Adaptive Feature Weighting and Hierarchical Fusion for Lightweight Personality Detection

Anonymous ACL submission

Abstract

In recent years, personality detection has attracted increasing attention in natural language processing. However, existing approaches often suffer from either high computational cost or limited performance, which restricts their practical applicability in real-time scenarios.

In this work, we study lightweight personality detection by integrating traditional NLP features with representations derived from large language models. Specifically, we extract textual representations at different levels of abstraction and propose an adaptive feature weighting strategy to dynamically balance their contributions. Furthermore, we introduce a hierarchical fusion mechanism that aligns handcrafted NLP features with layer-wise LLM representations.

Experiments on multiple public datasets demonstrate that the proposed approach achieves competitive or state-of-the-art performance while maintaining low computational overhead, making it suitable for efficient and real-time personality detection.

1 Introduction

Personality refers to consistent patterns in how people think, feel, and behave, and such patterns are often reflected in individuals' language use (Fang et al., 2023). Consequently, personality detection from text has attracted sustained interest in natural language processing, with applications in personalized human–computer interaction, customer relationship management, and online content moderation (Yang et al., 2023b). In interactive systems, such as conversational agents or service robots, predefined personality profiles are often embedded into system design, making reliable personality detection important for maintaining consistent and adaptive interactions (Saha, 2024). From an NLP perspective, existing approaches to personality detection can be broadly categorized into two paradigms. The first paradigm relies on psycholinguistic or statistical features extracted from text, which are then used to train classifiers based on traditional machine learning or lightweight neural models. These approaches are computationally efficient and interpretable but often struggle to capture deeper semantic representations. The second paradigm leverages large language models (LLMs), frequently using prompting strategies such as chain-of-thought (CoT) reasoning to infer personality traits in zero-shot or few-shot settings. While LLM-based approaches demonstrate strong performance and flexibility, they typically incur substantial computational overhead, making real-time deployment challenging in practical scenarios. This trade-off between efficiency and representational power raises an important question: can lightweight feature-based methods benefit from the representational strength of LLMs without inheriting their high computational cost? In this work, we revisit personality detection from a feature-based NLP perspective and propose a lightweight framework that integrates traditional statistical representations with features derived from large language models. Specifically, we extract multiple textual representations, including TF–IDF features, n-gram statistics, and semantic representations derived from latent semantic analysis. To effectively combine these heterogeneous features, we introduce an adaptive feature weighting mechanism that dynamically adjusts the contribution of different features. In addition, we propose a hierarchical fusion strategy that organizes handcrafted NLP features and representations extracted from different layers of an LLM according to their levels of abstraction. Level-specific fusion is then applied to integrate these representations in a structured manner. The resulting fused features are finally fed into a lightweight multi-layer perceptron clas-

sifier. Experiments on multiple public benchmarks demonstrate that the proposed method achieves state-of-the-art or competitive performance while maintaining significantly lower inference cost than LLM-based chain-of-thought approaches, making it suitable for real-time personality detection scenarios.

The main contributions of this work are summarized as follows:
(1) This work presents a lightweight personality detection framework that combines of NLP features and LLM representations
(2) This work proposes an adaptive feature weighting strategy to dynamically assign features weight.
(3) This work proposes a hierarchical fusion mechanism that aligns handcrafted NLP features with layer-wise LLM representations.

2 Related Work

Text-based personality detection has evolved through four main research streams: psycholinguistic with handcrafted feature engineering, traditional machine learning, Advanced deep learning models, and large language model (LLM)-based approaches. This section reviews representative studies in each stream, discusses their strengths and limitations, and identifies the research gaps addressed by our work.

2.1 Psycholinguistic Feature-based Methods

Early research on text-based personality detection was dominated by handcrafted feature engineering pipelines, Pennebaker et al. (2001) introduced the Linguistic Inquiry and Word Count (LIWC 2001) tool. This a lexicon-driven system that categorizes words into psychologically meaningful dimensions (e.g., emotional expression, cognitive processes) and quantifies their occurrence frequencies. By transforming unstructured text into interpretable numerical features, LIWC enabled empirical analysis of the relationship between linguistic patterns and personality traits (e.g., Openness and Agreeableness), substantially reducing reliance on manual annotation.

