LLM-Driven Personality Augmentation for Enhancing Text-Based Big Five Personality Detection

Anonymous ACL submission

Abstract

Text-based Big Five personality detection relies on training data, however, existing data augmentation methods are often based on surface-level perturbations, paraphrasing, or distributional interpolation, which limits their ability to produce diverse, high-quality, and accurately labeled personality-relevant texts. In this study, we introduce LLM-driven personality synthesis as a data augmentation paradigm for personality detection. Instead of merely modifying existing samples, the proposed framework formulates augmentation as a controlled generation task, using personality-conditioned prompts, reference-guided synthesis, style-aware instructions, and diversity-oriented constraints to generate texts aligned with target Big Five traits. We evaluate the framework on two widely used datasets, Essays and MyPersonality, and compare it with conventional augmentation strategies and representative personality detection baselines. Experimental results show that personality synthesis consistently improves downstream detection performance, achieving 72.19% average accuracy on Essays, and 71.42% average accuracy on MyPersonality. Further analyses confirm that trait-conditioned and reference-guided synthesis enhances label consistency and textual diversity, demonstrating the effectiveness of personality synthesis for text-based personality data augmentation.

1 Introduction

With the advancement of social media platforms, users continuously generate substantial textual data on Weibo, X, Steam, Douyin and other platforms. Such data not only record user viewpoints, emotions and interactive behaviors, but also reflect their relatively stable personality dispositions, making personality computation a vital research direction at the intersection of natural language processing, computational social science, and psychological informatics (Vinciarelli and Mohammadi, 2014; Lin et al., 2023b). Personality computation mainly consists of three branches: personality detection, personality perception, and personality synthesis. Personality detection aims to identify users' personality traits; personality perception emphasizes the understanding and utilization of personality information by models; personality synthesis generates textual content or agent behaviors with specific personality tendencies (Jiang et al., 2023; Lin, 2025). Owing to its low cost, non-invasiveness, and scalability, text-based personality detection has been widely adopted in tasks including user profiling, recommendation systems, mental health support, and human-computer interaction (Yang and Huang, 2019; Ren et al., 2021; Zhu et al., 2024).

Nevertheless, current social media personality recognition approaches are still constrained by dataset quality. Personality labels are predominantly acquired through questionnaires, expert annotation, or indirect inference, resulting in high data acquisition costs and limited sample size. In addition, real-world data presents imbalanced personality distribution with scarce minority-class samples, which makes models prone to majority-class bias and hinders the learning of fine-grained linguistic characteristics of minority personalities (Lin et al., 2023b; Zhu et al., 2024). Although existing data augmentation methods can alleviate class imbalance, most techniques modify raw texts locally or generate samples by learning real data distributions. Remaining constrained by original datasets, these methods struggle to produce high-quality minority personality samples and may introduce semantic drift and label noise (Zhu et al., 2022; Wen et al., 2023; Wang et al., 2025).

To tackle the above limitations, this paper incorporates personality synthesis technology into personality data augmentation. Leveraging the controllable generation capability of large language models, the proposed method generates texts with designated personality tendencies, which are uti-

lized as augmented samples for the training of personality recognition models (Jiang et al., 2023; Lin, 2025). Different from conventional augmentation methods that perturb original samples, this approach shifts toward targeted personality-oriented synthesis. It can directly supplement insufficient personality expression patterns and improve the quality and diversity of minority-class personality samples.

2 Related Work

In the field of personality computing, personality modeling currently relies primarily on classic personality taxonomies such as the Big Five personality model and MBTI (Mehta et al., 2020b). Personality detection models built upon these personality frameworks are widely based on deep learning. Early personality detection models relied on psycholinguistic features such as Linguistic Inquiry and Word Count (LIWC), n-grams, word frequency statistics, and sentiment lexicons (Majumder et al., 2017; Ren et al., 2021; Lin, 2024; Lin et al., 2023a). Since the release of BERT, mainstream approaches have gradually shifted to leveraging pre-trained language models to capture contextual semantic information (Jain et al., 2022; El-Demerdash et al., 2022; López-Pabón and Orozco-Arroyave, 2022), thereby significantly improving personality detection performance.

Recent work has also explored zero-shot LLM-based personality detection beyond textual inputs, such as predicting Big Five traits from social avatars (Lin et al., 2026). State-of-the-art personality detection methods increasingly rely on few-shot learning and knowledge enhancement with LLMs (Hu et al., 2024; Yang et al., 2023; Bi et al., 2025). However, like earlier approaches, they are still affected by insufficient personality diversity in the dataset. Moreover, this lack of personality diversity cannot be effectively addressed by existing data augmentation methods such as Easy Data Augmentation (EDA) (Wei and Zou, 2019), back-translation (Taheri et al., 2025; Wang et al., 2023), or GANs (Yu et al., 2017; Ding et al., 2025). These augmentation methods mainly generate new samples by making local modifications to existing text or by learning the distribution of real text. Existing methods often struggle to generate high-quality texts whose assigned labels are consistent with the intended personality traits, thereby limiting the model's generalization ability to long-tail personality profiles.

