

# Empaths at WASSA 2024 Empathy and Personality Shared Task: Turn-Level Empathy Prediction Using Psychological Indicators

Shaz Furniturewala<sup>1\*</sup>, Kokil Jaidka<sup>2</sup>,

<sup>1</sup>Birla Institute of Technology and Science, Pilani,

<sup>2</sup>NUS Center for Trusted Internet and Community, National University of Singapore,

## Abstract

For the WASSA 2024 Empathy and Personality Prediction Shared Task, we propose a novel turn-level empathy detection method that decomposes empathy into six psychological indicators: Emotional Language, Perspective-Taking, Sympathy and Compassion, Extroversion, Openness, and Agreeableness. A pipeline of text enrichment using a Large Language Model (LLM) followed by DeBERTa fine-tuning demonstrates a significant improvement in the Pearson Correlation Coefficient and F1 scores for empathy detection, highlighting the effectiveness of our approach. Our system officially ranked 7th at the CONV-turn track.

## 1 Introduction

Empathy, a critical construct in human social interaction, involves perceiving, understanding, and resonating with the emotional states and perspectives of others. This construct is essential in domains such as mental health support, customer service, and human-computer interaction (Paiva et al., 2014). Empathy comprises both cognitive and affective dimensions (Batson, 2009; Singer and Lamm, 2009), with the cognitive dimension involving the intellectual understanding of another’s psychological state and the affective dimension involving the emotional experience of another’s feelings.

Empathy detection in natural language processing (NLP) focuses on identifying and quantifying empathetic expressions in text. Accurate empathy detection can enhance the performance of automated systems in responding to human emotions appropriately (Shum et al., 2018). Applications include therapeutic conversational agents, customer service bots, and social robotics. Traditional empathy detection methods rely on

lexical and syntactic analysis (Provence and Forcehimes, 2024), using features such as sentiment polarity, emotion lexicons, and dialogue acts. These methods often fail to capture the nuanced and context-dependent nature of empathy, rooted in deeper psychological constructs. Effective empathy detection requires a sophisticated analytical framework to interpret underlying psychological indicators.

Our study employs GPT-4o (OpenAI et al., 2024) to evaluate six psychological indicators for each utterance in our dataset. GPT-4o’s advanced language understanding and generation capabilities allow it to assess and articulate the presence of these indicators, providing ratings and explanatory sentences. These enriched inputs are used to train a DeBERTa classifier (He et al., 2021), known for its superior performance in NLP tasks due to its enhanced attention mechanisms and optimized representation learning. Our empirical analysis demonstrates that incorporating the psychological indicators significantly enhances the performance of the empathy detection models, as evidenced by improvements in Pearson correlation, F1 scores, and accuracy metrics.

By integrating psychological indicators and leveraging advanced NLP models, our work offers a new pipeline for multi-task learning that relies on the cognitive underpinnings of human behavior. This contributes to developing more contextually aware and empathetic conversational agents, improving human-computer interaction, and enabling more emotionally intelligent automated systems.

## 2 Related Work

Simple approaches to empathy detection have employed rule-based systems and manually crafted features, leveraging predefined empathy-related keywords and patterns to identify empathetic ex-

\*Work done during internship at NUS Center for Trusted Internet and Community

Speaker	Utterance	Empathy
Person 1	What did you think about this article	0.6667
Person 2	It's definitely really sad to read, considering everything they're all going through. What did you think?	4.3333
Person 1	I think it's super sad... they seem to never catch a break, always struggling.	4.6667
Person 2	I can't imagine just living in an area that is constantly being ravaged by hurricanes or earthquakes. I take my location for granted.	4.6667

Table 1: A snippet of the dataset.

pressions (Provence and Forcehimes, 2024). With the advent of machine learning, statistical models such as support vector machines (SVMs) and random forests were utilized, which leveraged a broader set of features, including syntactic structures, word embeddings, and discourse markers (Chen et al., 2020; Mathur et al., 2021). However, these approaches are limited by their reliance on predefined patterns and surface-level features, which may not generalize well across different contexts and fail to capture the complexity and contextual nature of empathetic language.

