



Age Group Differences in Social Media Language Use: Identity Representation via Emojis and Abbreviations

Pratik Emmanuel^{1*}, Olufemi Isiaq²

¹ Creative Computing Institute, Postgraduate Research, University of the Arts London, London, United Kingdom

² Creative Computing Institute, Computer Data Science and AI, University of the Arts London, London, United Kingdom

Abstract

This paper aims to analyse how age groups impact the use of social media language (SL) in online textual conversations, focusing on the type, polarity, and subjectivity of SL to explore identity representation across different age groups. During data collection in 2023, there were 4.76 billion social media users worldwide; hence, understanding the use of SL is essential for analysing how individuals express their emotions and identities online. SL consists of emoticons/emojis, abbreviations, and mixed language in textual conversations. For this study, a qualitative approach was used, with 46 participants from three age groups: 25-34, 35-44, and 45-54 engaging in mock group conversations for one hour using WhatsApp. The participants were separated into six group conversations. The findings show that the older age groups (35-44 and 45-54) prefer to discuss technical or professional topics, whereas the younger age groups (25-34) are more interested in games, food, entertainment, and fun. Furthermore, for the age group 25-34 there is a high usage of emoticons/emojis, whereas the age group 45-54 uses them the least. These findings indicate that age groups have a significant influence on the use of SL and topic preferences in textual conversations, providing insights that will be used to develop a Natural Language Processing (NLP) tool for online identity classification. This study enhances human-computer interaction by investigating how different age groups use SL in digital environments.

Keywords: social media language; online identity management; emoticons/emojis; age groups.

1. Introduction

In recent times, the Internet has become such an integral part of people's lives that individuals and businesses cannot function without it. This is because the Internet facilitates almost everything, from retrieving information to global interaction. The Internet has also given rise to social media, which is used on a regular basis to connect with other individuals, organisations, or businesses, giving and receiving information via a variety of communication

channels ranging from visual to textual. When social media users interact with a post by a business or a friend, they leave a trail of text messages, comments, and reactions, making their online activity available for scrutiny and reflection. In 2023, there were 4.76 billion social media users worldwide, so it is essential to study social media language (SL), as it has become an important part of social media activity online, fulfilling the need to express the feelings and emotions of an individual in reaction to a post, comment, or message during online chat depending on the content (Kemp, 2023).

Social media users develop online identities to influence the opposite person with whom they are interacting. Online identity by Warburton and Hatzipanagos (2015) is defined as an individual's social identity, which actively constructs self-impressions based on the topic of discussion. Therefore, it is a fashion on social media for users to display a positive 'self' to gain more attention from friends and non-friends; this also occurs during conversations where individuals hide their negative 'self' to show their positive 'self' for having their presence felt amongst the online group to gain social approval (Walther 2007, p. 2555). Some studies on online dating sites addressing people's self-impression show that there were some variations in people's behaviour within open online platforms than in closed platforms (McCabe et al., 2005; Gibbs et al., 2006; Heino et al., 2006). This meant that in the online platforms, a person's self-impression is diverse based on the settings of that particular online platform. However, the impact of age groups on the use of SL, particularly on identity management in textual conversations, is the one factor that is left unexplored.

This study seeks to address this gap by examining how SL is utilised strategically and performatively by different age groups (25–54) within textual conversations to present an identity, focusing on the type, polarity, and subjectivity of SL use. Subsequently, the data obtained will be used to develop a Natural Language Processing (NLP) tool for understanding SL and classifying online identity. By studying the formation of online identity within textual conversations, this project improves human-computer interaction by enhancing NLP tools tailored to age-specific communication patterns and improving user experience across different digital platforms (Pietro, 2020). Moreover, this study contributes to a broader field of digital communication by providing empirical data on how age groups influence the use of SL, thereby increasing the understanding of identity management in the digital age.

The research question being tested in this study asks the following question: Can the use of social media language be impacted by age groups of individuals to present an 'identity' within a conversation? To answer this question, there is a null hypothesis that states: *social media language cannot be impacted by age groups to present an 'identity' within a conversation*. The alternate hypothesis states that: *social media language can be impacted by age groups to present an 'identity' within a conversation*.

