Generational Variations in Human–AI Conversational Styles: A Computational Linguistic Analysis

Dr Prakash Kuppuswamy[1*], Ms. Vijaya Ramineni[2], Dr Phaneendra Kanakamedala[3],
Dr Anuradha Anumolu[4], Senthil Kumar A.A.R[5], Ms. Mounika Samudrapu[6], Dr Indu Sharma[7], Ms. Preeti Nashier[8]

[*1,5]Department of Computer Science Engineering, KL University, Vadeeswaram, Andhra Pradesh, India
[2,4,6]Department of Information Technology, DVR & Dr. HS MIC College of Technology, A.P, India
[3]Department of Computer Science and Information Technology, Lakireddy Bali Reddy College of Engineering, A.P. India
[7,8]Department of Computer Science & Engineering, SRM University, Sonepat, Delhi, India

Corresponding Author
Prof (Dr) Prakash Kuppuswamy
Department of Computer Science Engineering, KL University,
Vadeeswaram, Andhra Pradesh, India
prakashcnet@gmail.com

Abstract

Background: While human–AI interaction has become a major part of modern digital communication, few studies have been conducted to examine intergenerational differences in users' conversational behavior. The relationship between communication among Generation X, Millennials, and Generation Z is explored through a computational linguistic analysis**. Materials and Methods:** Train.csv is a large conversational dataset that was analyzed to see if there were differences between the languages used, how they interacted, and the organization of their conversations across age groups. Because the dataset lacked explicit demographic information, synthetic generational labels were generated based on linguistic features and clustering techniques that were based on the textual characteristics. Text preprocessing, tokenization, normalization, TF–IDF vectorization, sentiment analysis, and stylistic feature extraction were all applied to the text using multiple natural language processing (NLP) methods. The factors studied included sentence length, words per sentence, the use of emojis, the use of slang, and the tendency to ask questions. Furthermore, statistical methods and machine learning algorithms were used to assess the predictiveness of the generational style when examining text-based communications.

Results and Discussion: The results show that there are certain differences in conversations between generations. Gen Z's communication is brief, casual, and simple, while Gen X's communication is formal, structured, and verbose. Millennials have a blend of both communication styles. In generational classification, the Random Forest model was the best machine learning model of the ones evaluated. **Conclusion:** The research emphasizes the need for AI systems to be adaptive and sensitive to the context, aiming to customize their tone and style to align with user preferences in specific sectors like healthcare, education, and digital services.

Keywords: Human–AI Communication, Generational Differences, Computational Linguistics, Conversational AI, Digital Communication, Linguistic Style Analysis.

Introduction

Artificial intelligence (AI) technology has changed human-digital interaction (HDI), particularly in the areas of chatbots, virtual assistants, and large language models [1]. These systems are used in social media, education, healthcare, and customer support, among other everyday applications [2]. Understanding the many forms of user contact with AI systems is one important research issue that has come to light in view of growing user-AI interactions [3, 4]. Among the many elements influencing interaction patterns and affecting the user experience are generational variances in language use and communication approaches [5, 6].

Linguistic behaviors, influenced by surrounding digital technologies, cultural shifts, and communication methods, are identified across generations from Gen Z through Millennials to Gen X [7, 8]. For instance, younger users tend to prefer short, casual phrases with emojis and other abbreviations and internet jargon, whereas older users prefer more formal, organized, and context-rich language [9, 10]. These disparities may affect the user's perception, trust, and engagement with AI systems, impacting the effectiveness of human-AI interaction. Although more and more research is examining differences between generations, there is little systematic research that has explored these differences computationally [11, 12].

Recent advances in machine learning (ML) and natural language processing (NLP) enable the processing of large volumes of conversation data and the identification of linguistic structures within the text [13]. Researchers can quantify users' communication styles and determine important variations in communication between users by applying sentiment analysis, topic modeling, and text categorization to the data. One problem is that there is a lack of clear demographic information in most publicly available datasets. To overcome this limitation, researchers are increasingly relying on proxy techniques, such as linguistic feature analysis and clustering methods, to gain insight into generational characteristics based on the textual data [16, 17].

Literature Review

Rashmi Nagpal, Sankung Fatty, and David Guy Brizan (2024) highlighted conversational AI as an important area within natural language processing research. Their study reviewed major conversational agent architectures, methodologies, and datasets, focusing on both rule-based and generative AI approaches. The work also analyzed widely used chatbot training datasets and discussed techniques employed in chatbot development. The authors emphasized current advancements, ongoing challenges, and emerging directions in conversational AI research. Through an extensive review of previous studies, the research provided valuable insights into the evolution and improvement of intelligent conversational systems across multiple application domains. The findings may support future developments aimed at improving the adaptability and effectiveness of conversational agents [18].