Recent advancements in deep learning have further propelled the field, with neural network architectures such as recurrent neural networks (RNNs) (Tavabi et al., 2019), LSTMs (Tan et al., 2019), and transformers (Guda et al., 2021) demonstrating significant improvements in capturing the contextual dependencies and semantic richness of empathetic language. Transformer-based models, particularly BERT (Devlin et al., 2019) and its variants, have shown remarkable performance in various NLP tasks, including empathy detection. These models are expected to provide a more nuanced understanding of empathetic expressions by leveraging self-attention mechanisms that model long-range dependencies and contextual relationships within the text. Nevertheless, these models would still need sufficient context to interpret empathetic behavior and, by themselves, do not offer a way to consider the specific psychological constructs that underpin empathetic behavior.

To address this gap, our approach for empathy detection focused on first enriching the data with more psychological indicators and then improving upon the design of the current best-performing model. Our objective was to focus on the psychological indicators underpinning empathy, leveraging our prior work in modeling cognitive appraisals of happiness (Liu and Jaidka, 2023) and purchase

behavior (Yeo and Jaidka, 2023; Yeo et al., Accepted) by translating a text classification paradigm into a multi-task classification problem. Our prior work successfully demonstrated the effectiveness of using psychological constructs to enhance predictive models, providing a foundation for our current approach. We believed that applying a similar framework to empathy detection would yield robust and interpretable models capable of capturing the nuanced and multifaceted nature of empathetic expressions in language.

### 3 Method

We decomposed the concept of "Empathy" into theory-inspired fundamental components of empathetic behavior (Batson, 2009; Singer and Lamm, 2009), focusing on six distinct psychological indicators:

- Emotional Language: Represented by the use of emotion-laden words (e.g., "sad," "happy," "worried") and descriptions of feelings or emotions, both personal and those of others.
- Perspective-Taking: Indicated by statements that show an understanding of another person's point of view.
- Sympathy and Compassion: Demonstrated by expressions of concern for another person's well-being.
- Extroversion: Reflected by signs of sociability, such as mentions of interactions with others, excitement about social events, or enjoyment of group activities.
- Openness: Indicated by signs of creativity, intellectual curiosity, or unconventional thinking, such as discussing diverse topics, exploring different perspectives, or expressing interest in novel ideas.
- Agreeableness: Shown by kindness, altruism, or cooperation in the text, for example, ex-

pressions of concern for others’ well-being, willingness to help, or avoiding conflict.

By enriching our data with information about these indicators, we aimed to provide a more comprehensive and interpretable framework for empathy detection. However, the challenge remains in accurately operationalizing these psychological constructs and ensuring that models can reliably differentiate and interpret these indicators within varied contexts and expressions of empathy. To do so, we closely relied on how these concepts are defined, worded, and measured in surveys to human participants.

### 3.1 Dataset

We use the dataset created by (Omitaomu et al., 2022) and provided as part of the WASSA 2024 Shared Task 2 (Giorgi et al., 2024; Barriere et al., 2023), which is an empathetic conversation dataset consisting of conversations in response to news articles. It consists of 500 conversations between AMT workers reacting to 100 articles about negative events from (Buechel et al., 2018). Each conversation is greater than 15 turns. This conversation data has been third-person annotated at the turn-level on a range of 0-5 for the level of empathy displayed in the text. Scores in the dataset, however, are also fractional, presumably due to averaging among reviewers. For training, we round the scores to the nearest integer; however, while computing the Pearson Correlation Coefficient, we use the original scores. We divide the 11059 utterances into a training set of 8294 and a test set of 2765. Table 1 contains a snippet of the Dataset. The results demonstrated in this paper are unofficial, based on the test dataset we created using a subset of the training data. The official result is also provided in Section 4.

### 3.2 Enrichment

For each utterance in the dataset, we used GPT-4o to detect the level of the psychological indicators described previously, rating them as Low, Medium, or High. Additionally, GPT-4o provided a sentence explaining the rating, highlighting words or phrases contributing to a high or low rating. Table 4 in Appendix A presents the generated ratings and explanations for a sample utterance from the dataset. The prompt used to generate them is also provided in Fig 1 in Appendix A. These ratings and explanations were used as additional context

and concatenated to the original utterance. Subsequently, a DeBERTa V3 classifier was trained on this new set of inputs.

