



## Social Media Data for Mental Health Screening: Perspectives on Age Bias

Hamdah Hamid<sup>1</sup>

**Abstract:** During the past few years, there have been a surge in the use of social media. Various social media platforms have emerged such as Twitter, WhatsApp, Facebook and LinkedIn. The existing studies reveal that there has been a relation between social media usage and mental health. The objective of this study is to explore age biasness related to mental health data on social media platforms. Some studies have already been done on analysis of the effect of social media platforms on mental health. However, it has been observed that there is an issue of lack of inclusiveness towards the usage of social media platforms by older people. In other words, people of all age groups have not been using social media platforms. Hence, it is very hard to devise any conclusive statements based on these studies. This research is aimed towards the study of the lack of inclusiveness and biasedness prevalent on usage of existing social media platforms by diverse people. The paper highlights that the existing studies that have been done on the impact of social media usage on mental health has inherent biasedness issues. To perform the study, the paper employed an online dataset (Zeybus 2024). By means of interactive data visualizations, correlation analysis and discussions, the hypothesis has been substantiated. It was therefore observed that an inherent biasedness exists in social media data usage. This is due to an imbalance in the participation of people of different sects, races, genders and age groups. The paper has been structured as follows. First, the analysis has been done based on demographics. Using various questions, we determine what are the usage patterns of social media by these prospective users. In the end, the indicators of mental health of respondents were analyzed. A data pipeline was used, comprising preprocessing, descriptive analysis and exploratory data analysis (EDA). The paper concludes by showing that social media-based mental health data is heavily skewed toward younger users and certain demographic groups, limiting its reliability for broader population-level mental health assessments.

**Keywords:** *Age Bias; Mental health; Social Media; Data Bias; Demographic Analysis*

### 1. INTRODUCTION

During the past few years, we have seen rapid expansion in internet usage. At the same time, various social media platforms and their acceptance by masses have been observed. The social media growth has been found to be greatly impacting end-users. This has raised concerns about the impact of social media on different age groups. This ranges from physical health as well as mental well-being. There have been several studies (as we will see) that attempted to analyze the impact on emotional lives of people. However, there is an imbalance in the people who are using social media. As a result, the data collected for various studies are biased. There are people from different age groups and genders that are not using social media with same proportion. Hence, they are under-represented. This study aims to explore age biases in social media data. The study analyzed mental health indicators through a dataset that was obtained through an online survey. We try to understand how social media usage varies with age and

---

<sup>1</sup> NED University of Engineering and Technology, Karachi

its impact on mental health. It was concluded that there is an imbalance in their frequency of usage that can lead to biasedness in analysis.

The dataset was obtained based on responses from a 20-question survey conducted by [12]. The dataset contains different questions about spatiotemporal details, social media usage and different mental health indicators. Once the dataset has been obtained, preprocessing was performed including renaming of some columns. Afterwards, different data visualization techniques were performed over the columns like age, gender, relationship such that hypothesis can be validated. In the end various types of correlation analysis were performed.

The objective of this research paper was to uncover patterns and correlations underlying the data fields such that one can ascertain about age biases and the corresponding impact on mental health. The outcomes of this study can be used as target interventions helping in treating old age mental health diseases that occur because of the use of social media. The major contributions of this paper are: a) to highlight the lack of inclusivity in social media usage by people of various age groups b) To pinpoint the bias in existing studies on impact of social media on mental health. The organization of this paper is as follows. In the next section, the literature review is provided. This follows the methodology and then we discuss the results. The paper ends with discussion on limitations and future work.

## **2. LITERATURE REVIEW**

There have been several studies reported on the topic related to age biases and mental health as far as social media usage is concerned. The next section presents the studies proposed in literature from two perspectives: studies related to mental wellbeing and studies related to biasedness in the data.

Kross et al. [1] discussed how different social media platforms affect different psychological processes. Berger et al. [2] explored associations between social media usage on different genders such as gays, lesbian, transgenders etc. A similar study for Asians during the COVID-19 period was done in [3]. Lambert et al. [4] concluded that taking a break of 7 days from social platforms may help in improving mental well-being, depression and anxiety.

To explore the age bias in social media data for mental health, various studies involved understanding how people of different age interact on social sites. In addition, how these interactions influence the mental health. Literature has highlighted that there are age-related differences and associated biasedness in social media use statistics and the associated effect on mental health and well-being. One of the studies highlighted that the sensitivity to the usefulness and harms of social media differs across old and young people[5]. Despite, other studies show that there are stronger correlations among social media usage, anxiety/depression, and poor mental health outcomes for girls compared to boys [6].