2. Method

For participant recruitment, a survey for participation in this study was designed and distributed across survey exchange groups on the social media platform Facebook alongside survey exchange websites (SurveyCircle, 2023; SurveySwap, 2023). Survey exchange is a method by which researchers gain real participants by completing the surveys of other researchers. Furthermore, the JISC Survey tool, which is safe concerning data protection, was adopted for survey design and recording survey responses (JISC, 2023). Informed consent was also presented to the participants at the beginning of the survey, so they could first read about the research project, then data safety assurance and the informed consent before deciding if they wanted to take part in this study. The contents of the informed consent were derived from the university's informed consent document (Solent, 2023). Overall, 46

participants agreed to participate and shared their email addresses for contact and further discussion by the researcher. Hence, the mock group conversation was conducted among 46 participants.

The data collection lasted for two months, keeping each participant’s time availability in mind. This study gathered mock group conversation data for one hour where each group had one topic of their choice, each participant was assigned to one of the six groups for the group conversation, and each group had up to eight to ten participants. The group conversation was held on WhatsApp for one hour. The participants were encouraged to be natural and open in their conversation as they saw fit to avoid any bias (Nyumba et al. 2018).

2.1 Data Analysis Process

The data analysis for this study gathered information on how social media language (SL) is being used in textual conversations. SpaCy, an advanced open-source library for NLP in Python, was used to carry out the data analysis and also for conducting the sentiment and subjectivity analysis (Table 1). SpaCy can process large volumes of text efficiently and provides tools for sentiment and subjectivity analysis, making it ideal for analysing human language (Bhandari, 2020). Python was implemented using Jupyter Notebook, a web-based multi-language programming platform (Jupyter, 2023) (Table 1).

Table 1: Phase two variables and data analysis

	Variables	Data analysis	Tools for analysis
Demography (Age Groups)	Age Groups Vs. Themes for each group conversation.	Thematic Analysis and Topic classification.	Jupyter Notebook using Python programming language.
Social media Language (SL).	Age Groups Vs. All Social media languages (emoticons/emoji, abbreviations and mixed-languages).	What SL is used the most, frequency of SL use, polarity sentiment, and subjectivity scores for contextual understanding.	Jupyter Notebook using Python programming language.

The data analysis involved preprocessing, sentiment analysis, and subjectivity analysis. Preprocessing involved cleaning and normalising the conversation data as well as removing irrelevant content. Sentiment analysis specifically identifying polarity (positive, neutral, negative) of the messages was performed using SpaCy, along with considering the complex sentence structure with SL use, their context, and idiomatic expressions. Subjectivity analysis determined whether the use of SL reflected personal opinions or objective facts, using both SpaCy’s machine learning model and subjectivity lexicons to classify the messages. Finally, a comparative analysis was conducted across the age groups (25-34, 35-44, 45-54) showing significant differences in emotional expression and identity representation, providing key insights that can be used for developing age-specific NLP tools.

3. Results

3.1 Age Groups and Conversation Themes

Tables 2 and 3 below show how various themes are split across different age groups in the group conversations. For group one, the medical science topics AML, IBMS, Allogeneic,

Neutrophil, and Anticoagulant are mostly addressed by the age groups 35-44 and 45-54, with the age group 35-44 exhibiting a considerable interest in Neutrophil (80%) and other medical-related themes. Furthermore, the age group 25-34 displays low engagement with these topics. Group two reveals that the age group 35-44 mostly converse about morphology, patients, blood, cells, and analysers, with percentages ranging from 66.7% to 100%, whereas the age group 45-54 is less engaged in these topics. Group three shows a considerable preference for food-related subjects, with themes being identified such as pasta, burgers, palace, and sushi among the age group 25-34, with percentages ranging from 64% to 100%. The age group 35-44 (66.7%) chose to converse about their favourite food in group three. In group four, the age group 25-34 is particularly interested in topics related to games; hence, they converse about subjects related to the game genre, such as horror (66.7%) being their preference, and talked about game sales (100%) on Steam, while the age group 35-44 discussed game genres related to RPG (85.7%) and strategy (80%). Group five shows that the age group 35-44 prefers discussing movie-related subjects such as action, scenes, likes or dislikes, movies, and characters, with significant numbers of respondents in all categories. Finally, group six shows that the age group 35-44 (100%) preferred to converse about subjects like games, work, sports, and walking. The results also show that the age group 45-54 engages only slightly in these topics. Tables 2 and 3 reveal varied preferences of themes throughout the age groups; it also shows that people with higher age groups prefer discussing more technical or professional-related subjects, while individuals with younger age groups prefer to discuss entertainment, games, and food-related subjects.

