

Social media political hate speech and its implications on mental health: The need for digital literacy

Hamed Hussain Shah¹, Ghulam Abbas Balti^{2**} and Haleema Sadia Sherazi³

Abstract

The study focuses on political hate speech and its implication on mental health. Considering this, social media posts and responses of social media users through open-ended questionnaires provide the data for thematic analysis. The main source of data is the text of posts on the specific pages of mainstream political parties: Pakistan Tehreek Insaf (PTI), Pakistan Muslim League(N), and Pakistan People's Party Parliamentarian (PPPP). Sampling is convenient, and only those posts are taken which are against other parties or leaders during the general election of 2024 in Pakistan. For validation of analysis of the posts, the survey shows how political hate speech effect mental health and helps in understanding the role of digital literacy. Thematic analysis of posts and responses are analyzed using the social identity theory and framing theory. The key findings related to mental health are increased anxiety, stress, social isolation, depression, decreased trust in institutions and media, and hopelessness. With platforms such as Twitter, Facebook, Instagram, and YouTube serving as a source of political information, these spaces have become increasingly polarized, often hosting divisive and hostile rhetoric. The prevalence of political hate speech, characterized by identity-based hostility, dehumanizing language, and moral accusations, poses challenges to individual's mental well-being and contributes to a culture of social intolerance and conflict. Participants express a sense of societal disillusionment, mentioning the way political hate speech encourages division and decreases trust in political processes and intergroup harmony. The findings also suggest that a multifaceted approach, combining personal resilience strategies with digital literacy and platform moderation policies, is essential for encouraging a healthier online environment. This study contributes to understanding the mental health implications of political hate speech, advocating for both individual and systemic solutions to reduce the spread and impact of divisive rhetoric.

Keywords: political hate speech, dehumanizing language, mental health, digital literacy

¹ Riphah International University, Islamabad – Pakistan

² University of Baltistan, Skardu – Pakistan

³ Beacon House School System, Islamabad – Pakistan

Introduction

The present study deals with the use of social media discourse related to political hate speech. Furthermore, it investigates the impact of political hate speech on the readers and the way they try to mitigate its impact for their better mental well-being. As social media is one of the powerful tools to disseminate political information and build a political discourse in the present era, political information is mostly used and spread by the social media users either in favor or against the political parties. Social media users, especially political workers, show their affiliation with certain political parties through the posts that represent their sentiments either in favor or against the political parties. Usually, the posts highlight the language that favors their own party and against the opponents. Such use of language in the posts often affects the mental health of the readers, especially when the post is against someone's beliefs and thoughts. In this situation, the use of ridiculing words or political hate speech emerges as a challenge and becomes a tool to impact the mental health of readers. This study explores such posts and impact of social media political hate speech on mental health, specifically focusing on Pakistan's political landscape.

Political hate speech has become an essential element of online discourse that carries implications for not only mental health, but also social stability. Taking the broader perspective, it promotes a language that has hostility, dehumanization, and discrimination against political groups or individuals (Davidson et al., 2017). Furthermore, political hate speech becomes a tool for social exploitation and political division bringing in 'us vs them' mentality (Bilewicz & Soral, 2020). Different studies highlight that such rhetoric of division can have detrimental effects on mental health which increase stress, anxiety and weakens social cohesion (Cramer et al., 2016). This study aims to explore the themes within political hate speech, examining its psychological impact on individuals and highlighting the role of digital literacy.

Political hate speech is often fueled by in-group and out-group dynamics, where individuals categorize themselves and others based on political affiliations, developing biases and prejudice (Tajfel & Turner, 1979). This concept relates to Social Identity Theory (SIT) that motivates self-identity relating to a group and leads to favoritism towards in-group and hostility towards out-group (Tajfel & Turner, 1979). In the political setting, hate speech frames political opponents not only as rivals, but also as a national threat to identity that intensifies negativity and mental

strain among both opponents and supporters (Bohner & Dickel, 2011). Additionally, hate speech aimed at political figures or groups employs dehumanizing language, reducing individuals to stereotypes or derogatory characters (Haslam, 2006). Use of such language desensitizes the audience to aggression and reduces empathy creating a hostile environment online and offline (Bandura, 1999).

Framing Theory (Goffman, 1974; Entman, 1993) suggests that the way hate speech is used influences how people interpret and internalize it. Frames that depict opponents as corrupt, threatening, or morally inferior evoke anxiety or hostility among readers, leading to cognitive and emotional stress (Matthes, 2012). Frequent exposure to political hate speech can reinforce negative perceptions that develop vulnerability among readers, especially among individuals who identify with the targeted group (Cikara & Fiske, 2013). For instance, individuals who perceive their political affiliation as under threat may experience chronic stress, social isolation, or even low self-esteem (Pietrabissa & Simpson, 2020). These factors emphasize the importance of digital literacy to counter the negative effects through critical thinking and resilience among social media users (Livingstone, 2007; Buckingham, 2015). Considering these points, social media is continuously triggering the mental state of individuals. This normally is in the form of negative emotional stress, resulting in low self-esteem, anxiety, and social isolation. Such a state mostly evolves into either a response in the form of political hate speech or an effort to find ways to mitigate it through different means to improve mental well-being. This study is considering all these points to highlight the perspective of respondents which can be generalized to broader population.