In addition, it has been found that the effect of social media on once psychological health is not totally conditioned on usage patterns. It also depends on the motives/ intention behind the social media platforms' usage. It was found in various studies that the usage of social sites for personal purposes has been associated with better health. On the contrary, using social media for entertainment purposes is linked with poorer health [7]. We must also talk about the intention factor. The intention behind social media use can play a very important role in the said study. It is to be realized that social media data is highly susceptible to different types of biases. This includes positivity bias, social desirability bias, and self-idealization. These can change the true picture of an individual's mental health status. This ultimately can lead to poor conclusions by various studies and researchers [8].

Now, let's discuss the various studies reported in literature. The association between social media usage and depression has been analyzed in [9]. Several miscellaneous studies have been conducted on the mental wellbeing of older adults [10]. Based on the literature review, it has been found that besides the explosive volume of the research on the topic, there is no single opinion, and the results are far from being straightforward [11]. The next section now presents the proposed methodology for exploring age bias in mental health screening. This is followed by conclusions.

Based on this literature, the reviewed studies collectively highlight gaps and inconsistencies particularly regarding under-representation of older age groups which directly aligns with and motivates our research question on exploring age bias in social media-based mental health data.

### **3. METHODOLOGY**

This study analyzes potential age biases in social media data concerning mental health. The objective is to determine whether age influences patterns of social media engagement and mental-health-related responses. For this purpose, a dataset was considered that is available at [12].

#### ***3.1. Processing pipeline***

The following is the methodology adopted for the study. Several types of pre-processing were performed. This includes the following activities:

1. Data acquisition >> 2. Cleaning >> 3. Preprocessing >> 4. Analysis

The following paragraphs discuss these steps:

***Data acquisition:*** The preprocessing starts with data acquisition. The data was gathered from the particular source. It was ensured that the dataset comprises demographic and mental-health indicators. By employing these factors, one can perform a direct comparison and establish age bias.

***Data cleaning:*** After preprocessing, the next step is data cleaning. The data cleaning enhances the data quality. The null values were also removed. Data de-duplication was performed. This addresses the data inconsistency issues. The data cleaning ascertains the fact that one can compare between age groups.

***Preprocessing:*** This is a very important step and comprises tokenization, stop word removal, stemming and lemmatization. These preprocessing activities normalize the data and ensure further analysis can be done smoothly. Therefore, any text-based responses are in standard form. As a result, meaningful comparison across demographic groups can be performed.

***Analysis:*** After data preprocessing step, an array of statistical techniques is employed. This includes the step to perform inferencing and using data mining to extract the meaning from the data. By means of this stage, it is ascertained if the relation between social media usage and mental-health indicators deviate extensively. Specifically, it is checked that between age groups, potential age bias exists between age groups or not.

#### ***3.2. Dataset details***

The dataset was collected from the Kaggle website. The dataset comprises twenty questions. Before proceeding further, we will discuss the dataset in detail. Almost all the questions are

related to three key aspects a) demographics of participants, b) their social media usage, and c) their mental health indicators. In the dataset, the various types of survey responses were recorded in a comma-separated values (CSV) format. These acquired data were then analyzed using python.

**Table 1: Questionnaire/ columns for the survey**

1. 'Timestamp'
2. 'What is your age?'
3. 'Gender'
4. 'Relationship Status'
5. 'Occupation Status'
6. 'What type of organizations are you affiliated with?'
7. 'Do you use social media?'
8. 'What social media platforms do you commonly use?'
9. 'What is the average time you spend on social media every day?'
10. 'How often do you find yourself using Social media without a specific purpose?'
11. 'How often do you get distracted by Social media when you are busy doing something?'
12. 'Do you feel restless if you haven't used Social media in a while?'
13. 'On a scale of 1 to 5, how easily distracted are you?'
14. 'On a scale of 1 to 5, how much are you bothered by worries?'
15. 'Do you find it difficult to concentrate on things?'
16. 'On a scale of 1-5, how often do you compare yourself to other successful people through the use of social media?'
17. 'Following the previous question, how do you feel about these comparisons, generally speaking?'
18. 'How often do you look to seek validation from features of social media?'
19. 'How often do you feel depressed or down?'
20. 'On a scale of 1 to 5, how frequently does your interest in daily activities fluctuate?'
21. 'On a scale of 1 to 5, how often do you face issues regarding sleep?'