Despite its widespread adoption, LIWC has notable limitations. Its static and large Western-centric lexicon limits its generalization to diverse domains such as informal social media text and multilingual corpora. Moreover, LIWC focuses primarily on word-level frequency statistics, overlooking contextual semantics and syntactic structures that are often crucial for capturing subtle personality-related signals.

2.2 Traditional Machine Learning for Personality Prediction

Building on handcrafted feature representations such as LIWC, subsequent studies explored traditional machine learning models for personality detection. A comprehensive survey by Cui and Qi (2017a) systematically analyzed machine learning approaches for MBTI personality prediction. Their work compared classic classifiers, including support vector machines (SVMs), naive Bayes, logistic regression, and shallow neural networks, across multiple textual datasets such as social media posts and self-reported narratives.

An important finding of Cui and Qi (2017a) was that feature quality, rather than model complexity alone, played a dominant role in prediction performance. SVMs consistently outperformed other traditional models due to their effectiveness in handling high-dimensional handcrafted feature spaces. However, this line of work also revealed key limitations: Feature design is subjective and task-specific, leading to poor generalization, and even the strongest traditional models struggled to capture fine-grained semantic nuances essential for personality inference. In addition,Traditional ML models have limited capacity to model long-range dependencies and high-level semantic abstractions.

2.3 Advanced Deep Learning Models

As deep learning gained traction in NLP, researchers explored more expressive architectures to address the limitations of classic machine learning models in personality detection, particularly their limited representational capacity and failure to model inherent trait correlations.

A representative subsequent work is Zhu et al. (2022), who proposed the Contrastive Graph Transformer Network (CGTN) that integrates graph neural networks (GNNs), transformers, and contrastive learning to tackle these challenges. CGTN has the following points: leveraging labeled and unlabeled data via self-supervised learning to mitigate data scarcity, constructing a fully-connected post graph for each user with two types of augmentations (incorporating LIWC psycholinguistic knowledge and post semantics as domain priors), and combining GNN encoders with transformer decoders (equipped with post-to-trait at-

tention) to capture trait correlations and generate traits sequentially.

Experimental results on standard datasets demonstrated that CGTN outperformed state-of-the-art methods at the time. However, the model's integration of graph structure modeling, contrastive learning, and transformer decoding introduces non-trivial computational complexity. This makes it less suitable for lightweight deployment scenarios (e.g., real-time human-computer interaction with service robots).

2.4 LLM-Based Personality Detection Approaches

With the rise of pre-trained language models, LLM-based methods have become the mainstream for personality detection, thanks to their strong automatic feature extraction capabilities. Kerz et al. (Kerz et al., 2022) proposed a hybrid framework that integrates transformer-based models with psycholinguistic text features, demonstrating that fusing LLMs' contextual semantic understanding with interpretable psycholinguistic features can effectively enhance the discriminability of personality traits. This work bridges the gap between handcrafted feature engineering and end-to-end LLM learning for personality detection. Jain et al. (Jain et al., 2022) proposed Personality BERT, a dedicated transformer-based model tailored for textual personality detection, which leverages the contextual feature extraction capability of BERT to capture personality-related semantic information from text data. Labeed and Liang (Labeed and Liang, 2024) conducted a comparative analysis of multi-task learning transformers for emotion classification and emotion intensity prediction in social media, providing empirical insights into the performance of transformer-based multi-task models on these two closely related tasks. Despite their high accuracy, LLM-based methods have two critical limitations that hinder practical deployment: (1) High computational complexity: Full fine-tuning of LLMs requires massive computational resources, making them unsuitable for lightweight deployment scenarios (e.g., edge devices or real-time applications); (2) Black-box interpretability: The feature extraction process of LLMs is opaque, and it is difficult to link the learned features to specific personality dimensions, which limits the model's reliability in psychological applications.

2.5 Summary of Research Gaps

In summary, existing research has progressed from handcrafted features to LLM-based models, yet three gaps remain: (1) traditional feature-based methods lack contextual semantic modeling; (2) advanced deep learning and LLM-based approaches incur high computational costs and are unsuitable for lightweight deployment; (3) few studies integrate handcrafted NLP features with hierarchical representations extracted from LLMs. Our work addresses these gaps by proposing a lightweight framework that adaptively fuses multi-level NLP features with representations from different LLM layers, targeting efficient and real-time personality detection.