Table 2: Group one to group three theme distribution

Group One Theme Distribution %						
Group One (10 participants)		AML	Neutrophil	Allogeneic	IBMS	Anticoagulant
	Age Groups					
	25-34	25.7%	0%	16.7%	0%	0%
	35-44	10.7%	80%	83.3%	100%	100%
45-54	85.0%	20%	0%	0%	0%	
Group one topic classification		Haematology, a field of Biomedical Sciences.				
Group Two Theme Distribution %						
Group Two (8 participants)		Morphology	Patient	Blood	Cell	Analysers
	Age Groups					
	35-44	75%	66.7%	80%	100%	71.4%
	45-54	25%	33%	20%	0%	28.6%
Group two topic classification		Blood Science, a field of Biomedical Sciences.				
Group Three Theme Distribution %						
Group Three (6 participants)		Burger	Pasta	Sushi	Palace	Favourite
	Age Groups					
	25-34	80%	65%	64%	100%	33.3%
	35-44	20%	35%	36%	0%	66.7%
Group three topic classification		Food.				

Table 3: Group four to group six theme distribution

Group Four Theme Distribution %						
Group Four (6 participants)		Sale	Strategy	Game	RPG	Horror
	Age Groups					
	25-34	100%	0%	44.4%	0%	66.7%
	35-44	0%	80%	55.6%	85.7%	33.3%
45-54	0%	20%	0%	14.3%	0%	
Group four topic classification		Video Games.				
Group Five Theme Distribution %						
Group Five (8 participants)		Scene	Action	Movie	Like	Character
	Age Groups					
	25-34	22.2%	20%	0%	33.3%	25%
	35-44	77.8%	80%	100%	66.7%	75%
Group five topic classification		Movies.				
Group Six Theme Distribution %						
Group Six (8 participants)		Work	Pet	Game	Walk	Sports
	Age Groups					
	35-44	100%	43.5%	100%	100%	100%
	45-54	0%	56.5%	0%	0%	0%
Group six topic classification		Sports, Work and Games.				

3.2 Social Media Language (SL) Use Across Age Groups

Tables 4 and 5 in Section 3.3 reveal that the use of SL significantly varies across the age groups. The use of emoticons/emojis is quite frequent by the age group 25-34 in contrast to the older age groups. In group three, there is a high use of emoticons/emojis (10 times) by the age group 25-34, revealing their preference for an informal and expressive style of communication. On the other hand, in group six there is a notable use of SL by the age group 35-44, where emoticons/emojis are used nine times. In this group, the participants preferred a balanced approach to their conversation, often balancing the informal use of SL with formal content related to professional discussions. However, it was observed across group conversations that the age group 45-54 used SL the least, except for group two and group six. Even when this age group used SL, their contributions to the discussions were significantly lower than the younger participants, demonstrating a more formal and reserved style of communication.

3.3 Sentiment and Subjectivity Analysis

Sentiment and subjectivity analyses were conducted to examine how age groups influence the expression of emotions and opinions. As seen in Tables 4 and 5, moderate polarity and subjectivity scores were observed for the age group of 25-34, suggesting they were slightly positive in their conversations and were subjective in their tone. According to Gaby (2023), polarity scores range from -1 to 1, where -1 represents a negative sentiment and 1 represents a positive sentiment. Whereas, the subjectivity scores range from 0 to 1, suggesting how much of the textual material has a personal opinion (closer to 1) and factual information

(closer to 0) (Gaby, 2023). In group three, there is a slight positive sentiment and subjectivity with a polarity score of 0.32 and a subjectivity score of 0.40, implying their conversations expressed personal opinions frequently within their group. Furthermore, in group six, the age group 35-44 had a range of sentiment and subjectivity; the lower polarity score of 0.08 for this age group suggests their tone was neutral in their conversation, though they were more objective rather than subjective, as shown by their subjectivity score of 0.23. However, this same age group (35–44) showed a more positive tone in other group conversations, like in group three. In group three, the polarity score of 0.69 and subjectivity score of 0.79 reveal a wide range of engagement based on the topic. Additionally, there were lower levels of sentiment and subjectivity among the participants of the age group 45-54, with slightly positive polarity but mostly preferring to be neutral. For example, in group six, the same age group displayed emotions less frequently, as seen by their subjectivity score of 0.36, and they also had a preference for a neutral tone, as observed by their polarity score of 0.23.