To facilitate analysis, the columns were renamed to more concise and descriptive names:

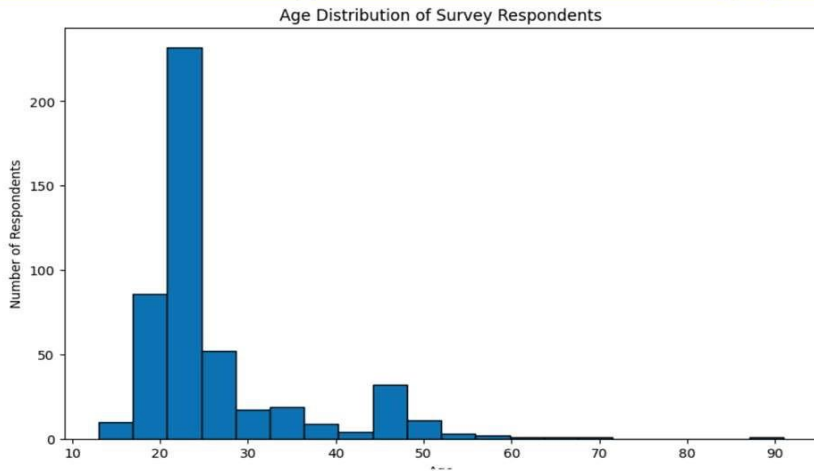
'timestamp', 'age', 'gender', 'relationship\_status', 'occupation\_status', 'affiliated\_organizations', 'use\_social\_media', 'social\_media\_platforms', 'daily\_social\_media\_time', 'frequency\_social\_media\_no\_purpose', 'frequency\_social\_media\_distracted', 'restless\_without\_social\_media', 'distractibility\_scale', 'worry\_level\_scale', 'difficulty\_concentrating', 'compare\_to\_successful\_people\_scale', 'feelings\_about\_comparisons', 'frequency\_seeking\_validation', 'frequency\_feeling\_depressed', 'interest\_fluctuation\_scale', 'sleep\_issues\_scale'

#### 4. DESCRIPTIVE ANALYTICS

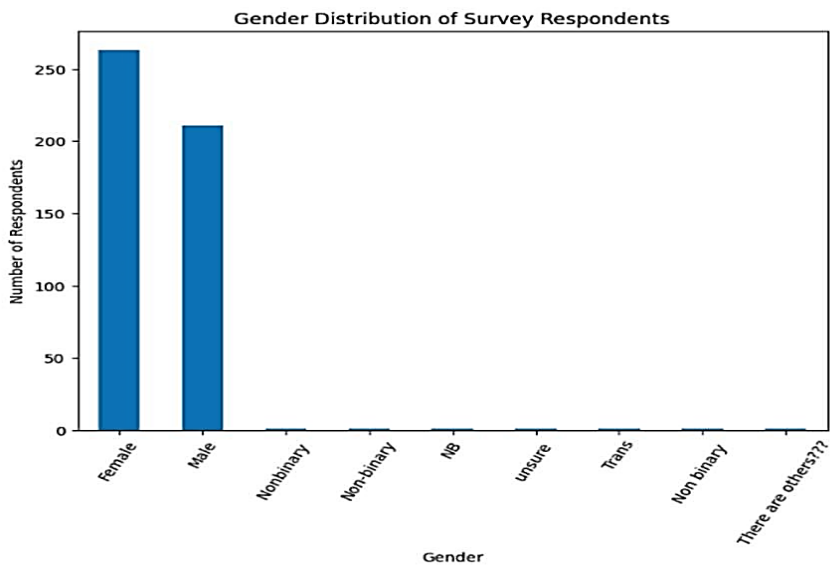
This section presents the descriptive analytics based on the survey results.

##### 4.1. Age distribution

Here the objective was to determine the average age of the survey respondents. The methodology used was that a histogram was created to visualize the age distribution. Fig 1 shows the results. As can be seen, most of the respondents were between 20-30 years old. This shows that there is a disparity in age for the usage of various social media platforms. People of old age are generally not using social media platforms. Hence, it can't be concluded that social media platforms can have an impact on mental well for all age groups. Fig 1 directly supports this by clearly showing a sharp concentration in younger age brackets, reinforcing the age imbalance in the dataset.



*Figure 1: Age distribution of respondents*



*Figure 2: Gender distribution of respondents*

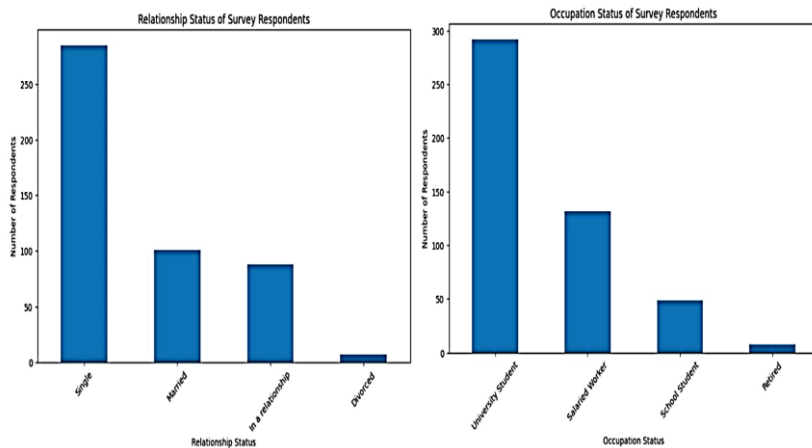
#### **4.2. Gender distribution**

Here, we analyzed gender distribution among the respondents. A bar chart was used to represent gender distribution. Most of the respondents were female (Fig 2). Again, there is a

