles.

Personality prediction from text has emerged as a prominent task in NLP (Stajner and Yenikent, 2020; Mehta et al., 2020; Amirhosseini and Kazemian, 2020). Extensive research has focused on enhancing the detection of personality traits in human-generated text using LLMs (Sorokovikova et al., 2024). For example, PADO (Yeo et al., 2025) introduces personality-induced agents that estimate OCEAN trait levels by using GPT-4o and LLaMA3-8B. Similarly, PsyCoT (Yang et al., 2023) employs LLMs as AI assistants, utilizing a CoT approach based on specially designed questionnaires to facilitate personality inference. Beyond identifying personality traits, it is crucial to understand their interplay with other cognitive functions, such as emotional processing, which is central to our work.

Emotion features have been shown to enhance personality prediction performance in LLMs (Li et al., 2025, 2022), while personality traits themselves serve as valuable features for emotion recognition, particularly in conversational scenarios; LaERC-S (Fu et al., 2025) exploits speaker characteristics to improve emotion prediction in dialogue and ERC-DP (Wang et al., 2024) proposes a dynamic personality detection module that extracts OCEAN traits of a speaker from conversations rather than assuming static traits, thereby improving conversational emotion recognition.

While prior work explores the interplay between emotions and personality and are often focus on either OCEAN or MBTI, we examine their combined impact on recognizing implicit emotional expressions, enabling more nuanced emotional inference through a fuller psychological profile.

2.2 Emotional Intelligence (EI)

EI, the ability to recognize, understand, and regulate emotions, is key in psychology and social computing (Salovey and Mayer, 1990). As LLMs enter emotionally sensitive domains, EI has gained prominence in AI. Early work (Schuller and Schuller, 2018) identified emotion recognition, generation, and augmentation as pillars of Artificial Emotional Intelligence (AEI).

LLMs have achieved high performance on emotion-related tasks in practical domains, such as emotion-cause pair extraction using CoT prompting (Wu et al., 2024), and emotionally supportive dialogue generation via explicit strategy modeling (Wan et al., 2025). Despite these promising results, rigorous evaluation of emotional reasoning in LLMs has remained limited. Recent EI benchmarks like EmoBench (Sabour et al., 2024) and EQ-Bench (Paech, 2023) were developed to assess deeper EI capabilities such as EU, management, and social reasoning, thus these benchmarks show LLMs lag behind humans on EI tasks.

However, existing EI benchmarks and systems often treat EI as a generic skill, overlooking individual-level factors that shape emotional responses. Personality traits, as defined by OCEAN and MBTI, are crucial in how emotions are perceived, interpreted, and expressed, yet remain underutilized in current LLM-based EI evaluations. We address this gap by introducing a personality-aware framework that promotes more personalized and psychologically grounded emotional reasoning. To enhance contextual grounding, we employ CL to build a retrieval system that surfaces emotionally and personally aligned scenarios.

3 Methodology

3.1 Problem Definition

Given a dataset of EU scenarios $\mathcal{S} = \{s_1, s_2, ..., s_N\}$, each scenario s_i is a natural language description involving a subject a_i who experiences one or more emotions in a specific context. The task is to infer the pair $(\hat{e}_i, \hat{r}_i) \in \mathcal{E} \times \mathcal{R}$, where $\hat{e}_i$ denotes the predicted emotion label from a predefined set of emotion categories $\mathcal{E}$ (e.g., *grateful*, *anxious*, *frustrated*) and $\hat{r}_i$ denotes the corresponding predicted cause or trigger. Each scenario is also assigned a category label $c_i \in \mathcal{C}$, specifying the reasoning type required for interpretation.

Our key innovation is conditioning inference on personality traits. The objective is to learn a function f_{EI} where $P(s_i, a_i) = (M_i, O_i)$ represent the personality profile for the subject a_i in scenario s_i, M_i is the MBTI type and O_i is the OCEAN profile.

$$f_{\text{EI}} : (s_i, a_i, P(a_i)) \mapsto (\hat{e}_i, \hat{r}_i) \quad (1)$$

Function f_{EI} maps a scenario s_i with subject a_i and its associated subject's personality profile $P(a_i)$ to its corresponding emotion-cause pair through context-aware and personality-informed reasoning.

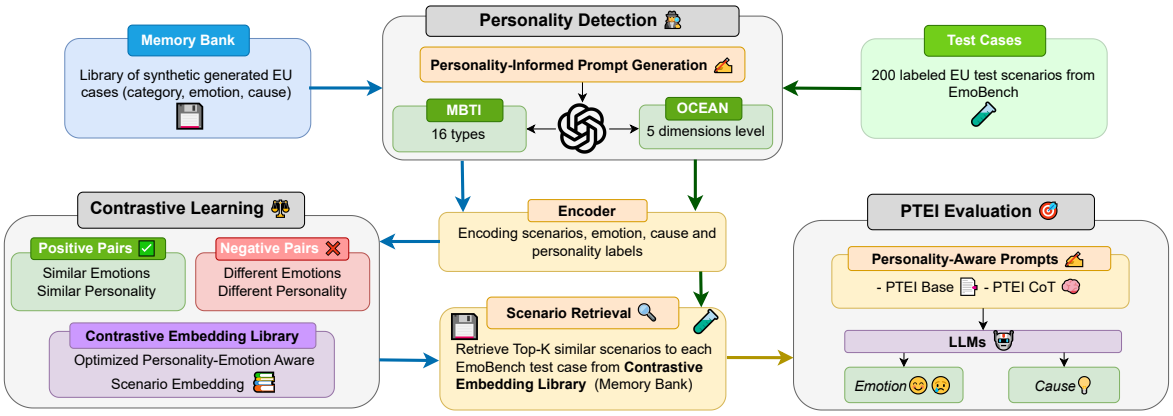

Figure 2: PTEI system architecture. First, the personality detection module extracts personality traits for both memory bank and test scenarios. Then, the memory bank scenarios are encoded and used in a Contrastive Learning setup to generate the contrastive embedding library [blue arrow]. For each test case, its encoded representation is used to retrieve similar scenarios from contrastive embedding [green arrow]. Finally, the test scenario is combined with its personality traits and the retrieved examples to evaluate PTEI's framework [yellow arrow].