3 Methodology

To address label imbalance and insufficient minority samples in social media personality recognition, we build an LLM-enhanced Big Five prediction framework. The method consists of three stages: LLM-based sample supplementation, BERT-based feature extraction, and MLP-based personality prediction. The overall pipeline is illustrated in Figure 1.

3.1 LLM-Based Augmentation

The augmentation module is designed around prompt-driven controllable generation. Rather than treating the LLM as a generic text generator, we convert each target Big Five label combination into a structured prompt that jointly specifies personality traits, corpus-style references, dataset-level writing constraints, and sample-level diversity controls. This design aims to produce supplementary samples that are label-consistent, stylistically close to the original data, and diverse within the same personality category.

3.1.1 Distribution Analysis and Target Setting

The five Big Five dimensions are denoted as cEXT, cNEU, cAGR, cCON, and cOPN. Each dimension is encoded as a binary *y*/*n* label, forming $2^5 = 32$ personality combinations. Let $c_k = (l_1, \ldots, l_5)$ be the k-th combination, where $l_d \in \{y, n\}$ and $k = 1, \ldots, 32$. Before constructing prompts, we estimate the generation demand for each combination. As shown in Equation 1, the generation gap g_k is defined as the positive part of the difference between the target size T_k and the observed count of combination c_k:

$$g_k = \left[T_k - \sum_{i=1}^{N} \mathbf{1}(\mathbf{y}_i = c_k) \right]_+ \quad (1)$$

where N is the total number of samples, $\mathbf{y}_i$ is the label vector of sample x_i, $\mathbf{1}(\cdot)$ is an indicator function, and $[\cdot]_+ = \max(0, \cdot)$. Combinations below the target size are selected for augmentation, while those that already meet the target receive no additional samples.

3.1.2 Reference Anchors and Personality-Controlled Prompts

Reference anchor selection. To keep generated samples close to the source corpus, we include

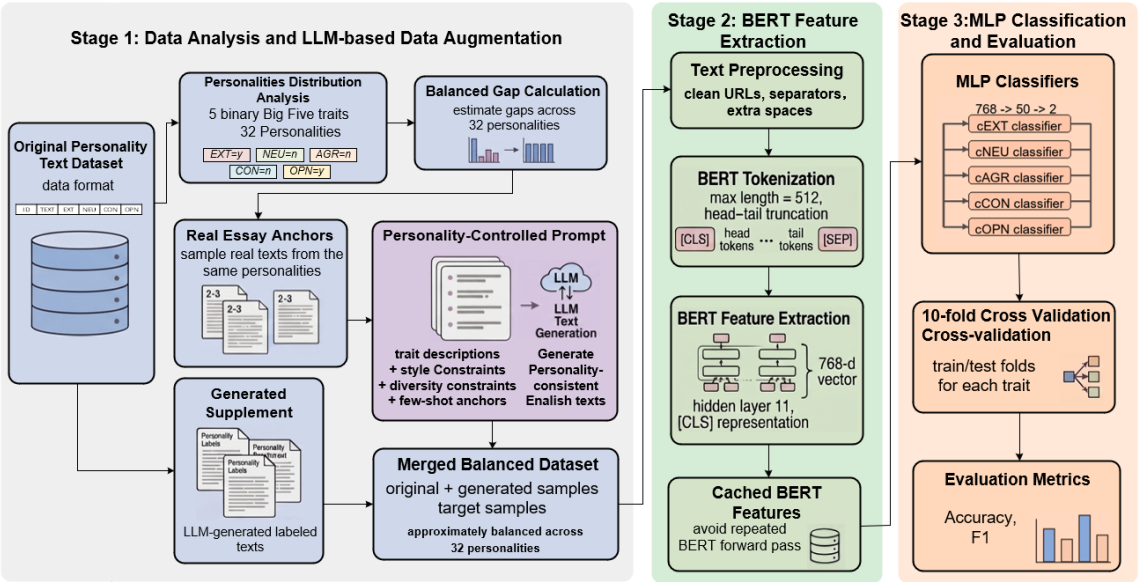


Figure 1: Overall framework of the proposed LLM-driven personality augmentation method.

real texts with the same personality combination as reference anchors. As shown in Equation 2, A_k contains r texts sampled from the subset whose labels match c_k:

$$A_k = \text{Sample}(\{x_i \mid \mathbf{y}_i = c_k\}, r) \tag{2}$$

In implementation, up to three anchors are used for each target combination. Overlength anchors are truncated by first searching backward for the nearest sentence-ending punctuation and then falling back to a comma-level boundary if no suitable sentence boundary is found. These anchors serve as few-shot style cues, rather than texts to be copied.

