International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 6, Nov-Dec 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Enhancing Empathy and Communication for Individuals with Autism Spectrum Disorder in Virtual Environments through AI-Generated Emotional Cues

Jeeho Lee*, Eunsung You**
*Los Altos High School, California, United States of America
Email: jeeholife2@gmail.com
**Cheongna Dalton School, Incheon, Republic of Korea
Email: esyou7210@gmail.com

----------**********----------

Abstract:

Many individuals with autism spectrum disorder (ASD) face challenges in communicating effectively due to difficulties in interpreting and empathizing with emotions. This issue is further exacerbated in the virtual realm, where comprehending others' emotions during conversations can be particularly challenging. This study investigates the potential of artificial intelligence (AI) to enhance virtual communication for individuals with ASD by training two distinct sentiment analysis machine learning models: one that includes neutral sentiments and another that excludes them. Both models are trained using human-tagged sentiments from a variety of texts. The ultimate goal of this study is to evaluate the feasibility of AI-generated sentiment indicators in supporting individuals with ASD in their virtual communication.

Keywords — **Artificial Intelligence, Machine Learning, Autism Spectrum Disorder, Virtual Communication, Sentiment Analysis**

----------**********----------

I. INTRODUCTION

A. Research Background

Artificial intelligence (AI) has progressively become an integral part of our lives, and its applications are shown in various fields, ranging from healthcare to education. Recently, it has gained interest from many through AI platforms, notably ChatGPT. AI's ability to recognize and identify patterns and quickly do tedious tasks makes it particularly appealing for psychological and behavioral studies. Particularly, when individuals are impacted by neurodevelopmental disorders such as autism spectrum disorder (ASD or autism), they may have difficulties interacting or communicating with others [1]. While only about 1 in 100 of the population is affected by this lifelong condition, easing life for potentially 800 million people is not to be looked down upon [2].

ASD is categorized with a "spectrum" of symptoms, from challenges to empathy to excelling at arts and sciences [3]. However, a crucial part of one's life that people affected with ASD often struggle with is communication, particularly pragmatic language - actions such as eye contact, humor/jokes, and body language - that is widely used in socialization [4]. Those who suffer from autism also have different pragmatic languages from others, making relations between them and others challenging. Furthermore, emotional empathy, the ability to understand and acknowledge the emotions of others, can be challenging for ASD patients. Results of a 2018 study on self-reported sympathy of participants with autism show that the autism group gave significantly lower sympathy and distress ratings when shown a photographic image [5].

To fix these issues, we have created natural language processing (NLP) models and used AI to help alleviate these issues. AI's applications in psychological and behavioral studies are particularly promising, offering solutions to challenges that have been difficult to overcome

through human intervention alone. Moreover, the rise of virtual communication platforms such as Instagram and Discord has introduced a new limitation: online interactions often lack the nonverbal cues that individuals with ASD already find difficult to interpret. However, as AI technologies continue to evolve, they offer the possibility to bridge these gaps and provide support to individuals affected by ASD. The use of AI in this context goes beyond creating functional tools; it raises questions about how technology can foster empathy. By integrating AI into communication, we can create more inclusive digital spaces where isolated individuals can express themselves more freely and connect with others.

B. Research Objectives

This research paper introduces sentiment analysis models developed to answer a critical question in the context of Autism Spectrum Disorder (ASD): "To what extent can AI-generated emotional context indicators enhance empathy and communication quality for individuals with ASD in a virtual communication environment?" By examining the effectiveness of sentiment-based AI tools providing more explicit emotional context indicators to texts, we seek to understand how these cues may impact the empathy and communication dynamics of individuals with ASD, particularly in online, non-contact environments where social cues are often limited.

Through this study, we aim to analyze whether real-time emotional context indicators can mitigate these challenges by improving the ability of individuals with ASD to recognize and respond to the underlying emotional tone in digital conversations. Using Discord as the virtual communication environment, we assess the influence of AI-generated sentiment indicators on conversation flow, empathy, and user engagement. Two sentiment analysis configurations are tested: one that includes neutrality as a distinct emotion and one that excludes neutrality to exaggerate other emotions.