Table 4: Group one to group three, Social media Language, Sentiment and Subjectivity

Social media Language (SL) Vs Sentiment and Subjectivity (Group One)					
Group One		What SL is being used the most?	Frequency of SL use	Polarity Score	Subjectivity Score
	Age Groups				
	25-34	None.	None.	0.00 (Neutral)	0.00 (Objective)
	35-44	Emoji and Abbreviation.	6(x) Emoji used, 3(x) Abbreviations.	0.29 (Slight positive)	0.38 (Objective)
	45-54	None.	None.	0.00 (Neutral)	0.00 (Objective)
Social media Language (SL) Vs Sentiment and Subjectivity (Group Two)					
Group Two		What SL is being used the most?	Frequency of SL use	Polarity Score	Subjectivity Score
	Age Groups				
	35-44	Emoticons/Emoji.	8(x) Emoticons/Emoji.	0.14 (Slight positive)	0.59 (Moderately Subjective)
	45-54	Emoticons/Emoji.	6(x) Emoticons/Emoji.	0.39 (Slight positive)	0.53 (Moderately Subjective)
Social media Language (SL) Vs Sentiment and Subjectivity (Group Three)					
Group Three		What SL is being used the most?	Frequency of SL use	Polarity Score	Subjectivity Score
	Age Groups				
	25-34	Emoticons/Emoji.	10(x) Emoticons/Emoji.	0.32 (Slight positive)	0.40 (Moderately Subjective)
	35-44	Emoticons/Emoji.	6(x) Emoticons/Emoji.	0.69 (Strong positive)	0.79 (Subjective)

Table 5: Group four to group six, Social media Language, Sentiment and Subjectivity

Social media Language (SL) Vs Sentiment and Subjectivity (Group Four)					
		What SL is being used the most?	Frequency of SL use	Polarity Score	Subjectivity Score
Group Four	Age Groups				
	25-34	Emoticons/Emoji.	4(x) Emoticons/Emoji.	0.08 (Slight Positive)	0.69 (Moderately Subjective)
	35-44	Emoticons/Emoji.	4(x) Emoticons/Emoji.	0.24 (Moderately Positive)	0.60 (Moderately Subjective)
	45-54	None	None	0.00 (Neutral)	0.00 (Neutral)
Social media Language (SL) Vs Sentiment and Subjectivity (Group Five)					
		What SL is being used the most?	Frequency of SL use	Polarity Score	Subjectivity Score
Group Five	Age Groups				
	25-34	Emoticons/Emoji.	4(x) Emoticons/Emoji.	0.45 (Moderately Positive)	0.61 (Moderately Subjective)
	35-44	Emoticons/Emoji.	6(x) Emoticons/Emoji.	0.23 (Moderately Positive)	0.63 (Moderately Subjective)
Social media Language (SL) Vs Sentiment and Subjectivity (Group Six)					
		What SL is being used the most?	Frequency of SL use	Polarity Score	Subjectivity Score
Group Six	Age Groups				
	35-44	Emoticons/Emoji.	9(x) Emoticons/Emoji.	0.08 (Slight Positive)	0.23 (Relatively Objective)
	45-54	Emoticons/Emoji.	6(x) Emoticons/Emoji.	0.23 (Moderately Positive)	0.36 (Moderately Subjective)

4. Discussion

The two hypotheses for this study are about understanding if age groups impact the use of social media language for identity representation online; therefore, it was critical to prove one of the hypotheses correct using supporting data. As a result, the data analysis supports the alternate hypothesis, which states that social media language can be impacted by age groups to present an 'identity' within a conversation; this answers the critical research question for this project.