D. Aneja, R. Hoegen, Daniel McDuff, and Mary Czerwinski (2021) explored how embodied conversational agents can improve human-computer interaction through expressive and engaging communication. They introduced SIVA, a socially intelligent virtual agent designed to identify conversational behaviors and adapt its responses based on user expectations. Using multimodal inputs, SIVA generated synchronized speech and realistic facial expressions during interactions. A user study involving thirty participants demonstrated that SIVA was perceived as more empathetic, believable, and realistic than a non-adaptive control agent. The study emphasized the importance of adaptive conversational behavior in enhancing user engagement and interaction quality [19].

M. Bronikowski (2025), In their paper noted that while AI-generated text is grammatically correct, it often fails to exhibit human-like traits like lexical diversity, stylistic variation, and idiomatic language.

To overcome these issues, the study proposed Stylistic Contrastive Learning (SCL). The use of GPT-5 enhanced the diversity of discourse, the inclusion of idiomatic expressions, and the naturalness of AI-generated articles, conversations, and news-style content. Human assessments showed that the text produced was more realistic and accurate in terms of content. Another part of the study also showed that idiomatic phrases and discourse markers were factors that significantly contributed to the quality of communication [20].

J. Wang et al. (2024) Examined gender differences in the initial communication used when interacting with an AI conversational assistant. In addition to recommendation and text polishing, the system allowed users to choose from several communication styles, from humorous to respectful, and to select different degrees of extroversion for the AI. The system helped create a conversational interaction and led to positive user ratings for usability. Men normally had a higher level of satisfaction compared to the female participants. The study found that adaptive AI conversational widgets can help create a more comfortable and stable experience in the early stages of digital interactions, suggesting the potential of adaptive conversational systems in improving online communication [21].

Q. Xu and C. Chen (2026) In their study with 137 participants, they looked at how user interactions with Microsoft Copilot differ when it is conversant in different AI styles and has different roles to play. The study found that there was a significant impact on the readability, analytical language, and perceived authenticity of the AI-generated responses. Interaction modes also had an impact on user writing behavior, as there were strong relationships between emotional tone, conversational authority, AI outputs, and user inputs. Furthermore, the combination of conversational style and the role of the AI influenced the user's perceptions about the reliability and usefulness of AI responses. The results further point to the significance of conversational design for influencing how people communicate and perceive in an AI-supported problem-solving task [22].

Research Objectives

This study aims to investigate generational variations in human–AI conversational styles using a computational linguistic framework. By analyzing a conversational dataset, the research employs both rule-based and machine learning techniques to simulate generational categories and extract relevant linguistic features. The study focuses on identifying differences in sentiment, message length, lexical choices, and stylistic elements such as emoji usage and interrogative patterns. Furthermore, statistical and predictive models are used to validate the significance of these variations and assess the extent to which generational traits can be inferred from conversational text. By understanding how different generations communicate, developers can create more inclusive and user-centric conversational agents that dynamically adjust their language style according to user preferences. Such flexibility is especially vital in sensitive domains like healthcare, where effective communication directly influences user trust and decision-making. Ultimately, this study contributes to bridging the gap between computational linguistics and human-centered AI design, offering insights into building more responsive and context-aware communication systems.

Text Preprocessing and Feature Extraction

TF–IDF (Term Frequency–Inverse Document Frequency) is a text processing technique applied in Natural Language Processing (NLP) to convert textual data into numerical feature vectors. It is a measure that considers the significance of a word in a document when compared with the entire data set; it is a combined of term frequency and inverse document frequency. Common words in the document are weighted more and common words across documents are weighted less. These TF–IDF features are then used in sentiment analysis models to determine emotional tone. Sentiment scores are calculated using weighted word polarity values, and machine learning classifiers predict whether the sentiment stated in the text is constructive or harmful.

4.1. TF–IDF (Term Frequency – Inverse Document Frequency)

To convert text into numerical features, TF-IDF is used to measure the importance of a term in a document relative to the entire corpus.