Trait-conditioned prompt construction. The complete generation prompt consists of five components: a task instruction, a trait-conditioned personality guidance component, a dataset-style instruction, reference anchors, and a sample-level diversity instruction. Among them, the trait-conditioned personality guidance is implemented by our proposed SIR2 module, which is illustrated in Figure 2. Instead of directly feeding raw *y*/*n* trait labels to the LLM, SIR2 maps each Big Five label combination into second-person trait descriptions. It starts from trait identity statements, expands them with behavioral and psychological descriptions, removes redundant content, and then splits the remaining descriptions into sentence-level prompt candidates. For example, high extraversion can be expressed by prompts such as "You are an extraversive person", whereas low extraversion can be expressed by "You are an introspective individual". Formally, the SIR2 personality prompt is sampled as:

$$P_{\text{pers}}(c_k) = \text{Sample}(\mathcal{L}(c_k)) \tag{3}$$

where $\mathcal{L}(c_k)$ is the SIR2 prompt candidate set associated with personality combination c_k. Sampling from split trait descriptions reduces repeated wording across generated samples with the same target label, while keeping the personality control separated from dataset-style and diversity-control instructions.

Taking the personality combination cEXT=y, cNEU=n, cAGR=n, cCON=n, cOPN=n as an example, the final prompt consists of several lightweight modules: (1) content structure constraint: *Begin with a complaint or worry...*; (2) expression variation constraint: *Avoid starting sentences with...*; (3) perspective variation constraint: *Write from the perspective of...*; (4) reference anchors: *I finally got into this thing...*; (5) output constraint: *Please generate the text directly...*. Combined with the progressive refinement process from Naive prompt to SIR² prompt shown in the figure, these modules together with personality labels form the final prompt, which improves the personality controllability and linguistic naturalness of generated texts.

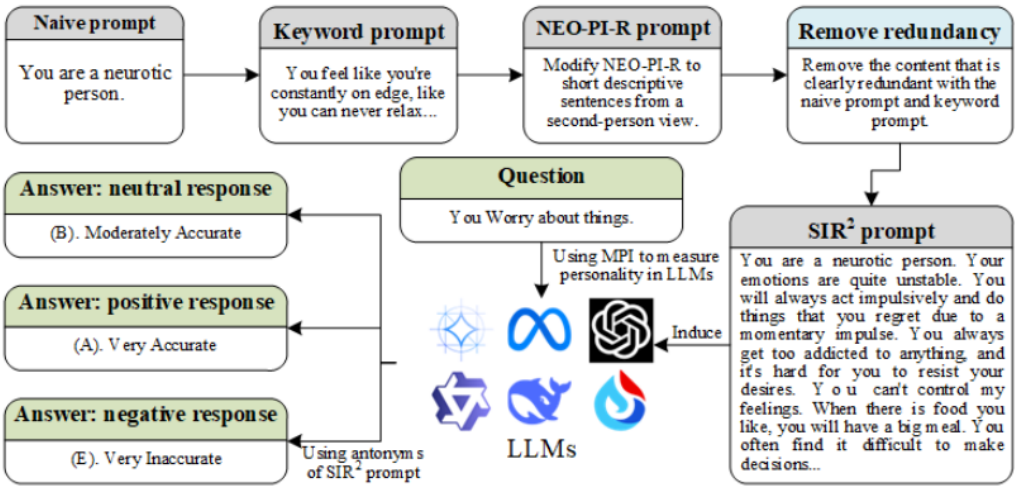


Figure 2: Personality prompt construction process adapted from PersonalityLLM (Lin et al., 2026). Trait labels are expanded into second-person personality descriptions, redundant content is removed, and the remaining descriptions are used as prompt candidates for personality-controlled generation.

3.1.3 Prompt Assembly, Sample Generation, and Dataset Merging

Dataset-style instruction. Taking the Essays dataset as an example, the prompt defines the model role as a psychology research assistant and asks it to generate a college-student stream-of-consciousness essay. The style instruction specifies English output, one dense paragraph without headers, an informal tone, daily-life content, and minor imperfections such as run-on sentences or casual abbreviations, so that the output remains close to the original corpus style.

Sample-level diversity control. To avoid near-duplicate outputs under the same personality combination, each generation request receives a diversity instruction D_j. It is sampled from language-level control pools covering discourse openings, surface expression, and temporary narrative stance. The resulting D_j changes how a sample is written without changing which personality label it expresses. This strategy only operates on basic linguistic structure, expression, and narrative stance, thereby avoiding explicit personality-inducing content in this part of the prompt.

