Big Five Personality Prediction through Emotion-Conditioned Representations and Learnable Psycholinguistic Mapping

Lorenzo Zangari, Antonin Schnyder, Davide Picca
SLI, University of Lausanne
Lausanne, Switzerland
{lorenzo.zangari,davide.picca}@unil.ch, antonin.schnyder.unil@mailfence.com

Abstract

Personality traits influence human behavior and social interactions, making their accurate prediction essential across multiple domains. The Big Five Model, a widely recognized framework in psychological science for assessing personality traits, has become the foundation for different computational approaches to personality prediction. In recent years, a growing body of research has highlighted the dynamic interplay between emotions and personality, as individuals navigate diverse emotional experiences that evoke distinct responses and ultimately shape their behavioral patterns. In this work, we present a novel framework that systematically integrates affective information into Pre-trained Language Models for Big Five Personality trait prediction. Our framework leverages text-based embeddings, emotion-conditioned features, and learnable psycholinguistic information that bridges affective dimensions with personality traits. This design preserves established psycholinguistic knowledge while enabling adaptive refinement through data-driven learning. Our experiments showed that our framework outperformed sentence embedding-based methods and Large Language Models across various datasets from different domains, achieving an average F1-score improvement of at least 15% in out-of-domain scenarios.

Keywords: PLMs, Personality, Emotions

1. Introduction

Personality traits are patterns of thinking, feeling, and behaving that are shaped by a person's genetic makeup and experiences in various emotional settings (Revelle and Scherer, 2009). These traits contribute to individual differences and influence a wide array of life outcomes, including decision-making (Russell et al., 2017; Yılmaz et al., 2019), mental health (Kang et al., 2023), occupational performance (He et al., 2019), knowledge processing (Taramigkou et al., 2018), and individual attitudes (Aslan and Güdükbay, 2019). Given their pervasive influence, accurate personality prediction is critical across multiple domains (Oh and Tong, 2020).

The Big Five Model (Digman, 1990), also known as the *OCEAN* model, is among the most widely used frameworks for personality assessment in psychological science (John et al., 2010), recognized for its strong empirical foundation and cross-cultural validity (Ching et al., 2014; John et al., 1999). It encompasses five broad dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. In recent years, it has proven useful for operationalizing traits as predictive features across different applications, including personnel selection (He et al., 2019), health intervention design (Molloy et al., 2014), and political attitudes (Gerber et al., 2010). For instance, Openness consistently predicts liberal ideologies, while Conscientiousness aligns with conservatism across different cultural contexts. In the health domain, low Neuroticism is strongly associated with enhanced emotional regulation and better long-term health outcomes (Kang et al., 2023), while high Conscientiousness predicts academic and occupational success (He et al., 2019).

An emerging research direction leverages *Pre-trained Language Models (PLMs)*, such as SentenceTransformers (Reimers and Gurevych, 2019), to predict Big Five personality traits from text data (Habib et al., 2024). This is motivated by the intrinsic relationship between language and personality expression: linguistic patterns—including lexical choices, syntactic structures, and semantic content—provide observable indicators of underlying emotional states, cognitive processes, and social behaviors that characterize personality traits (Revelle and Scherer, 2009). Through exposure to extensive training corpora, PLMs acquire lexical, semantic, and factual knowledge, while also internalizing cultural patterns inherent in the data, positioning them as promising tools for personality prediction. In fact, individual differences in linguistic style are reliable indicators of personality (Pennebaker and King, 1999).

Furthermore, personality is also significantly influenced by emotions (Akber et al., 2024), which can be defined as complex states that produce physical and psychological reactions, shaping both cognition and behavior (Cambria et al., 2012). In fact, personality profiles emerge from the interaction between genetic predispositions and emotional experiences across different contexts, with individual differences in emotional reactivity leading to distinct personality configurations.

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 8162–8173
11–16 May 2026.

Since linguistic expressions often carry emotional content that reflects personality, computational models can leverage affective information extracted from text. Specifically, this information can serve as auxiliary features alongside lexical and syntactic signals, enabling computational models to predict personality traits. Moreover, prior studies indicate that integrating emotion signals with linguistic markers improves prediction of Big Five traits, though gains are task- and domain-dependent (Akber et al., 2024; Yash et al., 2020). To computationally extract affective information, SenticNet provides a widely adopted framework for concept-level emotion modeling, mapping textual concepts to affective dimensions, such as Introspection, Temper, Aptitude, and Sensitivity (Cambria et al., 2012).

In this work, we propose an emotion-aware PLM-based framework for Big Five Personality trait prediction. Our approach leverages the systematic relationship between language use and personality: word choice, emotional expression, and linguistic style serve as reliable indicators of individual differences, known as *psycholinguistic features* (Yash et al., 2020). To this end, we integrate emotional information through two complementary mechanisms. First, we introduce a learnable psycholinguistic prior that encodes the structural relationship between SenticNet emotions and OCEAN personality traits. This prior combines a SenticNet-derived mapping with data-driven information, allowing the model to preserve established psycholinguistic knowledge while adapting to domain-specific patterns. Second, we employ a Feature-wise Linear Modulation (FiLM) (Perez et al., 2018; de Vries et al., 2017) component to directly condition the embeddings generated by a PLM on affective dimensions, enabling the model to modulate its text representations according to the emotional context. This dual strategy ensures that emotional information influences predictions both through explicit, theoretically grounded psycholinguistic information and implicit feature transformation. We summarize our contributions as follows:

1. We propose a novel framework for Big Five Personality trait prediction that enables parameter-efficient fine-tuning of PLMs by systematically integrating emotional information.

2. We designed two complementary components to integrate emotional information for personality trait prediction: (i) learnable psycholinguistic features derived from SenticNet, and (ii) a feature-wise modulation mechanism that conditions PLM-based embeddings on emotional context.

3. Our experimental evaluation on popular datasets annotated with Big Five labels showed the effectiveness of our framework compared with different methods for personality prediction, including Sentence Embedding-based approaches and Large Language Models, with an average percentage increase in F1-score of at least 15%.