This research also aims to provide a comparative analysis of the two sentiment configurations—neutrality-included and neutrality-excluded—to evaluate how these models influence the interaction patterns of individuals with ASD. Through experimental testing, we aim to present the potential of AI not only as a tool for communication but also as a medium for fostering understanding and reducing barriers in virtual spaces.

II. Data Description

C. Neutral Included

The data for the ML model predicting emotions, including neutrality, was taken from Kaggle [6]. The dataset totaled more than 393k data points, each consisting of pairs of text taken from tweets from Twitter and human-tagged emotions. These emotions, sorted from the most common to the least common, are: [Neutral, love, happiness, sadness, relief, hate, anger, fun, enthusiasm, surprise, emptiness, worry, and boredom]. However, for the actual model, boredom was removed due to insufficient data to generate reliable patterns.

D. Neutral Excluded

The dataset used in this study was sourced from the Hugging Face platform, specifically the "dair-ai/emotion" dataset, which contains over 417k text entries [7]. The dataset listed six emotions: sadness, joy, love, anger, fear, and surprise. They are each associated with a label from 0 to 5, respectively. Each text entry represents a single emotional state, and a large portion consists of sentences starting and/or using "I," meaning emotions revolve around oneself.

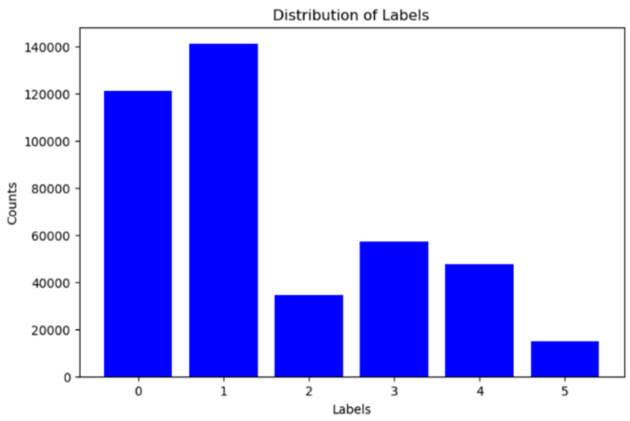

Fig. 1 Data distribution of the dataset

III. METHODOLOGIES

E. AI Models

1) ***Neutral Included :*** The initial approach to creating a sentiment analysis module, including neutrality, took a logistic regression approach with TF-IDF (Term Frequency-Inverse Document Frequency) vectorization, representing words based on their occurrence frequency and factoring in their rarity across documents. While this was effective for simple sentiment analysis, this approach did not fulfill the nuanced requirements of recognizing and identifying more subtle and indirect emotions. A bit of testing with the model revealed inherent biases within the mode: it over-relied on obvious, emotion-specific keywords (e.g., "love" for love, "happy" for happiness), which resulted in poor generalization with more complex or subtly expression emotions, like emptiness, fun, relief, and sadness. For example, the emotion of "relief" may be conveyed with a phrase like, "It's finally over," which does not contain explicit emotional keywords but still hints at a release from stress. Similarly, "anger" might be expressed through phrases like, "That wasn't what I expected from him," where frustration may be implied through the tone and context rather than direct words such as "angry" or "mad." Given these limitations, it became clear that the simple approach of a logistic regression model lacked the necessary depth for such subtle emotion recognition, as it exhibited excessive reliance on certain words and failed to view the texts in a broader context.

