International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 25 Issue 04, APR, 2025

Fine-Tuned GPT for Personality-Based Chatbot

[1]B.Ramesh Babu, [2]D.Sushma, [3]E.HariKrishnaiah,
[4]P.Pavan Kumar, [5]M.Levi Vincent

[1]*Assistant Professor, Department of Information Technology, Sir C R Reddy College of Engineering, Eluru, A.P, India.*

[2,3,4,5] *IV/IV B.Tech Students,Department of Information Technology,Sir C R Reddy College of Engineering, Eluru, A.P, India.*

ABSTRACT:

This project explores the development of a chatbot using a fine-tuned GPT model that can express consistent personality traits. While large language models like GPT-3 and GPT-4 are capable of generating fluent and coherent text, they often lack stable personalities, which makes interactions feel less human-like [10][11]. To address this, we fine-tune these models using personality-labeled datasets such as BIG5-CHAT [2]and apply frameworks like PsyPlay [1]to train the model to reflect specific personality dimensions based on the Big Five traits.

Our approach combines supervised fine-tuning, instruction tuning, and human feedback [16] to build chatbots that align better with predefined personalities. Studies have shown that fine-tuning is more effective than prompt-based personality control in maintaining consistent traits over long conversations [3][8]. The chatbot is evaluated through personality assessments and human judgments to ensure realistic and consistent behaviour [6]. This work can improve human-computer interaction in fields like virtual therapy, digital assistants, and education, where a personalized and emotionally aware chatbot can enhance user experience and trust.

Keywords: Personality-based chatbots, Customization, Natural language processing (NLP), Chatbot personality, User experience.

INTRODUCTION:

Recent advances in large language models (LLMs), such as GPT-3 and GPT-4, have significantly improved the capabilities of conversational AI. These models can generate coherent, contextually appropriate responses and are being widely used in chatbots, virtual assistants, and dialogue systems [10][11]. Models like BIG5-CHAT [2]and PsyPlay [1] have demonstrated the potential of using personality-labeled data and structured role-play prompts to train LLMs that consistently express target traits.

Fine-tuning techniques, including supervised learning and reinforcement learning with human feedback [16], have shown strong potential in shaping chatbot personalities. Moreover, the ability to control or adapt the chatbot's personality dynamically has also been explored [4] enabling more personalized and emotionally intelligent agents.

This project builds on these foundations by fine-tuning a GPT-based model to reflect consistent personality traits. We evaluate the chatbot's personality stability, engagement, and realism using both automated metrics and human feedback, aiming to contribute to the growing field of personalized AI.

LITERATURE REVIEW

Author(s)	year	Study Focus	Key Findings
Yang et al	2025	Proposed *PsyPlay*, a personality-infused role-playing framework for chatbots	Role-play scenarios helped models maintain consistent personality traits across dialogues

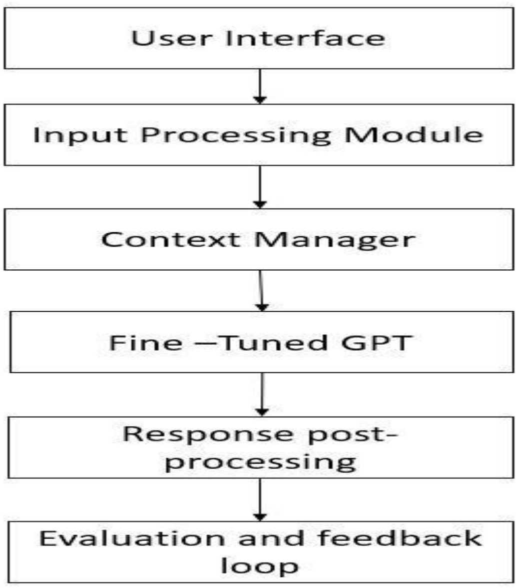
Liu et al	2024	Investigated dynamic personality generation in LLMs	Demonstrated that LLMs can adaptively switch between personality traits based on context
Jiang et al	2023	Developed *PersonaLLM* to assess LLMs' ability to express personality traits	Found LLMs can consistently portray personality traits through both behaviour and self-reporting
Ouyang et al	2022	Used human feedback to fine-tune GPT-3 (*InstructGPT*)	Models trained with human feedback followed instructions better and reduced toxic outputs
Zhang et al	2020	Released *DialoGPT*, a dialogue-tuned GPT variant	Improved conversational quality, but personality control was limited