We make the source code publicly available.[1]

2. Preliminaries

Big Five Model. Our work is grounded in the Big Five Personality Model (Digman, 1990), which posits that the personality of an individual is characterized through five traits: *Openness (O)*, reflecting intellectual curiosity, imagination, and a desire for new experiences; *Conscientiousness (C)*, capturing diligence and self-control; *Extraversion (E)*, characterized by sociability, assertiveness, talkativeness, and positive affect; *Agreeableness (A)*, denoting trust, cooperation, and compassion; and *Neuroticism (N)*, reflecting emotional instability and susceptibility to negative affect. These traits are often described with the acronym OCEAN and emerged from the statistical analysis of thousands of adjectives used to describe people. The model's strength lies in its ability to capture personality differences concisely and across different cultures (Ching et al., 2014; John et al., 2010).

SenticNet affective model. SenticNet (Cambria et al., 2012) is a concept-level affective knowledge base built upon the *Hourglass of Emotions*, a biologically inspired model aligned with canonical frameworks such as Plutchik's wheel of emotions (Plutchik, 2001). In this work, we employ the revisited Hourglass of Emotions model (Susanto et al., 2020), which represents affective states through four independent yet concomitant dimensions: *Introspection (Intr.)*, capturing the contrast between positive self-reflection and melancholic states; *Temper (Temp.)*, measuring the degree of emotional composure versus reactive irritability; *Aptitude (Apt.)*, representing the evaluation of stimuli along a spectrum from acceptance to aversion; and *Sensitivity (Sens.)*, quantifying the anticipatory orientation ranging from proactive engagement to apprehensive withdrawal. Concepts in SenticNet are mapped to vectors in a four-dimensional space, formalizing affect semantics at the concept level (Cambria et al., 2020). We employed SenticNet for its structured affective representation that is well aligned with the Big Five model (Yash et al., 2020).

Problem definition: Big Five Personality prediction. We define the task of Big Five personality prediction *(hereafter referred to as OCEAN prediction)* as inferring personality traits from text in

[1] https://github.com/lorenzozangari

accordance with the Big Five model of personality. We are given a dataset $\mathcal{D} = \{(d_i, \mathbf{y}_i)\}$, where d_i represents a text and $\mathbf{y}_i \in \{0,1\}^{|M|}$ is a binary ground-truth vector associated with the personality traits, with $M = \{O, C, E, A, N\}$ being the set of OCEAN traits. The goal is to learn a function $f : \boldsymbol{\mathcal{T}} \rightarrow \{0,1\}^{|M|}$ that predicts the personality traits for any $d_i \in \boldsymbol{\mathcal{T}}$, where $\boldsymbol{\mathcal{T}}$ is the space of all possible input texts. We denote $\mathbf{o}_i \in \{0,1\}^{|M|}$ as the predicted personality trait vector produced by our model for d_i. In this work, we learn f by leveraging emotional information. To this end, we denote by $\mathbf{e}_i \in \mathbb{R}^{|E|}$ the vector representing the SenticNet affective features associated with d_i, where E denotes the set of SenticNet affective dimensions, with $|E| = 4$. Note that we frame the problem of OCEAN prediction as a multi-label classification task. This aligns with most existing personality datasets, which are binary-labeled (Pennebaker and King, 1999). Nevertheless, our framework is not inherently limited to discrete labels. Following (Shum et al., 2025), trait-specific classification heads can be seamlessly replaced with regression heads in continuous-score settings.

Psycholinguistic features for OCEAN prediction. Psycholinguistic features represent theory-grounded textual descriptors derived from linguistic and cognitive theories that capture systematic patterns in language use associated with psychological traits. The work by Yash et al. (2020) demonstrated the effectiveness of leveraging such features for OCEAN prediction. They computed a prior matrix encoding the relationship between SenticNet affective dimensions and personality traits. Specifically, they extracted SenticNet features at the document level and computed a correlation matrix $\mathbf{W} \in \mathbb{R}^{|E|\times|M|}$ using point-biserial correlations between each continuous SenticNet dimension and the binary personality labels across the training corpus. Their empirical analysis on specific datasets revealed statistically significant correlations between certain affective dimensions and personality traits. For instance, they found that Aptitude positively correlated with Conscientiousness and Agreeableness.

In this work, we employ this correlation-based matrix as a learnable prior that provides a psycholinguistic bridge between affective features and OCEAN traits. This design allows the model to retain established psycholinguistic knowledge while adapting to domain-specific data patterns. By balancing theory-driven structure with data-driven learning, the model can refine emotion-trait associations beyond fixed linear correlations. Additionally, we employed a more recent and comprehensive version of SenticNet (Cambria et al., 2020), which provides improved concept coverage and refined affective annotations, compared with SenticNet 5 (Cambria et al., 2018) used by Yash et al. (2020). Moreover, our framework is agnostic to the SenticNet releases available at the time of writing, and additional versions can be seamlessly integrated.

Feature-wise Linear Modulation. Feature-wise Linear Modulation (FiLM) (Perez et al., 2018) modulates the internal representations of a neural network by applying an affine transformation conditioned on an external input $\mathbf{c}$. Let $\mathbf{h}_i \in \mathbb{R}^D$ be the hidden representation of a neural model. FiLM consists of a modulator that receives a conditioning signal $\mathbf{c} \in \mathbb{R}^K$ and produces two modulation vectors $\gamma(\mathbf{c}), \beta(\mathbf{c}) \in \mathbb{R}^D$. The FiLM transformation is then defined as:

$$\mathrm{FiLM}(\mathbf{h}_i \mid \mathbf{c}) = \mathbf{z}_i = \gamma(\mathbf{c}) \odot \mathbf{h}_i + \beta(\mathbf{c}), \quad (1)$$

where $\odot$ denotes element-wise multiplication. This operation performs a feature-wise affine transformation on each element of the representation, leaving the shape of the input unchanged. FiLM has proven effective for conditioning neural networks on auxiliary information across various domains, including visual question answering (Perez et al., 2018), visual reasoning (de Vries et al., 2017), and multimodal learning (Houlsby et al., 2019). In our work, $\mathbf{c}$ is represented by $\mathbf{e}_i$.

3. Related Work

Over the past few years, a growing body of work has investigated text-based automatic personality prediction (Naz et al., 2025; Sorokovikova et al., 2024). Early studies correlated lexical, syntactic, stylistic, and topical features with trait scores. Pennebaker et al. (2001) identified a correlation between linguistic features—including word frequency and grammatical categories—and personality traits. Park et al. (2015) showed that language-based assessment represents an effective measure for evaluating personality traits. Majumder et al. (2017) and Yash et al. (2020) leveraged SenticNet to demonstrate that incorporating common-sense knowledge with psycholinguistic features enhances the accuracy of existing personality detection frameworks.