4.1.1 TF-IDF Formula

$$TF - IDF(t,d) = TF(t,d) \times IDF(t)$$

Where:

- t= term (word)
- d= document
- TF = term frequency
- IDF = inverse document frequency

4.1.2 Term Frequency (TF)

$$TF(t,d) = \frac{f_{t,d}}{\sum_{t' \in d} \quad f_{t',d}}$$

- $f_{t,d}$= number of times term $t$appears in document d

4.1.3 Inverse Document Frequency (IDF)

$$IDF(t) = log\left(\frac{N}{df_t}\right)$$

- N= total number of documents
- df_t= number of documents containing term t

TF-IDF gives **higher weight to important words** and reduces the impact of common words.

4.2. Sentiment Analysis Model

To measure emotional tone, sentiment polarity is computed.

4.2.1 Basic Sentiment Polarity Score

$$Sentiment(d) = \frac{\sum_{i=1}^{n} \quad w_i \cdot s_i}{\sum_{i=1}^{n} \quad w_i}$$

Where:

- s_i= sentiment score of word i
- w_i= weight of word (importance)
- n= total number of words

4.2.2 Classification Function (Optional ML Model)

For classification (positive/negative):

$$y = sign(w \cdot x + b)$$

Where:

- x= feature vector (TF-IDF)
- w= learned weights
- b= bias
- y= predicted sentiment class

Research Methodology

5.1 Dataset Description

This study employed conversational text data from the dataset called train.csv (Table 1). It includes several dialogue instances, each of which is a textual dialogue, such as a user query, a response, and a conversational utterance. The dataset is table 2 and usually contains columns including text/message content, speaker identities, and optionally metadata, depending on the source.

Demographic data, such as age or generation, is not readily available; therefore, this research uses a synthetic labeling strategy to identify generations (Gen Z, Millennials, and Gen X). These labels are generated according to linguistic indicators (Table 3) (such as slang, the use of emojis, sentence structure, etc.) and clustering methods based on text features. This can be useful in examining generational communication styles without the need for personally identifying information.

Table 1. Research dataset

Dataset	Samples	Purpose
train.csv	11118	Model training
validation.csv	1000	Hyperparameter validation
test.csv	1000	Final evaluation

The data undergoes common natural language processing (NLP) pre-processing steps, such as text cleaning, tokenization, and normalization. Other features, such as sentiment polarity, word count, emoji rate, and question patterns, are captured to include the stylistic and emotional aspects of communication. These characteristics form the basis for comparison and statistics between generational groups.

Table 2. Feature Table

Feature Name	Type	Description
text	Text	Raw conversational message or utterance
generation	Categorical	Assigned label (Gen Z, Millennial, Gen X)
speaker_type	Categorical	Indicates Human or AI
word_count	Numeric	Number of words in the message
length	Numeric	Total number of characters
sentiment	Numeric	Sentiment polarity score (-1 to +1)
emoji_count	Numeric	Number of emojis in the text
question_flag	Binary	1 if message contains a question, else 0

Table 3. Linguistic Feature Categories

Category	Features Included	Purpose
Lexical Features	word_count, length	Measure verbosity and complexity
Sentiment Features	sentiment	Capture emotional tone
Style Features	emoji_count, question_flag	Identify informal vs formal style
Structural Features	sentence patterns, punctuation usage	Analyze communication structure
Derived Features	generation label	Enable generational comparison

First, the data is loaded and analyzed for attributes suitable for the conversation, such as conversational text and information about the speakers. The raw text data source may include noise; thus, a raw text data pipeline is used. It includes changing the case of the text, removing punctuation, removing stop words, and conducting tokenization. The form of the words is also normalized using other forms of normalization, such as stemming or lemmatization. These actions confirm the quality and transparency of the given dataset for later linguistic analysis.
The one problem with this study is that it does not explicitly describe the demographics. To remedy this, a hybrid approach is taken to develop synthetic generational labels.

The first is a rule-based approach that depends on linguistic markers, such as slang (e.g., lol, emojis), and informal language used by younger generations (Gen Z). The older generations (Millennials and Gen X) have been correlated with more formal and structured expressions. To boost robustness, an unsupervised learning approach, K-means clustering, is used for text representation with TF-IDF. The resulting clusters are then "read" and mapped to the generational categories, thus fine-tuning the labeling process in a data-driven way.

5.2 Feature Extraction Model

This research relies especially on a feature extraction framework through which unstructured conversational text is transformed into numerical information that can be analyzed using computational strategies. After preprocessing the text, the cleaned and standardized text is fed as input to the model. The framework, demonstrated in figure 1 and table 4, captures a number of different linguistic characteristics that reflect various aspects of the lexical, semantic, syntactic, structural, and stylistic aspects of human-AI communication. The complexity and verbosity of languages belonging to the same generational group are measured by lexical properties like total word count, length of the text, average word size, etc.