Considering these aspects, this study uses thematic analysis to identify patterns within political hate speech, focusing on how language reflects and reinforces in-group/out-group divisions, frames political opponents, and impacts mental health. Braun and Clarke (2006) discuss thematic analysis as a qualitative method to examine textual data in detail, allowing researchers to uncover underlying meanings and themes within hate speech. The analysis is grounded in social identity theory and framing theory to interpret the themes within political hate speech and the themes of mental health highlighted by the respondents of open-ended questionnaire. Iordache and Marien (2017) highlight that digital literacy is an important skill for navigating and critically assessing harmful rhetoric online. Considering this as part of the study, the findings related to digital literacy are based on the themes from the responses of individuals. By exploring the themes

and psychological impact of political hate speech, the researcher tries to understand the complex relationship between divisive rhetoric and mental health. Furthermore, it highlights the pressing need for digital literacy initiatives that empower individuals to recognize and resist the manipulative framing within hate speech. Taking this into account, the researcher has selected Facebook pages of major political parties in Pakistan: Pakistan Tehreek-e-Insaf (PTI), Pakistan Muslim League-Nawaz (PML-N), and Pakistan Peoples Party Parliamentarian (PPPP). By analyzing the textual content of these parties' social media posts and conducting a mental health survey, this study focuses on the psychological impact of such political hate speech.

In recent years, Pakistan has witnessed a surge in political polarization on social media to use politically motivated hate speech. This rhetoric not only affects political discourse, but also has significant implications for mental health, especially when individuals face identity-based hostility, dehumanization through language, toxicity, and accusations based on morality. Despite political hate speech having wide impact, there is limited research related to its impact on mental health, the strategies individuals use to mitigate the impact, and effective digital literacy strategies to minimize its impact. The study helps to explore the themes of polarization, dehumanization, and victimization while examining user-driven strategies to reduce the impact. By examining the content of social media posts by major political parties (PTI, PML-N, and PPP) and analyzing the psychological impact on their audience, this research aims to contribute to a growing body of literature on the implications of political hate speech. Taking these concepts into consideration, this study seeks to address the following research questions:

- How does the language of posts in different social media groups by individuals contribute to the spread of political hate speech that impacts audience perception and intergroup relations?
- What are the emotional and psychological effects of political hate speech on social media users?
- How do individuals exposed to political hate speech minimize its impact and what digital strategies do they use?

Review of the Literature

The relationship between political hate speech, mental health, and social media usage has gained academic attention as the digital public sphere has taken over as

major platform for political discourse. This review critically examines existing literature on political hate speech, its psychological impacts, and strategies for mitigating its effects, contextualizing the findings of the current study within broader theoretical frameworks.

Political hate speech has been widely studied as a manifestation of digital polarization. Researchers argue that hate speech often serves as a rhetorical tool to delegitimize opponents by employing identity-based hostility and dehumanizing language (Citron & Norton, 2011). For example, online political discourse uses derogatory terms to belittle opponents, reinforcing in-group/out-group divisions (Brown, 2015). The current study builds on these findings by identifying dehumanization, ridicule, and moral superiority as key rhetorical devices in political hate speech. This concept aligns with Social Identity Theory (Tajfel & Turner, 1979), which explains that individuals develop self-esteem from group affiliations, often leading to belittling of out-groups to strengthen in-group identity. Rafi and Shafiq (2024) argue that the followers of political parties recycle hateful language in verbal confrontation for the defense of their political ideology, which reflects that they want to strengthen their group. Political hate speech, as observed in the analyzed posts, reflects this phenomenon, where opposing parties are framed as national threats or morally inferior.

Exposure to hate speech has been linked to psychological effects, including stress, anxiety, and depression (Saha et al., 2019). Repeated exposure to hostile political rhetoric increases these effects, creating a toxic environment that undermines individual well-being. The concept of online harassment suggests that such exposure can result in psychological distress and less trust in social systems (Hinduja & Patchin, 2012). The findings of the current study corroborate these results, showing that political hate speech fosters feelings of frustration, hopelessness, and emotional strain. However, the study also highlights individual resilience strategies, such as selective engagement and emotional detachment, which are less explored in existing literature. This suggests a gap in research on how coping mechanisms mitigate the mental health impacts of hate speech.

Social media platforms play a pivotal role in facilitating hate speech due to their algorithm-driven content visibility (Gillespie, 2018). Social media groups often prioritize engagement over content moderation, allowing divisive rhetoric. While some scholars advocate for stricter content regulation and community guidelines

(Ganesh & Bright, 2020), few other highlight potential implications for free speech (Livingstone, 2008). However, this study highlights a need for user-driven initiatives, such as promoting respectful dialogue and digital literacy, offering a balanced perspective between systemic and individual-level solutions. Gillespie (2020) suggests that content moderation can be used for bigger data. For that, media literacy has been proposed to counter online hate speech by developing critical thinking and digital resilience (Livingstone & Helsper, 2007). In this context, Rheingold (2012) suggests media literacy programs can empower users to identify misinformation, resist hateful content, and engage constructively in political discourse. The current study reinforces the importance of media literacy, identifying it as a key strategy for mitigating the effects of political hate speech. The existing literature addresses the prevalence and impact of political hate speech, the psychological and behavioral strategies which individuals use to cope with hate speech are less explored. The current study explores individual strategies, such as emotional detachment and media literacy, and user driven responses for digital literacy.