More recently, the deep learning revolution has opened two principal research directions: (i) supervised fine-tuning of encoder-only PLMs, and (ii) application of generative *Large Language Models (LLMs)* in zero-shot or few-shot settings.

Considering the first category, sentence embedding-based methods have been widely adopted (Ren et al., 2021; Jain et al., 2022). Habib et al. (2024) showed that compact models like ALBERT (Lan et al., 2019) can match larger architectures such as RoBERTa (Liu et al., 2019) with sufficient data. Shum et al. (2025) fine-tuned

BERT (Devlin et al., 2019a) and RoBERTa variants on the Reddit corpus PANDORA (Gjurković et al., 2021). It has also been shown that augmenting PLMs with affective information is beneficial for personality prediction. Ren et al. (2021) concatenated SenticNet-derived emotional features with BERT embeddings, while Yash et al. (2020) combined transformers with handcrafted linguistic features. Li et al. (2022) proposed jointly predicting Big Five traits and emotional states, showing that multi-output models improve performance by exploiting inter-trait correlations.

Regarding the second category, Ganesan et al. (2023) found that GPT-3's zero-shot accuracy declined for fine-grained trait scoring. Piastra and Catellani (2025); Peters and Matz (2024) revealed that ChatGPT-4 struggled with uncertainty awareness when textual evidence was insufficient. Zhu et al. (2025) observed weak alignment between GPT-4's predictions and validated trait scores, with minimal improvement from chain-of-thought prompting. Li et al. (2025) further showed that integrating emotion regulation knowledge enhances trait inference.

Comparison with our framework. The above findings indicate that Transformer-based models (Vaswani et al., 2017) specifically fine-tuned for personality prediction represent the dominant approach, while LLMs in zero- or few-shot settings show limited effectiveness. Our work falls within the first category as an encoder-only method relying on SentenceTransformers, fine-tuned for OCEAN prediction. Unlike existing methods that incorporate emotions either through simple feature concatenation (Ren et al., 2021) or static correlation-based mappings (Yash et al., 2020), our method holistically exploits textual and affective information through three complementary parallel components: the first leverages text embeddings alone, the second dynamically conditions text representations on emotional context via Feature-wise Linear Modulation, and the third directly maps emotions to OCEAN traits through learnable psycholinguistic information. These parallel components enable the model to leverage both implicit affect-modulated representations (via FiLM) and explicit emotion-trait mappings (via the learnable psycholinguistic mapping), balancing interpretable psycholinguistic structure with adaptive context-aware learning for personality trait prediction.

4. Methodology

The overall architecture of our framework is shown in Fig. 1. The proposed methodology consists of two stages. In the *Pre-processing Stage*, we extract SenticNet features and compute a prior matrix that provides psycholinguistic information. In the *Model Forward Stage*, we perform OCEAN prediction by injecting emotional information into the learning process through two components. In the following, we detail each stage and its components.

4.1. Pre-processing Stage

In this stage, we first extract SenticNet emotional features for each text in every dataset using the SenticNet API.[2] Following (Yash et al., 2020), we then compute a correlation matrix $\mathbf{W} \in \mathbb{R}^{|E|\times|M|}$ between the extracted SenticNet dimensions and the personality trait labels from the annotated training data via point-biserial correlation (cf. Sec. 2). This matrix encodes the initial psycholinguistic information exploited by our framework.

4.2. Model Forward Stage

In this stage, we first encode the input text through an Encoder-only PLM, then we apply three parallel components. Specifically, given a PLM Emb$(\cdot)$ and a text instance d_i, the PLM encodes it by generating an embedding $\mathbf{h}_i = \text{Emb}(d_i)$, where $\mathbf{h}_i \in \mathbb{R}^D$ and D denotes the embedding dimension of Emb$(\cdot)$. Then, three components operate in parallel: (i) an embedding-based head that predicts OCEAN traits directly from $\mathbf{h}_i$; (ii) an emotion-conditioned head that leverages SenticNet features $\mathbf{e}_i$ to modulate the embedding $\mathbf{h}_i$ through a lightweight Feature-wise Linear Modulation generator (Perez et al., 2018); and (iii) a prior-informed head that projects emotions $\mathbf{e}_i$ into OCEAN predictions through a learnable psycholinguistic prior derived from the correlations between SenticNet dimensions and personality traits initialized from the training data (Yash et al., 2020). The final personality prediction $\mathbf{o}_i$ is then obtained through a convex combination of the three heads.

Embedding-based head. This module predicts the OCEAN traits from the text embedding encoded by the PLM, i.e., $\mathbf{h}_i = \text{Emb}(d_i)$. Given the textual representation $\mathbf{h}_i \in \mathbb{R}^D$, the text-based OCEAN prediction is obtained as follows:

$$\mathbf{o}_i^{(t)} = g^{(t)}(\mathbf{h}_i), \tag{2}$$

where $g^{(t)}$ is a classification head (a two-layer MLP), and $\mathbf{o}_i^{(t)}$ denotes the text-based prediction.

Emotion-conditioned head. This component is designed to dynamically inject emotional context into text representations through FiLM (cf. Sec. 2). Unlike methods relying on simple feature concatenation, FiLM enables adaptive modulation by selectively amplifying or attenuating dimensions relevant

[2] https://sentic.net/api/

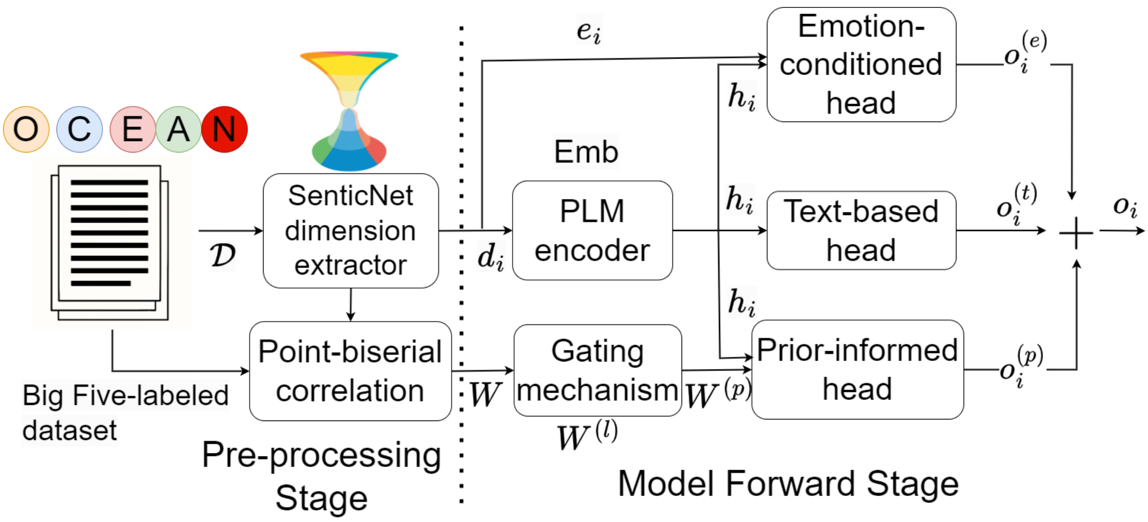

Figure 1: Framework overview. Our approach predicts OCEAN traits through three components: one text-based head and two heads that leverage SenticNet, including a prior-informed component for psycholinguistic structure encoding and a FiLM component for dynamic embedding modulation.