Table 4. Linguistic Feature Categories

Category	Features Included	Purpose
Lexical Features	word_count, length	Measure verbosity and complexity
Sentiment Features	sentiment	Capture emotional tone
Style Features	emoji_count, question_flag	Identify informal vs formal style
Structural Features	sentence patterns, punctuation usage	Analyze communication structure
Derived Features	generation label	Enable generational comparison

Semantic information is produced by techniques like TF-IDF, which converts text into vectors with weights indicating important terms in the dataset. Sentiment analysis is used to recognize the polarity of emotions and the tone of communication. Stylistic features include punctuation marks, abbreviations, emojis, and interrogative expressions. Furthermore, structural features, such as sentence length and the frequency of questions, are included. Together, these characteristics provide a holistic portrayal that enables statistical analyses and classification of generational conversational patterns using machine learning techniques.

5.3 Structure Diagram of Feature Extraction Model

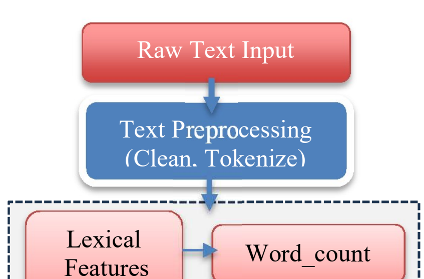


Figure 1. Conversational Extraction Model

5.4 Feature Extraction

The linguistic aspects retrieved for generational communication analysis are shown in this table 5. Features, including word count, sentiment score, emoji frequency, and TF-IDF vectors, enable the identification of structural, emotional, and linguistic variations across generations. These characteristics provide the foundation for statistical comparison and machine learning classification in the methodology.

Table 5. Description of Feature

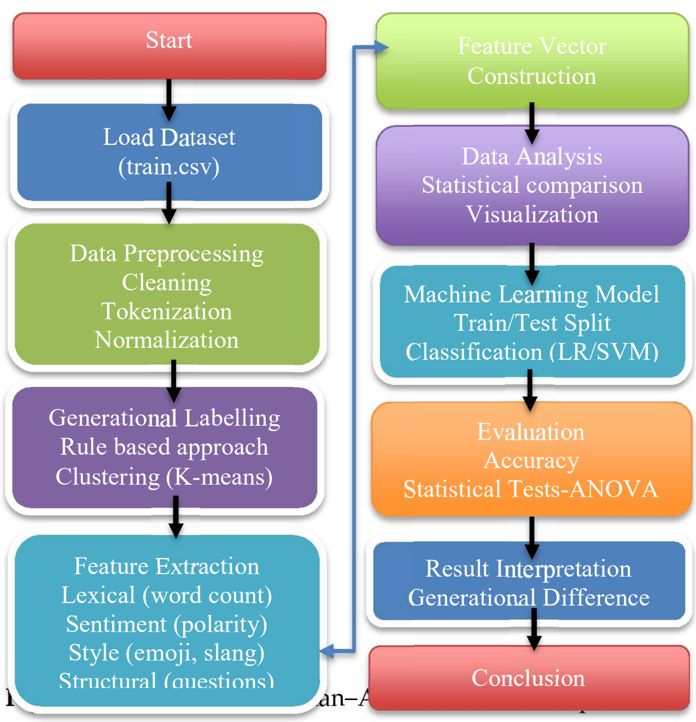
Feature Name	**Description**
word_count	Number of words
text_length	Total characters
sentiment_score	Emotional polarity
emoji_count	Emoji frequency
question_flag	Presence of questions
TF-IDF vectors	Text representation

This figure 2 depicts characteristics retrieved from processed example conversation records. Every conversation has a generational label, as well as numerical characteristics like word count, sentiment polarity, and question frequency. Extracted features table 6, dataset shows how conversational traits differ among Generation X, Generation Z, and Millennial communication patterns.