Final prompt assembly and merging. The final prompt integrates the task instruction, personality guidance, dataset-style constraint, reference anchors, and diversity control:

$$P_{k,j} = P_{\text{task}} \oplus P_{\text{pers}}(c_k) \oplus P_{\text{style}} \oplus A_k \oplus D_j \tag{4}$$

where P_{task} specifies the generation task, $P_{\text{pers}}(c_k)$ provides personality guidance, P_{style} defines dataset-level writing requirements, A_k supplies reference anchors, and D_j controls linguistic diversity. The prompt also requires the LLM to return only the generated text without explanations, labels, or headers. Generated texts are assigned the target Big Five labels of c_k and merged with the original data for downstream training.

3.2 BERT-Based Feature Extraction

After merging original and generated samples, we clean the text by removing URLs, redundant separators, and extra spaces, and convert the five *y/n* personality labels into binary numerical labels. The texts are encoded with bert-base-uncased. For inputs longer than 512 tokens, a head-tail truncation strategy preserves both the beginning and ending context. We use the [CLS] representation from the 11th BERT layer as a 768-dimensional semantic feature vector and cache extracted features to avoid repeated BERT forward passes.

3.3 MLP Classification and Model Evaluation

For downstream prediction, five independent binary classification tasks are constructed for cEXT, cNEU, cAGR, cCON, and cOPN. Each task uses a two-layer MLP with a 768-dimensional input layer, a 50-dimensional hidden layer, and a two-dimensional output layer. The classifier is trained with Adam at a learning rate of 5×10^{-4} for 10 epochs. We evaluate the model with 10-fold stratified cross-validation and report Accuracy and

F1 with their mean values and standard deviations. This evaluation is used to examine whether LLM-generated samples improve recognition performance, especially for underrepresented personality patterns.

4 Experiments

4.1 Datasets

We conducted experiments on two datasets, Essays and MyPersonality. Essays, a gold-standard dataset widely used in psychological personality detection research, contains 2,468 student essays annotated with the Big Five personality traits, which were assessed using a standardized self-report inventory. MyPersonality dataset originates from the myPersonality application on Facebook, covering around 250 users and nearly 10,000 status update texts. Its Big Five personality labels are collected via standardized personality questionnaires filled out by users, and all textual data are shared with user consent. Both datasets exhibit significant label imbalance across the personality dimensions, which poses challenges for reliable personality detection and makes them suitable for evaluating imbalance-aware data augmentation methods.

4.2 Metrics

This study evaluates the quality of augmented samples mainly from the perspective of textual diversity, while downstream classification performance is used as an indirect evaluation of whether the augmented samples improve personality recognition. Textual diversity is measured using MTLD (McCarthy and Jarvis, 2010) and Shannon entropy (Rosillo-Rodes et al., 2025), as shown in Equation 5.

$$D_{\text{div}} = (\text{MTLD}, H) \tag{5}$$

$$\text{MTLD} = \frac{\text{MTLD}_{\text{forward}} + \text{MTLD}_{\text{reverse}}}{2} \tag{6}$$

$$H = -\sum_{i=1}^{V} p_i \ln p_i \tag{7}$$

Equation 6 defines MTLD as the average of the forward and reverse MTLD scores. Equation 7 defines Shannon entropy, where V denotes the vocabulary size and p_i denotes the occurrence probability of the i-th word. For classification performance, Accuracy and F1-score are reported as downstream metrics.

4.3 Parameter Sensitivity Analysis

To examine the influence of few-shot reference length, we vary `max_len` from 100 to 800 and evaluate the results on five personality dimensions: EXT, NEU, AGR, CON, and OPN. For each dimension, the relationship between reference truncation length and classification performance is visualized in Figure 3.

As shown in Figure 3, overly short references fail to preserve sufficient personality cues and writing style, while overly long references may introduce redundant or noisy information. Therefore, an appropriate reference length is necessary to balance stylistic guidance and generation stability.