to the conditioning signal. In this case, the modulation enables the model to adjust text encoding based on emotional context, as expressions reflected in language vary significantly with affective state (Akber et al., 2024).

Formally, given the SenticNet vector $\mathbf{e}_i$ associated with text d_i, our framework learns feature-wise affine parameters through a neural network $g^{(e_1)}$ that generates scale $\boldsymbol{\gamma}_i$ and shift $\boldsymbol{\beta}_i$ parameters:

$$(\boldsymbol{\gamma}_i, \boldsymbol{\beta}_i) = g^{(e_1)}(\mathbf{e}_i), \tag{3}$$

where $\boldsymbol{\gamma}_i \in \mathbb{R}^D$ and $\boldsymbol{\beta}_i \in \mathbb{R}^D$ enable per-feature adaptation. The vector $\boldsymbol{\gamma}_i$ amplifies or attenuates affect-relevant dimensions, while $\boldsymbol{\beta}_i$ injects emotion-specific bias. The base embedding $\mathbf{h}_i$ is then modulated as:

$$\mathbf{z}_i^{(e)} = \mathbf{h}_i \odot \boldsymbol{\gamma}_i + \boldsymbol{\beta}_i, \tag{4}$$

where $\odot$ denotes the element-wise multiplication, and $\mathbf{z}_i^{(e)}$ is the emotion-conditioned feature vector that preserves base text content while adapting to emotional information. Finally, $\mathbf{z}_i^{(e)}$ is fed to a two-layer MLP ($g^{(e_2)}$) for OCEAN prediction:

$$\mathbf{o}_i^{(e)} = g^{(e_2)}\left(\mathbf{z}_i^{(e)}\right), \tag{5}$$

where $\mathbf{o}_i^{(e)}$ is the OCEAN prediction under the emotion-conditioned component.

Prior-informed head. This module projects emotions directly onto OCEAN traits through learnable psycholinguistic information. Specifically, we leverage the prior matrix $\mathbf{W}$ computed in the Pre-processing Stage, which encodes theoretically-grounded associations between emotions and personality traits. To enable adaptive refinement of the prior, we introduce a learnable matrix $\mathbf{W}^{(l)} \in \mathbb{R}^{|E|\times|M|} \sim \mathcal{N}(0, \mathbf{I})$ initialized from a Gaussian distribution, which captures data-driven patterns beyond fixed correlations. Then, a gating mechanism with trainable per-trait weights $\mathbf{z} \in \mathbb{R}^{|M|}$ dynamically balances the contribution of the psycholinguistic prior with the learned data-driven patterns. The final emotion-to-OCEAN mapping matrix, $\mathbf{W}^{(p)}$, is thus defined as:

$$\mathbf{W}^{(p)} = \mathbf{W}\,\text{Diag}(\boldsymbol{\sigma}(\mathbf{z})) + \mathbf{W}^{(l)}\,\text{Diag}(\mathbf{1} - \boldsymbol{\sigma}(\mathbf{z})), \tag{6}$$

where $\boldsymbol{\sigma}(\mathbf{z})$ controls the per-trait balance between psycholinguistic knowledge and learned representations; σ is the sigmoid function and Diag$(\cdot)$ creates a diagonal matrix.

Finally, we leverage the psycholinguistic structure encoded in $\mathbf{W}^{(p)}$ to transform SenticNet emotions into personality trait scores:

$$\mathbf{o}_i^{(p)} = g^{(p)}\left(\mathbf{e}_i^\top \mathbf{W}^{(p)}\right), \tag{7}$$

where $g^{(p)}$ is an MLP that controls the scale of the representation and captures potential non-linear relationships between emotions and personality.

OCEAN trait prediction. The final OCEAN prediction is given by a convex combination of the scores provided by the three components of our framework:

$$\mathbf{o}_i = \lambda_1 \sigma(\mathbf{o}_i^{(t)}) + \lambda_2 \sigma(\mathbf{o}_i^{(e)}) + \lambda_3 \sigma(\mathbf{o}_i^{(p)}), \quad \text{s.t.} \sum_{k=1}^{3} \lambda_k = 1 \tag{8}$$

where λ is a learnable parameter, and $\sigma(\cdot)$ is the sigmoid activation function.

Training. Since we formulated the OCEAN prediction problem as a multi-label classification task, we employed the Binary Cross-Entropy loss as the objective function:

$$\mathcal{L}_o = -\sum_{i=1}^{N}\sum_{j=1}^{|M|}\left[y_{ij}\log(o_{ij}) + (1-y_{ij})\log(1-o_{ij})\right], \tag{9}$$

where N is the number of training samples, $y_{ij} \in \{0,1\}$ is the ground truth for the j-th label of the i-th sample, and o_{ij} is the corresponding predicted score. Additionally, to maintain psycholinguistic grounding and prevent unconstrained drift, we add the following regularization term:

$$\mathcal{L}_{\mathrm{p}} = \beta^{(p)}\,\|\mathbf{W}^{(p)} - \mathbf{W}\|_F^2 + \beta^{(l)}\,\|\mathbf{W}^{(l)}\|_F^2, \tag{10}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, $\beta^{(\mathrm{p})}$ anchors $\mathbf{W}^{(\mathrm{p})}$ to its psycholinguistic initialization and $\beta^{(\mathrm{l})}$ penalizes arbitrary deviations in the data-driven component. Thus, the final loss function is:

$$\mathcal{L} = \mathcal{L}_o + \mathcal{L}_{\mathrm{p}}. \tag{11}$$

Computational complexity. We analyze the time complexity of the Model forward stage assuming a Transformer-based encoder with maximum sequence length T and embedding dimension D, where all neural components are two-layer MLPs with hidden dimension $D^{(b)}$.