Table 6. Extracted Features

	clean_dialog	**generation**	**word_count**	**text_length**	**sentiment_score**	**emoji_count**	**question_flag**

0	Say Jim, how about going for a few beers af...	Gen X	160	713	0.197768	0	1
1	Can you do push-ups? Of course I can. Its a ...	Gen Z	59	244	-0.158125	0	1
2	Can you study with the radio on? No, I liste...	Gen Z	41	185	0.425000	0	1
3	Are you all, right? I will be all right soon	Gen Z	33	141	0.285714	0	1
4	Hey John, nice skates. Are they new? Yeah,	Gen X	128	506	0.186317	0	1
5	Hey Lydia, what are you reading? I’ m looki...	Gen X	274	1248	0.260964	0	1
6	Frank’ s getting married, do you believe thi...	Gen Z	56	223	0.236667	0	1
7	I hear you bought a new house in the northern ...	Gen X	161	689	0.377522	0	1
8	Hi , Becky , whats up ? Not much , except that...	Gen X	253	1189	0.112011	0	1
9	How are Zinas new programmers working out ? I ...	Millen.	82	370	0.156061	0	1



Result Analysis

For generational classification, three machine learning models were assessed: logistic regression, support vector machine (SVM), and random forest. With a test accuracy of 73.1%, logistic regression set a good foundation for conversational text classification. SVM's superior ability to process high-dimensional TF-IDF feature representations increased the performance to 76.6%. With 86.7% test accuracy, random forest outperformed all other models, proving its ability to identify intricate linguistic and stylistic patterns within conversational data.

Average feature values for every generation are shown in Table 5. Generation X has longer text and word counts, suggesting that they communicate in detail. Generation Z, on the other hand, shows short and direct interactions. Millennials have somewhat conventional values. The comparison points out quantifiable linguistic variances across generations in human-AI discussions.

Table 7. Generational Feature summary

	word_count	**text_length**	**sentiment_score**	**emoji_count**	**question_flag**
generation					
Gen X	180.937874	820.572605	0.167386	0.0	0.970309
Gen Z	46.016129	198.252688	0.139702	0.0	0.808602
Millennial	85.093215	368.319469	0.153980	0.0	0.948378

The following bar graph illustrates the average number of words for each generation. Gen X has the longest and most detailed conversations, with the highest average word count. Millennials demonstrate moderate-length communication, whereas Gen Z has the shortest communication. The figure shows that there are important differences between generations in their preferred communication style and verbosity in conversation with AI when communicating with AI.

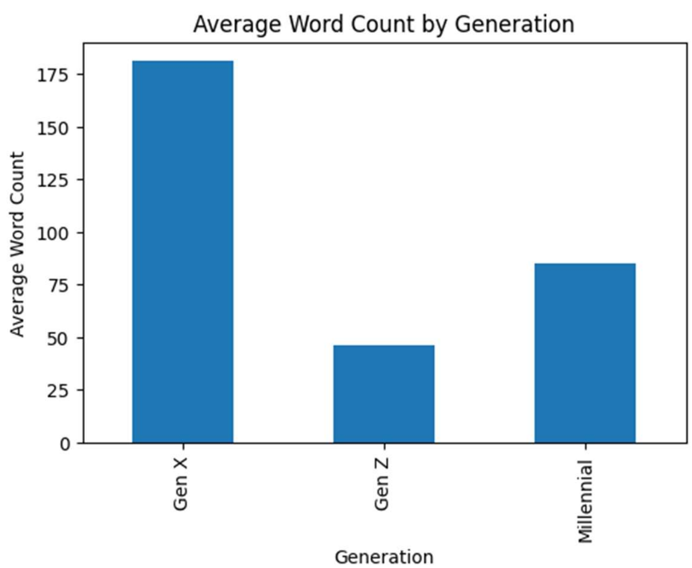


Figure 3. Word count Generation

The chart below is a bar chart showing the average sentiment scores by generation. Gen X's positive sentiment is the highest, followed by that of Millennials, while Gen Z has comparatively lower sentiment values. The numbers show nuanced differences in emotional expressions in conversational interactions between humans and AI, such as emotional tone and expression style.

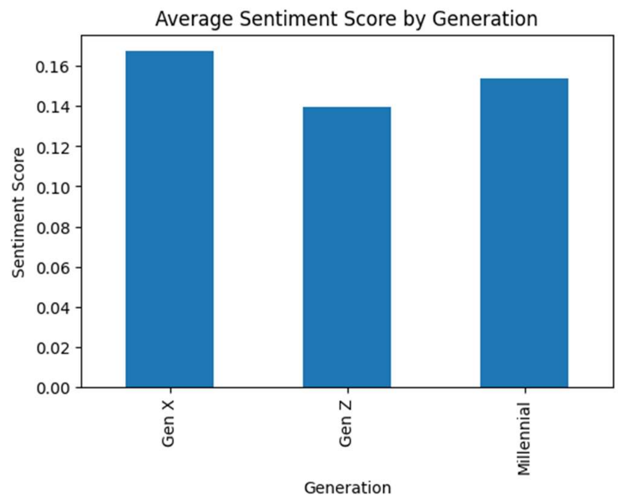


Figure 4. Sentiment score

This bar chart compares the average frequency of questions used within each generation. It is clear that Gen X is the most vocal user of questions in conversations, suggesting more interactive and inquiry-based communication. Millennials also use questions with great frequency, while Gen Z demonstrates comparatively less frequency in using questions. It's a generational shift in engagement and conversational interaction styles.