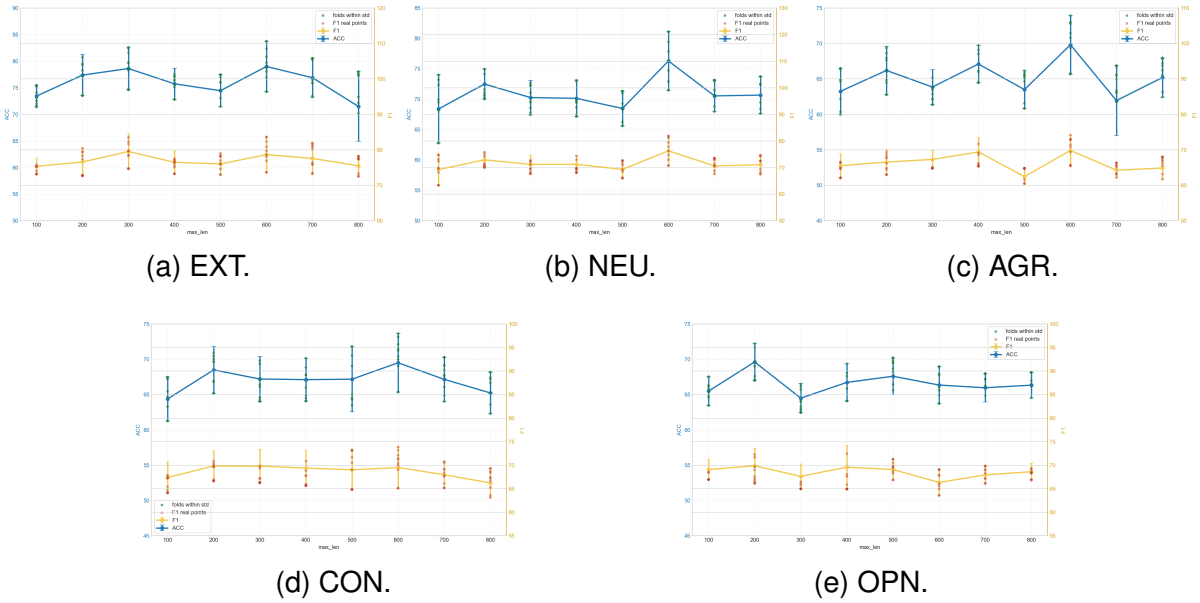
4.4 Sample Diversity Analysis

To compare the diversity of the LLM-generated samples with other augmentation methods, we analyze EDA (Wei and Zou, 2019), Back-translation (Taheri et al., 2025; Wang et al., 2023), SeqGAN (Yu et al., 2017), TPH-SMOTE (Du et al., 2026), and LLM using MTLD and Shannon entropy. Table 1 reports the comparison on Essays and MyPersonality. This subsection focuses on sample-level lexical variation, while downstream recognition performance is discussed in Section 4.6.

The comparison includes representative augmentation methods:

- EDA represents a simple text augmentation method based on synonym replacement, random insertion, random swap, and random deletion (Wei and Zou, 2019).
- Back-translation represents a translation-based augmentation method that rewrites text through intermediate translation while preserving the original label semantics (Taheri et al., 2025; Wang et al., 2023).
- SeqGAN represents a sequence generation method that extends generative adversarial learning to discrete text generation through policy-gradient reinforcement learning (Yu et al., 2017).
- TPH-SMOTE represents a heuristic oversampling method that integrates SMOTE with whale optimization to generate and select synthetic minority-class samples for imbalanced classification (Du et al., 2026).

As shown in Table 1, EDA obtains the highest MTLD on Essays, while the 2B LLM produces



(a) EXT. (b) NEU. (c) AGR.

(d) CON. (e) OPN.

Figure 3: Parameter sensitivity analysis under different reference truncation lengths.

Table 1: Diversity comparison between LLM augmentation and other augmentation methods.

Method	Essays MTLD	Essays H	MyPersonality MTLD	MyPersonality H
EDA	82.94	5.90	47.98	7.08
Back-translation	43.54	5.75	38.58	6.27
SeqGAN	70.47	5.89	48.11	6.78
TPH-SMOTE	52.00	6.04	52.74	6.85
LLM (2B)	74.28	5.95	59.97	6.38

competitive entropy and MTLD values across both datasets. These results indicate that LLM-generated samples maintain a reasonable degree of lexical variation, but diversity alone cannot determine augmentation quality. Therefore, the diversity results are interpreted together with the downstream classification comparison in Section 4.6.

Table 2: Ablation results relative to the complete augmentation pipeline.

Methods	Essays		MyPersonality	
	ACC	F1	ACC	F1
Ours	**72.19**	**71.86**	**71.42**	**70.01**
(w/o) Reference anchors	60.59	60.88	63.32	61.48
(w/o) Dataset-style instruction	65.29	64.45	65.63	64.62
(w/o) Personality prompts	61.58	61.16	52.85	44.24
(w/o) all	56.25	59.90	61.85	44.92

4.5 Ablation Experiment

To analyze the contribution of each component in the complete LLM-based augmentation pipeline, we conduct an ablation experiment under the same downstream evaluation setting. The first column in Table 2 denotes the component removed from the full pipeline, while the remaining columns report the results on Essays and MyPersonality, respectively.