Generating the text embeddings requires $\mathcal{O}(n^{(b)}LT^2D)$ (Vaswani et al., 2017), where L is the number of Transformer layers, $n^{(b)}$ is the batch size, and D is the hidden dimension. We separately compute the complexity of the three components of our framework. The text-based head has complexity $\mathcal{O}(n^{(b)}DD^{(b)})$, while the emotion-conditioned head has complexity $\mathcal{O}(n^{(b)}D^{(b)}D)$. For the prior-informed head, the correlation matrix $\mathbf{W}$ can be pre-computed offline and is therefore excluded from our complexity analysis. The gating operation in Eq. 6 costs $\mathcal{O}(n^{(b)}|E||M|)$, the emotion projection in Eq. 7 costs $\mathcal{O}(n^{(b)}(|E|D^{(b)} + D^{(b)}|M|))$.

Since $|E|$ and $|M|$ are constants, the three components simplify to $\mathcal{O}(n^{(b)}D^{(b)}D)$. Thus, the overall complexity is $\mathcal{O}(n^{(b)}LT^2D + n^{(b)}D^{(b)}D)$. Note that the three components can be executed in parallel, making the computational overhead minimal compared to the encoder.

5. Experimental Evaluation

Data. We evaluated our framework on corpora labeled with Big Five traits in the out-of-domain setting, i.e., given a set of K datasets, we fine-tune on the concatenation of $K-1$ datasets and evaluate on the remaining one. We used the following English-language datasets spanning different domains: (i) The **Essay** (Pennebaker and King, 1999) dataset comprises 2468 essays authored by students. Each text is annotated with binary indicators corresponding to the Big Five personality traits, as determined through a standardized self-report personality assessment; (ii) The **FriendsPersona** (Jiang et al., 2020) dataset comprises 711 short multi-party conversations sampled from Seasons 1–4 of the TV show Friends. For each conversation, three crowd annotators label the main speaker on the Big Five personality traits; (iii) The **Backstage** (Tiuleneva et al., 2024) dataset contains 3265 text samples, each formed by concatenating the spoken lines of a single fictional character extracted from English theatre plays. Big Five traits are annotated per character using GPT-3.5 and then checked on a 10% stratified subsample by two human annotators; (iv) The **Pandora** (Gjurković et al., 2021) dataset contains Reddit comments. Big Five personality labels were derived directly from self-reported test results voluntarily shared by Reddit users. We used a subset of 10,000 examples in our experiments, which we named **Pandora**$_{sub}$.

The statistics of the data are reported in Tab. 1

Dataset	Rows	O	C	E	A	N
Pandora$_{sub}$	10000	4284	7239	3288	3690	5196
Essay	2468	1271	1254	1275	1309	1234
Backstage	3265	959	1254	1096	2026	1361
FriendsPersona	711	462	330	312	405	332

Table 1: Dataset statistics showing the total number of samples (Rows) and the distribution of personality traits. Each trait column indicates the number of samples labeled as positive for that trait.

Competing methods. As a baseline, we employed a two-layer MLP applied to the correlation-based mapping between OCEAN traits and SenticNet emotions computed via point-biserial correlation, as proposed by Yash et al. (2020). We also compared our framework against two classes of competing methods: (i) *Sentence embedding-based methods* specifically fine-tuned for OCEAN prediction, including BERT (Devlin et al., 2019b), *all-mpnet-base-v2*, *all-MiniLM-L6-v2*, and *all-distilroberta-v1*, all equipped with a classification head for OCEAN prediction; (ii) *LLMs with zero-shot inference*, which included *Llama-3.1-8B-Instruct* and *GPT-OSS-20B*.

Experimental setting. We used SenticNet 6 for computing affective information (Cambria et al., 2020), and we instantiated our framework with *all-mpnet-base-v2* as the sentence encoder.

Sentence embedding-based methods, including ours, were fine-tuned on the training domains, while the test domain was used exclusively for evaluation. We employed LoRA (Hu et al., 2022) for parameter-efficient fine-tuning of the last 4 encoder layers while keeping the base model frozen. Specifically, we used rank $r = 8$, scaling factor $\alpha = 16$, and dropout rate of 0.3. We set a maximum sequence length of 100 tokens with padding to the maximum length. Each neural component beyond the PLM consisted of a two-layer MLP with a hidden dimension of 100. We set $\beta^{(l)} = \beta^{(p)} = 0.1$, and we used the AdamW optimizer (Loshchilov and Hutter, 2019) for training, with weight decay equal to 1×10^{-4}. We performed a grid search to select the optimal learning rate and dropout values, yielding 1×10^{-4} and 0.3, respectively. For each method, we report the average performance over 5 runs.

Considering LLMs, we prompted each model to predict the OCEAN traits of the given sentence, instructing it to assume the *role of an expert in the Big Five personality model*. Following prior work demonstrating that incorporating domain-specific knowledge enhances personality trait inference (Li et al., 2025), we included the description of each OCEAN trait in the prompt to improve prediction performance. The prompt used is shown as follows:

Prompt used by LLMs

[ROLE]
You are an expert in the Big Five Personality Model. From the given TEXT, infer Big Five (OCEAN) traits expressed by the speaker/writer.

[TRAIT CRITERIA]
Agreeableness (A): warmth, empathy, co-operation, politeness, helpfulness, conflict avoidance; Conscientiousness (C): orderliness, reliability, planning, self-discipline, goal focus, rule-following; Extraversion (E): sociability, talkativeness, assertiveness, enthusiasm, excitement-seeking, and positive affect; Neuroticism (N): anxiety, irritability, moodiness, self-doubt, stress reactivity, emotional volatility; Openness (O): curiosity, imagination, creativity, intellectual/artistic interest, preference for novelty.

[TEXT]
`{<target text>}`

[OUTPUT FORMAT]
Your response must contain just the list of the detected OCEAN traits. Do not add any additional words or introductions.

We set the models' *temperature* to 1.0×10^{-2} to reduce randomness and ensure the selection of the most probable tokens for more precise and deterministic responses. We ran each model five times and reported the average performance.