Methodology

❖ Research Design

A qualitative research design is used to examine the complex and nuanced ways in which political hate speech is constructed and its effects on mental health. Thematic analysis, as outlined by Braun and Clarke (2006), is employed to systematically identify patterns or themes within hate speech and in the responses of individuals. His method is flexible in nature and focuses on familiarization with data, generating initial codes, searching for themes, reviewing themes, defining and naming these, and producing the report. For the present study initial codes and themes were generated from the selected political posts and the responses of participants, and then both were aligned in a way that there is a connection between the posts and mental health. The subjectivity of the researcher remains in the process of coding. In this scenario, Brain & Clarke (2021) emphasizes the reflexive thematic analysis, which values the subjectivity of a researcher as an interpretative story rather than just patterns. The researcher considers that the whole patterns, however, provide a story in the form of interpretation, which is vivid in the analysis.

❖ Data Collection

Posts are collected from social media platforms i.e., Facebook groups keeping in mind that political hate speech is prevalent in the posts. The data reflects diverse language and tone used by the members of various political groups. The posts are usually in favor of their political party, but the posts against the political parties are purposively found. For data collection, I have chosen the months of January and February 2024. This period reflects the sentiments of political workers in Pakistan as the general election were held in the month of February. To undermine the opponents, group members of political parties represented their own parties with positive sentiments, however, opponents were labeled with negative sentiments. A purposive sampling method is used to select content from separate groups of political parties: Pakistan Tehreek-e-Insaaf (PTI), Pakistan Muslim League Nawaz [PML(N)] and Pakistan Peoples Party (PPP). At minimum five posts were chosen from each group of political parties. After understanding the trend, I have chosen seventeen posts as primary data. The collected data is categorized based on the type of hate speech employed.

To understand the implication of political hate speech, a qualitative open-ended survey is conducted that has five items focusing on four aspects: social media usage, exposure to political hate speech, mental health impact, and social media literacy.

The questionnaire was shared through Google forms with the participants who aged between 18 to 25 coming from diverse backgrounds and areas of Pakistan. 49 participants answered the questionnaire.

❖ Data Analysis

For familiarization with data, a thorough reading of the posts is done. The researcher identified the tone of political hate speech in the text. I took the initial notes on prevalent expressions, sentiments, and apparent themes. A systematic thematic analysis is conducted on social media posts from the official accounts of the three major political parties in Pakistan: PTI, PML-N, and PPP. For generating codes, words, phrases, and sentences are identified which reflect hostility, dehumanization, and perceived threats. Dominant codes are threat, moral superiority, and dehumanizing language. These codes are based on the theoretical concepts of SIT that capture group distinctions and framing theory that highlight framing techniques to shape perception. Based on similar codes that form identity-based targeting theme category patterns of traitor, enemy, and immoral are

identified that are related to political hate speech. The theme of dehumanizing language portrays opponents as inferior.

All the themes are distinct and cover the entire data set. To avoid the overlap, themes are refined with a focus on capturing the nuances of hate speech rhetoric. All the themes are clearly defined e.g., Us vs Them narrative identifies division. On the other hand, moral and ethical condemnation frames opponents as morally inferior. Social identity theory examines such themes which reinforces out group and in group identity in a hostile situation. However, framing theory reflects such themes that shapes political perception of opponents in the form of threat or morally corrupt. For validation secondary, secondary data of survey questionnaire is used to ensure that themes of political hate speech have relation to the mental health issues. The relation of hostile framing with mental health is created through open ended questionnaire which reflects the impact of political hate speech and its implication on the mental health of individual who are exposed to it.

Theoretical Framework

To understand the nuances of political hate speech and its mental health implications along with the digital literacy, social identity theory (Tajfel et al., 1979) helps to explore the in group and out group favoritism and discrimination. Political hate speech often reflects such division where in group supporters vilify out group opponents to strengthen group identity and cohesion. On the other hand, social identity theory is used to understand the hostility of a group towards the out-group and reinforces in-group identity. This helps to align the derogatory and dehumanizing language with framing the opponents as threats. In such a scenario, when individual feel that they or his/her party is targeted as out group, it may impact the mental well-being that can lead to anxiety, stress, and social isolation.

Framing Theory, proposed by Goffman (1974) and further developed by Entman (1993), suggests that how information is presented (or "framed") influences audience perception. Political hate speech often employs specific frames, such as "us vs. them" or moral degradation, to shape public opinion and rally support. Framing theory helps in the analysis of language used in political hate speech to identify common frames. Understanding these frames provides a way to explore the psychological impact of repeated negative portrayals. Analyzing the frames in hate speech and

responses gives crucial understanding for digital literacy considering the critical evaluation and mitigation of psychological impacts.