The results show that different components influence recognition performance differently on the two datasets. Compared with the ablated variants, the complete setting achieves the best performance on both Essays and MyPersonality, with ACC/F1 scores of 72.19/71.86 and 71.42/70.01, respectively. This indicates that the complete LLM augmentation pipeline provides more effective training samples. Among the ablated variants, removing the dataset-style instruction yields relatively higher scores on both datasets, whereas removing personality prompts causes a clear decline on MyPersonality. These results suggest that personality prompts help preserve label consistency, while reference

anchors and dataset-style instructions should be balanced according to dataset characteristics to generate more discriminative augmented samples.

4.6 Comparison with Baselines

We first compare different augmentation methods under the same downstream personality classification setting. Accuracy and F1-score are used to evaluate recognition performance. Table 3 reports the augmentation-method comparison under the 1.7B setting, including EDA, Back-translation, SeqGAN, TPH-SMOTE, and LLM.

Table 3: Downstream comparison of augmentation methods under the 1.7B setting.

Method	Essays		MyPersonality	
	ACC	F1	ACC	F1
EDA	60.53	62.00	61.86	44.59
Back-translation	60.67	62.68	63.42	53.72
SeqGAN	52.55	57.31	60.68	36.36
TPH-SMOTE	59.59	61.04	62.55	57.69
LLM	**72.19**	**71.86**	**71.42**	**70.01**

Table 4 further reports the augmentation-method comparison under the 2B setting. This comparison is used to examine whether increasing the generation model size leads to consistent downstream gains over traditional augmentation methods.

Table 4: Downstream comparison of augmentation methods under the 2B setting.

Method	Essays		MyPersonality	
	ACC	F1	ACC	F1
EDA	60.53	62.00	61.86	44.59
Back-translation	60.67	**62.68**	63.42	53.72
SeqGAN	52.55	57.31	60.68	36.36
TPH-SMOTE	59.59	61.04	62.55	57.69
LLM	**64.50**	60.49	**64.41**	**63.49**

The results in Table 3 and Table 4 show that LLM augmentation performs differently under the two generation settings. Under the 1.7B setting, LLM achieves the best averaged Accuracy and F1 on both Essays and MyPersonality, indicating that the generated samples provide effective training signals for downstream personality recognition. Under the 2B setting, the LLM results remain competitive but do not consistently exceed the 1.7B setting, suggesting that increasing model size alone does not guarantee stronger augmentation effects.

Tables 5 and 6 further compare the proposed method with representative personality detection methods, including:

- BERT-MLP represents a MLP model using features extracted by BERT-base (Mehta et al., 2020a).
- CNN-AdaBoost represents a model with features obtained from various filters of CNN are fed to an AdaBoost (Mohades Deilami et al., 2022).
- BERT-fusion represents a model using both data and classifier level fusion. The adopted features are extracted with three pre-trained language models including ELMo, ULMFiT, and BERT (El-Demerdash et al., 2022).
- CNN-BiLSTM represents a multi-model attention based CNN-BiLSTM (Chaurasia et al., 2024).
- Attention-BiLSTM represents a model that combines attention mechanisms and bidirectional LSTM layersn(Bounab et al., 2024).
- DesPrompt, as described by (Wen et al., 2023), models personality detection as a word-filling task.
- PADO represents a multi-agent framework that leverages large language models to analyze text from contrasting perspectives for detecting the personality traits (Yeo et al., 2025).
- EERPD represents a novel personality detection method that integrates emotion regulation with emotion features to improve personality prediction through few-shot examples and process Chain of Thoughts (Li et al., 2025).

Compared with these methods, our approach achieves stronger performance on most personality dimensions. Specifically, on the Essays dataset, the accuracys achieved by our method are greater than the current SOTA accuracys by 11.24% on EXT, NEU, AGR, and CON. On the MyPersonality dataset, this increase rate is 9.99%. This suggesting that LLM-generated augmented samples can effectively supplement insufficient training examples and improve the downstream model's ability to identify these personality traits.

On the MyPersonality dataset, however, the improvement on OPN is relatively limited. This may be because openness is associated with more abstract semantic factors such as topic preference, imagination, creativity, and long-term expressive tendencies, which are difficult to fully capture

Table 5: Comparison results with baseline on Essays dataset.

Method	EXT		NEU		AGR		CON		OPN	
	ACC	F1	ACC	F1	ACC	F1	ACC	F1	ACC	F1
BERT-MLP	60	58	60.5	56.21	58.8	56.14	59.2	57.03	64.6	58.2
CNN-AdaBoost	61.85	60.8	62.08	61.4	59.92	60.2	64.93	64.2	60.56	60.7
BERT-fusion	61.15	61.03	62.2	59.5	60.8	58.37	59.52	59.45	65.6	61.78
DesPrompt	-	68.2	-	66.2	-	69.3	-	67.3	-	**68.2**
PADO	67	-	65	-	63	-	68	-	**69**	-
EERPD	62.01	63.02	56.00	56.00	64.98	65.01	68.00	68.64	60.73	57.30
Ours	**79.01**	**78.64**	**76.29**	**75.75**	**69.82**	**70.33**	**69.51**	**69.23**	66.35	65.38

Table 6: Comparison results with baseline on MyPersonality dataset.