disparity as far as people of various genders are concerned. In addition, people of different genders tend to use specific social media platforms as we will see later. This again highlights the bias in various studies on mental well-being. Figure 2 illustrates this imbalance visually, showing a noticeably higher number of female respondents compared to other genders, which shapes the interpretation of later results.

#### 4.3. Relationship and occupation status

Here, we identified the most common relationship statuses and occupation statuses among the respondents. Bar charts were created to illustrate the distribution of relationship and occupation statuses. Most of the participants were single, and university students (Fig 3). Figure 3 strengthens this finding by showing a clear dominance of single respondents and students, indicating that the dataset largely represents young academic populations.



*Figure 3: Relationship status and occupation of respondents*

#### 4.4. Social media platform usage

Here, we determined the most used social media platforms. A bar chart was generated to show the distribution of social media platform usage. Youtube and Facebook are the platforms used mostly by respondents (Fig 4). However, there are representation on other social media platforms such as Instagram, discord, snapchat, twitter etc. Figure 4 visually highlights these platform preferences, confirming that usage patterns cluster around a few major platforms rather than being evenly distributed.

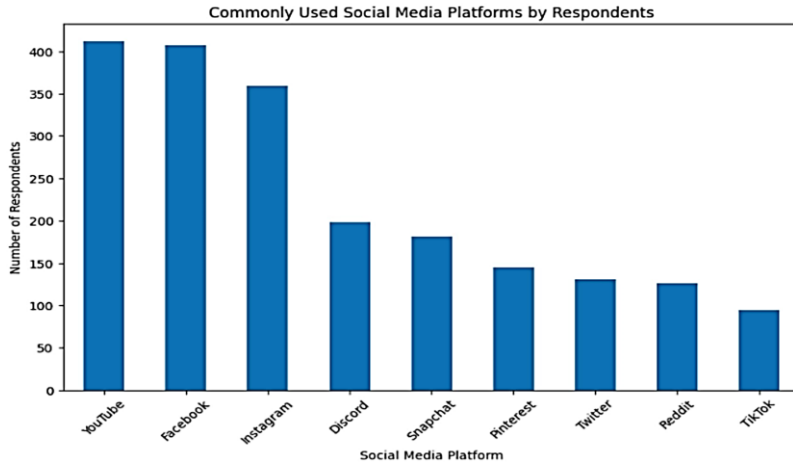


Figure 4: Social media platform usage of respondents

#### 4.5. Purposeless usage of social media usage

Here, we examined how often respondents use social media purposelessly and how this varies by age and gender. A heat map was generated to show the frequency of purposeless social media usage by age and gender (Fig 5). Social media usage without specific objective is widely seen in females and people between 13-19 ages. Figure 5 reinforces this conclusion by showing darker intensity in younger age ranges—especially among females—indicating higher purposeless usage.

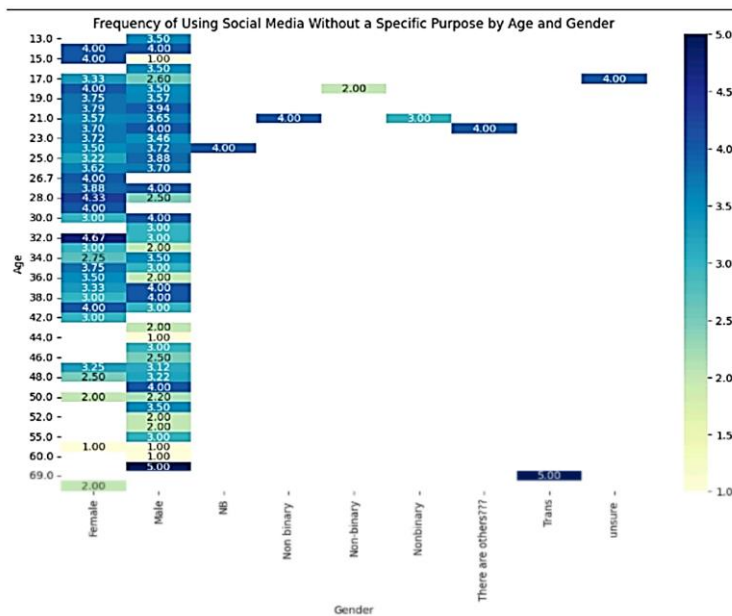


Figure 5: Social media usage and gender heatmap

## 5. EXPLORATORY DATA ANALYSIS

### 5.1. Correlation analysis

The correlation analysis investigates the correlation between the time spent on social media and feelings of restlessness when not using it. For this purpose, a joint plot is drawn in Figure 6 between *daily\_social\_media\_time* and *restless\_without\_social\_media*. There is trend observed in the joint plot that the high social media usage is associated with restlessness thus substantiating the hypothesis.