	O	C	E	A	N	AVG
FriendsPersona						
Baseline	0.575	0.472	0.503	0.526	0.500	0.515
BERT	0.746	0.630	0.574	0.600	0.232	0.556
All-MiniLM	0.227	0.525	0.607	0.626	0.631	0.523
All-distilroberta	0.467	0.597	**0.695**	0.458	0.533	0.550
All-mpnet-base	0.431	0.566	0.545	0.712	0.273	0.505
Llama 3.1	0.225	0.063	0.545	0.482	0.425	0.348
GPT-OSS-20B	0.219	0.168	0.451	0.335	0.411	0.317
Ours	**0.789**	**0.636**	0.610	**0.726**	**0.637**	**0.680**
Essay						
Baseline	0.478	0.534	0.475	0.467	0.471	0.485
BERT	0.675	0.663	0.640	0.511	0.323	0.562
All-MiniLM	0.587	0.674	0.618	0.624	0.616	0.624
All-distilroberta	0.556	0.662	0.650	0.686	0.542	0.619
All-mpnet-base	0.557	0.653	0.622	0.672	0.633	0.627
Llama 3.1	0.657	0.225	0.527	0.594	0.655	0.531
GPT-OSS-20B	0.611	0.316	0.466	0.314	0.659	0.473
Ours	**0.680**	**0.675**	**0.681**	**0.694**	**0.677**	**0.681**
Backstage						
Baseline	0.440	0.373	0.428	0.402	0.543	0.454
BERT	0.350	0.504	0.469	0.740	0.162	0.445
All-MiniLM	0.454	0.554	0.448	0.667	0.444	0.513
All-distilroberta	0.253	0.547	0.457	0.760	0.446	0.493
All-mpnet-base	**0.491**	**0.556**	0.481	0.737	0.418	0.537
Llama 3.1	0.363	0.211	0.429	0.641	0.587	0.446
GPT-OSS-20B	0.302	0.264	0.384	0.373	0.514	0.367
Ours	0.489	**0.556**	**0.504**	**0.767**	**0.589**	**0.581**
Pandora$_{sub}$						
Baseline	0.477	0.593	0.393	0.420	0.495	0.476
BERT	0.578	0.757	0.462	0.263	0.476	0.507
All-MiniLM	0.586	0.725	0.467	0.531	0.515	0.565
All-distilroberta	0.566	0.804	0.436	0.444	0.581	0.566
All-mpnet-base	0.576	0.764	0.396	0.473	0.594	0.561
Llama 3.1	0.321	0.126	0.240	0.343	0.230	0.252
GPT-OSS-20B	0.199	0.211	0.171	0.187	0.214	0.196
Ours	**0.685**	**0.840**	**0.495**	**0.540**	**0.685**	**0.649**

Table 2: Comparative evaluation: F1-scores. Best results in bold, second-best results underlined.

5.1. Results

Comparative evaluation. Table 2 reports the F1-score values achieved by our framework and the competing methods on each dataset. Our method stands out as the best approach, yielding an average F1-score of 0.648 across all datasets. The sentence embedding-based methods fine-tuned for OCEAN prediction achieve average F1-scores of 0.561 (*All-mpnet-base*), 0.557 (*All-distilroberta*) and 0.556 (*All-MiniLM*). By contrast, our emotion-aware framework achieves an average improvement of +15.5% over the best baseline, demonstrating the effectiveness of our approach in enhancing personality trait prediction.

Although LLMs benefit from extensive pre-training on diverse text sources, the bestperforming LLM (Llama-3.1-8B) yields substantially lower performance than our framework across all datasets (0.394 on average). GPT-OSS-20B performs even worse, achieving only 0.338 on average. These results align with previous research highlighting the difficulties of LLMs in detecting personality traits (Ganesan et al., 2023; Piastra and Catellani, 2025). The performance gap is particularly pronounced on the Pandora$_{sub}$ dataset, where Llama-3.1-8B achieves only 0.252 compared to our 0.649, suggesting that LLMs struggle especially with certain types of textual content, like Reddit posts.

Notably, encoder-only methods consistently outperform LLMs across all datasets, showing that task-specific models fine-tuned for OCEAN prediction are more effective than general-purpose LLMs with significantly more parameters for this task.

At the trait level, our framework consistently performs well across all OCEAN dimensions on each dataset. Specifically, it achieves the best F1-score in 18 out of 20 trait-dataset combinations. The most significant improvements are observed for Openness on Pandora$_{sub}$ (0.685 vs. 0.586 for the second-best model) and Neuroticism on FriendsPersona (0.637 vs. 0.631), demonstrating the robustness of our emotion-based approach across diverse personality traits and textual domains.

Qualitative inspection of the learned prior matrix. Figure 2 shows the prior matrix $\mathbf{W}^{(p)}$ (cf. Eq. 6) learned by our framework after training on the Essay (left) and Friends (right) datasets using an 80/20% train/test split.

On the Essay dataset, the largest entry is between Aptitude and Agreeableness. The relatively low entry values are plausible for long, reflective texts where competence, calmness (low anger), and positive mood co-occur with higher Agreeableness and Conscientiousness. This pattern aligns with prior point-biserial correlation findings on the Essay dataset (Yash et al., 2020), indicating that the gating mechanism is capable of preserving interpretable emotion-personality priors when needed.

Similarly, on the FriendsPersona dataset, most coefficients converge toward 0 except for a clear negative Temper–Agreeableness association (-0.132), consistent with anger-colored utterances lowering Agreeableness in short, multi-speaker dialogues. This shrinkage likely reflects the characteristics of brief utterances, sarcasm, topic volatility, and sparser concept coverage typical of conversational data (Jiang et al., 2020).

Qualitative analysis of the responses. Figure 3 illustrates predictions of our framework on some instances of the datasets. The top-left image is an excerpt from the Essay dataset, describing a college student's diary-style reflection expressing anxiety about examinations, excitement, and general frustration regarding college life. In this case, the model correctly predicts all ground truth OCEAN traits, namely Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

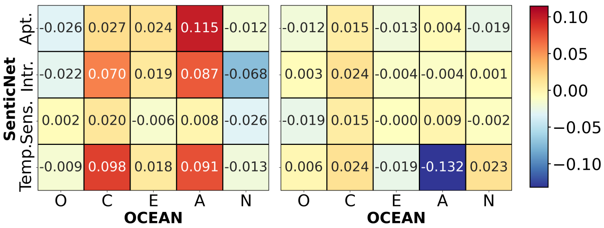

Figure 2: Learned prior matrices. Left. Essay dataset. Right. FriendsPersona dataset.