3 Method

The task of personality detection is to classify an input text into predefined personality traits (Yang et al., 2023b),
There are two widely used taxonomies for personality trait: Myers-Briggs Type Indicator (MBTI) and the Big Five (OCEAN) model (Randall et al., 2017; Sleep et al., 2021)
The MBTI taxonomy characterizes personality along four binary dimensions: Extraversion/Introversion, Sensing/Intuition, Thinking/Feeling, and Judging/Perceiving.
The Big Five model defines personality as five continuous and independent dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

Formally, given an input text sequence $T = \{t_1, t_2, \ldots, t_n\}$, the objective is to predict a personality label vector $Y = \{y_1, y_2, \ldots, y_k\}$, where k denotes the number of personality dimensions (4 for MBTI and 5 for the Big Five). Each element y_i represents the predicted personality trait along the i-th dimension.

3.1 Overview

Figure 1 presents an overview of the proposed framework. The model integrates heterogeneous features extracted from two sources: basic NLP feature representations and intermediate-layer outputs from a large language model (LLM). The left dashed box illustrates three types of NLP features, including TF-IDF, n-gram, and latent semantic features. The right dashed box shows the representations extracted from different abstraction layers of the LLM. The module labeled "FC" denotes

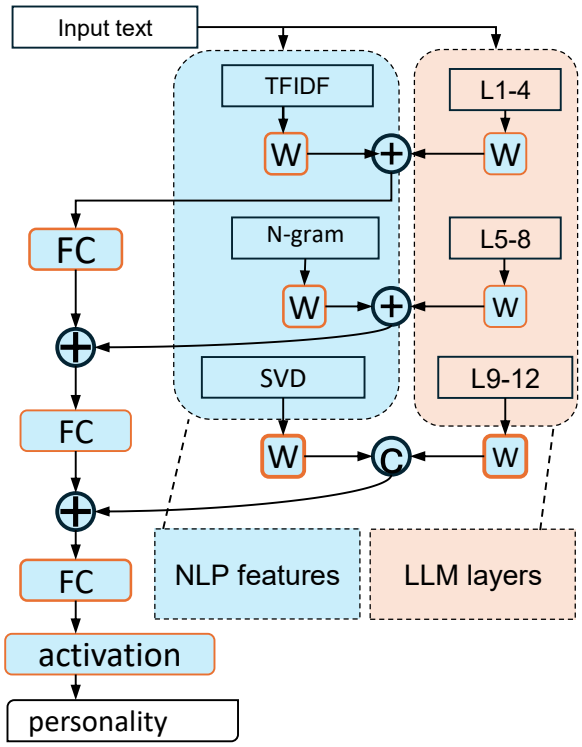

Figure 1: Overview of the proposed framework.

the Fully Connected Layer The module labeled "W" denotes the adaptive feature weighting component proposed in this work (Section 3.2). The circle marked with "+" indicates element-wise addition, while "C" denotes feature concatenation. Through adaptive weighting and hierarchical fusion, the framework produces a compact yet informative representation for downstream MLP-based classifier.

3.2 Feature Engine

Text feature extraction is a core component of personality detection. In this work, we employ three widely adopted and computationally efficient NLP feature extraction methods: TF-IDF, n-gram, and SVD-based Latent Semantic Indexing (LSI).

TF-IDF (Sparck Jones, 1972) quantifies term importance by jointly considering term frequency and inverse document frequency, providing a robust measure of lexical salience. This representation is effective for capturing global word usage patterns associated with personality traits. To model local contextual information, we further extract 2-gram and 3-gram features, which capture short-range word co-occurrence patterns commonly observed in personality-related expressions (Brown et al., 1992).

However, TF-IDF and n-gram features are limited to low-level lexical feature expression and lack the ability to capture deep textual semantics due to insufficient abstraction. For this reason, we introduce Singular Value Decomposition (SVD)-based Latent Semantic Indexing (LSI) to achieve more effective semantic representation.

3.2.1 Adaptive Feature Weighting

Figure 2 illustrates the proposed adaptive feature weighting module. Given an input feature tensor of shape $(batch_size, dim)$, an average pooling operation is applied to aggregate global feature statistics, resulting in a tensor of shape $(1, dim)$. This aggregated representation is then passed through a fully connected layer followed by a activation layer to generate learnable weight coefficients.

The proposed feature-level weighting module is designed to be computationally lightweight, which assigns adaptive importance weights to the extracted textual features with minimal additional overhead.