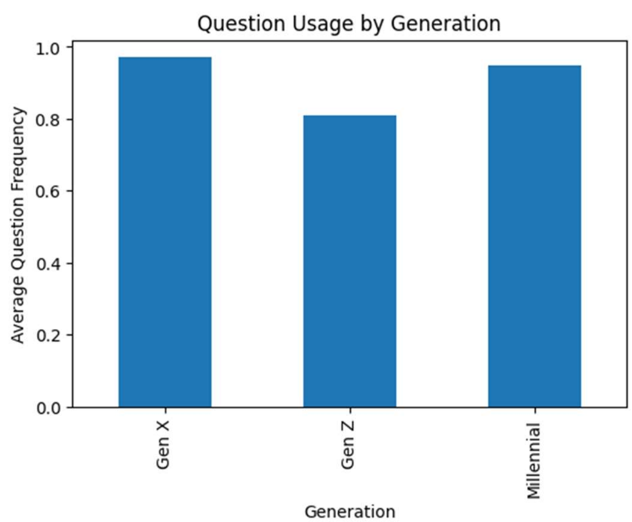


Figure 5. Question Usage

6.1 Generational Distribution Table

The number of dialogue samples available for each generation is shown in Table 6. The dataset is balanced, covering all of Generation X, Generation Z, and Millennials, which allows for fair training and evaluation of the models. The distribution also helps to ensure reliable synthetic labeling and minimize bias in classification experiments.

Table 8. Distribution Table

Generation	Count
Gen X	4008
Gen Z	3720

Millennial	3390

It supports balanced dataset and synthetic labelling validity.

a. Model Comparison Table (MOST IMPORTANT)

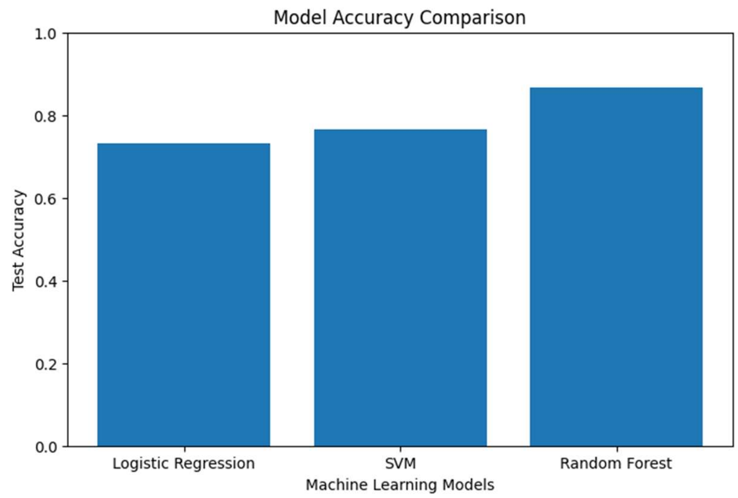


Figure 6. Accuracy Comparison

This table and figure show the validation and test accuracy of machine learning models. Random Forest exposed the best performance, with 86.7% test accuracy, followed by Logistic Regression and SVM. The results show that ensemble learning techniques work better at capturing generational communication patterns from conversational data sets.

Table 9. Accuracy comparison

Model	Validation Accuracy	Test Accuracy
Logistic Regression	73.4%	73.1%
SVM	75.2%	76.6%
Random Forest	84.1%	86.7%

6.3 Classification Report Summary

This table shows the results of precision, recall, and the F1-score for each classification model. Random Forest had the best-balanced performance, with scores of 0.86 on all evaluation measures. The results show its excellent performance in correctly categorizing generational groups and in avoiding classification errors.