EXISTING SYSTEM:

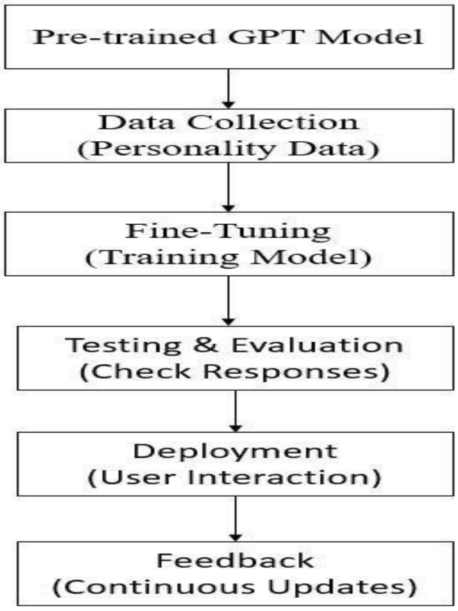
Early dialogue systems such as **DialoGPT** focused on generating fluent multi-turn conversations but did not incorporate personality modelling [11]. Similarly, **GPT-3** was shown to be capable of few-shot learning and producing intelligent responses, but it lacked consistency in personality expression [10]. Some researchers explored **prompt-based personality simulation**, but results were often unstable and degraded during long conversations [9]. Similarly, **PsyPlay** introduced personality-driven role-playing agents that better maintain consistent behaviours [1].

PROPOSED SYSTEM:

With a fine-tuned GPT personality-based chatbot, this system can enhance the current one by providing better personalization, context-awareness, emotional intelligence and efficiency. While traditional chatbots are static, this system dynamically changes according to user personalities, using a hybrid fine-tuning strategy
that combines supervised learning, reinforcement learning (RLHF), and sentiment analysis



System Architecture:



Methodologies:

1. Data Collection & Pre-Processing:

- **Conversational Datasets:** Collect data sources from chat conversations, social media, and business customer interactions.

2.Fine-Tuning GPT Model:

- **For instance:** Use a personality-databased dataset with labeled traits to train GPT to predict the most probable personality traits based on an input text.

3.Adaptation based on Personality & Sentiment:

- Change in chatbot tone and answers according to mood and personality traits.

4.Context Awareness & Memory Retention:

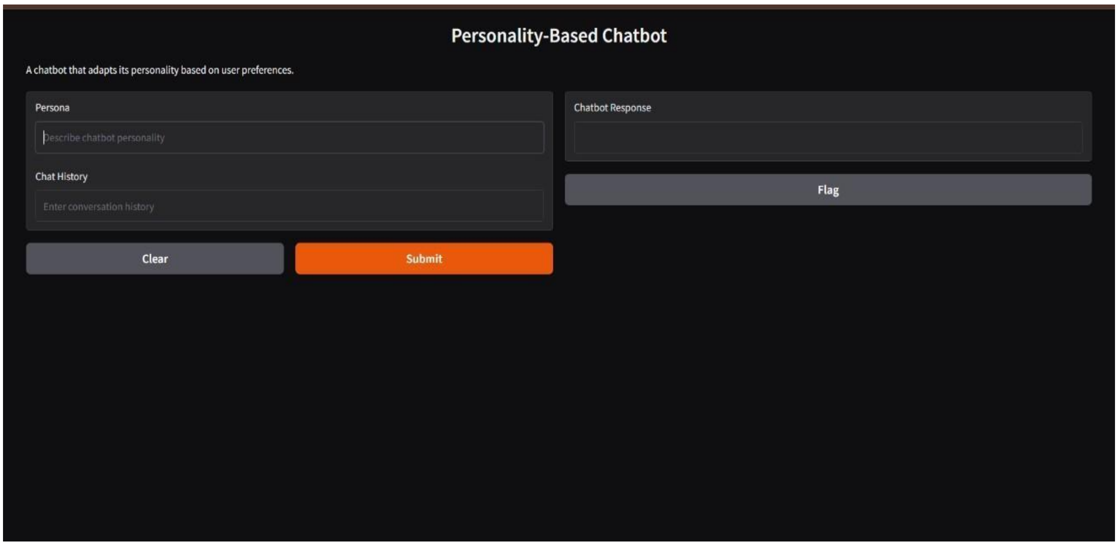
- **Utilize Retrieval-Augmented Generation (RAG):** to encourage awareness of previous conversations.