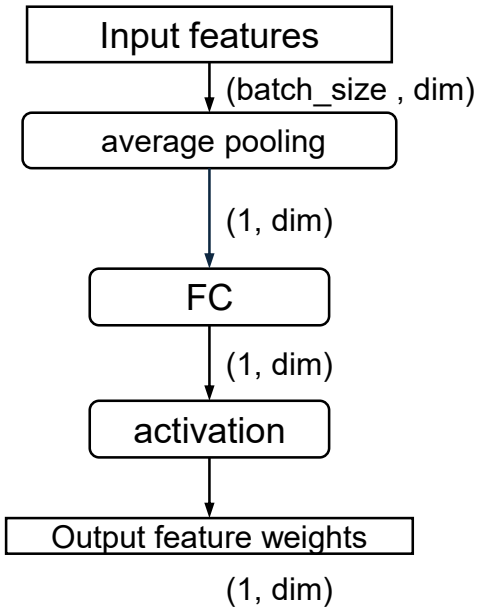

Figure 2: Adaptive feature weighting module.

3.2.2 Hierarchical Feature Fusion

From an abstraction hierarchy perspective, both NLP features and LLM representations exhibit progressively increasing semantic complexity. Specifically, TF-IDF captures surface-level lexical statistics, n-gram features model local contextual patterns, and SVD-based features encode higher-level latent semantics. Similarly, lower layers of

the LLM tend to represent syntactic and lexical information, while higher layers capture more abstract semantic representations.

Based on this observation, we adopt level-specific fusion strategies for features with different abstraction levels. As illustrated in Figure 1, TF-IDF features are fused with LLM Layers 1-4 via element-wise addition, and n-gram features are fused with Layers 5-8 in the same manner. For high-level semantic features, SVD-based representations are concatenated with LLM Layers 9-12 to preserve complementary information. This hierarchical fusion strategy enables effective integration of heterogeneous features while maintaining computational efficiency.

4 Experiments

4.1 Datasets

We evaluate the proposed method on three public datasets: Essays (Pennebaker and King, 1999), Kaggle[1], and Pandora (Gjurković et al., 2021)[2].

The Essays dataset contains 2,467 English essays annotated with the Big Five personality dimensions. The Kaggle dataset consists of 8,675 textual samples labeled with MBTI personality types.

The Pandora dataset contains approximately 12,000 users with multimodal posts, and each user contributing a large number of posts collected from multiple social media platforms, where MBTI personality types are annotated at the user level. Since our work focuses on text-based personality detection, we extract and preprocess the textual modality only. After noise reduction (e.g., removing usernames and special symbols), approximately 9,000 textual samples are retained. The extracted text is processed using the same preprocessing pipeline as the Essays and Kaggle datasets, including special character removal, URL masking, lowercasing and stemming. All datasets are randomly split into training, validation, and test sets with a ratio of 8:1:1 to ensure consistent evaluation.

4.2 Experimental Setup

All experiments are conducted on a single NVIDIA L40 GPU (46 GB memory) under a Linux environment (Ubuntu 24.04.3 LTS), with CUDA 12.8 and NVIDIA driver version 570.144. The software stack includes Python 3.12, PyTorch 2.8 (CUDA 12.8), Hugging Face Transformers 4.56.1, scikit-learn 1.7.2, and NumPy 2.3.3.

For layer-wise LLM representations, we employ the pre-trained sentence embedding model `paraphrase-mpnet-base-v2` from the Sentence-Transformers library, which is loaded via the Hugging Face Transformers interface. The embedding model is fine-tuned jointly with the downstream classifier.

The batch size is set to 16. The learning rate is set to 1×10^{-4} for the LLM-based encoder and 1×10^{-3} for the MLP classifier. Models are trained for up to 80 epochs with early stopping (patience = 5). All experiments are repeated five times using fixed random seeds, and the average results are reported.

4.3 Baselines

We compare the proposed method with representative baselines spanning different paradigms of personality detection:

(1) Traditional machine learning methods. TF-IDF+SVM (Cui and Qi, 2017a) serves as a lightweight baseline based on handcrafted lexical features.

(2) Advanced deep learning models. CGTN (Zhu et al., 2022) (Contrastive Graph Transformer Network) is a representative graph-based model that captures semantic correlations via graph transformers.

(3) LLM-based approaches. This category includes hybrid models that integrate pre-trained language models with task-specific architectures, such as BERT (Devlin et al., 2019), BERT+LSTM (Kerz et al., 2022), and TAE (Hu et al., 2024).