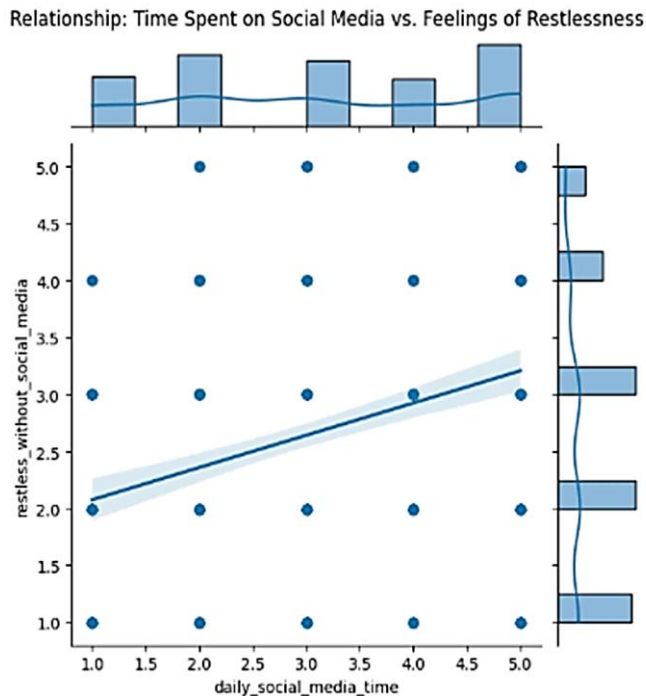


Figure 6: Social media usage and restlessness

### 5.2. Gender-based differences

To analyze the gender-based differences, box plot has been drawn. The objective is to assess what the impact of social media usage is on mental health. The box plot in Figure 7 shows that gender analytics highlighting distractibility and worry levels. The figure substantiates the hypothesis by showing visual evidence having variations in median values across genders.

### 5.3. Age-based differences

In this analysis, it was analyzed whether younger people report distractions more often as compared to older people while using social media. For this paper, scatter plot was drawn. As can be seen in Figure 8, x-axis represents 'age' and on the y-axis represent 'frequency\_social\_media\_distracted'. It can be observed that there is a downward trend with

respect to age. This validates that younger people are distracted more due to social media usage.