3.2 Architecture Overview

Figure 2 illustrates the overall architecture of our proposed PTEI framework. The system is designed to improve EI in LLMs by systematically incorporating psychological and contextual signals. It consists of 3 primary components: (1) a personality detection module, (2) CL for embedding optimization and (3) scenario retrieval and final inference. The pipeline begins by inferring the subject's personality traits $P(a_i)$ (both OCEAN and MBTI) from input scenarios, which are used to construct personality-aware prompts for the final inference LLM. Moreover, a CL module optimizes the scenario embedding space from a memory bank by aligning similar emotion/personality pairs and separating dissimilar ones. In parallel, top-k similar scenarios are retrieved to support our base and CoT prompting setup. Finally, these components feed into an LLM that performs joint emotion and cause prediction, enabling the model to reason across diverse and psychologically grounded contexts.

3.3 Memory Bank Construction

To support few-shot learning and enrich EU via contextual analogies, we construct a memory bank $\mathbb{B}$ of $N_{\mathbb{B}} = 500$ diverse EU scenarios, each annotated with an emotion label, its corresponding cause, and inferred personality traits. These instances are synthesized using a GPT-4 model (OpenAI, 2023) inspired by EmoBench EU test cases scenarios and structure, as manual creation is costly and requires expert annotation (see Appendix A).

We define the memory bank $\mathbb{B}$ as:

$$\mathbb{B} = \{(s_i, a_i, e_i, r_i, P(a_i))\}_{i=1}^{N_{\mathbb{B}}} \quad (2)$$

where s_i is a generated scenario, a_i is the subject, e_i denotes emotion labels, r_i denotes cause labels, and $P(a_i)$ denotes personality profile extracted for the subject a_i in the scenario using the personality detection module. This memory bank $\mathbb{B}$ serves as a reference library from which top-k similar examples are retrieved based on proximity in the personality-aware embedding space to a given test scenario. The retrieved examples $S_{\text{retrieved}}$ are then used to construct few-shot prompts, enabling the LLM to perform personality-aware emotional inference with improved contextual grounding.

$$S_{\text{retrieved}}(s_j) = \{(\hat{s}_j, \hat{a}_j, \hat{e}_j, \hat{r}_j, P(\hat{a}_j))\}_{j=1}^{k} \quad (3)$$

More results of the memory bank analysis are available in Appendix B.

3.4 Personality Detection Module

To provide structured psychological context for each EU scenario, we implement a personality detection module that infers both MBTI and OCEAN profiles directly from scenario text s_i. For subject a_i in s_i, the module outputs: (1) an MBTI type $M_i \in \mathcal{M}_{\text{MBTI}}$, where $\mathcal{M}_{\text{MBTI}}$ is a set of 16 Myers-Briggs personality types (Myers, 1987), and (2) a set of Big Five (OCEAN) trait levels (McCrae and John, 1992) $o_i = \{o_i^{(1)}, ..., o_i^{(5)}\}$, where each $o_i^{(j)} \in \{low, medium, high\}$.

We adopt a prompt-based annotation strategy using GPT-4o-mini, where each prompt is designed

to elicit structured MBTI and OCEAN predictions based on the subject's inferred behavior and contextual cues in the scenario. This approach allows for efficient and scalable personality annotation without requiring manually labeled personality data for every scenario. Full prompt templates and examples are provided in Appendix C.1.

3.5 Personality-Aware Contrastive Learning

To enhance the ability of our framework to understand and reason about emotions within individual psychological contexts, we propose a personality-aware contrastive learning approach. It learns a robust scenario embedding space $E : s \mapsto z \in \mathbb{R}^d$ such that scenarios are embedded based on both emotional content and personality traits.

Pair construction. We construct positive and negative pairs from our memory bank $\mathbb{B}$. Given a pair of scenarios $((s_i, a_i), (s_j, a_j))$:

- Positive pair: if their emotion labels match $(e_i = e_j)$ and their personality profiles are similar $(S_{\text{personality}}(P(a_i), P(a_j)) \geq \theta_s)$, where the similarity threshold $\theta_s = 0.7$.
- Negative pair: if their emotion labels differ $(e_i \neq e_j)$ or they share the same emotion label but have dissimilar personalities $(S_{\text{personality}}(P(a_i), P(a_j)) < \theta_d)$, with dissimilarity threshold $\theta_d = 0.3$.

Personality similarity metric. To quantify personality similarity $S_{\text{ps}} = S_{\text{personality}}(P_1, P_2)$ between two profiles $P_1 = (M_1, O_1)$ and $P_2 = (M_2, O_2)$, we use the following composite metric:

$$S_{\text{ps}} = \alpha \cdot S_{\text{MBTI}}(M_1, M_2) + (1-\alpha) \cdot S_{\text{OCEAN}}(O_1, O_2) \quad (4)$$

where α balances the influence of MBTI and OCEAN similarities (default $\alpha = 0.5$). S_{MBTI} is computed as the fraction of matching MBTI dimensions, and S_{OCEAN} is the cosine similarity between trait vectors, with each trait mapped to a numeric scale: *high* = 1.0, *medium* = 0.5, and *low* = 0.0.

Contrastive training. Scenarios are encoded into vector representations using a pretrained sentence encoder (`all-mpnet-base-v2`[1]) from the SentenceTransformers library to generate semantically and personally aligned embeddings from our training data in the memory bank. (Reimers and Gurevych, 2019). We fine-tune the encoder on scenario pairs using a contrastive objective, normalized temperature-scaled cross-entropy (NT-Xent) loss (Chen et al., 2020):

$$\mathcal{L}_{\text{contrastive}} = -\log \frac{\exp(\text{sim}(i,j)/\tau)}{\sum_{k=1}^{N} \exp(\text{sim}(i,k)/\tau)} \quad (5)$$

Here, $\text{sim}(i, j)$ is the cosine similarity between positive pair embeddings, and τ (typically 0.07) is the temperature hyperparameter.

Scenario retrieval. After training, the learned embedding space enables efficient retrieval of psychologically aligned examples through k-nearest neighbor (KNN) search (Johnson et al., 2021) at inference. These retrieved examples are then incorporated into few-shot prompts to enhance the personality-aware EU capability of the framework.

4 Experiments

4.1 Experimental Setup

All experiments were conducted using an NVIDIA A100 GPU with 40GB VRAM. We used a combination of open and closed source LLMs, including GPT-4 (OpenAI, 2023), LLaMA-3 (Grattafiori et al., 2024), and Qwen (Bai et al., 2023), accessed through API interfaces.