To address these limitations of the initial approach, we transitioned to a more complex neural network model using Kears's Sequential framework, leveraging the complexity and contextual processing capabilities of a BiLSTM (Bidirectional Long Short-Term Memory) layer. We also added multiple Dropout layers to reduce bias on a few specific words. To run this more complex model, we were forced to use a high-performance virtual machine to configure and train this model. The following are the specs of the virtual machine used in this model:

TABLE I
VIRTUAL MACHINE SPEC USED FOR NEUTRAL-INCLUDED MODEL

Operating System	Ubuntu 20.04 + TensorFlow + Jupyter
GPUs	1x NVIDIA H100 SXM5 80GB
vCPUs	16
RAM	128GB
Local Storage	64GB

The final model structure was as follows:

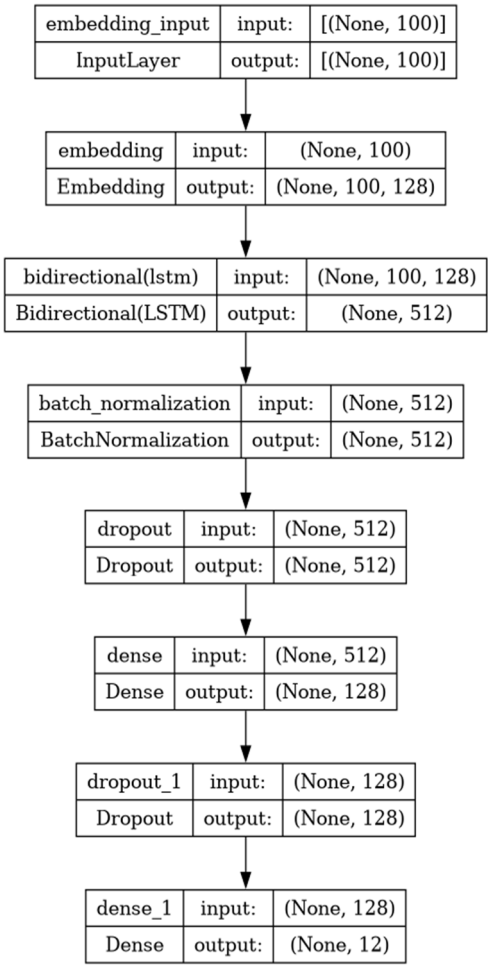

Fig. 2 The model architecture of neutral-included AI model

This final model architecture was purposefully structured to balance robustness and generalizability while addressing the ASD-specific challenges of recognizing diverse emotional tones. The critical layers to doing such are the following:

- BiLSTM (Bidirectional Long Short-Term Memory) layer – Processes the input sequences in both forward and backward directions, capturing dependencies from both past and future contexts. Using a BiLSTM, the model can better understand the context of each word in a sentence rather than isolating each word, which is essential for accurately determining the sentiment and emotional cues of the text.
- Dropout Layer—Randomly sets a fraction of input units to 0 during training to prevent overfitting. This process helps generalize the model, ensuring that it performs well on unseen or unusual data. This is crucial for the model to catch onto more subtle and minor emotional indicators in texts.

To ensure accurate and reliable sentiment classification, we trained the BiLSTM on the dataset described in 2.1, with texts labeled with emotional categories. We employed a total

of 20 training epochs, using EarlyStopping to halt training when the model's validation loss plateaued, preventing overfitting and maximizing generalizability. The model's performance was saved at checkpoints based on validation loss, allowing retrieval of the best-performing version.

The BiLSTM approach demonstrated significant improvements over the logistic regression approach, particularly in its ability to accurately classify a diverse range of emotions, including the more subtly expressed or implied emotions. To evaluate this model's performance and understand how well it addresses the needs specific to ASD communication, we examined the confusion matrix of accuracy and the top words that were associated with each emotion.