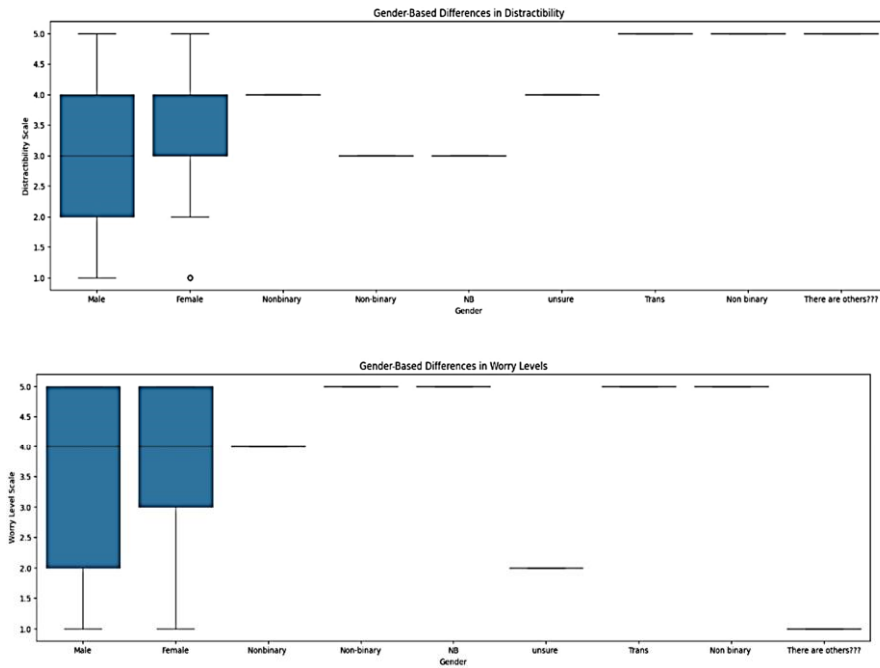


Figure 7: Gender based differences

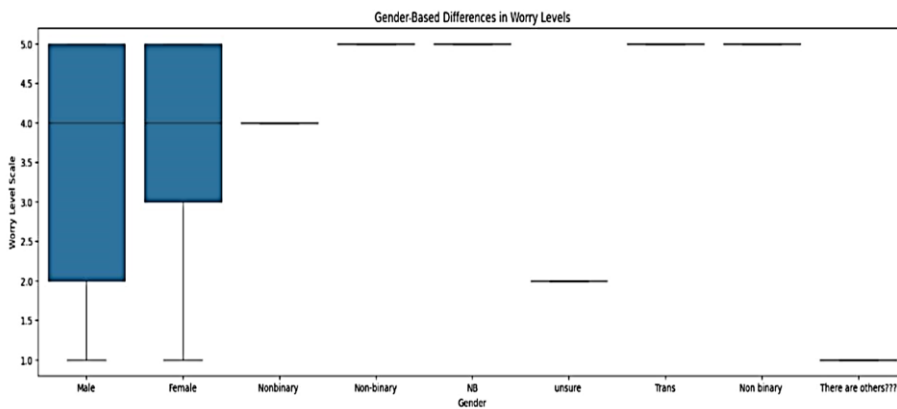


Figure 8: Age-based differences

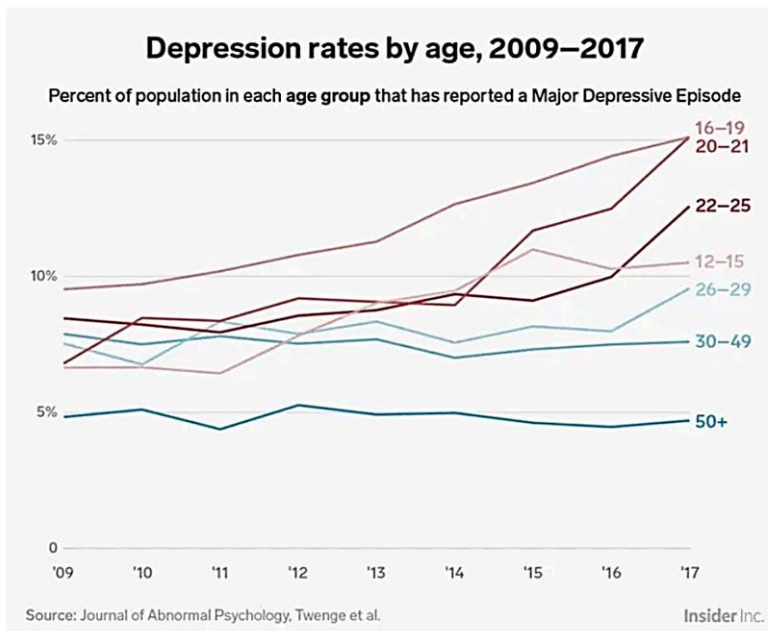
The analysis performed above explored the age-biasedness in social media usage pertaining to mental health. The exploratory data analysis helped in analyzing the underlying relationships present the data. It was concluded that there is a significant relation between age, usage of social media data and the underlying mental health implications.

## 6. DISCUSSIONS

In this section, discussions are provided based on the results presented in previous section. It will be discussed what are patterns uncovered related to age distribution, the various types of users active on social media, the corresponding impact on mental health. Finally, several ethical considerations are reported.

### 6.1. Patterns of age distribution

Figure 9 reports that the age distribution of people active on social media is not uniform. It has been found that people of young ages spanning from 18 to 29 years old are found to be using social media actively. These results are same for various social media platforms such as Instagram, Snapchat, and TikTok. In contrast, older people, over 65, are not using social media as actively as young people. Hence, the data is not available for people of this age to analyze impact on mental health. They are not well represented, and they use social media very rarely. It can be concluded that mental health screening data that is acquired from social media to analyze age trends are highly biased towards younger adults. Hence, most of these analytics are not reliable.



*Figure 9: Depression rates by age (Ibrahim et al. 2022)*

### 6.2. Variation of users based across social media platforms and demographics

Different social media platforms are appropriate / liked by different age groups and demographics. Take for example, the TikTok social media platform and Instagram have a

younger user base. As compared to this, Facebook has a more balanced age distribution. Still, there are fewer older ones on Facebook. Finally, twitter has a more serious userbase. In addition, the demographic factors such as sex/ gender, socioeconomic status, financial/ family status, and cultural background can also influence the usage frequency / trends of social media.