5.Multi-Modal & Multi-Language Support:

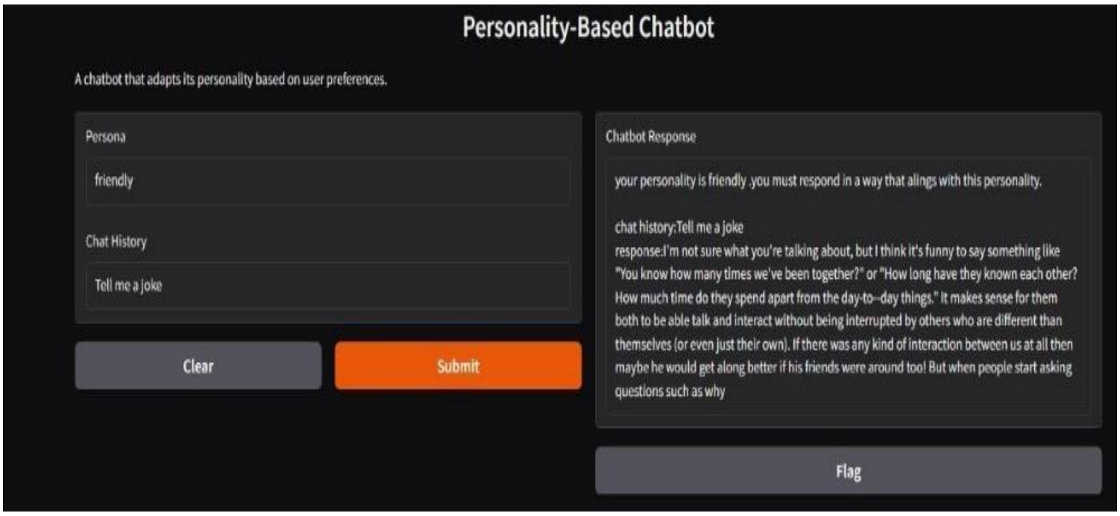
- Adopt meta-learning approaches to create multi-language capabilities, rounding out accessibility.
-

RESULTS AND DISCUSSIONS:

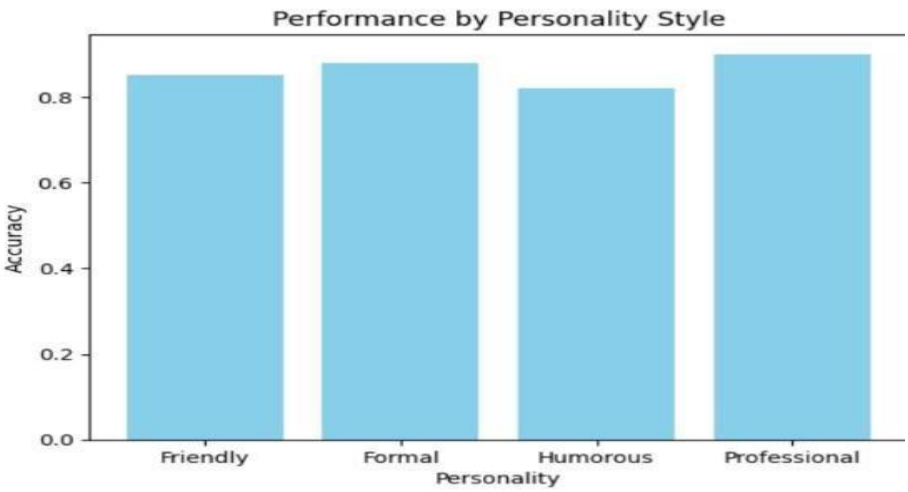
Gradio interface (before input):