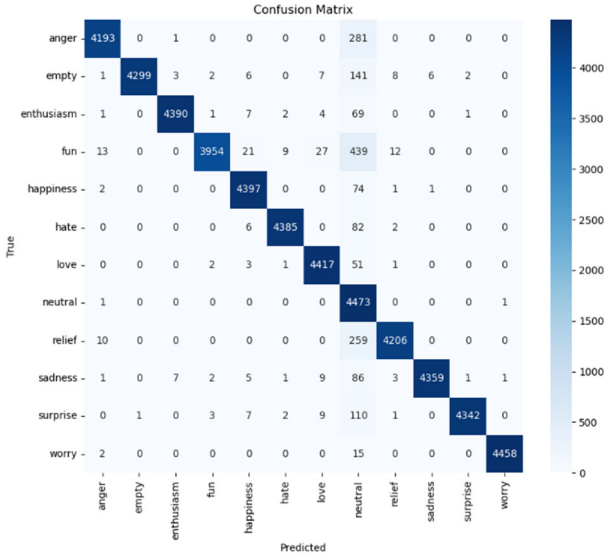

Fig. 3 Classification accuracy of the neutral-included model represented through a confusion matrix

The diagonal of the confusion matrix represents the correctly classified instances for each emotion, while other cells show the misclassification of emotions. As can be seen, the diagonal is where most of the predictions were centered around. However, the model struggled relatively more with emotions like "relief," "anger," "empty," and "fun," as evidenced by many of the data values being misclassified as neutral or other emotions. This shows how more nuanced and subtle emotions like such were harder for the model to catch.

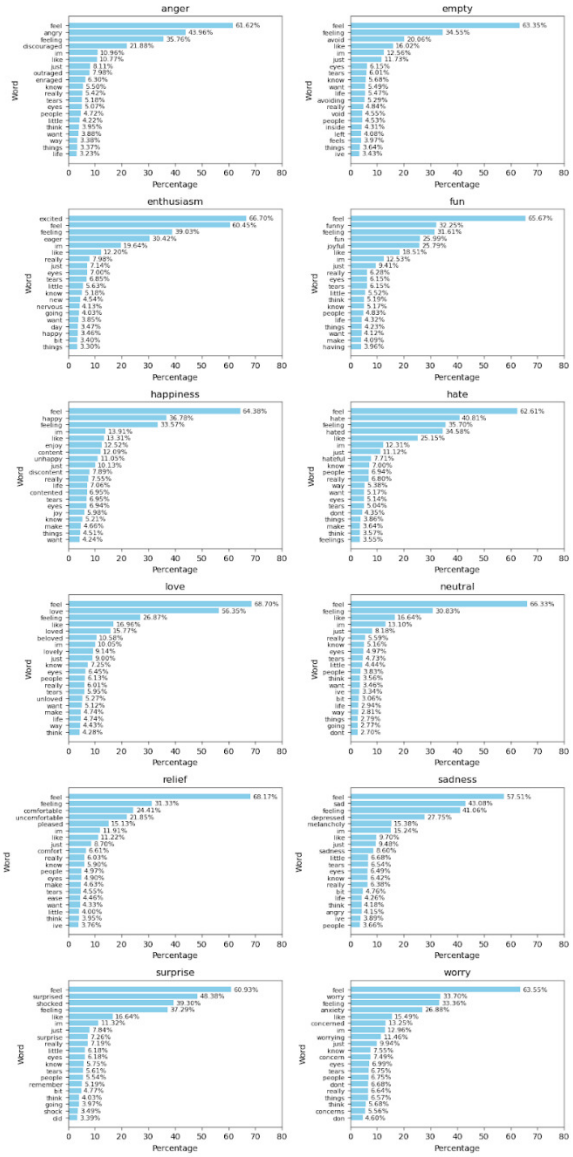

Fig. 4 Most associated words for each emotion in percentages

Even for those lower accuracy labels, our model was able to achieve a 96% or above accuracy for all of those labels. In this neutral-included model, we can see that the neutrality has introduced a balancing effect, reducing misclassifications between emotions with similar valence (e.g., "sadness" and "anger") and instead pointing those towards neutrality, accounting for a large number of misclassifications to neutrality while other misclassifications only rarely occurred.

The relatively high significance of less obvious emotional context words illustrated reveals the model's interpretive flexibility, showing that it can better capture emotional nuance and implied sentiment.