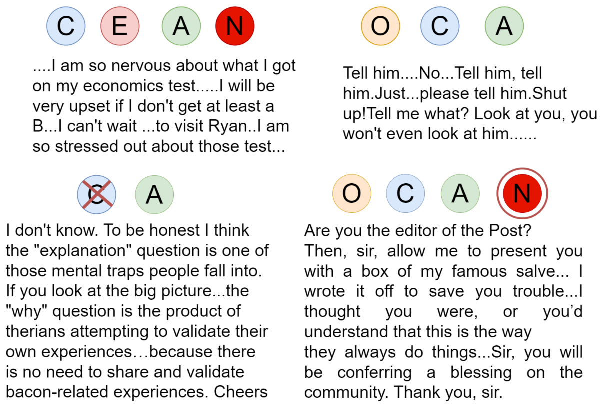

Figure 3: Examples of OCEAN predictions by our framework. Red circles indicate false negatives, and red crosses indicate false positives. Top left: Essay. Top right: FriendsPersona. Bottom left: Pandora$_{sub}$. Bottom right: Backstage.

The top-right image, extracted from the FriendsPersona dataset, depicts a tense dialogue in which one individual urges another to reveal information. Here again, the model accurately predicts all ground truth labels.

The bottom-left image, drawn from the Pandora$_{sub}$ dataset, contains a philosophical text about individuals with animal identities. In this case, the model incorrectly predicts high Conscientiousness, likely because the text exhibits highly structured, methodical thinking with academic references, careful logical progression, and thorough systematic analysis—all markers of Conscientiousness.

In the bottom-right image, extracted from the Backstage dataset, the model fails to predict Neuroticism, likely due to the stylized and performative nature of theatrical dialogue, which differs significantly from the conversational or reflective texts in the training data.

6. Conclusion

In this work, we presented a novel emotion-aware framework for Big Five Personality prediction. Our approach integrates affective information through two complementary mechanisms: first, emotion-conditioned embedding modulation via FiLM; second, interpretable psycholinguistic features obtained through learnable mappings between SenticNet dimensions and OCEAN personality traits. Our experimental evaluation across different datasets from different domains showed the effectiveness of our framework. Nevertheless, further investigation is needed to better understand the multifaceted emotion-personality relationship and develop more effective, interpretable LLM-based frameworks for OCEAN prediction. In future work, we aim to extend the approach by incorporating temporal dynamics to model personality trait evolution over time.

7. Limitations

Our framework is designed to be independent of the particular Pre-trained Language Model employed in the encoding phase. Therefore, alternative or more advanced models can be seamlessly integrated for computing sentence embeddings. Nonetheless, the reliance on SenticNet for emotion extraction introduces certain limitations. For example, the prior-informed head is specifically designed to leverage the relationships between SenticNet emotions and the Big Five traits (Yash et al., 2020). Additionally, the SenticNet lexicon may not adequately capture domain-specific or emerging emotional expressions, particularly in informal online discourse, potentially affecting coverage. Thus, adopting alternative personality or emotional models may require architectural modifications.

However, the proposed framework is agnostic to the specific SenticNet release, provided that the components relying on SenticNet outputs are retrained on the chosen version. Accordingly, it can be seamlessly integrated with newer SenticNet versions, benefiting from their improvements and potentially yielding better results. We leave a comparison across SenticNet versions for future work.

Furthermore, the evaluation is restricted to English text, and the framework's effectiveness for languages with different linguistic structures or cultural personality expression patterns remains unexplored.

Finally, note that the learned prior $\mathbf{W}^{(p)}$ represents corpus-specific associations rather than causal relationships between emotions and personality traits. Accordingly, it should be treated as an interpretable, domain-adaptive prior that regularizes predictions.

8. Ethical remarks

The outputs generated by the model do not reflect the authors' views or beliefs and are solely intended for research purposes. Nonetheless, biases embedded in the training data may propagate through the model, making it essential to carefully consider its deployment to prevent the reinforcement of harmful stereotypes or biases. Ensuring ethical alignment requires ongoing evaluation.

Moreover, we acknowledge that our framework employs psychological theories that are not immune to limitations in terms of cross-cultural generalizability. We also hypothesize that the challenges regarding the universality and unambiguous detection of emotions, as discussed in previous work (Stark and Hutson, 2021; Stark, 2023; Stark and Hoey, 2021), might similarly arise in personality detection and, more broadly, in AI systems designed to integrate psychosocial dimensions.

9. Bibliographical References

Md Ali Akber, Tahira Ferdousi, Rasel Ahmed, Risha Asfara, Raqeebir Rab, and Umme Zakia. 2024. Personality and emotion—a comprehensive analysis using contextual text embeddings. *Natural Language Processing Journal*, 9:100105.

Süleyman Aslan and Uğur Güdükbay. 2019. Multimodal video-based apparent personality recognition using long short-term memory and convolutional neural networks. *arXiv preprint arXiv:1911.00381*.

Erik Cambria, Yang Li, Frank Z Xing, Soujanya Poria, and Kenneth Kwok. 2020. Senticnet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 105–114.

Erik Cambria, Andrew Livingstone, and Amir Hussain. 2012. The hourglass of emotions. In *Cognitive behavioural systems: COST 2102 international training school, dresden, Germany, February 21-26, 2011, revised selected papers*, pages 144–157. Springer.

Charles M Ching, A Timothy Church, Marcia S Katigbak, Jose Alberto S Reyes, Junko Tanaka-Matsumi, Shino Takaoka, Hengsheng Zhang, Jiliang Shen, Rina Mazuera Arias, Brigida Carolina Rincon, et al. 2014. The manifestation of traits in everyday behavior and affect: A five-culture study. *Journal of Research in Personality*, 48:1–16.

Harm de Vries, Florian Strub, Jérémie Mary, Hugo Larochelle, Olivier Pietquin, and Aaron C. Courville. 2017. Modulating early visual processing by language. In *Advances in Neural Information Processing Systems (NeurIPS)*. Conditional Batch Normalization.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019a. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019b. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

John M Digman. 1990. Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1):417–440.

Adithya V Ganesan, Yash Kumar Lal, August Håkan Nilsson, and H Andrew Schwartz. 2023. Systematic evaluation of gpt-3 for zero-shot personality estimation. *arXiv preprint arXiv:2306.01183*.

Alan S Gerber, Gregory A Huber, David Doherty, Conor M Dowling, and Shang E Ha. 2010. Personality and political attitudes: Relationships across issue domains and political contexts. *American Political Science Review*, 104(1):111–133.