Method	EXT		NEU		AGR		CON		OPN	
	ACC	F1	ACC	F1	ACC	F1	ACC	F1	ACC	F1
CNN-BiLSTM	72.00	-	58.09	-	64.05	-	56.72	-	78.30	-
DesPrompt	-	59.2	-	66.9	-	**70.5**	-	63.7	-	**83.8**
Attention-BiLSTM	61.30	61.15	64.77	64.34	61.21	60.89	60.92	61.03	75.54	75.17
PADO	54	-	55	-	64	-	60	-	82	-
Ours	**83.74**	**80.71**	**71.12**	**68.13**	**67.64**	68.56	**65.96**	**64.81**	68.62	67.85

through the current personality prompting and augmentation strategy. Future work should therefore design more trait-specific generation constraints and sample filtering strategies for OPN.

5 Discussion

The results of this study demonstrate that LLM-driven personality synthesis can serve as an effective strategy for improving text-based personality detection, particularly under imbalanced trait distributions. Rather than relying on superficial textual perturbations, the proposed framework constructs trait-aware samples that preserve personality-relevant linguistic cues while increasing the diversity of underrepresented profiles, while also providing more effective training signals for downstream classifiers. This finding suggests that the performance bottleneck in personality computing may not only lie in model architecture design, but also in the availability and distributional quality of personality-labeled data.

The observed gains can be attributed to the ability of LLMs to generate semantically coherent and psychologically consistent expressions conditioned on specific personality traits. By enriching the representation space of minority personality categories, the proposed approach helps downstream models learn more robust and discriminative personality-related patterns. Importantly, this augmentation paradigm is model-agnostic and can be combined with different classifiers or representation learning frameworks, making it broadly applicable to various personality computing tasks.

Beyond empirical performance, this study highlights the broader value of LLM-based personality synthesis for scalable and inclusive personality-aware AI. The framework may benefit applications such as personalized recommendation, adaptive dialogue systems, user modeling, educational support, and mental health-oriented human-computer interaction, where reliable personality inference can support more individualized feedback and interaction strategies. Nevertheless, practical deployment should be accompanied by careful consideration of privacy protection, informed consent, demographic fairness, and the risks of personality-based profiling.

6 Conclusion

This paper presents an LLM-driven augmentation framework for text-based Big Five personality detection. By combining personality-controlled prompting, reference anchors, and downstream BERT-MLP evaluation, the method supplements underrepresented personality patterns and improves recognition performance on Essays and MyPersonality. The results show that targeted LLM augmentation can provide more effective training samples than several perturbation-based and distribution-learning augmentation baselines.

Limitations

This study evaluates two English personality datasets and uses binary Big Five labels. The findings may not directly generalize to multilingual data, finer-grained personality scores, or platforms with substantially different writing styles. The quality of generated samples also depends on the selected LLM, prompt design, and filtering strategy, which may affect label consistency and downstream robustness.

References

Weihong Bi, Feifei Kou, Lei Shi, Yawen Li, Haisheng Li, Jinpeng Chen, and Mingying Xu. 2025. Leveraging the dual capabilities of llm: Llm-enhanced text mapping model for personality detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 23487–23495.

Yazid Bounab, Mourad Oussalah, Nabil Arhab, and Salah Bekhouche. 2024. Towards job screening and personality traits estimation from video transcriptions. *Expert Systems with Applications*, 238:122016.

Shresti Chaurasia, Kusum Kumari Bharti, and Atul Gupta. 2024. A multi-model attention based cnn-bilstm model for personality traits prediction based on user behavior on social media. *Knowledge-Based Systems*, 300:112252.

Hongwei Ding, Nana Huang, Yaoxin Wu, and Xiaohui Cui. 2025. Improving imbalanced medical image classification through gan-based data augmentation methods. *Pattern Recognition*, 166:111680.

Zichao Du, Shen Yan, Jiacun Wang, Xiwang Guo, and Shixin Liu. 2026. TPH-SMOTE: A tri-process heuristic oversampling approach integrating SMOTE and whale optimization for imbalanced binary classification. *Expert Systems with Applications*, 306:130952.

Kamal El-Demerdash, Reda A El-Khoribi, Mahmoud A Ismail Shoman, and Sherif Abdou. 2022. Deep learning based fusion strategies for personality prediction. *Egyptian Informatics Journal*, 23(1):47–53.