Results and Discussion

❖ Political Hate Speech

Thematic analysis is used to identify key themes and patterns in the primary data of social media posts for political hate speech, a qualitative open-ended survey is conducted through google sheet to understand the participants responses after reading the politically polarized posts. The survey also identifies the possible mental issues that result due to political hate speech; and understand what the participants suggest for digital literacy. Participant data is kept confidential and anonymized. This study aims to provide a comprehensive understanding of the complex relationship between social media, political hate speech, and mental health.

❖ Political Polarization Based on Identity

Political hate speech in the data highlights intense polarization, where political identities are framed as inherently harmful or unpatriotic. Statements often equate political loyalty to moral and national values, portraying opposition as a betrayal to their own ideology. In "Mujhe PTI sey shadeed nafrat hey, kyun ke myjhe Pakistan sy muhabbat hey" (I hate PTI very much because I love Pakistan - PML 2), hate for a political entity is directly tied to patriotism, suggesting that opposition to the individual belonging to PML, and equates it to disloyalty to the nation. In other words, he aligns his identity to the party and to the national ideology by negating the the ideology of PTI. In the post "PTI aur Muslim league ko vote denay walay Pakistan ki tabahi main brabar ky shareek hain" (Those who voted for PTI and Muslim League are equally involved in the destruction of Pakistan - PPP 6) blames the voters that their right to vote is under question, and it would destroy Pakistan if they vote for PTI and PML. This rhetorical strategy is used to attract the voters for his/her party aligning it to national and political identity. These textual examples are related to identity-based hostility. The use of words 'nafrat'(hate) and tabahi (destruction) are tagged with out-group parties. Such hate speech portrays opponents as a threat and develops intolerance and hostility in political discourse. In other words, it is used as a weapon to deepen the division between the people who consume such posts.

❖ **Dehumanizing Language**

One of the prominent aspects in the data is the use of dehumanizing language. This is used to belittle political opponents with derogatory terms to challenge the presence and authority of political parties. In "Nawaz Sharif ki apne dono halqon main kuttay wali ho rahi hey" (Nawaz Sharif is being treated like a dog in both of his constituencies - PTI 2), animalistic language is used to humiliate a prominent political figure with the use of word 'kuttay wali' (like a dog). This expression is commonly used in Pakistani culture to dehumanize a person and equating it to a dog as dogs are considered stranded and of no use in the context. The sole purpose is to humiliate the political figure. In "Lahore highcourt faisla karey tu theek, Peshawar highcourt karey tu ghalat, shabash chotay khotay per" (The Lahore High Court made a judgment, so it is right, the Peshawar High Court made a judgment, so it is wrong. Well done little donkey! - PTI 4) metaphors are implied to undermine the decision of judiciary and relating it to the political figure to 'khotay' (donkey). The term is used to ridicule the opponent party's head as donkey is considered as an animal having no brain which implies that he/she may have no brains to critically analyze the situation and decision of two different courts in favor or against a party. Such dehumanizing and derogatory rhetoric shows hostility and alienation and respect for other party members which creates a hostile environment leading to deepening the division.

❖ **Moral and Ethical Superiority**

In the data, moral and ethical superiority is evident where the political opponents are framed immoral and corrupt, and prioritized one's own party morally and ethically upright. This is done to delegitimize the others. In "Kaptan ka jail sey paigham: Mere Pakistaniyo, kabhi chori na karna, mulk sy ghadari na karna aur zana sey bachna, warna mere jaisa haal ho ga" (Captain's message from jail: My Pakistanis, never steal, never do treachory the country, and avoid adultery, otherwise you will end up like me - PML 3), morality is evoked using the terms 'chori' (steal), 'zana' (adultery), ghadari (treachory), and relating them to the political figure without any legitimate evidence. Posting these vices, the person is highlighting the consequences for him that political opponent has all these vices. By doing this, he is putting him and his party at the high pedestal of morality. Such rhetoric framing leads to hate speech for the political opponents without evidence and showing intolerance towards others. In "Jo jitna zyada badtameez ho ga, PTI mai sab sy barray ohday ka haqdaar ho ga" (Whoever is more ill-mannered will be entitled to a position in PTI - PPP 3), the writer used the word badtameez (ill-mannered) for the

opponent political workers by dismissing their ethical standards for a position within a party. This reflects strong ethical opposition, and a hostile frame to discredit opponents by moral and ethical condemnation of others and making him/herself morally and ethically superior.