(4) End-to-end standalone large language models. The large language models operating in an end-to-end manner, including GPT-4o (OpenAI, 2025) and DeepSeek-V3 (DeepSeek AI, 2024).

4.4 Main Results

The main results on the Essays, Kaggle, and Pandora datasets are reported in Tables 1, 2, and 3, respectively.

Notably, the proposed method achieves competitive or superior results on each datasets. Compared with traditional machine learning and advanced deep learning baselines (e.g., TF-IDF+SVM and CGTN), our method benefits from

[1] https://www.kaggle.com/datasnaek/mbti-type
[2] https://psy.takelab.fer.hr/datasets/all

incorporating representations derived from pre-trained language models. In contrast to purely LLM-based approaches, our framework additionally integrates handcrafted NLP features through adaptive weighting and hierarchical fusion.

As a result, the proposed method achieves the best performance on the Essays and Kaggle datasets. On the Pandora dataset, our method performs competitively but is slightly inferior to TAE. We attribute this difference primarily to the character of Pandora dataset which is particularly noisy and often contains very short sentences. This represents a drawback of our method

Methods	F1	Accuracy
TF-IDF+SVM	50.20	51.09
CGTN	58.35	59.45
BERT	57.15	58.23
BERT+LSTM	57.64	58.71
TAE	60.55	61.61
DeepSeek-V3	59.04	60.11
GPT-4o	59.53	60.61
NLP+LLM library	**62.24**	**63.33**

Table 1: Results on the essays dataset(Macro-F1 (%))

Methods	F1	Accuracy
TF-IDF+SVM	61.04	62.08
CGTN	67.65	68.74
BERT	63.10	64.22
BERT+LSTM	63.23	64.37
TAE	72.07	73.37
DeepSeek-V3	68.56	69.91
GPT-4o	71.31	72.58
NLP+LLM library	**72.19**	**73.42**

Table 2: Results on the Kaggle dataset(Macro-F1 (%))

Methods	F1	Accuracy
TF-IDF+SVM	54.16	55.13
CGTN	59.89	60.98
BERT	56.52	57.51
BERT+LSTM	57.65	58.69
TAE	**63.05**	**64.13**
DeepSeek-V3	61.20	62.31
GPT-4o	61.42	62.53
NLP+LLM library	62.26	63.34

Table 3: Results on the Pandora dataset(Macro-F1 (%))

4.5 Ablation Study

We conduct an ablation study on the essays dataset to evaluate the contribution of each component. The results are shown in Table 4. In this table:

No NLP features uses only the LLM-derived representations.

NLP (no AW) incorporates NLP features without adaptive weighting.

NLP+AW applies adaptive feature weighting, where NLP features are projected to 768 dimensions and fused with LLM outputs via addition.

NLP+AW+HF further introduces hierarchical fusion using addition or concatenation depending on abstraction levels.

Using only LLM outputs yields an F1 score of 55.56%. Incorporating NLP features improves performance to 57.61%, while adaptive weighting further increases the score to 60.51%. The full model achieves the best performance (62.24%), demonstrating that NLP features are most effective when combined with adaptive weighting and hierarchical fusion.

Methods	F1	Accuracy
No NLP features	55.56	56.65
NLP(noAW)	57.61	58.69
NLP+AW	60.51	60.59
NLP+AW+HF	**62.24**	**63.33**

Table 4: Results of ablation study on the essays dataset(Macro-F1 (%))

4.6 Efficiency Analysis

We compare the proposed method with GPT-4o and DeepSeek-V3 in terms of model scale and inference latency on the Kaggle dataset. As shown in Table 5, our model requires significantly fewer parameters and achieves faster inference.

Overall, the efficiency results demonstrate that the proposed framework achieves a favorable trade-off between performance and computational cost. By combining lightweight NLP features with selectively extracted LLM representations, the model significantly reduces inference latency compared to end-to-end LLM-based approaches, while maintaining competitive or superior accuracy. This balance between effectiveness and efficiency makes the proposed method particularly suitable for real-time and resource-constrained personality detection scenarios.

Metric	Ours	GPT-4o	DeepSeek-V3
Total parameters	**115M**	200B	671B
Inference time (s)	**19**	39	42

Table 5: Comparison of model size and inference latency.

5 Conclusion

Summary and Contributions.