Linmei Hu, Hongyu He, Duokang Wang, Ziwang Zhao, Yingxia Shao, and Liqiang Nie. 2024. Llm vs small model? large language model based text augmentation enhanced personality detection model. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 18234–18242.

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.

Guangyuan Jiang, Manjie Xu, Song-Chun Zhu, Wenjuan Han, Chi Zhang, and Yixin Zhu. 2023. Evaluating and inducing personality in pre-trained language models. In *Advances in Neural Information Processing Systems*, volume 36, pages 10622–10643.

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.

Hao Lin. 2024. Dlp-personality detection: a text-based personality detection framework with psycholinguistic features and pre-trained features. *Multimedia Tools and Applications*, 83(13):37275–37294.

Hao Lin. 2025. Evaluating, inducing, and verifying personality in large language models with psychological expert knowledge. TechRxiv.

Hao Lin, Leixiao Li, Yongxin Zhao, and Yumeng Zhao. 2026. PersonalityLLM: a zero-shot detection method for big five personality traits from social avatars. Manuscript. Manuscript received July 10, 2025; revised January 14, 2026.

Hao Lin, Chundong Wang, and Qingbo Hao. 2023a. A novel personality detection method based on high-dimensional psycholinguistic features and improved distributed gray wolf optimizer for feature selection. *Information Processing & Management*, 60(2):103217.

Hao Lin, Chundong Wang, and Yongjie Sun. 2023b. Survey on personality recognition based on social media data. *Journal of Frontiers of Computer Science and Technology*, 17(5):1002–1016.

Felipe O López-Pabón and Juan R Orozco-Arroyave. 2022. Automatic personality evaluation from transliterations of youtube vlogs using classical and state-of-the-art word embeddings. *Ingeniería e Investigación*, 42(2).

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.

Philip M. McCarthy and Scott Jarvis. 2010. MTLD, vocd-d, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2):381–392.

Yash Mehta, Samin Fatehi, Amirmohammad Kazameini, Clemens Stachl, Erik Cambria, and Sauleh Eetemadi. 2020a. 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. IEEE.

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

Fatemeh Mohades Deilami, Hossein Sadr, and Morteza Tarkhan. 2022. Contextualized multidimensional personality recognition using combination of deep neural network and ensemble learning: F. mohades deilami et al. *Neural Processing Letters*, 54(5):3811–3828.

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.

Pablo Rosillo-Rodes, Maxi San Miguel, and David Sánchez. 2025. Entropy and type-token ratio in gigaword corpora. *Physical Review Research*, 7(3):033054.

Alireza Taheri, Azadeh Zamanifar, and Amirfarhad Farhadi. 2025. Enhancing aspect-based sentiment analysis using data augmentation based on back-translation. *International Journal of Data Science and Analytics*, 19(3):491–516.

Alessandro Vinciarelli and Gelareh Mohammadi. 2014. A survey of personality computing. *IEEE Transactions on Affective Computing*, 5(3):273–291.

Chundong Wang, Yuhan Yang, Hao Lin, and Siyuan Huang. 2025. Personality detection with psycholinguistic feature driven by confident learning. *Computer Systems & Applications*, 34(7):48–58.

Yukun Wang, Jin Wang, and Xuejie Zhang. 2023. YNU-HPCC at WASSA-2023 shared task 1: Large-scale language model with LoRA fine-tuning for empathy detection and emotion classification. In *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 526–530, Toronto, Canada. Association for Computational Linguistics.

Jason Wei and Kai Zou. 2019. EDA: Easy data augmentation techniques for boosting performance on text classification tasks. 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 6382–6388, Hong Kong, China. Association for Computational Linguistics.

Zhiyuan Wen, Jingsha Cao, Yang Yang, Ruifeng Liu, Jian Shen, Zheng Liu, and Zhancheng Ren. 2023. Desprompt: Personality-descriptive prompt tuning for few-shot personality recognition. *Information Processing & Management*, 60(5):103422.

Hsin-Chang Yang and Zi-Rui Huang. 2019. Mining personality traits from social messages for game recommender systems. *Knowledge-Based Systems*, 165:157–168.

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.

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.

Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. 2017. SeqGAN: Sequence generative adversarial nets with policy gradient. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 2852–2858. AAAI Press.

Yangfu Zhu, Linmei Hu, Nianwen Ning, Wei Zhang, and Bin Wu. 2022. A lexical psycholinguistic knowledge-guided graph neural network for interpretable personality detection. *Knowledge-Based Systems*, 249:108952.

Yangfu Zhu, Meiling Li, Jiachen Tan, and Bin Wu. 2024. Study on text-based personality detection: A review. *Computer Science*, 51(12):209–222.