### ***6.3. Impact of Underrepresentation of Older Age Groups Accuracy and Reliability***

We can conclude that the low representation of older age groups in social media data can have biased mental health screening algorithms. In other words, these algorithms will not reflect the mental health issues of older adults. These algorithms would be trained in data of younger users. Their influences will be biased towards the younger generation. This can result in models that can't identify or properly evaluate mental health issues in older populations.

### ***6.4. How to deal with unavailability of metadata?***

For improving the accuracy of mental health studies and the associate data, it is very important to holistically acquire data about various ages and genders. For this purpose, the following paragraph highlights some important steps:

- Encouraging all sorts of users to provide their demographic information voluntarily.
- Having public profiles where age-related information is readily available.
- Applying data augmentation techniques/ generating synthetic data for underrepresented groups.

Machine learning (ML) and natural language processing (NLP) techniques can be adapted to reduce age bias. The way in which machine learning can help include:

- Implementation of fairness-aware algorithms. The algorithms can be adapted to weigh differently for different demographic
- Use of transfer learning algorithms: Transfer learning algorithms are based on pretrained models. Hence, knowledge about different demographics can be transferred to handle disparity issues
- Updating models more frequently or employing continuous training such that analysis and prediction can be performed in real-time and accurately

### ***6.5. Ethical Considerations Addressing Age Bias and Privacy Concerns***

Before proceeding towards conclusion, several ethical considerations are discussed here. To reduce age biasedness in analytics, following are some recommendations:

- Transparency: It is very important to clearly report how the data acquisition was performed, how they are processed and what are the protection mechanisms?
- Consent: Before the data can be used, consent must be taken from the user and they are clearly informed about the use in research and the implications
- Privacy: The identity of the user must be protected. The credentials, tokens, keys and personal data of the user must not be disclosed under any circumstances
- Inclusivity: To make any concrete decisions requires study on diverse age groups. As a result, a balanced and fair analysis can be performed.

Based on above ethical considerations implemented in true spirit, one can address age biasedness issues and we can observe a coherent and reliable analysis on mental health impact because of social media usage.

## **7. CONCLUSION**

This paper reported age biases on analytics performed over social media data related to mental health. Several studies reported this issue. For instance, [18] hypothesized that online behavior can be linked to mental health. For this purpose, a survey data set was employed. An interesting study has been reported in [19] where the concept of social virtual reality has been explored. The authors analyzed relationships between social media, social virtual reality and mental health. Another interesting study for a specific region has been explored in [20] where authors analyzed the relation between digital technology usage and depression among older adults specifically for Chinese region.

Carrying forward the existing study, this research provides direct evidence of demographic variations in both social media usage patterns and their associations with mental health indicators. The following are some key empirical findings extracted based on the analysis performed in this study:

- The females are generally involved in the usage of social media platforms. A causal effect analysis must be done to identify the reasons such as personal life, family issues, customs etc.
- A high proportion of the users are involved in non-purposeful or unstructured social media engagement.
- Younger individuals reported more frequent distractions and restlessness related to social media usage.
- On the other hand, gender differences were quite evident in and found to be distractibility and high worry levels.
- These findings reflect age-related patterns in social media behavior and the corresponding mental health effects.

The following are a few recommendations/ future works:

- Targeted interventions to evolve. These can address specific age group concerns.
- Future research should further be done on these patterns and their impact on mental health support strategies.
- To identify correlations between specific social media platform usage and gender, age groups.

## **REFERENCES**

- [1] E. Kross, P. Verduyn, G. Sheppes, C. K. Costello, J. Jonides, and O. Ybarra, "Social Media and Well-Being: Pitfalls, progress, and next steps," *Trends in Cognitive Sciences*, vol. 25, no. 1, pp. 55–66, Nov. 2020, doi: 10.1016/j.tics.2020.10.005.
- [2] M. N. Berger, M. Taba, J. L. Marino, M. S. C. Lim, and S. R. Skinner, "Social media use and health and well-being of lesbian, gay, bisexual, transgender, and queer Youth: Systematic

review,” *Journal of Medical Internet Research*, vol. 24, no. 9, p. e38449, Aug. 2022, doi: 10.2196/38449.

[3] S. Pan, C.-C. Yang, J.-Y. Tsai, and C. Dong, “Experience of and worry about discrimination, social media use, and depression among Asians in the United States during the COVID-19 pandemic: Cross-sectional survey study,” *Journal of Medical Internet Research*, vol. 23, no. 9, p. e29024, Aug. 2021, doi: 10.2196/29024.

[4] O. E. Smith, J. S. Mills, and L. Samson, “Out of the loop: Taking a one-week break from social media leads to better self-esteem and body image among young women,” *Body Image*, vol. 49, p. 101715, Apr. 2024, doi: 10.1016/j.bodyim.2024.101715.