In this work, we revisit text-based personality detection from a feature-oriented NLP perspective and study how lightweight statistical features can be effectively integrated with representations derived from large language models. We propose an adaptive feature weighting mechanism to dynamically balance the contributions of different features, together with a hierarchical feature fusion strategy that aligns NLP features with LLM representations at different levels of abstraction. Extensive experiments on three public benchmarks demonstrate that the proposed framework achieves strong and competitive performance while maintaining high efficiency, making it suitable for real-time and resource-constrained scenarios.

Future Work.

Future work includes exploring the integration of psychological knowledge into the feature extraction process, developing more principled strategies for feature parameter optimization, and investigating more robust fusion mechanisms to better handle noisy and heterogeneous text sources.

6 Limitations

This work focuses on text-based personality detection. Potential risks include misuse of personality predictions for profiling individuals or making unfair decisions.

This work has several limitations. First, the proposed method relies on NLP features and does not explicitly incorporate psychological knowledge due to the limited availability of psychological resources.

Second, the hyper parameters of feature extraction methods (e.g., TF-IDF, n-gram, and SVD) are selected empirically, which may not be optimal across all datasets. Third, the proposed model shows reduced robustness on noisy or short texts, as observed on the Pandora dataset.

Data Ethics. The datasets used in this work are publicly available and consist of user-generated text from online platforms. Such data may contain potentially sensitive or offensive language. However, we only use publicly available data that has been widely adopted in prior personality detection research. No attempt is made to identify individual users.

Acknowledgments

AI assistants were used only for minor language editing and proofreading. The authors take full responsibility for the content of the paper.

References

Eneko Agirre, Oier Lopez de Lacalle, and David Martınez. 2005. Exploring feature spaces with svd and unlabeled data for word sense disambiguation. In *Proceedings of RANLP*, pages 32–38.

Peter F Brown, Vincent J Della Pietra, Peter V Desouza, Jennifer C Lai, and Robert L Mercer. 1992. Class-based n-gram models of natural language. *Computational linguistics*, 18(4):467–480.

Hangcheng Cao, Guowen Xu, Ziyang He, Xuan Zhang, Shengmin Xu, Xinyuan Qian, Anjia Yang, and Jianting Ning. 2025. The lives of others: Snooping on smartphone usage behaviors via attention-enabled multi-channel spatiotemporal information fusion. *IEEE Transactions on Information Forensics and Security*, 20:12680–12692.

Guanming Chen, Lingzhi Shen, Xiaohao Cai, Imran Razzak, and Shoaib Jameel. 2025. Hippd: Brain-inspired hierarchical information processing for personality detection. *arXiv preprint arXiv:2510.09893*.

Brandon Cui and Calvin Qi. 2017a. Survey analysis of machine learning methods for natural language processing for mbti personality type prediction. Final report.

Brandon Cui and Calvin Qi. 2017b. Survey analysis of machine learning methods for natural language processing for mbti personality type prediction. *Final Report Stanford University*.

Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 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.

Qixiang Fang, Anastasia Giachanou, Ayoub Bagheri, Laura Boeschoten, Erik-Jan van Kesteren, Mahdi Shafiee Kamalabad, and Daniel Oberski. 2023. On text-based personality computing: Challenges and future directions. In *Findings of the association for computational linguistics: ACL 2023*, pages 10861–10879.

Adithya V Ganesan, Yash Kumar Lal, August Nilsson, and H Andrew Schwartz. 2023. Systematic evaluation of gpt-3 for zero-shot personality estimation. In *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 390–400.

Matej Gjurković, Vanja M Karan, Iva Vukojević, Mihaela Bošnjak, and Jan Šnajder. 2021. Pandora talks: Personality and demographics on reddit. In *Proceedings of the ninth international workshop on natural language processing for social media*, pages 138–152.

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*, 16, 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.

Xiaoqi Jiao, Yichun Yin, Lifeng Shang, Xin Jiang, Xiao Chen, Linlin Li, Fang Wang, and Qun Liu. 2020. Tinybert: Distilling bert for natural language understanding. In *Findings of the association for computational linguistics: EMNLP 2020*, pages 4163–4174.

Elma Kerz, Yu Qiao, Sourabh Zanwar, and Daniel Wiechmann. 2022. Pushing on personality detection from verbal behavior: A transformer meets text contours of psycholinguistic features. *arXiv preprint arXiv:2204.04629*.