❖ Injustice and Sarcasm as Political Discourse

The data shows the narrative of injustice and the tone of sarcasm in the statements emphasizing humiliation, biased political support, and suffering faced by political supporters and the leaders. Humor and sarcasm are employed to demean the opponents' actions to create a dismissive and contemptuous political discourse. The data highlights unjust targeting and grievance. In "Raat ke faisley ny youthiyon ko sonay nahi diya, isliye faisla mehfooz ker liya, kabhi bhi sunaya ja sakta hey" (The youthias were not allowed to sleep at night, hence the decision was reserved, it can be announced at any time - PML 6) indicates the use of words 'faisla mehfooz' (decision reserved), which is perceived as a tactic by the judiciary to overturn the actual decision. The rhetoric is to undermine the expectation of opponents thus provoking injustice and illustrating victimhood. In other words, the speaker is satisfied with the victimization of opponents with a sarcastic tone of uncertainty for opponents that they could not sleep. The tag of 'youthias' reflects is also to demean their existence to have solace and justice. Furthermore, in "Maulana sahb ka gussa bajaa hey, maulana ko dhandli mai matlooba hissa nahi mila" (The anger of Maulana Saheb is relevant; Maulana did not get his desired share through rigging - PPP 5) provides an indication of feelings through the words 'gussa bajaa' (anger is relevant), and sarcastically tagging them victim with 'hissa nahi mila' (not got the share). This victimhood, injustice, and unfair treatment is portrayed with a sarcastic tone to exclude the opponents from the judicial and electoral system rather than inclusion. These categories display the perceived injustice and deepen the political divide based on the tone of hate speech in the rhetoric. In "Social media per Tehreek e Insaaf ke numbers pooray ho gaye. Umer Ayub Facebook, Twitter, aur Tik Tok per jald halaf uthayen gey" (Tehreek-e-Insaaf's numbers on social media are complete. Umer Ayub will soon take oath on Facebook, Twitter, and Tik Tok - PML 1) The speaker mocks the opponent party and its leader with a sarcastic use of social media as the electoral college for PTI and they do not have the basis in actual electoral system. Hence, their leader would be the prime minister of social media rather than in real life. Oath taking is a solemn activity to be a legitimate ruler of a country. However, in this case, the opponents are ridiculed for undermining this activity for the opponents. In "Patwariyon ka naara: Mian jadoon aaway ga tey lag pata jaavay

ga, Naara al jadeed: Miyan jadoon nasay ga tey kissi nu nai dasay ga" (Slogan of Patwaris: When Mian comes, you will get to know. New Slogan: When Mian runs, he will not tell anyone - PTI 1) the speaker satirizes the slogan of opponent party by creating a new slogan. The original slogan eulogizes the party head as the savior with the lag pata jaavay ga (you will get to know) by announcing his return from foreign as political turn in the country. However, the new slogan created by opponents relates it to history of the leader that he will run from country and kissi nu nai dasay (will not tell anyone). Same is the case in Nawaz Sharif Aae ga, tey yadda javey ga. (When Nawaz Sharif comes, he will be fucked- PTI 5), where the return of the leader is mocked with the toxic word of yadda (fucked), which is an unwanted word in the cultural and political setting of Pakistan. The word itself is used to hurt the sentiments of the opposition negatively. The mockery is to reduce the opponents and undermining them with sarcastic tone of mockery at their leader. The statements show that humor is used to delegitimize the opponents and their political discourse.

❖ Threat to National Cohesion

The data set highlights that political opponents are framed as a threat to the nation. Their political actions are associated with betrayal and harm to national cohesion and integrity. The opponents are portrayed as disloyal to the nation for their political activities even though they remain in the legitimate electoral system. In "PTI ke azaad umeedwaar vote maang ker sharmina na karain, kyun ky hum ny israeli products ka boycott kiya hua hey" (PTI's independent candidates should not embarrass themselves by asking for votes, because we have boycotted Israeli products - PML 5) opponent candidates are associated with the foreign elements which are perceived as a threat to Pakistan. The perception is foregrounded in the historical context of non-acceptance of Israel as a country. The alignment of public boycott against the Israeli products reflects the public sentiments. However, the speaker attaches these sentiments to the opposing candidates that they need not ask for vote as they have already boycotted them. This highlights that the opponents are considered a threat to national loyalty. Hence, the public sentiments are evoked through this post against political opposition through hate speech. In "Pakhtun Khwa pichlay 15 saal sey tosha chorr ko vote detay hain. Ilaj aur rozgaar key liye Punjab aur Sindh ka rukh kartay hain" (Pakhtun Khawa has been voting for Tosha Khana thief for the last 15 years. People turn to Punjab and Sindh for treatment and employment - PPP 4) regional division is created to undermine the opposition who has government in one province calling them Tosha Khana thief. This division is

embedded in the rhetoric of no deliverance in one province which makes them ineligible for rule as they could not provide facilities to public in one province. Contrary to that article 15 and article 26 of constitution of Pakistan allows every citizen to have free movement and access to public place, respectively. Therefore, such hate speech is harming the national cohesion, but the speaker blames this on the opposition posing them as a threat by undermining the voters and the opposition.

The textual analysis elaborates that the data has the elements of political hate speech. For this, content creators use the rhetorical strategies of identity-based hostility, dehumanizing language, moral condemnation, ridiculing the opponents with sarcasm, and framing them as a threat to national cohesion. These strategies develop political polarization and an environment of hostility and toxicity. On the one hand the speakers want to emotionally appeal to their own party but hurt the emotional sentiments of opponents. These themes reflect the political hate speech in their rhetoric which has effects on mental wellbeing of opponents. Through such speech public narratives are developed and recycled to undermine the opposition and unconsciously a polarized political environment is created.