For the contrastive learning module (Section 3.5), we fine-tuned a sentence encoder with a lightweight projection head using the NT-Xent loss ($\tau = 0.07$). Scenario pairs were sampled from our personality-enriched memory bank $\mathbb{B}$ (Section 3.3), with $\alpha = 0.5$, a positive similarity threshold = 0.7 and a negative threshold = 0.3. We trained the model with a batch size of 32, using the Adam optimizer (learning rate = 2e-5), selecting the checkpoint with the best validation retrieval accuracy.

4.2 Dataset

We primarily evaluated our PTEI framework on the EmoBench benchmark (Sabour et al., 2024). EmoBench features emotionally complex scenarios for EU and Emotional Application (EA) tasks. We use 200 English multiple-choice scenarios from the EU task, selected for their focus on inferring emotions and causes from rich textual descriptions. Each scenario includes two questions: one on the subject's primary emotion and another on its cause. To support few-shot prompting and retrieval, we also generated 500 synthetic scenarios using GPT-4 (Section 3.3). These synthetic examples were

[1] https://www.sbert.net/docs/pretrained_models.html

LLM	Method	CE	PBE	PT	EC	Emotion	Cause	Overall
Qwen-7B	Base	28.06	21.88	16.42	**28.57**	29.13	57.38	22.5
	PTEI-Base	**29.59**↑	**23.21**↑	**16.79**↑	27.68↓	**30.50**↑	**57.75**↑	**23.25**↑
	CoT	25.51	21.88	16.67	26.79	29.63	55.13	21.38
	PTEI-CoT	23.98↓	20.09↓	16.04↓	**28.57**↑	29.25↓	55.00↓	20.88↓
Llama3.1-8B	Base	**18.37**	14.73	17.91	**14.29**	26.25	53.37	16.62
	PTEI-Base	17.86↓	**16.07**↑	**20.15**↑	**14.29**	**27.12**↑	**54.12**↑	**17.63**↑
	CoT	14.80	14.29	9.70	9.82	22.50	34.75	12.25
	PTEI-CoT	17.86↑	15.18↑	11.94↑	10.71↑	24.25↑	36.12↑	14.13↑
Qwen-14B	Base	46.94	**35.27**	26.12	**38.39**	42.13	62.88	35.50
	PTEI-Base	47.45↑	**35.27**	**27.61**↑	**38.39**	42.75↑	**63.38**↑	**36.12**↑
	CoT	43.37	25.45	22.76	33.93	40.62	58.12	30.12
	PTEI-CoT	**49.49**↑	29.46↑	25.75↑	**38.39**↑	**45.00**↑	60.38↑	34.38↑
GPT-4o	Base	**78.57**	43.75	**55.22**	73.21	63.63	79.62	60.25
	PTEI-Base	73.47↓	57.14↑	52.99↓	74.11↑	65.50↑	80.75↑	62.12↑
	CoT	69.90	54.02	49.25	72.32	62.50	79.75	58.88
	PTEI-CoT	74.49↑	**58.04**↑	**55.22**↑	**75.89**↑	**66.88**↑	**82.38**↑	**63.62**↑

Table 1: Evaluation results on the Emotional Understanding (EU) task across LLMs of different sizes. We compare four prompting methods: Base (no personality or retrieval), PTEI-Base (personality-informed prompting), CoT (standard chain-of-thought), and PTEI-CoT (our full framework combining personality, CoT, and retrieval). Results cover four reasoning categories (CE, PBE, PT, EC), along with Emotion, Cause, and Overall accuracy. ↑ and ↓ indicate PTEI changes relative to Base or CoT. **Bold** denotes the best score per model.

annotated with emotion labels, causes, and inferred MBTI and OCEAN personality traits, forming a personality-enriched memory bank $\mathbb{B}$ to improve emotional reasoning during inference.

4.3 Baselines

To evaluate the effectiveness of PTEI framework, we analyse across a selection of LLMs (see Section 4.1). For each model, we implement two configurations: *PTEI-Base*, which applies personality-aware few-shot prompting using retrieved examples, and *PTEI-CoT*, which extends this with CoT reasoning. We compare these with corresponding non-personality-aware variants: *Base* (zero-shot) and *CoT* (zero-shot with CoT), both of which exclude personality conditioning. Our comparisons include the best-performing LLMs from the benchmark as personality-agnostic baselines, representing small-scale (<14B), mid-scale (14B), and large-scale (>14B) models, providing a robust reference for quantifying the added value of incorporating personality traits.

We evaluate across LLMs used in EmoBench, including **Qwen-7B** and **Qwen-14B**. While **GPT-4** was reported as the best-performing model in EmoBench, we conduct all experiments on **GPT-4o**, a more recent variant, to assess PTEI's effectiveness under current state-of-the-art conditions.

For **LLaMA 3.1 8B**, which was not originally covered in Emobench, we replicate the benchmark setup to generate our own *Base* and *CoT* results, ensuring consistency. We then apply our proposed personality-injected approach (PTEI-Base and PTEI-CoT) to this model for direct performance comparison.

5 Result and Analysis

In this section, we present a comprehensive evaluation of the proposed PTEI framework through a series of experiments, including main performance results (Section 5.1), ablation studies (Section 5.2), robustness analysis (Section 5.3), and qualitative case studies (Section 5.4).

5.1 Main Results

Table 1 presents evaluation results (accuracy) for the EU task across various LLMs and prompting configurations. **GPT-4o achieves the highest combined accuracy (63.62%)** under the personality-informed CoT configuration (PTEI-CoT), significantly outperforming all other models. In contrast, smaller models such as *Qwen-7B* and *LLaMA3.1-8B* generally struggle to surpass majority-class heuristics, particularly under standard zero-shot or CoT prompting.