Lingzhen Kong[1], Chuhao Jin[1], Ruihua Song, and Xiting Wang. 2025. Persuasive dialogue models. In *Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2025 Workshops, ADUR, FairPC, GLFM, PM4B and RAFDA, Sydney, NSW, Australia, June 10–13, 2025, Proceedings*, page 298. Springer Nature.

Niraj Kumar and Kannan Srinathan. 2008. Automatic keyphrase extraction from scientific documents using n-gram filtration technique. In *Proceedings of the eighth ACM symposium on Document engineering*, pages 199–208.

Qasid Labeed and Xing Liang. 2024. Multi-task learning transformers: comparative analysis for emotion classification and intensity prediction in social media. In *2024 14th International Conference on Pattern Recognition Systems (ICPRS)*, pages 1–7. IEEE.

Feihu Ma, Xuechen Liang, and Meiling TAo. 2024. Linguashrink: Reducing token overhead with psycholinguistics. *IEEE Access*.

Julina Maharjan, Ruoming Jin, Jianfeng Zhu, and Deric Kenne. 2025. Psychometric evaluation of large language model embeddings for personality trait prediction. *Journal of Medical Internet Research*, 27:e75347.

Maryam Nadhim and Salam Al-augby. 2024. A comprehensive review of machine and deep learning for personality detection. *Journal of Kufa for Mathematics and Computer*, 11(2):121–126.

James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. *Linguistic Inquiry and Word Count: LIWC 2001*. Lawrence Erlbaum Associates, Mahway.

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.

Juan Ramos and 1 others. 2003. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, 1, pages 29–48. New Jersey, USA.

Ken Randall, Mary Isaacson, and Carrie Ciro. 2017. Validity and reliability of the myers-briggs personality type indicator: A systematic review and meta-analysis. *Journal of Best Practices in Health Professions Diversity*, 10(1):1–27.

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.

Sougata Saha. 2024. *Persuasive Dialogue Systems for Social Good*. Ph.D. thesis, State University of New York at Buffalo.

Kourosh Shahnazari and Seyed Moein Ayyoubzadeh. 2025. Who are you behind the screen? implicit mbti and gender detection using artificial intelligence. *arXiv preprint arXiv:2503.09853*.

Chelsea E Sleep, Donald R Lynam, and Joshua D Miller. 2021. A comparison of the validity of very brief measures of the big five/five-factor model of personality. *Assessment*, 28(3):739–758.

R Sowmiya, P Revathi, D Ragunath, P Gokila, T Kalaivani, and 1 others. 2024. Multi-modal llm driven computer interface. In *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, pages 484–489. IEEE.

Karen Sparck Jones. 1972. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1):11–21.

Edward P Tighe, Jennifer C Ureta, Bernard Andrei L Pollo, Charibeth K Cheng, and Remedios de Dios Bulos. 2016. Personality trait classification of essays with the application of feature reduction. In *SAAIP@ IJCAI*, pages 22–28.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, and 1 others. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Tao Yang, Jinghao Deng, Xiaojun Quan, and Qifan Wang. 2023a. Orders are unwanted: dynamic deep graph convolutional network for personality detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 11, pages 13896–13904.

Tao Yang, Tianyuan Shi, Fanqi Wan, Xiaojun Quan, Qifan Wang, Bingzhe Wu, and Jiaxiang Wu. 2023b. Psycot: Psychological questionnaire as powerful chain-of-thought for personality detection. *arXiv preprint arXiv:2310.20256*.

Tao Yang, Feifan Yang, Haolan Ouyang, and Xiaojun Quan. 2021. Psycholinguistic tripartite graph network for personality detection. *arXiv preprint arXiv:2106.04963*.

Yiren Zhao, Chen Zhang, Ming Li, and Hao Wang. 2023. Do emergent abilities exist in quantized large language models: An empirical study. OpenReview. *Preprint*, OpenReview:2306.08977. Submitted to the International Conference on Machine Learning (ICML) 2023 Workshop.

Yangfu Zhu, Linmei Hu, Xinkai Ge, Wanrong Peng, and Bin Wu. 2022. Contrastive graph transformer network for personality detection. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI)*, pages 4559–4565.

Yangfu Zhu, Yue Xia, Meiling Li, Tingting Zhang, and Bin Wu. 2024. Data augmented graph neural networks for personality detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 1, pages 664–672.