❖ **Exposure to Political Hate Speech and Mental Health**

Social media is used to have political information and awareness. For that, open ended responses of 50 participants are used to understand their usage of social media, exposure to political hate speech, impact on their mental health, and digital strategies they used to reduce this impact. Responses show that they use different platforms: X, Facebook, Instagram, and YouTube, for political discussion, awareness, and getting information about political events. The participants highlight that exposure to political content has a relation with mental health reporting it as the source of frustration and stress. Their responses highlight different patterns of exposure to political hate speech that are related to mental health.

❖ **Emotional Toll of Toxic Speech and Resilience**

Participants report that continuous exposure to political hate speech develops negative sentiments. In case the speech is against the party, which they like or support, mostly develop the negative sentiments and feelings. The sentiments are reported as an emotional toll with the words – frustration, sadness, anger, defensive, anxious, and demotivation. These words show that the consumer of hate speech has

embedded hostility and toxicity which heightens the emotional reaction. The respondents also report that they carry this emotional toll for a longer time which outburst in the form a response against the political party they do not like which works as a catharsis. Such catharsis evokes hate sentiments and frames the opponents as enemies further creating the speech of hate. In other words, it affects psychological well-being and hate speech continues to remain in a cyclic situation, where there is no end to it. They also report that exposure to HS causes strong negative emotions that develop hopelessness and pessimism for the country's future. Such negative emotions often heightened with the political situation of the country. Along with hopelessness and pessimism, they used the words of stress, anxiety, depression for their own mental health which developed due to continuous exposure of political hate speech against their party from different social media platforms. On the other hand, some of the respondents show resilience. They report the words- unaffected, indifferent, avoidance which highlight that either they are not related to any political party, or they avoid such speech. Their response also reflects that their selective engagement or disengagement is to manage any emotional impact. However, they feel emotionally disturbed if they engage in hate speech.

❖ **Reinforcement of Belief**

Respondents report that they follow or support a political party based on their awareness and knowledge from political discourse on different platforms. Social media as the main source of political awareness is common in their responses. Belief and political ideology are reinforced through hate speech. They report words like political belief, loyalty, and existing knowledge. The beliefs and existing knowledge are either in favor of their own party or against the opposition party. The hate speech against the opponents reinforces their belief of disliking and hatred. However, the same reinforces the existing patterns in the favor of their own party. In both cases, reinforcement works as a mental mechanism of defense for own party and hatred for other parties. Positive and negative emotional frames are prevalent that have a relation with the mental implications as they do not change the views remain rigid after exposure to political hate speech. The rigid and biased approach of political ideas based on hate speech works as a strategy to minimize mental strain.

❖ **Strategies to Reduce Political Hate Speech**

Considering the importance of digital literacy, the participants were asked about the possible strategies they use, and think can be effective to educate the social media users about hate speech and ways to reduce the mental health implications. The responses highlight the themes from regulatory policies to collective and individual responsibilities.

❖ **Policy Enforcement and Social Media Education**

Participants who are university students have a great deal of knowledge about the usage of social media at local and international levels. Taking into their background knowledge, the patterns from their responses focused on the responsibility of government. Their perspective highlights that government has different rules and regulations for all the media bodies though there must be freedom of speech under the obligatory article 19 of constitution of Pakistan, which also impose reasonable restriction under law e.g., for print media: Press, Newspaper, News Agencies and Books Registration (PNNBR) Ordinance, 2002 is in place which gives The press Council of Pakistan an authority to ensure and monitor ethical standards of print media. The continuous checks regulate the print media to avoid unethical and unlawful practices. Similarly, there is an authority Pakistan Electronic Media Regulatory Authority (PEMRA) Ordinance, 2002 to regulate electronic media. There are laws to regulate social media including Prevention of Electronic Crime Act, 2016, and CP (Against Online Harm), 2020; however, there are no guidelines for social media users. These laws focus on criminal activities and have no mechanism for the moderation of social media. For moderation and regulation respondents use the words: filter, reporting tools, government regulations, social media syllabus, educate, and grievance reporting. Their responses suggest that hate speech is a new phenomenon which requires regulation at platform level and after implementation of social media policy, there is a need to enforce it as well. For the guidelines, there is a need for digital literacy. The most important part of this is the suggestion of adding a social media syllabus to the national curriculum. Stressing on the importance of media literacy, they suggest online educational programs, and awareness campaigns to empower the users to identify and resist hate speech, discouraging extremist views, and divisive rhetoric. Taking these concepts into consideration, it becomes necessary that government make policies and develop digital literacy programs related to hate speech

❖ **Selective Engagement and Respectful Dialogue**

Considering the relation of mental health with political hate speech, the participants are aware of its long term and short-term consequences. The early discussion also highlights that participants apply individual strategies to avoid hate speech for the sake of better mental health. They also advocate the concept of self-regulation. Their responses show the expressions of avoiding negative content, limiting social media use, and unfollowing toxic accounts. These expressions refer towards individual responsibility and selective engagement with political posts. Considering these measures, individuals can prevent or reduce the mental health implications from the exposure of political hate speech. Furthermore, the participants use expressions like culture of respect, tolerance, and positive discussion, which shift the self-responsibility towards collective responsibility of all the social media users. This shift is for the positive and constructive political dialogue on the social media by all the political workers and supporters. Either written or spoken content, it requires ethical and social check, which can help counteract political hate speech at mass level. This strategy can also help restore the environment of tolerance, which reduces implications of hate speech on mental health. These responses contribute to the question of media literacy. Impact of political hate speech on mental health can be reduced by critical individual strategies, but society can also play a key role to reduce the impact. This analysis highlights how participants experience, manage, and seek solutions for political hate speech on social media. The emotional and psychological toll of political hate speech, alongside strategies for resilience and media literacy, reveals the need for both individual and platform-level interventions.