From the analysis of Tables 4 and 5, different patterns can be observed in terms of the usage of social media language, such as emoticons/emojis and abbreviations used among all the age groups who participated in the group conversations. The findings reveal that there is a high use of emoticons/emojis by the age group 25-34 as opposed to the age group 45-54, which uses them the least. This outcome agrees with an article by Alshenqeti (2016) that focused

on emojis and new generations and their socio-semiotic study, which found that youths tend to use more emoticons/emojis in their textual conversations, and this is because of their increased amount of exposure to social media platforms and how quick they are to adapt to new forms of digital communication tools. The fact that the age groups 25-34 and 35-44 use more emoticons/emojis during textual conversations also agrees with the findings of another study by Cavalheiro *et al.* (2023), where not only the same observations were made, but it was also observed that these age groups find visual communication tools such as emoticons/emojis more expressive of their feelings and emotions, which helps make the atmosphere in their conversations more engaging. Furthermore, the high subjectivity score for the age groups 25-34 and 35-44 also confirms the argument by Tang and Hew (2019) that young adults utilise emoticons/emojis to express their feelings and emotions more clearly. However, Tang and Hew (2019) also mention that alongside the expressive purposes of using social media language, young adults also use it to perform pragmatic functions, including preventing misunderstandings, assisting with message interpretation, and using other social media languages like stickers and gifs, which helps enhance amusement and individualisation.

Additionally, a study by Boutet *et al.* (2021) also said that digital natives mostly use positive emoticons/emojis to enhance the emotional tone of their textual conversations; this is also supported by the positive polarity ratings throughout the age groups in Tables 4 and 5, suggesting an overall positive tone in the conversation. Tables 4 and 5 also show that the age group 45-54 use emoticons/emojis less, even though the subjectivity scores are higher for younger age groups; however, a study by Cui (2022) shows that older age groups can have high subjectivity scores, especially when emoticons/emojis are used in their communication group. This demonstrates that, even if older individuals use social media less frequently, the emotional expressions they provide may still be used to study generational differences. As the age groups 25-34 and 35-44 use emoticons/emojis more frequently than the age group 45-54, they establish an identity of emotional expressiveness in a positive context on the digital platform. Individuals in the age group 45-54 displayed a more reserved style of communication, which might be attributed to their lack of knowledge or awareness of social media language; it could also be that they did not contemplate using social media language in their conversation (Chen *et al.*, 2024).

Previous studies on age differences in online communication and social media language use found a significant difference across age groups. According to research, younger individuals in the age group 25-34 are more likely to use informal language such as emoticons/emojis and abbreviations in their online conversations (Varnhagen *et al.*, 2010). This indicates a trend towards a more casual and expressive digital interaction style. Whereas the age group 35-44 takes a more balanced approach, which combines casual and formal language and uses social media for both personal and professional use (Quan & Young, 2010). Meanwhile, the individuals belonging to the older age groups 45-54 or even higher tend to lean towards using more formal languages and place a greater emphasis on politeness and clarity in their digital interactions; this could be due to their less frequent use of digital communication tools and their preference for a traditional method of communication (Herring & Androustopoulos, 2015). These variations highlight the importance of filling the gap in the lack of digital experience that exists between younger and older generations; this indicates the necessity for specialised techniques in digital literacy instruction to suit different preferences and practices (Poushter & Chwe, 2018).

The findings of this study also have some essential implications for online communication strategies, meaning that the communication preferences of different age groups must be kept in mind by the developers of digital platforms for increased acceptance. The results show that

individuals in the age group 25-34 and the age group 35-44 often use emoticons/emojis in their respective digital environments, indicating a demand for more engaging and expressive communication tools in their textual conversations; this could even be reflected in their social media posts (Berg, 2020). These younger age groups also have a higher subjectivity rating, which indicates that their communication is emotionally expressive and intimate (Tang & Hew, 2019). As a result, digital platforms attempting to engage users in these age groups should integrate tools that enable the usage of emoticons/emojis and other visual communication aids to improve the emotional tone and participation in online textual conversation environments. Furthermore, the study's alternate hypothesis, which supports the strategic use of social media language for identity representation, suggests that online platforms can facilitate identity expression by providing diverse and rich communication tools that meet the expressive needs of younger individuals.

In contrast, individuals belonging to the age group 45-54 use fewer emoticons/emojis and are more reserved in their conversations (Cui, 2022). Although this age group (45–54) uses fewer emoticons/emojis, as shown by the findings, their use of social media language when used expresses their emotional tone (Chen *et al.*, 2024). This suggests that the design of the digital platforms for this age group (45–54) and ever higher age groups should prioritise simplicity and usability, making these types of digital platforms and social media language accessible and clear to increase their acceptance. Online communication strategies aimed at older age groups may benefit from the simplicity and clarity of digital platforms, which might even include small tutorials or prompts, as well as gradually introducing emoticons/emojis and other social media language. Understanding and responding to generational communication preferences enables digital platforms to provide a more inclusive and effective user experience that appeals to a wider range of age groups (Singer & Zalmanson, 2013).