### 3.3 Models and baselines

We tested two classification models: DeBERTa-v3-Large finetuned and GPT-4o for zero-shot classification. For each model, we tested its performance on just the utterance and the utterance combined with additional context provided by the six psychological indicators. For the DeBERTa classifier, we concatenated the rating and explanation for each indicator to the original utterance, separated by [SEP] tokens. For GPT-4o, we crafted an instructional prompt, providing all the information in bullet points.

## 4 Results and Discussion

In Table 3, we demonstrate the Pearson Correlation Coefficient of the psychological indicator scores with the annotated empathy ratings. We converted the levels predicted by GPT-4o (Low, Medium, High) to integers (-1, 0, 1) and computed the coefficients of each feature with the empathy ratings. Emotional Language exhibited the highest positive correlation (0.481\*), underscoring the significance of emotion-laden words in conveying empathy. Sympathy and Compassion also showed a strong positive correlation (0.437\*), validating the role of compassionate expressions in empathetic communication. Perspective-Taking had a moderate positive correlation (0.186\*), suggesting that understanding another person’s point of view contributes to empathy but is less influential than direct emotional expressions. Interestingly, Extroversion had a negative correlation (-0.152\*), implying that sociability may not align with empathetic responses in these conversations. Openness showed a very weak correlation (0.010), indicating minimal impact on empathy perception, while Agreeableness had a modest positive correlation (0.120\*), reflecting a mild association with empathetic responses through expressions of kindness and cooperation. Table 2 reports the performance of the four classification methods in terms of their Pearson Correlation Coefficient, Accuracy, and F1 Score at empathy detection. While the coefficient is computed with the fractional empathy labels, the accuracy and F1 score are computed after rounding those labels to the nearest integer. This rounding is done because 6-label classification (0 to 5 in increments

Model	Input	Pearson R	F1 (Rounded)	Accuracy (Rounded)
GPT-4o	Utterance Only	0.38	0.24	0.29
GPT-4o	Utterance + Indicators	0.41	0.20	0.26
DeBERTa	Utterance Only	0.65	0.32	0.52
DeBERTa	Utterance + Indicators	<b>0.68</b>	<b>0.35</b>	<b>0.55</b>

Table 2: The Pearson Correlation Coefficient and F1 scores for each of the four classification methods.

Psych. Indicator	Pearson R
Emotional Language	0.481*
Perspective-Taking	0.186*
Sympathy and Compassion	0.437*
Extroversion	-0.152*
Openness	0.010
Agreeableness	0.120*

Table 3: The Pearson Correlation Coefficient of the GPT-4o scores (converted to integers) for each psychological indicator with the annotated Empathy Rating.

of 1) is much simpler than 16-label classification (0 to 5 in increments of 0.33). Further, we found that the rounded labels have a correlation coefficient of 0.96 with the original labels, demonstrating minimal knowledge loss. There is a large gulf between the performance of GPT-4o and the trained DeBERTa classifier for both input formats, reiterating the necessity of relying on attention mechanisms for interpreting implicit concepts like empathy from dialogic data.

The results demonstrate the efficacy of incorporating psychological indicators into empathy detection models. The baseline DeBERTa model trained on utterances alone achieved a Pearson correlation of 0.65, an F1 score of 0.32, and an accuracy of 0.52. When augmented with the additional context from the six psychological indicators, the model’s performance improved, achieving a Pearson correlation of 0.68, an F1 score of 0.35, and an accuracy of 0.55. This indicates that the enriched input provides more comprehensive information, allowing the model to understand and predict empathy levels.

In contrast, the zero-shot classification using GPT-4o showed more modest improvements. The model’s Pearson correlation increased from 0.38 to 0.41 when augmented with the psychological indicators, although the F1 score and accuracy slightly decreased. This suggests that while the additional

context benefits GPT-4o, the model may require further fine-tuning to leverage the enriched input fully.