Notably, **CoT prompting alone yields limited or negative effects** on smaller-scale models. For instance, accuracy for *LLaMA3.1-8B* decreases notably from the Base (16.62%) to CoT (12.25%). This suggests constrained structured reasoning abil-

Setup	GPT-4o	Qwen 7B	Qwen 14B	LLaMA 3.1 8B
PTEI-Base (Zero-shot)	59.44	20.63	33.11	16.82
PTEI-Base (Two-shot)	62.12	23.25	36.12	17.63
PTEI-CoT (Zero-shot)	59.20	16.28	31.87	13.79
PTEI-CoT (Two-shot)	63.62	20.88	34.38	14.13

Table 2: Robustness analysis of our method under different shot settings. The table reports the **overall average accuracy (emotion + cause)** for each LLM.

ities without sufficient grounding. However, **integrating personality context via PTEI-CoT considerably enhances CoT effectiveness**, particularly for larger models such as *GPT-4o*, which improves by +3.3 points over its base and +4.7 points over standard CoT.

Performance across the four EU categories: *Complex Emotions (CE)*, *Personal Beliefs and Experiences (PBE)*, *Perspective-Taking (PT)*, and *Emotional Cues (EC)*, demonstrates that **PT and PBE scenarios consistently pose the greatest challenges**. These tasks demand reasoning about mental states and personal beliefs. Conversely, scenarios categorized as CE and EC, which contain more explicit emotional indicators, yield comparatively higher accuracy.

Lastly, model scale notably impacts performance: **larger models like GPT-4o** not only achieve higher baseline accuracy but also **show greater relative improvements from personality-aware prompting**. This emphasizes the synergy between increased model capacity, structured reasoning (CoT), and psychologically grounded personality context in enhancing emotional reasoning capabilities.

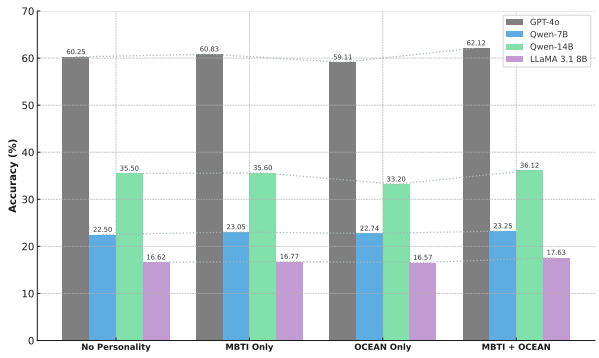

Figure 3: Average accuracy for each LLM under different personality input configurations. Injecting both MBTI and OCEAN traits leads to the highest performance across models.

5.2 Ablation Study

To assess the impact of personality conditioning in our framework, we perform an ablation study focusing on the Personality Detection Module (Section 3.4). We evaluate four LLMs: *GPT-4o*, *Qwen-7B*, *Qwen-14B*, and *LLaMA 3.1-8B*, under four personality input strategies: (1) no personality, (2) MBTI-only, (3) OCEAN-only, and (4) MBTI + OCEAN (full PTEI). Each model is evaluated on emotion and cause prediction, and we report the average accuracy across both sub-tasks. To isolate the effect of each personality trait type, we vary the weighting parameter α in the personality similarity function: $\alpha = 1.0$ (MBTI-only), $\alpha = 0.0$ (OCEAN-only), and $\alpha = 0.5$ (combined). This weighting influences both contrastive training (for constructing scenario pairs) and retrieval during inference. A higher α emphasizes MBTI alignment, while a lower value emphasizes OCEAN-based similarity.

As shown in Figure 3, incorporating personality information consistently improves performance across all models compared to the no-personality baseline. *GPT-4o* achieves the highest accuracy with the combined MBTI+OCEAN setting, outperforming the MBTI-only and no-personality setups by +1.29% and +1.87%, respectively. *Qwen-14B* similarly benefits from personality input, with MBTI+OCEAN exceeding the OCEAN-only and no-personality variants by +2.92% and +0.62%. While MBTI-only and OCEAN-only settings yield competitive results, the combined strategy consistently leads to the best performance across mid- and large-scale models.

5.3 Robustness Analysis

To evaluate the stability of model predictions, we prompt each LLM three times per question and apply majority voting over the responses. To mitigate sensitivity to answer ordering, we additionally permute the multiple-choice options three times, yielding four total permutations (original + 3). The final accuracy is computed as the average over these permutations, inspired by Emobench's robustness evaluation strategy.

We also test the consistency of our personality-aware methods (*PTEI-Base* and *PTEI-CoT*) under both **zero-shot** and **two-shot** prompting conditions across all LLMs used in our system. As shown in Table 2, the two-shot setup consistently leads to better performance, demonstrating the benefit

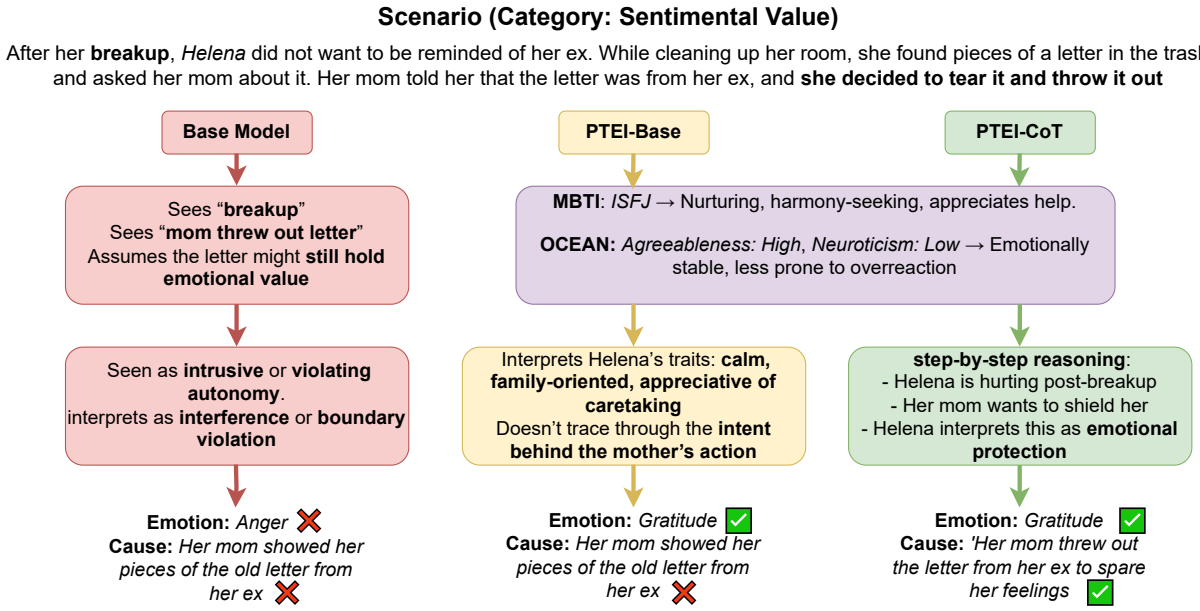

Figure 4: Case study comparison for GPT-4o, showing how personality-aware prompting improves emotional prediction accuracy.

of limited context augmentation across different model scales.