While the study's findings give useful insights into how different age groups utilise social media language to express their identities, there are some limitations of this study. One significant limitation is the small sample size; with only 46 participants spread across three age groups, the generalisation of the findings can be limited. A larger sample size would be able to provide a more robust basis for comparison and reveal more similarities and differences in the use of social media language across different demographics. Furthermore, this study focuses solely on age groups, ignoring other influential factors such as educational qualification, cultural background, and social media literacy, all of which may influence the use of social media language. As a result, future research will take a multidimensional approach, investigating how these additional variables interact with age groups to influence social media language use.

This study had another limitation, which is that it relied on mock group conversations as the primary source of data. While this strategy allows for controlled data collection, it might not accurately represent real-world social media interactions. Participant language or behaviour may have altered as a result of being aware that they were part of the research, thereby leading to a potential bias. This problem can be alleviated by collecting real-life data to gain a more realistic understanding of the usage of social media language online, but this strategy raises privacy issues. Despite these limitations, the current study provides fundamental knowledge of the age-related differences in social media language use and emphasises the need for more research in this digital field.

5. Conclusion

The study aimed at investigating the use of social media language, frequency, and the context of its use by examining the polarity and subjectivity scores focusing on age groups in each of

the group conversations. The findings of the mock group conversations obtained through qualitative data analyses provide enough evidence to support the alternate hypothesis, which states that 'social media language can be impacted by age group to present an 'identity' within a conversation'. This conclusion was achieved by comparing age groups and themes being discussed in each group conversation and conducting topic classification along with studying the use of social media language to understand its impact on identity representation. However, one of the limitations of this study is that although these findings can be used to formulate a theory, the effectiveness of the theory can only be tested using theory triangulation, which will also further reveal the effectiveness of the results as well.

The results of this study contribute to expanding the knowledge of digital communication by focusing on how age groups influence the use of social media language and identity representation online. This study demonstrates that age groups influence how individuals express themselves on social media, providing new insights into developing identities in digital environments. This has implications for theories related to digital identity and communication, as it emphasises the need to focus on age-related differences when developing models for online behaviour and interactions. In terms of the practical implications, especially for the development of NLP tools, the findings of this study suggest that age-specific social media language features and patterns should be introduced and integrated into algorithms to improve content moderation, sentiment analysis, and user engagement strategies. Understanding these factors can help improve the design of digital platforms to better accommodate varied user demographics, resulting in more personalised and effective communication tools and platforms. Furthermore, this can assist in enhancing social media content strategies by recognising the varied communication preferences of different age groups.

For future work, the outcome of this study will be compared to the phase one data for conducting data triangulation to better understand the impact of age groups, personality, and social media language on identity representation in textual conversations. The data triangulation will be used to formulate a theory and communication model headed in the path of social media communication using textual conversations and online identity. The theory that will be developed from the data triangulation will be used to conduct a theory triangulation to give a more holistic understanding of the research topic as well as the results.

References

- Alshenqeeti, H, (2006). Are Emojis Creating a New or Old Visual Language for New Generations? A Socio-semiotic Study. *Advances in Language and Literary Studies*, vol. 7, pp. 56-69.
- Boutet, I., LeBlanc, M., Chamberland, J. and Collin, C. (2021). Emojis influence emotional communication, social attributions, and information processing. *Computers in Human Behavior*, vol. 119, pp. 106722.
- Berg, A., Giest, S., Groeneveld, S., Kraaij, W, (2020). Inclusivity in Online Platforms: Recruitment Strategies for Improving Participation of Diverse Sociodemographic Groups. *Research Symposium: Transformation of Government in the Era of Smart Technology*, vol. 80, pp. 989-1000.
- Bhandari, A. (2023, June 12). Using SpaCy for Natural Language Processing [Online]. Available from: <https://medium.com/swlh/using-spacy-for-natural-language-processing-3d127d5ca5ba>