Table 10. Accuracy comparison

Model	Precision	Recall	F1-score
Logistic Regression	0.72	0.72	0.72
SVM	0.76	0.76	0.76
Random Forest	0.86	0.86	0.86

The confusion matrix shows the classification performance of the Random Forest model for the generational categories. The diagonal values are the largest in most samples, which indicates high predictive accuracy. The primary misclassifications occur between Millennials and other groups. Figure

6 illustrates the model's success in separating generational communication patterns from conversational features.

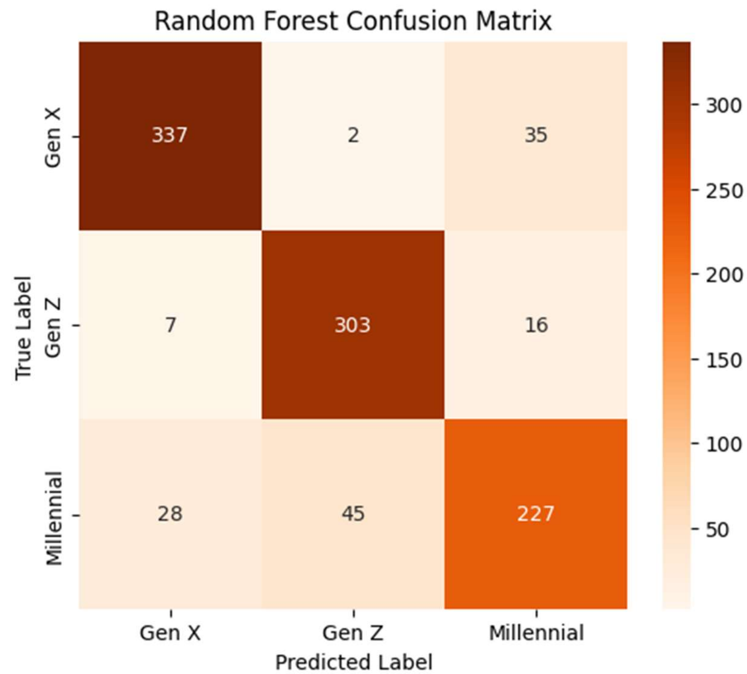


Figure 7. Random Forest Confusion Matrix

In general, the results demonstrate that computational linguistic analysis is effective for modeling generational conversational behavior in human-AI interaction. The results underscore the possibility of creating adaptive conversational AI systems that tailor their communication style based on users' preferences and interactions.

Research Contribution

The experimental study found that there are significant intergenerational differences in human-AI conversational behavior. Gen X participants tended to talk longer, be very organized and formal in their conversations, and preferred to communicate in detail. On the other hand, Gen Z users communicated more informally, with short and concise messages. The results validate the effectiveness of computational linguistic characteristics in detecting and categorizing generational communication styles in human-AI interaction.

The study makes a number of important contributions, including the development of a synthetic generational labeling framework, the application of natural language processing-based conversational analysis, and the discovery of linguistic and stylistic traits from conversation data. The study also looked at a number of machine learning algorithms and found that the Random Forest method had a classification accuracy of 86.7%. The study also plainly revealed intergenerational differences in human-AI communication and provided insights into the creation of adaptive conversational AI systems.

Conclusion

This study proposes a computational linguistic approach to studying generational differences in human-AI conversations using the DailyDialog dataset. Direct demographic data were not available, and synthetic generational categories were artificially generated based on heuristic rules and linguistic patterns. Several NLP techniques, such as preprocessing, TF-IDF representation, feature extraction, and classification by machine learning, were applied to the study of conversational behavior among Gen Z, Millennials, and Gen X. The results revealed differences between the generations in terms of communication preferences. Generally speaking, the communication style of Gen Z users was brief, informal, and direct, while Gen X users' communication was more structured, detailed, and formal. Millennials had a hybrid communication style, incorporating elements from both groups. In terms of classification models, the best model is Random Forest, with 86.7% accuracy, followed by Logistic Regression and SVM. Overall, it shows that language can be a key component in creating adaptive and personalized conversational AI systems.

Acknowledgement

The authors acknowledge the academic and infrastructural support provided by KL University, Vijayawada, DVR & Dr. HS MIC College of Technology, Vijayawada, A.P, India, SIR M Visvesvaraya Institute of Technology, Bangalore, Karnataka, India.

Data availability

Data set available openly in UCI machine learning sources, that are utilized in this work.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declares that they have no competing interests.