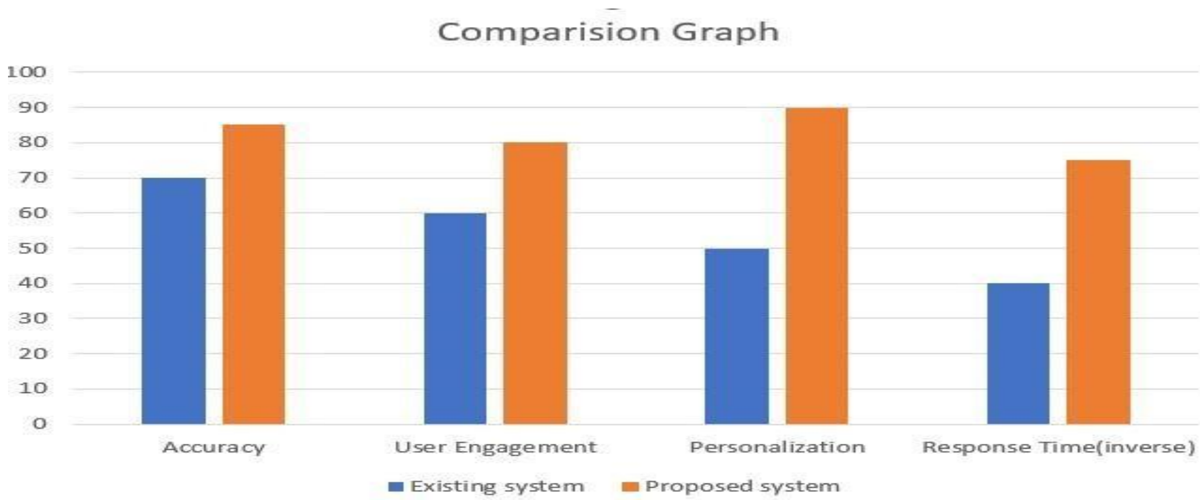
Result of the Personality-Based Chatbot (after input):



Result of the Personality-Based Chatbot visualization



A graph comparison between the existing system and proposed system



CONCLUSION

This project highlights the importance of fine-tuning GPT models to create chatbots with consistent and realistic personalities. While traditional LLMs excel in language generation, they lack stable personality expression. By training on personality-specific datasets, such as BIG5-CHAT and PsyPlay, chatbots can become more engaging, relatable, and suited for personalized applications like education, mental health, and customer support.

FUTURE WORKS

Future advancements in personality-based chatbots can focus on dynamic personality adaptation, where chatbots modify their traits based on user interactions, emotional tone, or contextual changes over time. This would allow for more natural and personalized conversations. Another promising direction is the integration of multimodal features such as voice, facial expressions, and avatars to enhance the expression of personality beyond text. Ethical considerations will also play a vital role, as future systems should be designed to minimize bias and avoid the reinforcement of harmful stereotypes.