It is to be noted that these results are on the test set generated as a subset of the provided training data. We were unable to conduct these experiments on the official test data as the labels for that were not available to us. The Pearson correlation coefficient for the Utterance + Indicators DeBERTa classifier on the official test dataset is 0.534.

## 5 Discussion and Conclusion

Our method for empathy detection relied on enriching the available data with more psychological indicators that could help support the ultimate Empathy label. We show that our approach boosts performance and provides interpretable AI insights, which can be crucial for applications requiring transparency and trust.

The significant improvement observed with the DeBERTa model underscores the importance of considering psychological components in empathy detection. However, it is the role and performance of GPT-4o, the leading LLM, that provides the most interesting insights. On the one hand, the nuanced explanations provided by GPT-4o offer valuable context that enhances the model’s ability to detect empathy. On the other hand, its relatively poor performance in empathy prediction indicates that LLMs cannot yet effectively extract and understand all the underlying information in a dialogic exchange, even in a few-shot manner, highlighting the need for further exploration of reasoning-based approaches.

## 6 Limitations

We acknowledge one particular limitation of our work. The reliance on GPT-4o for both enriching the data and attempting to label it may lead to concept drift, where the interpretation of the labels relies heavily on prompt sensitivity and ad-

herence, and ultimately digresses from the original definition. To address this, our future work will involve obtaining additional expert annotations and conducting a thorough inspection of GPT-4o’s reasoning. This will ensure that our data enrichment is valid, nuanced, and accurately reflects human intent.

## References

- Valentin Barriere, João Sedoc, Shabnam Tafreshi, and Salvatore Giorgi. 2023. Findings of wassa 2023 shared task on empathy, emotion and personality detection in conversation and reactions to news articles. In *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 511–525.
- C. Batson. 2009. *These things called empathy: Eight related but distinct phenomena*. *The Social Neuroscience of Empathy*.
- Sven Buechel, Anneke Buffone, Barry Slaff, Lyle Ungar, and João Sedoc. 2018. *Modeling empathy and distress in reaction to news stories*. *Preprint*, arXiv:1808.10399.
- Zhuohao Chen, James Gibson, Ming-Chang Chiu, Qiaohong Hu, Tara K Knight, Daniella Meeker, James A Tulsy, Kathryn I Pollak, and Shrikanth Narayanan. 2020. Automated empathy detection for oncology encounters. In *2020 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 1–8. IEEE.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. *Bert: Pre-training of deep bidirectional transformers for language understanding*. *Preprint*, arXiv:1810.04805.
- Salvatore Giorgi, João Sedoc, Valentin Barriere, and Shabnam Tafreshi. 2024. Findings of wassa 2024 shared task on empathy and personality detection in interactions. In *Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*.
- Bhanu Prakash Reddy Guda, Aparna Garimella, and Niyati Chhaya. 2021. Empathbert: A bert-based framework for demographic-aware empathy prediction. *arXiv preprint arXiv:2102.00272*.
- Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021. *Deberta: Decoding-enhanced bert with disentangled attention*. *Preprint*, arXiv:2006.03654.
- Xuan Liu and Kokil Jaidka. 2023. I am psyam: Modeling happiness with cognitive appraisal dimensions. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1192–1210.
- Leena Mathur, Micol Spitale, Hao Xi, Jieyun Li, and Maja J Matarić. 2021. Modeling user empathy elicited by a robot storyteller. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–8. IEEE.
- Damilola Omitaomu, Shabnam Tafreshi, Tingting Liu, Sven Buechel, Chris Callison-Burch, Johannes Eichstaedt, Lyle Ungar, and João Sedoc. 2022. *Empathic conversations: A multi-level dataset of contextualized conversations*. *Preprint*, arXiv:2205.12698.
- OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altmenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O’Keefe, Jakub Pachocki, Alex

Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. 2024. *Gpt-4 technical report. Preprint*, arXiv:2303.08774.

Ana Paiva, Iolanda Leite, and Tiago Ramos Ribeiro. 2014. *Emotion modelling for social robots*.

Scott Provence and Alyssa Forcehimes. 2024. *Algorithms for empathy: Using machine learning to categorize common empathetic traits across professional and peer-based conversations. Cureus*, 16.