Findings

The analysis of social media posts and the responses of participants reveal several key findings which are related to the mental health implications of political health speech, patterns of emotional and mental engagement, and strategies to reduce the impact of hate speech through media literacy. Identification of political hate speech through the patterns in the posts highlight political polarization, identity-based hate speech, discourse of national threat and dehumanizing language. Furthermore, participants report negative emotions such as frustration, anxiety, stress, sadness, and anxiety when exposed to political hate speech. Psychological strain intensifies with continuous exposure to political hate speech which often leads to hopelessness, helplessness, and disappointment. Hate speech appears to evoke emotional distress due to the divisive, hostile nature of the discourse, which is often perceived as

targeting not only political figures but also the values and identities of their supporters. This theme explains how political hate speech contributes to mental strain and encourages a sense of disillusionment and instability, impacting individual's mental well-being and social cohesion.

Political hate speech strengthens political identities, portraying opposition groups as threats to social and national values or moral integrity. This polarization is deeply rooted in identity-based hostility with loyalty to one's own party often seen as defining an individual's character or patriotism. Such framing heightens divisions, as participants express a sense of moral superiority for their group and view opposing parties as harmful or corrupt. In the context of elections, political hate speech intensifies the use of toxic expressions against the opponents. The analysis shows that this polarization based on identity intensifies feelings of hatred and heightens emotional defense, creating an environment which reinforces existing beliefs of social media users. Despite the strong emotional toll of political hate speech, some participants show resilience by practicing selective engagement, limiting their exposure, and employing strategies of reframing the speech and critical thinking. These individuals avoid deep emotional impact by ignoring political rhetoric, choosing not to engage in hostile discussions, and setting personal boundaries on social media. This indifference and detachment provide a shield against the negative impact of hate speech, enabling individuals to protect their mental health amidst a divisive political environment. The theme suggests that personal control over social media use and selective engagement are effective means to minimize the adverse effects of hate speech.

The data show that toxic and dehumanizing language, including derogatory and animalistic comparisons, is used to delegitimize political opponents. This form of rhetoric frames the opposition as inferior or morally corrupt, furthering the "othering" process and creating an environment where hostility and aggression are normalized. Such language not only promotes alienation but can also negatively affect mental health by fostering feelings of shame, anger, or alienation in those who identify with the targeted group. The participants also emphasize media literacy which can develop critical thinking and resilience to hate speech. Media literacy education is seen as essential for empowering users to recognize manipulative rhetoric and reduce engagement with hostile content. Additionally, there is a strong call for social media moderation to manage hate speech, with suggestions for stricter content filters, reporting tools, and regulatory policies. Participants believe that

enhanced moderation could reduce the prevalence of divisive rhetoric on social media, that can provide protective measures to reduce the mental health toll associated with exposure to political hate speech.

Conclusion

The findings highlight that political hate speech on social media has a multifaceted impact on mental health, affecting users both emotionally and cognitively. The divisive nature of political discourse reinforces polarized identities, increasing hostility and emotional distress, especially for those who are exposed to political content continuously. Furthermore, the study also demonstrates that resilience strategies, such as selective engagement and media literacy, are valuable tools for managing the negative impact of political hate speech. By promoting critical thinking, digital literacy, and respectful dialogue, individuals can better navigate and reduce the psychological impact of divisive rhetoric. For that, mutual respect to opinions and ideas is the most important aspect, which is lacking in social media usage. The study emphasizes the need for both individual responsibility and collective accountability in reducing the mental health impact of hate speech. While individuals can practice selective exposure and critical engagement, social media policies play an important role in moderating content and establishing a safe digital environment. This combined approach can help users engage with political discourse in a way that respects diversity and protects mental health, encouraging a healthier and more constructive online space.

Though the study provides a comprehensive understanding of political hate speech and its impact on mental health which can be reduced through digital literacy, it does not study the long-term psychological impact of sustained exposure to political hate speech. A study can be conducted to examine the chronic exposure impacts such as stress, anxiety, and other mental health issues over time using the experimental approach. Investigation into effectiveness of media literacy programs for resilience through emotional techniques and critical thinking training can provide better understanding of different interventions for the reduction of mental health issues resulting from political hate speech. Moderation and feedback policies as effective strategies are yet to be explored through responses qualitatively and quantitatively for the improvements to create more supportive digital spaces.