[5] M. Politte-Corn, E. A. Nick, and A. Kujawa, “Age-related differences in social media use, online social support, and depressive symptoms in adolescents and emerging adults,” *Child and Adolescent Mental Health*, vol. 28, no. 4, pp. 497–503, Feb. 2023, doi: 10.1111/camh.12640.

[6] L. Winstone, B. Mars, C. M. A. Haworth, J. Heron, and J. Kidger, “Adolescent social media user types and their mental health and well-being: Results from a longitudinal survey of 13–14-year-olds in the United Kingdom,” *JCPP Advances*, vol. 2, no. 2, Mar. 2022, doi: 10.1002/jcv2.12071.

[7] H. Thygesen *et al.*, “Social media use and its associations with mental health 9 months after the COVID-19 outbreak: A Cross-National study,” *Frontiers in Public Health*, vol. 9, Jan. 2022, doi: 10.3389/fpubh.2021.752004.

[8] M. Tušl *et al.*, “Opportunities and challenges of using social media big data to assess mental health consequences of the COVID-19 crisis and future major events,” *Discover Mental Health*, vol. 2, no. 1, Jun. 2022, doi: 10.1007/s44192-022-00017-y.

[9] W. Xiao, J. Peng, and S. Liao, “Exploring the Associations between Social Media Addiction and Depression: Attentional Bias as a Mediator and Socio-Emotional Competence as a Moderator,” *International Journal of Environmental Research and Public Health*, vol. 19, no. 20, p. 13496, Oct. 2022, doi: 10.3390/ijerph192013496.

[10] H. Kang and H. Kim, “Ageism and Psychological Well-Being among Older Adults: A Systematic review,” *Gerontology and Geriatric Medicine*, vol. 8, Jan. 2022, doi: 10.1177/23337214221087023.

[11] S. Ahmed, S. Hina, R. Asif, S. Ahmed, and M. Ahmed, “Development and evaluation of Gold Standard Dataset for Sentiment analysis of tweets,” *Pakistan Journal of Engineering and Technology*, vol. 6, no. 4, pp. 7–12, Jan. 2024, doi: 10.51846/vol6iss4pp7-12.

[12] Zeybus, “GitHub - zeybus/Social-Media-Usage-and-Mental-Health,” *GitHub*. <https://github.com/zeybus/Social-Media-Usage-and-Mental-Health>

- [13] I. Song, S. R. Pendse, N. Kumar, and D. C. Munmun, “The Typing Cure: Experiences with Large Language Model Chatbots for Mental Health Support,” *arXiv (Cornell University)*, Jan. 2024, doi: 10.48550/arxiv.2401.14362.
- [14] Guo, Z., et al., “Large language model for mental health: A systematic review”. [Preprint]. 2024. <https://doi.org/10.2196/preprints.57400>
- [15] T. F. Heston, “Evaluating Risk Progression in Mental Health Chatbots Using Escalating Prompts,” *medRxiv (Cold Spring Harbor Laboratory)*, Sep. 2023, doi: 10.1101/2023.09.10.23295321.
- [16] A. H. Ibrahim, Z. C. Cob, S. M. Drus, A. A. Latif, H. M. Radzi, and R. Md. Anwar, “Using text analytics on social media posts to identify cues or features of depressive behavior,” *AIP Conference Proceedings*, vol. 2644, p. 030048, Jan. 2022, doi: 10.1063/5.0104446.
- [17] D. Bill and T. Eriksson, “Fine-tuning a LLM using reinforcement learning from human feedback for a therapy chatbot application”. *Bachelor Thesis, School of Electrical Engineering and Computer Science, Royal Institute of Technology, Stockholm, Sweden, 2023*
- [18] N. D. Nguyen, N.-A. Truong, P. Q. Dao, and H. H. Nguyen, “Can Online Behaviors be linked to Mental Health? Active versus Passive Social Network Usage on Depression via Envy and Self-Esteem,” *Computers in Human Behavior*, p. 108455, Sep. 2024, doi: 10.1016/j.chb.2024.108455.
- [19] S. X. Yao, J. Lee, R. M. Reynolds, and M. E. Ellithorpe, “Problematic social media use in 3D? Relationships between traditional social media use, social virtual reality (VR) use, and mental health,” *PLoS ONE*, vol. 20, no. 1, p. e0314863, Jan. 2025, doi: 10.1371/journal.pone.0314863.
- [20] Q. Wu *et al.*, “The association between digital technology use and depression among older people in China: a moderated mediation model,” *Frontiers in Psychiatry*, vol. 16, Feb. 2025, doi: 10.3389/fpsy.2025.1457967.