2) ***Neutral Excluded:*** This sentiment analysis model was designed to identify a set of emotional tones, now excluding the neutral tone. Excluding neutrality, this model aimed to emphasize the expression of other emotions,

allowing the model to better capture more subtle differences in emotional tone without returning to neutrality as a safe ground.

The prototype Stochastic Gradient Descent (SGD) analysis module was optimized to recognize explicit and implicit sentences and expressions; however, it achieved an accuracy of only 83%, sometimes less than 40%, depending on which data points were used in the training and testing.

There remained much room for improvement, particularly in the accuracy of classifications for more complex emotions like love and fear. Incorporating a more dynamic n-gram tuning would further enhance contextual understanding, adding dropout layers could prevent reliance on keywords, and these advancements would not only improve accuracy but also make the model more robust in other applications.

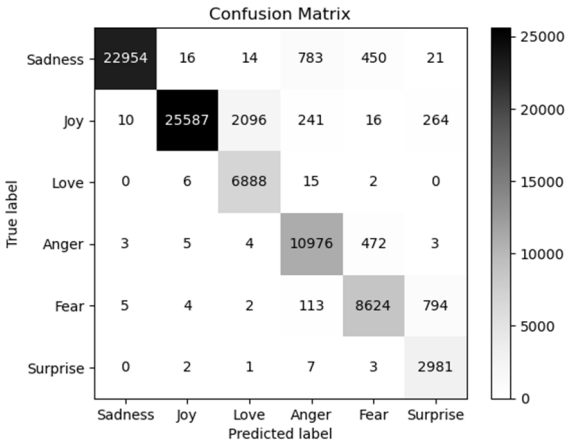

Fig. 5 Classification accuracy of the neutral-excluded model represented through a confusion matrix

As a result, we created a new, much more improved model using Keras's Sequential API. This model significantly outperformed the previous SGD approach, improving at least 12% overall and significantly on all of the emotions specifically. Most emotions were accurate at rates over 99%. We used the BiLSTM mentioned above (Long Short-Term Memory) layers to handle sequential data and dropout layers to prevent the overfitting problem.

Despite the very uneven data distribution shown in Figure 2, the model was able to make very accurate predictions about the emotions of love and surprise.