5.4 Case Study: Qualitative analysis

To better understand the reasoning capabilities of PTEI framework and the impact of personality conditioning in emotional inference, we analyze a representative scenario case, shown in Figure 4 involving a subject, 'Helena', who recently went through a breakup and discovers that her mother has torn up a letter from her ex. The emotional reasoning in this scenario hinges on factors such as sentimental value, perceived intrusion, and the emotional intent behind the mother's action.

Standard LLMs without personality awareness (e.g., Base models) tend to interpret this act as a boundary violation, resulting in an incorrect emotion label (*Anger*) and a cause that emphasizes intrusion. In contrast, our personality-aware models (PTEI-Base and PTEI-CoT) leverage Helena's inferred traits, ISFJ personality type, high Agreeableness, and low Neuroticism, to contextualize her reaction as emotionally stable and family-oriented. PTEI-CoT successfully traces the emotional motivation behind the mother's behavior: protecting her daughter from painful memories. This step-by-step reasoning yields the correct prediction of *Gratitude* and an accurate cause aligned with the emotional intent, not just a surface interpretation.

This example highlights how incorporating personality traits and structured reasoning enables the model to interpret emotionally complex and ambiguous situations more human-like, improving both emotional accuracy and cause identification.

6 Conclusion

This paper addressed the challenge that current LLMs face in handling complex EI tasks, particularly due to the neglect of critical psychological variables such as personality traits. To bridge this gap, we proposed PTEI, a personality-aware emotional inference framework that systematically integrates MBTI and OCEAN traits into EU scenarios. Our method uses structured prompting for personality detection, builds a synthetic memory bank enriched with personality annotations, and applies contrastive learning to optimize scenario retrieval and inference. Experiments show that integrating personality traits through few-shot and CoT prompting significantly enhances LLMs' emotion and cause reasoning, outperforming personality-agnostic baselines. Our contrastive learning approach shapes an embedding space sensitive to both emotional content and psychological profiles, enabling more targeted contextual retrieval. These results underscore the value of psychological modeling for more human-aligned emotional reasoning. Future work will extend PTEI to dynamic personality and multi-turn EI tasks.

Limitations

While our PTEI framework demonstrates significant improvements in EU through personality-aware reasoning, several limitations remain.

First, our experiments are limited to English-language, text-based scenarios. Emotional reactions and personality interpretations may vary across cultures and languages, which constrains the generalizability of our framework to multilingual or multimodal settings. Additionally, emotional inference is inherently subjective. While we adopt multiple-choice evaluation formats with predefined correct answers, some scenarios may reasonably allow for multiple plausible interpretations.

Second, due to the limited number of test cases (200 EU scenarios) in the EmoBench benchmark, the reported results may not fully capture the generalizability and impact of our method across the full spectrum of emotionally intelligent behavior. We plan to expand our evaluation to a broader range of benchmarks and real-world EI tasks in future work.

Third, despite careful prompt engineering for personality-aware reasoning and CoT prompting, model performance remains sensitive to prompt phrasing and structure. The prompt templates used may not be optimal for all LLM architectures or tasks.

Finally, our use of MBTI and OCEAN frameworks offers only a coarse approximation of personality. These static trait models do not capture dynamic, context-sensitive aspects of personality. Future work could explore adaptive or conversational personality modeling to more accurately reflect real-world psychological variability.

Ethical Statement

This work investigates the use of personality traits to improve large language models' (LLMs) reasoning about emotionally complex situations. We emphasize that our framework, PTEI, focuses on modeling *perceived* emotional intelligence through structured prompts and retrieval-based reasoning, rather than suggesting that LLMs possess genuine emotions or self-awareness. The goal is to study how personality knowledge can inform more human-aligned predictions in emotion and cause inference tasks.

While our method involves predicting MBTI and OCEAN personality traits based on textual descriptions, we do not use real user data, nor do we attempt to profile individuals in real-world settings. All personality information is synthetic and used strictly for academic experimentation in controlled benchmark scenarios.

We acknowledge that the use of personality modeling in NLP may raise ethical concerns, particularly around privacy, profiling, and fairness in downstream applications. Appropriate safeguards must be ensured in future deployments, including transparency, user consent, and bias auditing. Our current system is intended solely for research purposes, and we advocate for careful consideration of the psychological and social implications when applying similar methods in real-world contexts.

References

Mohammad Hossein Amirhosseini and Hassan Kazemian. 2020. Machine learning approach to personality type prediction based on the myers–briggs type indicator®. *Multimodal Technologies and Interaction*, 4(1):9.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, and et al. 2023. Qwen technical report. *arXiv preprint arXiv:2309.16609*.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*, pages 1597–1607. PMLR.

Zohar Elyoseph, Dorit Hadar-Shoval, Kfir Asraf, and Maya Lvovsky. 2023. Chatgpt outperforms humans in emotional awareness evaluations. *Frontiers in Psychology*, 14:1199058.

Yumeng Fu, Junjie Wu, Zhongjie Wang, Meishan Zhang, Lili Shan, Yulin Wu, and Bingquan Liu. 2025. LaERC-S: Improving LLM-based emotion recognition in conversation with speaker characteristics. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 6748–6761, Abu Dhabi, UAE. Association for Computational Linguistics.

Daniel Goleman. 1996. Emotional intelligence. why it can matter more than iq. *Learning*, 24(6):49–50.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, and et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

James D Hess and Arnold C Bacigalupo. 2011. Enhancing decisions and decision-making processes through the application of emotional intelligence skills. *Management decision*, 49(5):710–721.

Carroll E. Izard, Deborah Z. Libero, Priscilla Putnam, and O. Maurice Haynes. 1993. Stability of emotion experiences and their relations to traits of personality. *Journal of Personality and Social Psychology*, 64(5):847–860.

Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2021. Billion-scale similarity search with gpus.