- Cavalheiro, B., Rodrigues, D. and Prada, M. (2023). Who (and with whom) uses more emoji? Exploring individual, relational, and motivational characteristics driving emoji use. *Telematics and Informatics*, vol. 83, pp. 01-08.
- Cui, J. (2022). Respecting the Old and Loving the Young: Emoji-Based Sarcasm Interpretation Between Younger and Older Adults. *Personality and Social Psychology*, vol. 13, pp. 01-16.
- Chen, Y., Yang, X., Howman, H. and Filik, R. (2024). Individual differences in emoji comprehension: Gender, age, and culture. *PLoS*, vol. 19, pp. 01-20.
- Gibbs, J., Heino, R. and Ellison, N. (2006). Managing impressions online: Self-presentation processes in the online dating environment. *Journal of Computer-Mediated Communication*, vol. 11, pp. 415-441.
- Gaby, A. (2023, July 17). Python Sentiment Analysis using TextBlob and VADER for Glassdoor Reviews [Online]. Available from: <https://medium.com/@gabya06/python-sentiment-analysis-using-textblob-and-vader-for-glassdoor-reviews-cc9632babb73>
- Heino, D., Ellison, B. and Gibbs, J. (2006). Self-presentation in online personals: The role of anticipated future interaction, self-disclosure, and perceived success in Internet dating. *Communication Research*, vol. 33, pp. 152-177.
- Herring, S. C., and Androutsopoulos, J. (2015). *Computer-mediated discourse 2.0*. 2nd ed. New Jersey: Wiley-Blackwell.
- JISC, (2023, October 10). Online Surveys (formerly BOS) [Online]. Available from: <https://www.onlinesurveys.ac.uk/>
- Jupyter, (2023, August 21). Free software, open standards, and web services for interactive computing across all programming languages [Online]. Available from: <https://jupyter.org/>
- Kemp, S. (2023, January 26). Digital 2023: Global Overview Report [Online]. Available from: <https://datareportal.com/reports/digital-2023-global-overview-report#:~:text=There%20are%20now%204.76%20billion,growth%20of%20just%203%20percen>
- McCabe, D., Yurchisin, J. and Watchravesringkan. (2005). An exploration of identity re-creation in the context of Internet dating. *Social behaviour and Personality*, vol. 33, pp. 735-750.
- Nyumba, T., Wilson, K., Christina, D and Mukherjee, N. (2018). The use of focus group discussion methodology: Insights from two decades of application in conservation. *Special Feature: Qualitative methods for eliciting judgements for decision making*, vol. 9, pp. 20-32.
- Pietro, M. (2020, June 09). Text Analysis & feature engineering with NLP [Online]. Available from: <https://medium.com/towards-data-science/text-analysis-feature-engineering-with-nlp-502d6ea9225d>
- Poushter, J and Chwe, A. (2018, June 19). Social media use continues to rise in developing countries but plateaus across developed ones [Online]. Available from: <https://www.pewresearch.org/global/2018/06/19/social-media-use-continues-to-rise-in-developing-countries-but-plateaus-across-developed-ones/>

- Quan, A., and Young, L. (2010). Uses and gratifications of social media: A comparison of Facebook and instant messaging. *Bulletin of Science, Technology & Society*, vol. 30, pp. 350-361.
- Solent, (2023, April 12). Solent Informed Consent [Online]. Available from: <https://www.solent.ac.uk/research-innovation-enterprise/documents/informed-consent-form.pdf>
- Singer, G and Zalmanson, L. (2013). Content or Community? A Digital Business Strategy for Content Providers in the Social Age. *Digital Business Strategy*, vol. 37, pp. 591-616.
- SurveyCircle, (2023, April 14). Find Survey Participants with Survey Circle [Online]. Available from: <https://www.surveycircle.com/en/>
- SurveySwap, (2023, April 14). Find survey respondents now [Online]. Available from: <https://surveyswap.io/>
- Tang, Y and Hew, K. (2019). Emoticon, Emoji, and Sticker Use in Computer-Mediated Communication: A Review of Theories and Research Findings. *International Journal of Communication*, vol. 13, pp. 2457–2483.
- Varnhagen, C., McFall, G., Pugh, N., Routledge, L., Sumida, H., and Kwong, T. (2010). lol: New language and spelling in instant messaging. *Reading and Writing*, vol. 23, pp. 719-733.
- Warburton, S. and Hatzipanagos, S. (2015). *Digital Identity and Social-Media*. 1st ed. Pennsylvania: Information Science Reference.
- Walther. J. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behaviour*, vol 23, pp. 2538-2557.