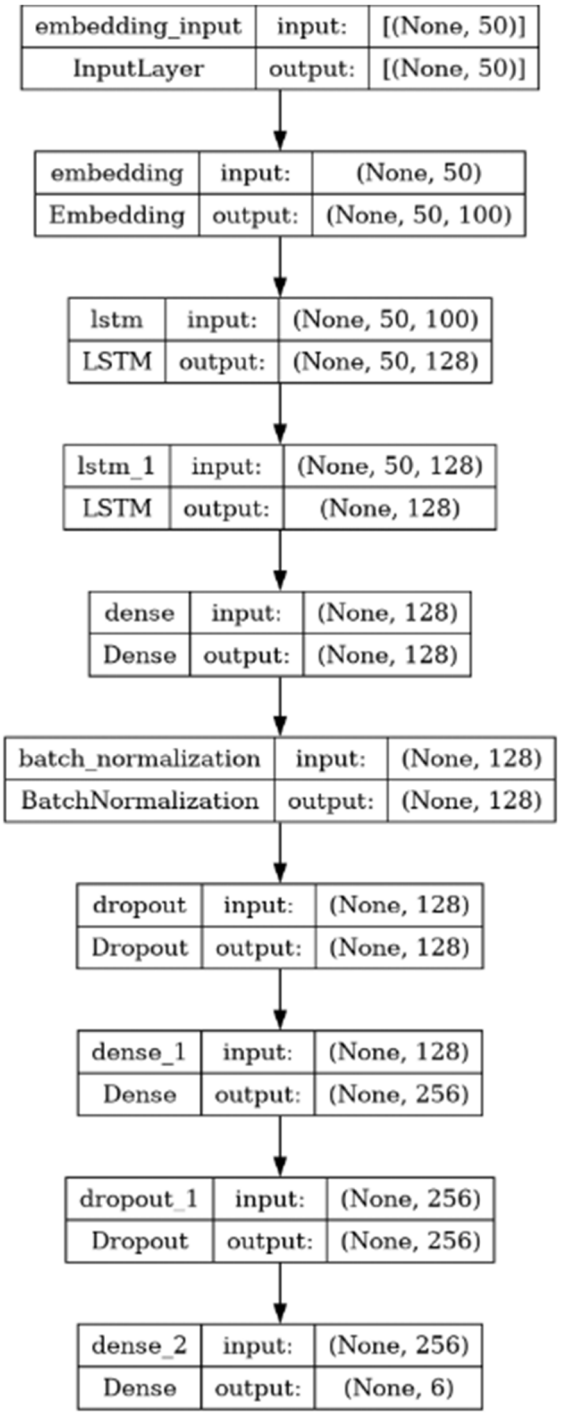

Fig. 6 The model architecture of neutral-excluded AI model

IV. Results & Discussion

F. Results of the Data Analysis

The results of this study highlight the possibility of the efficacy of AI-generated emotional context indicators in enhancing empathy and communication in virtual environments for individuals with autism spectrum disorder (ASD). The study created two sentiment analysis models:

one including neutrality as an emotional category and another excluding neutrality to emphasize non-neutral emotions. Below is a discussion of the key findings from our analysis.

The neutral-included model demonstrated high accuracy, achieving over 99.3% classification accuracy overall. This model's performance was particularly robust when handling emotions with clear indicators, such as "joy" and "sadness." However, it also showed limitations in distinguishing nuanced emotions like "relief" and "fun," which were often misclassified as "neutral." This indicates that neutrality served as a "safe fallback," reducing the model's ability to differentiate between subtle emotional states. Some key advantages of this approach are the balancing effect and the enhanced generalizability.

Neutrality primarily reduces the risk of misclassifying emotions with similar valances, which can be important, especially in conversations with individuals with ASD, where those kinds of misclassifications could potentially be a problem. Also, the inclusion of neutrality allowed the model to handle ambiguous texts with less emotional clarity, aligning well with real-world situations where messages may sometimes lack overt emotional tones.

G. Future Experimentation with Individuals with ASD

Future research endeavors will focus on refining methodologies and going through experimentation to achieve meaningful, generalizable insights. Below is an in-depth exploration of future pathways.

We have devised a structure for our experimentation process to examine the impact of AI-generated emotional context indicators on communication for individuals with ASD in a virtual environment.

1) **Virtual Environment and AI Tools:** *For the virtual conversation environment, we will use Discord due to its flexible platform that supports add-ons and bots, making it suitable for integrating our AI-based sentiment analysis modules. The modules will analyze text in real time and provide emotional content indicators in the form of emojis or text labels (e.g., "happy," "sad," and "angry") associated with each message. Two versions of the sentiment analysis will be implemented: one version with neutral emotions excluded and another without them.*

2) ***Participants:*** The study will involve 3–4 adults with ASD and 3–4 adults without ASD (ages 18 and older) recruited through outreach to ASD support organizations, online communities, or local centers. Individuals under adult guardianship will require their guardian's consent before participation. As this is an independent student-led research project, no monetary compensation will be provided; however, participants will be fully informed of the study's goals and their importance in advancing research on virtual communication for individuals with ASD.

3) ***Data Collection Sessions:*** The data collection will take place over four structured sessions, each lasting approximately 20 minutes. Adapting the approach of Peters et al. (2014), each participant will engage in a 20-minute conversation with a designated researcher in the virtual environment. The designated researcher will ask the participant at least ten open-ended questions about neutral, age-appropriate topics as necessary, and the researcher will wait approximately 3-5 seconds after any text for the participant to respond if desired [8]. To complement the virtual data, a concurrent Zoom meeting will record physical indicators such as eye movement, facial expressions, vocal tone changes, body gestures, and response times. The sessions are outlined below:

- Session 1 (baseline) - Participants converse with the researcher without any AI-generated emotional context indicators.
- Session 2 (without indicators) - Participants converse with the researcher on topics similar to session 1. The researcher will stimulate different emotional contexts to observe participants' reactions to varying conversational tones
- Session 3 (with neutrality indicator) - Identical to Session 2, but with the sentiment analysis module including a neutrality indicator
- Session 4 (without neutrality indicator) - Identical to Session 2, but with the sentiment analysis module that excludes the neutrality indicator

4) ***Data Security and Participant Privacy:*** All recordings and data collected will be securely stored on two password-protected computers accessible only to the research team. Participants will be assigned unique identifiers to anonymize their data, and personal identifiers will be excluded from all recorded data to ensure confidentiality. Additionally, all data handling will adhere to institutional ethics and data privacy guidelines. A mental health hotline number will be provided to all participants to ensure mental well-being following each session.

5. CONCLUSION

Our project, which investigated the potential of AI-generated emotional context indicators to support empathy and communications for individuals with autism spectrum disorder (ASD),

has set the stage for innovative applications of network models, such as BiLSTM, trained on extensive datasets. We explored how AI tools can detect nuanced emotional tones and provide actionable insights into virtual conversational dynamics. The results highlighted promising directions for addressing communication barriers, particularly by including and excluding neutrality as an emotion category.

Looking ahead, we aim to build on this foundation by conducting direct experiments with individuals with ASD, as outlined in our proposed experimentation methods. These studies will test the effectiveness of AI-generated emotional indicators in controlled virtual environments, such as Discord, to better understand how these tools impact empathy and conversational fluidity in real-time interactions. The structured sessions will provide insights into how different configurations influence communication dynamics, including response times and engagement. Additionally, we plan to expand our analysis by integrating physiological data, such as facial expressions and eye movement, to explore multimodal approaches for decoding emotions.

This project demonstrates AI's potential to enhance digital communication for neurodiverse populations. Advancing research through direct experimentation and targeted refinement contributes to an evolving understanding of how virtual environments can become more supportive and inclusive for individuals with ASD.

REFERENCES

[1] National Institute of Mental Health. (2024, February). Autism spectrum disorder. National Institute of Mental Health. https://www.nimh.nih.gov/health/topics/autism-spectrum-disorders-asd

[2] WHO. (2023, November 15). Autism. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders

[3] Hughes, J. E. A., Ward, J., Gruffydd, E., Baron-Cohen, S., Smith, P., Allison, C., & Simner, J. (2018, October 12). Savant syndrome has a distinct psychological profile in autism. Molecular autism. https://pmc.ncbi.nlm.nih.gov/articles/PMC6186137/

[4] Martin, G. E., Lee, M., Bicknell, K., Goodkind, A., Maltman, N., & Losh, M. (2023, July 10). A longitudinal investigation of pragmatic language across contexts in autism and related neurodevelopmental conditions. Frontiers in Neurology. https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2023.1155691/full

[5] Holt, R., Upadhyay, J., Smith, P., Allison, C., Baron-Cohen, S., & Chakrabarti, B. (2018, July 27). The Cambridge sympathy test: Self-reported sympathy and distress in autism. PLOS ONE. https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0198273

[6] Anjali, S. (2024, March 5). Emotion analysis based on text. Kaggle. https://www.kaggle.com/datasets/simaanjali/emotion-analysis-based-on-text/data

[7] Saravia, E., Liu, H.-C. T., Huang, Y.-H., Wu, J., & Chen, Y.-S. (2018, October-November). Dair-ai/emotion · datasets at hugging face. Datasets at Hugging Face. https://huggingface.co/datasets/dair-ai/emotion

[8] Koegel, L. K., Park, M. N., & Koegel, R. L. (2015, May 1). Using self-management to improve the reciprocal social conversation of children with autism spectrum disorder. Journal of autism and developmental disorders. https://pmc.ncbi.nlm.nih.gov/articles/PMC3981935/