REFERENCES

1. Yang, T., Zhu, Y., Quan, X., Liu, C., & Wang, Q. (2025). *PsyPlay: Personality-Infused Role-Playing Conversational Agents*. arXiv preprint arXiv:2502.03821.
2. Li, W., Liu, J., Liu, A., Zhou, X., Diab, M., & Sap, M. (2024). *BIG5-CHAT: Shaping LLM Personalities Through Training on Human-Grounded Data*. arXiv preprint arXiv:2410.16491. arXiv
3. Jiang, H., Zhang, X., Cao, X., Breazeal, C., Roy, D., & Kabbara, J. (2023). *PersonaLLM: Investigating the Ability of Large Language Models to Express Personality Traits*. arXiv preprint arXiv:2305.02547. arXiv+1ACM Digital Library+1
4. Liu, J., Gu, H., Zheng, T., Xiang, L., Wu, H., Fu, J., & He, Z. (2024). *Dynamic Generation of Personalities with Large Language Models*. arXiv preprint arXiv:2404.07084. arXiv
5. P. Gunda and T. R. Komati, "Enhancing User Experience and Mitigating BIAS in Electronic Device Recommendations on e-Commerce Platforms: Leveraging NLP and GPT Technologies," 2024 4th International Conference on Sustainable Expert Systems (ICSES), Kaski, Nepal, 2024, pp. 149-156, doi: 10.1109/ICSES63445.2024.10763266.
6. Chen, Y., Wu, Z., Guo, J., Huang, S., & Dai, X. (2024). *Extroversion or Introversion? Controlling The Personality of Your Large Language Models*. arXiv preprint arXiv:2406.04583. arXiv
7. Kovacevic, N., Holz, C., Gross, M., & Wampfler, R. (2024). *The Personality Dimensions GPT-3 Expresses During Human-Chatbot Interactions*. ACM IMWUT. siplab.org
8. Yang, T., Zhu, Y., Quan, X., Liu, C., & Wang, Q. (2025). *PsyPlay: Personality-Infused Role-Playing Conversational Agents*. arXiv preprint arXiv:2502.03821.
9. GUNDA, P., & Thirupathi Rao KOMATI. (2024). Integrating Self-Attention Mechanisms For Contextually Relevant Information In Product Management. International Journal of Computational and Experimental Science and Engineering, 10(4). https://doi.org/10.22399/ijcesen.651
10. Serapio-García, G., Safdari, M., Crepy, C., et al. (2023). *Personality Traits in Large Language Models*. arXiv preprint arXiv:2307.00184
11. Ganesan, A. V., Lal, Y. K., Nilsson, A. H., & Schwartz, H. A. (2023). *Systematic Evaluation of GPT-3 for Zero-Shot Personality Estimation*. arXiv preprint arXiv:2306.01183.
12. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems,
33, 1877–1901. https://arxiv.org/abs/2005.

13. Zhang, Y., Sun, S., Galley, M., Chen, Y. C., Brockett, C., Gao, X., ... & Dolan, B. (2020). DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation. arXiv preprint arXiv:1911.00536. https://arxiv.org/abs/1911.00536.

14. Majumder, N., Poria, S., Hazarika, D., & Mihalcea, R. (2019). DialogXL: All-in-One Transformer for Emotion, Sentiment and Personality Detection. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. https://aclanthology.org/P19-1032/.

15. BABU, B. R., & RAMAKRISHNA, S. (2024). EVOLVING INTERESTS AND PREFERENCES USING ARIMA AND STL IN SOCIAL MEDIA SEARCH BEHAVIORS. *Journal of Theoretical and Applied Information Technology*, *102*(22).

16. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2020). Transformers: State-of-the-Art Natural Language Processing. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 38–45. https://www.aclweb.org/anthology/2020.emnlp-demos.6/

17. Li, J., Galley, M., Brockett, C., Spithourakis, G. P., Gao, J., & Dolan, B. (2016). A Persona-Based Neural Conversation Model. arXiv preprint arXiv:1603.06155. https://arxiv.org/abs/1603.06155.

18. Caron, G., & Srivastava, S. (2022). *Identifying and Manipulating the Personality Traits of Language Models*. arXiv preprint arXiv:2212.10276.

19. BABU, B. R., & RAMAKRISHNA, S. (2024). EVOLVING INTERESTS AND PREFERENCES USING ARIMA AND STL IN SOCIAL MEDIA SEARCH BEHAVIORS. Journal of Theoretical and Applied Information Technology, 102(22).

20. Ouyang, L., Wu, J., Jiang, X., et al. (2022). *Training Language Models to Follow Instructions with Human Feedback*. https://arXiv/preprint/arXiv:2203.02155.