Yang Li, Amirmohammad Kazemeini, Yash Mehta, and Erik Cambria. 2022. Multitask learning for emotion and personality traits detection. *Neurocomputing*, 493:340–350.

Zheng Li, Sujian Li, Dawei Zhu, Qilong Ma, and Weimin Xiong. 2025. EERPD: Leveraging emotion and emotion regulation for improving personality detection. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 7721–7734, Abu Dhabi, UAE. Association for Computational Linguistics.

Abdullah Al Maruf, Fahima Khanam, Md. Mahmudul Haque, Zakaria Masud Jiyad, M. F. Mridha, and Zeyar Aung. 2024. Challenges and opportunities of text-based emotion detection: A survey. *IEEE Access*, 12:18416–18450.

Robert R. McCrae and Paul T. Jr. Costa. 1997. Personality trait structure as a human universal. *American Psychologist*, 52(5):509–516.

Robert R McCrae and Oliver P John. 1992. An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215.

Yash Mehta, Navonil Majumder, Alexander Gelbukh, and Erik Cambria. 2020. Recent trends in deep learning based personality detection. *Artificial Intelligence Review*, 53(4):2313–2339.

Isabel Briggs Myers. 1987. *Introduction to type: A description of the theory and applications of the Myers-Briggs Type Indicator*. Consulting Psychologists Press.

OpenAI. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Samuel J Paech. 2023. Eq-bench: An emotional intelligence benchmark for large language models. *arXiv preprint arXiv:2312.06281*.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Michael D Robinson and Gerald L Clore. 2002. Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, 128(6):934–960.

Sahand Sabour, Siyang Liu, Zheyuan Zhang, June Liu, Jinfeng Zhou, Alvionna Sunaryo, Tatia Lee, Rada Mihalcea, and Minlie Huang. 2024. EmoBench: Evaluating the emotional intelligence of large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5986–6004, Bangkok, Thailand. Association for Computational Linguistics.

Mustafa Safdari, Greg Serapio-García, Clément Crepy, Stephen Fitz, Peter Romero, Luning Sun, Marwa Abdulhai, Aleksandra Faust, and Maja J. Mataric. 2023. Personality traits in large language models. *CoRR*, abs/2307.00184.

Peter Salovey and John D. Mayer. 1990. Emotional intelligence. *Imagination, Cognition and Personality*, 9(3):185–211.

Maarten Sap, Ronan Le Bras, Daniel Fried, and Yejin Choi. 2022. Neural theory-of-mind? on the limits of social intelligence in large LMs. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 3762–3780, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Dagmar Schuller and Björn W. Schuller. 2018. The age of artificial emotional intelligence. *Computer*, 51(9):38–46.

Aleksandra Sorokovikova, Natalia Fedorova, Sharwin Rezagholi, and Ivan P Yamshchikov. 2024. Llms simulate big five personality traits: Further evidence. *arXiv preprint arXiv:2402.01765*.

Sanja Stajner and Seren Yenikent. 2020. A survey of automatic personality detection from texts. In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8 13,2020*, pages 6284–6295. International Committee on Computational Linguistics.

Chenwei Wan, Matthieu Labeau, and Chloé Clavel. 2025. EmoDynamiX: Emotional support dialogue strategy prediction by modelling MiXed emotions and discourse dynamics. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 1678–1695, Albuquerque, New Mexico. Association for Computational Linguistics.

Xuena Wang, Xueting Li, Zi Yin, Yue Wu, and Jia Liu. 2023. Emotional intelligence of large language models. *Journal of Pacific Rim Psychology*, 17.

Yan Wang, Bo Wang, Yachao Zhao, Dongming Zhao, Xiaojia Jin, Jijun Zhang, Ruifang He, and Yuexian Hou. 2024. Emotion recognition in conversation via dynamic personality. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 5711–5722, Torino, Italia. ELRA and ICCL.

Jialiang Wu, Yi Shen, Ziheng Zhang, and Longjun Cai. 2024. Enhancing large language model with decomposed reasoning for emotion cause pair extraction. *arXiv preprint arXiv:2401.17716*.

Tao Yang, Tianyuan Shi, Fanqi Wan, Xiaojun Quan, Qifan Wang, Bingzhe Wu, and Jiaxiang Wu. 2023. PsyCoT: Psychological questionnaire as powerful chain-of-thought for personality detection. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 3305–3320, Singapore. Association for Computational Linguistics.

Haein Yeo, Taehyeong Noh, Seungwan Jin, and Kyungsik Han. 2025. PADO: Personality-induced multi-agents for detecting OCEAN in human-generated texts. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 5719–5736, Abu Dhabi, UAE. Association for Computational Linguistics.

A Human Evaluation

To evaluate the quality of the generated scenarios used in our memory bank for contrastive learning and few-shot retrieval, we conducted a human assessment on a randomly selected 10% subset spanning various scenario categories. Human annotators assessed each scenario based on three criteria: (1) the overall coherence and realism of the narrative, (2) the alignment between the scenario and its assigned category (e.g., *mixture of emotions*, *false belief*), and (3) the appropriateness and clarity of the annotated emotion and cause labels in context.

The evaluation results indicate that the generated scenarios are generally well-formed and exhibit high semantic alignment with their designated categories. Annotators showed strong agreement with the original emotion and cause labels, supporting the use of the memory bank as a high-quality resource for retrieval-based prompting.

For the EmoBench test set used in the final evaluation, we rely on the gold-standard labels provided by the original benchmark authors. These labels for both emotion and cause were validated by human experts, as documented in the benchmark release.

B Memory Bank Analysis

B.1 Emotion Label Distribution

We analyzed the normalized distribution of emotions of the memory bank. Labels were grouped into high-level categories and split if mixed.

B.2 Emotion-Personality Correlation

We computed Pearson correlations between normalized emotion categories and personality traits using both MBTI and OCEAN frameworks. Emotion labels were one-hot encoded, and personality traits were encoded either as binary (MBTI) or ordinal (OCEAN).

MBTI Correlation: Each of the 16 standard MBTI types (e.g., INFP, ESTJ) was represented as a binary feature. We then calculated the Pearson correlation between these types and each individual emotion category. This analysis highlights patterns such as higher emotional resonance of Joy with ENFP and stronger ties between Apprehension and ISFP types.