References

1) Sumon, R. I., Uddin, S. M. I., Akter, S., Mozumder, M. A. I., Khan, M. O., & Kim, H. C. (2024). Natural language processing influence on digital socialization and linguistic interactions in the integration of the metaverse in regular social life. *Electronics*, *13*(7), 1331.
2) Haase, J., Kremser, W., Leopold, H., Mendling, J., Onnasch, L, Plattfaut, R. (2024). Interdisciplinary directions for researching the effects of robotic process automation and large language models on business processes. *Communications of the Association for Information Systems*, *54*(1), 579-604.
3) Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 chi conference on human factors in computing systems* (pp. 1-13).
4) Abedin, B., Meske, C., Junglas, I., Rabhi, F., & Motahari-Nezhad, H. R. (2022). Designing and managing human-AI interactions. *Information Systems Frontiers*, *24*(3), 691-697.
5) Azad, I., Chhibber, S., & Tajhizi, A. (2023). How do different generations communicate on social media? A comparative analysis of language styles, emoji usage, and visual elements. *Language, Technology, and Social Media*, *1*(2), 86-97.
6) Poivet, R., Lopez Malet, M., Pelachaud, C., & Auvray, M. (2023). The influence of conversational agents' role and communication style on user experience. *Frontiers in psychology*, *14*, 1266186.
7) Bratina, D., & Faganel, A. (2024). Understanding Gen Z and Gen X responses to influencer communications. *Administrative Sciences*, *14*(2), 33.

8) Wandhe, D. P. (2024). The new generation: Understanding millennials and Gen Z. *Available at SSRN 4716585*.
9) Wang, Q., Gao, Y., & Wang, X. (2025). Exploring engagement, self-efficacy, and anxiety in large language model EFL learning: A latent profile analysis of Chinese university students. *International Journal of Human–Computer Interaction*, *41*(12), 7815-7824.
10) Ke, Y. H., Jin, L., Elangovan, K., Abdullah, H. R., Liu, N., Sia, A. T. H., Ting, D. S. W. (2025). Retrieval augmented generation for 10 large language models and its generalizability in assessing medical fitness. *npj Digital Medicine*, *8*(1), 187.
11) Abedin, B., Meske, C., Junglas, I., Rabhi, F., & Motahari-Nezhad, H. R. (2022). Designing and managing human-AI interactions. *Information Systems Frontiers*, *24*(3), 691-697.
12) Chou, C. M., Shen, T. C., Shen, T. C., & Shen, C. H. (2022). Influencing factors on students' learning effectiveness of AI-based technology application: Mediation variable of the human-computer interaction experience. *Education and Information Technologies*, *27*(6), 8723-8750.
13) Sharma, R., Agarwal, P., & Arya, A. (2022). Natural language processing and big data: a strapping combination. In *New trends and applications in internet of things (IoT) and big data analytics* (pp. 255-271). Cham: Springer International Publishing.
14) Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia tools and applications*, *82*(3), 3713-3744.
15) Cai, J., Hadjinicolaou, A. E., Paulk, A. C., Soper, D. J., Xia, T., Wang, A. F., Cash, S. S. (2025). Natural language processing models reveal neural dynamics of human conversation. *Nature Communications*, *16*(1), 3376.
16) Wani, A. A. (2024). Comprehensive analysis of clustering algorithms: exploring limitations and innovative solutions. *PeerJ Computer Science*, *10*, e2286.
17) Antons, D., Grünwald, E., Cichy, P., & Salge, T. O. (2020). The application of text mining methods in innovation research: current state, evolution patterns, and development priorities. *R&D Management*, *50*(3), 329-351.
18) Nagpal, Rashmi, Sankung Fatty, and David Guy Brizan (2024). "A Survey of Conversational Styles and Systems." In *2024 16th International Conference on Human System Interaction (HSI)*, pp. 1-4. IEEE.
19) Aneja, D., Hoegen, R., McDuff, D., & Czerwinski, M. (2021). Understanding conversational and expressive style in a multimodal embodied conversational agent. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1-10).
20) Bronikowski, M. Stylistic Contrastive Learning for Human-Like AI Text Generation. In *Open Conference of AI Agents for Science 2025*.
21) Wang, J., Ivrissimtzis, I., Li, Z., & Shi, L. (2024). Impact of personalised AI chat assistant on mediated human-human textual conversations: exploring female-male differences. In *Companion Proceedings of the 29th International Conference on Intelligent User Interfaces* (pp. 78-83).
22) Xu, Q., & Chen, C. (2026). Conversation styles and role expectations in a large language model: effects on readability, linguistic features, and users' perceptions in a Copilot experiment. *Behaviour & Information Technology*, 1-15.