Heung-Yeung Shum, Xiaodong He, and Di Li. 2018. *From eliza to xiaoice: Challenges and opportunities with social chatbots. Preprint*, arXiv:1801.01957.

Tania Singer and Claus Lamm. 2009. *The social neuroscience of empathy. Annals of the New York Academy of Sciences*, 1156.

Zhi-Xuan Tan, Arushi Goel, Thanh-Son Nguyen, and Desmond C Ong. 2019. A multimodal lstm for predicting listener empathetic responses over time. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pages 1–4. IEEE.

Leili Tavabi, Kalin Stefanov, Setareh Nasihati Gilani, David Traum, and Mohammad Soleymani. 2019. Multimodal learning for identifying opportunities for empathetic responses. In *2019 International Conference on Multimodal Interaction*, pages 95–104.

Gerard Yeo, Shaz Furniturewala, and Kokil Jaidka. Accepted. Beyond text: Leveraging multi-task learning

and cognitive appraisal theory for post-purchase intention analysis. In *Findings of the Association for Computational Linguistics: ACL (2024)*.

Gerard Yeo and Kokil Jaidka. 2023. The peace-reviews dataset: Modeling cognitive appraisals in emotion text analysis. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 2822–2840.

## A Appendix

Table 4 contains an instance of the dataset and the ratings and explanations provided by GPT-4o on the six psychological indicators.

Figure 1 displays the prompt used to generate the ratings and explanations from GPT-4o.

I will provide you with a text. You have to rate the text as LOW, MEDIUM, or HIGH for each of the following five psychological dimensions and explain your score in a single sentence.

Psychological Dimensions:

- Emotional Language: HIGH emotional language contains emotion-laden words (e.g., "sad," "happy," "worried") and descriptions of feelings or emotions, both their own and those of others. LOW emotional language would NOT have this.

- Perspective-Taking: HIGH perspective taking statements indicate an understanding of another person's point of view. LOW perspective taking statements DO NOT.

- Sympathy and Compassion: HIGH sympathy and compassion is represented by statements showing concern for another person's well-being. LOW sympathy and compassion is represented by statements that DO NOT.

- Extroversion: HIGH extroversion is indicated by signs of sociability, such as mentioning interactions with others, excitement about social events, or enjoyment of group activities.

- Openness: HIGH extroversion is indicated by signs of creativity, intellectual curiosity, or unconventional thinking. This might include discussing diverse topics, exploring different perspectives, or expressing interest in novel ideas. LOW openness would not include these.

- Agreeableness: HIGH agreeableness is indicated by kindness, altruism, or cooperation in the text. For example, expressions of concern for others' well-being, willingness to help, or avoiding conflict. LOW agreeableness is not indicated by these.

Report the result in JSON format with the text, the ratings, and the single sentence explanation for each of the five psychological dimensions. The JSON string should have the keys 'Text', 'Scores', 'Explanations'.

Be very strict while giving ratings and don't give HIGH ratings unnecessarily. Also, highlight factors that contributed both positively and negatively to your rating in the single sentence explanation.

Here is the text: [TEXT]

Figure 1: Prompt given to GPT-4o.

Utterance	You've nailed it. Bringing children into the mix definitely amplifies the desire for privacy and protection. It's all about creating a safe and stable environment for them.
Emotional Language	<b>Medium.</b> The text contains words like 'desire' and implies concern for children's well-being, but lacks varied emotion-laden language.
Perspective-Taking	<b>High.</b> It acknowledges the point of view of someone concerned about privacy and protection for children, representing strong perspective-taking.
Sympathy and Compassion	<b>High.</b> The concern for creating a safe and stable environment for children shows significant sympathy and compassion towards their well-being.
Extroversion	<b>Low.</b> There is no mention of social interactions, excitement about social events, or enjoyment of group activities, indicating low extroversion.
Openness	<b>Low.</b> The text does not display signs of creativity, intellectual curiosity, or unconventional thinking, which reflect low openness.
Agreeableness	<b>Medium.</b> It shows some concern for others but does not explicitly demonstrate strong kindness, altruism, or cooperative behavior.

Table 4: Ratings and explanations provided by GPT-4o for a sample utterance from the dataset.