Fatima Habib, Zeeshan Ali, Akbar Azam, Komal Kamran, and Fahad Mansoor Pasha. 2024. Navigating pathways to automated personality prediction: a comparative study of small and medium language models. *Frontiers in Big Data*, 7:1387325.

Yimin He, M Brent Donnellan, and Anjelica M Mendoza. 2019. Five-factor personality domains and job performance: A second order meta-analysis. *Journal of Research in Personality*, 82:103848.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter-efficient transfer learning for nlp. In *International conference on machine learning*, pages 2790–2799. PMLR.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations (ICLR)*.

Dipika Jain, Akshi Kumar, and Rohit Beniwal. 2022. Personality bert: A transformer-based model for personality detection from textual data. In *Proceedings of international conference on computing and communication networks: ICCCN 2021*, pages 515–522. Springer.

Oliver P John, Richard W Robins, and Lawrence A Pervin. 2010. *Handbook of personality: Theory and research*. Guilford Press.

Oliver P John, Sanjay Srivastava, et al. 1999. The big-five trait taxonomy: History, measurement, and theoretical perspectives.

Weixi Kang, Francois Steffens, Sònia Pineda, Kaya Widuch, and Antonio Malvaso. 2023. Personality traits and dimensions of mental health. *Scientific Reports*, 13(1):7091.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

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.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*.

Navonil Majumder, Soujanya Poria, Alexander Gelbukh, and Erik Cambria. 2017. Deep learning-based document modeling for personality detection from text. *IEEE intelligent systems*, 32(2):74–79.

GJ Molloy, RE O'carroll, and Eamonn Ferguson. 2014. Conscientiousness and medication adherence: A meta-analysis. *Annals of Behavioral Medicine*, 47(1):92–101.

Anam Naz, Hikmat Ullah Khan, Amal Bukhari, Bader Alshemaimri, Ali Daud, and Muhammad Ramzan. 2025. Machine and deep learning for personality traits detection: a comprehensive survey and open research challenges. *Artificial Intelligence Review*, 58(8):239.

Vincent Oh and Eddie MW Tong. 2020. Negative emotion differentiation and long-term physical health—the moderating role of neuroticism. *Health Psychology*, 39(2):127.

Gregory Park, H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Michal Kosinski, David J Stillwell, Lyle H Ungar, and Martin EP Seligman. 2015. Automatic personality assessment through social media language. *Journal of personality and social psychology*, 108(6):934.

James W Pennebaker, Martha E Francis, Roger J Booth, et al. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. 2018. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*. ArXiv:1709.07871.

Heinrich Peters and Sandra C Matz. 2024. Large language models can infer psychological dispositions of social media users. *PNAS nexus*, 3(6):pgae231.

Marco Piastra and Patrizia Catellani. 2025. On the emergent capabilities of chatgpt 4 to estimate personality traits. *Frontiers in Artificial Intelligence*, 8:1484260.

Robert Plutchik. 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 89(4).

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)*. Association for Computational Linguistics.

Zhancheng Ren, Qiang Shen, Xiaolei Diao, and Hao Xu. 2021. A sentiment-aware deep learning approach for personality detection from text. *Information Processing & Management*, 58(3):102532.

William Revelle and Klaus R Scherer. 2009. Personality and emotion. *Oxford companion to emotion and the affective sciences*, 1:304–306.

Justin D Russell, Carl F Weems, Irfan Ahmed, and Golden G Richard III. 2017. Self-reported secure and insecure cyber behaviour: factor structure and associations with personality factors. *Journal of Cyber Security Technology*, 1(3-4):163–174.

Kit-May Shum, Michal Ptaszynski, and Fumito Masui. 2025. Big five personality trait prediction based on user comments. *Information*, 16(5):418.

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

Luke Stark. 2023. Artificial intelligence and the conjectural sciences. *BJHS Themes*, 8.

Luke Stark and Jesse Hoey. 2021. The ethics of emotion in artificial intelligence systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery.

Luke Stark and Jevan Hutson. 2021. Physiognomic artificial intelligence. *Fordham Intell. Prop. Media & Ent. LJ*, 32.

Yosephine Susanto, Andrew G Livingstone, Bee Chin Ng, and Erik Cambria. 2020. The hourglass model revisited. *IEEE Intelligent Systems*, 35(5):96–102.

Maria Taramigkou, Dimitris Apostolou, and Gregoris Mentzas. 2018. Leveraging exploratory search with personality traits and interactional context. *Information Processing & Management*, 54(4):609–629.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*.

Mehta Yash, Samin Fatehi, Amirmohammad Kazameini, Clemens Stachl, Erik Cambria, and Sauleh Eetemadi. 2020. Bottom-Up and Top-Down: Predicting Personality with Psycholinguistic and Language Model Features. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 1184–1189, Sorrento, Italy. IEEE. 25 citations (Crossref) [2023-05-02].

Tuncay Yılmaz, Abdullah Ergil, and Bahar İlgen. 2019. Deep learning-based document modeling for personality detection from turkish texts. In *Proceedings of the future technologies conference*, pages 729–736. Springer.

Jianfeng Zhu, Julina Maharjan, Xinyu Li, Karin G. Coifman, and Ruoming Jin. 2025. Evaluating llm alignment on personality inference from real-world interview data.

10. Language Resource References

Erik Cambria and Yang Li and Frank Z. Xing and Soujanya Poria and Kenneth Kwok. 2020. *SenticNet 6: Ensemble Application of Symbolic and Subsymbolic AI for Sentiment Analysis*.

Erik Cambria, Soujanya Poria, Devamanyu Hazarika, and Kenneth Kwok. 2018. Senticnet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 1795–1802.

Gjurković, Matej and Karan, Vanja Mladen and Vukojević, Iva and Bošnjak, Mihaela and Snajder, Jan. 2021. *PANDORA Talks: Personality and Demographics on Reddit*. Association for Computational Linguistics.

Jiang, Hang and Zhang, Xianzhe and Choi, Jinho D. 2020. *Automatic text-based personality recognition on monologues and multiparty dialogues using attentive networks and contextual embeddings (student abstract)*.

James W Pennebaker and Laura A King. 1999. Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, 77(6):1296.

Tiuleneva, Marina and Porvatov, Vadim A. and Strapparava, Carlo. 2024. *Big-Five Backstage: A Dramatic Dataset for Characters Personality Traits & Gender Analysis*. ELRA and ICCL.