OCEAN Correlation: OCEAN traits were originally qualitative and mapped numerically: Low = 1, Medium = 2, High = 3. We computed correlations between each trait and each emotion category. For instance, Neuroticism shows a positive correlation with emotions like Fear and Nervousness, while Agreeableness aligns more strongly with Trust and Caring emotions.

C Prompt Templates

This section outlines the prompt configurations used in our PTEI framework. Prompts are grouped by their corresponding modules.

C.1 Prompts for Personality Detection Module

These prompts are used to infer personality traits for each subject within the scenario.

- MBTI Personality Prompt: Table 3
- OCEAN Personality Prompt: Table 4

C.2 Prompts for EU Task

These prompts are designed to guide LLMs in predicting emotions and their causes for each scenario, with or without step-by-step reasoning. They incorporate contextual retrieval and personality traits.

- PTEI-Base Prompt: Table 5
- PTEI-CoT Prompt: Table 6

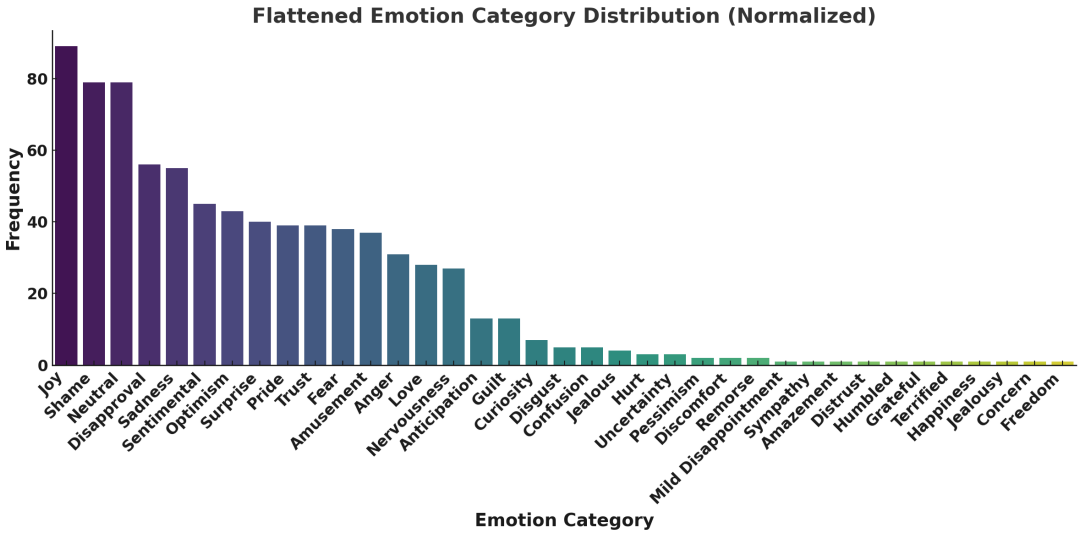

Figure 5: Flattened Emotion Category Distribution (Normalized)

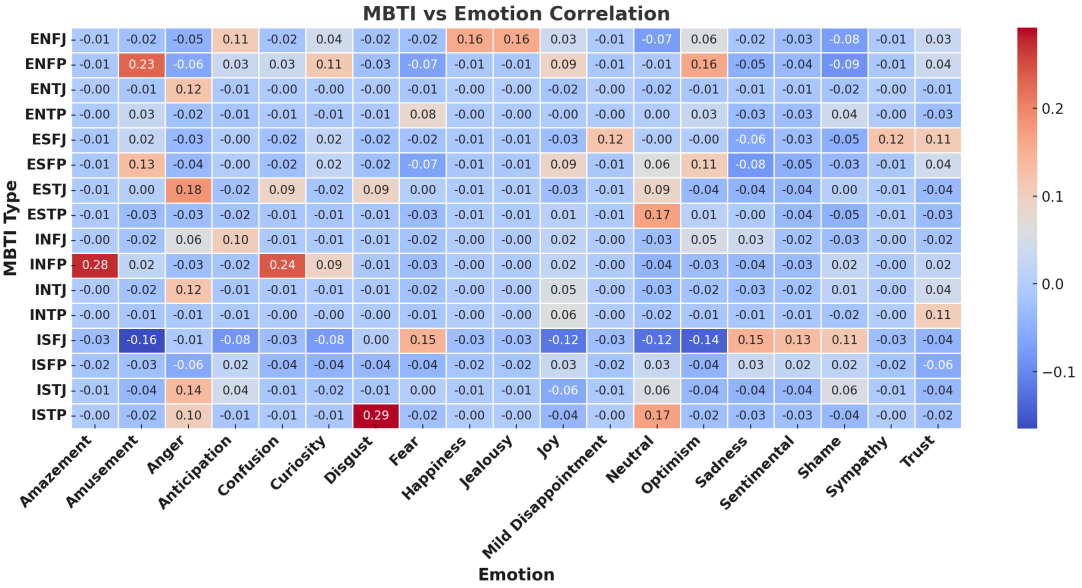

Figure 6: MBTI vs Emotion Correlation

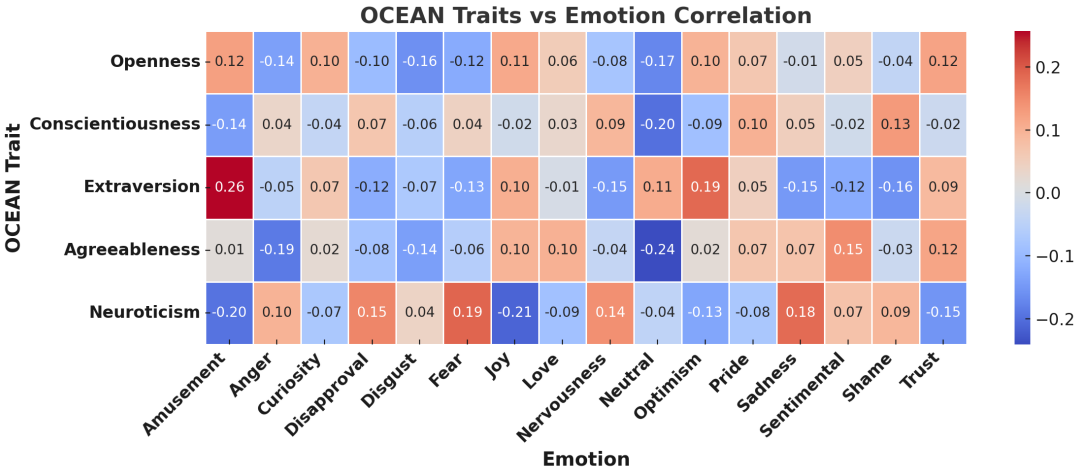

Figure 7: OCEAN Traits vs Emotion Correlation

MBTI Personality Prompt Template (with 2-Shot Demonstrations)

You are a personality assessment expert trained to analyze behavioral patterns and assign MBTI personality types.

MBTI Dimensions:
– I vs. E: alone/quiet vs. social/active
– S vs. N: practical/details vs. abstract/ideas
– T vs. F: logic/objectivity vs. emotions/values
– J vs. P: structured/planned vs. flexible/spontaneous

Scenario Context:
Category: `{category_name}`
Explanation: `{category_explanation}`

Task: Analyze the scenario, infer the subject's MBTI type, and explain the reasoning.

Input:
Scenario: `{scenario}`
Subject: `{subject}`

Expected Output (in JSON format):

```
{
   "MBTI": "XXXX",
   "Explanation": "Reasoning based on traits observed in the scenario."
}
```

Few-shot Examples:
Two illustrative demonstrations (scenario + subject + MBTI + explanation) are included before the test case to guide prediction.

Table 3: MBTI prompt template with structured instructions and two-shot demonstration format.

OCEAN Personality Prompt Template (with 2-Shot Demonstrations)

You are a personality assessment expert trained to analyze behavioral patterns and assign OCEAN personality traits.

OCEAN Dimensions:
– Openness: curiosity, creativity, interest in new experiences
– Conscientiousness: organization, responsibility, reliability
– Extraversion: sociability, talkativeness, assertiveness
– Agreeableness: kindness, trust, cooperation
– Neuroticism: emotional instability, anxiety, moodiness

Scenario Context:
Category: `{category_name}`
Explanation: `{category_explanation}`

This context helps interpret how the subject expresses themselves and reacts emotionally.

Task: Read the scenario and predict the subject's OCEAN traits. You should generate "Low", "Medium" or "High" for each trait. Then explain the reasoning based on the observed behaviors.
Input:
Scenario: `{scenario}`
Subject: `{subject}`

Expected Output Format (JSON):

```
{
   "Openness": "High",
   "Conscientiousness": "Medium",
   "Extraversion": "Low",
   "Agreeableness": "High",
   "Neuroticism": "Medium",
   "Explanation": "Brief justification of the above traits based on behavioral cues in the scenario."
}
```

Few-shot Examples:
Two example scenarios with annotated trait levels and explanations are included before the test case to guide prediction.

Table 4: OCEAN prompt template used for personality trait inference, including structured trait outputs, reasoning explanation, and a two-shot demonstration format.

PTEI-Base Prompt Template (with Memory Retrieval and Personality Context)

Instruction:

```
In this task, you are presented with a scenario, a question, and multiple choices. Please carefully
analyze the scenario and take the perspective of the individual involved.
Provide only one single correct answer to the question and respond only with the corresponding
letter. Do not provide explanations for your response.
```

Few-shot Memory Retrieval:

If retrieval is enabled, the prompt includes a few top-k scenarios retrieved from the memory bank. Each includes:
– Scenario description
– Annotated emotion and cause labels
– MBTI type
– OCEAN trait levels (High, Medium, Low)

Main Prompt Body (Emotion Task Example):

```
You are a personality and emotion analyst.
First, carefully read and understand the following similar situations. Pay attention to how the
individuals reacted emotionally, their causes, and their personalities.
{{Retrieved Scenarios}}
—
After analyzing these similar cases, consider the new situation below carefully.
Scenario: {{scenario}}
Personality Information: {{personality string}}
Question: What emotion(s) would {{subject}} ultimately feel in this situation?
Choices: A. ..., B. ..., C. ..., ...
```

Personality Context:

If available, the personality string is appended in the form:

```
...considering the MBTI personality is ESTJ and the levels of OCEAN personalities are Openness:
Medium, Conscientiousness: High, ...
```

Main Prompt Body (Cause Task Example):

```
Question: Why would {{subject}} feel {{emotion}} in this situation?
```

Table 5: PTEI-Base prompt template for emotion and cause prediction, incorporating memory-based retrieval and structured personality conditioning (MBTI + OCEAN).

PTEI-CoT Prompt Template (with Memory Retrieval, Personality, and Chain-of-Thought Reasoning)

Instruction:

```
In this task, you are presented with a scenario, a question, and multiple choices. Please carefully
pay close attention to the emotions and intentions. Analyze the scenario and take the perspective
of the individual involved. Reason step by step by exploring each option's potential impact on
the individual(s) in question
Think step-by-step to identify the correct answer. Provide both the selected answer (as a single
letter) and a brief explanation justifying your choice.
```

Few-shot Memory Retrieval:

If retrieval is enabled, the prompt includes top-k examples from the memory bank. Each includes:
– Scenario description
– Annotated emotion and cause labels
– MBTI type
– OCEAN trait levels (High, Medium, Low)

Main Prompt Body (Emotion Task Example):

```
You are a personality and emotion analyst.
First, carefully read and understand the following similar situations. Pay attention to how the
individuals reacted emotionally, their causes, and their personalities.
{{Retrieved Scenarios}}
—
After analyzing these similar cases, consider the new situation below carefully.
Scenario: {{scenario}}
Personality Information: {{personality string}}
Question: What emotion(s) would {{subject}} ultimately feel in this situation?
Choices: A. ..., B. ..., C. ..., ...
```

Personality Context:

If provided, personality cues are appended:

```
...considering the MBTI personality is INFP and the levels of OCEAN personalities are Openness:
High, Conscientiousness: Medium, ...
```

Main Prompt Body (Cause Task Example):

```
Question: Why would {{subject}} feel {{emotion}} in this situation?
```

Expected Output Format:

```
Answer: B
Explanation: The subject is likely to feel this way because ... (based on emotional cues,
personality traits, and scenario context).
```

Table 6: PTEI-CoT prompt template for emotion and cause prediction, combining retrieval, personality traits, and step-by-step reasoning. The model is expected to output both a selected answer and an explanation.