References

- Bandura, A. (1999). Moral Disengagement in the Perpetration of Inhumanities. *Personality and Social Psychology Review*, 3(3), 193-209. https://doi.org/10.1207/s15327957pspr0303_3
- Bilewicz, M. and Soral, W. (2020), Hate Speech Epidemic. The Dynamic Effects of Derogatory Language on Intergroup Relations and Political Radicalization. *Political Psychology*, 41: 3-33. <https://doi.org/10.1111/pops.12670>
- Bohner, Gerd & Dickel, Nina. (2011). Attitudes and attitude change. *Annual Review of Psychology*. 62. 391-417. [10.1146/annurev.psych.121208.131609](https://doi.org/10.1146/annurev.psych.121208.131609).
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp0630a>
- Braun, V., & Clarke, V. (2021). One size fit all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3), 328-352. <https://doi.org/10.1080/14780887.2020.1769238>
- Brown, Alexander (2015). *Hate Speech Law: A Philosophical Examination*. New York, NY: Routledge
- Buckingham, D. (2015) Defining Digital Literacy—What Do Young People Need to Know about Digital Media? *Nordic Journal of Digital Literacy*, 10, 21-35. <https://doi.org/10.18261/ISSN1891-943X-2015-Jubileumsnummer-03>
- Cikara, M., & Fiske, S. T. (2013). Their pain, our pleasure: stereotype content and schadenfreude. *Annals of the New York Academy of Sciences*, 1299, 52-59. <https://doi.org/10.1111/nyas.12179>
- Citron, D. K., & Norton, H. (2011). Intermediaries and Hate Speech: Fostering Digital Citizenship for Our Information Age. *Boston University Law Review*, 91(4), 1435-1484.
- Cramer, E. M., Song, H., & Drent, A. M. (2016). Social comparison on Facebook: Motivation, affective consequences, self-esteem, and Facebook fatigue. *Computers in Human Behavior*, 64, 739-746. <https://doi.org/10.1016/j.chb.2016.07.049>
- Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated Hate Speech Detection and the Problem of Offensive Language. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 512-515. <https://doi.org/10.1609/icwsm.v11i1.14955>
- Ganesh, B., & Bright, J. (2020). Countering Extremists on Social Media: Challenges for Strategic Communication and Content Moderation. *Policy & Internet*, 12(1), 6-19. <https://doi.org/10.1002/poi3.236>
- Gillespie, Tarleton. (2018). Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. [10.12987/9780300235029](https://doi.org/10.12987/9780300235029).
- Gillespie, Tarleton. (2020). Content moderation, AI, and the question of scale. *Big Data & Society*. 7. [2053951720943234](https://doi.org/10.1177/2053951720943234). [10.1177/2053951720943234](https://doi.org/10.1177/2053951720943234).
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harvard University Press
- Haslam, N. (2006). Dehumanization: An Integrative Review. *Personality and Social Psychology Review*, 10(3), 252-264. https://doi.org/10.1207/s15327957pspr1003_4
- Hinduja, S., & Patchin, J. W. (2012). Bullying beyond the schoolyard: preventing and responding to cyberbullying. *Security Journal*, 25, 88-89.
- Iordache, C., Mariën, I., & Baelden, D. (2017). Developing Digital Skills and Competences: A QuickScan Analysis of 13 Digital Literacy Models. *Italian Journal of Sociology of Education*, 9(1), 6-30. doi: 10.14658/pupi-ijse-2017-1-2
- Livingstone, Sonia & Helsper, Ellen. (2007). Gradations in digital inclusion: Children, young people and the digital divide. *New Media and Society*. 9. [10.1177/1461444807080335](https://doi.org/10.1177/1461444807080335).

- Livingstone, Sonia. (2008). Digital literacies: tracing the implications for learners and learning. <http://eprints.lse.ac.uk/id/eprint/33824>
- Matthes, J. (2012). Framing Politics: An Integrative Approach. *American Behavioral Scientist*, 56, 247-259. [10.1177/0002764211426324](https://doi.org/10.1177/0002764211426324).
- Pietrabissa, G., & Simpson, S. G. (2020). Psychological Consequences of Social Isolation During COVID-19 Outbreak. *Frontiers in Psychology*, 11, 2201. <https://doi.org/10.3389/fpsyg.2020.02201>
- Rafi, M. S., & Shafiq, Z. (2024). The role of political hate speech in promoting extreme ideologies and manipulating public discourse for social approval in Pakistan. *International Social Science Journal*, 74(253), 911-927. <https://doi.org/10.1111/issj.12488>
- Rheingold, H. (2012). *Net Smart: How to Thrive Online*. Cambridge, MA: The MIT Press
- Robert M. Entman, (1993). Framing: Toward Clarification of a Fractured Paradigm, *Journal of Communication*, 43, 51-58, <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Saha, K., Chandrasekharan, E., & De Choudhury, M. (2019). Prevalence and Psychological Effects of Hateful Speech in Online College Communities. *Proceedings of the ... ACM Web Science Conference*. ACM Web Science Conference, 2019, 255-264. <https://doi.org/10.1145/3292522.3326032>
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of inter-group conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of inter-group relations* (pp. 33-47). Monterey, CA: Brooks/Cole.

Article Information:

<i>Received</i>	13-Mar-2025
<i>Revised</i>	15-May-2025
<i>Accepted</i>	1-Jun-2025
<i>Published</i>	15-Jun-2025

Declarations:

Authors' Contribution:

- All authors **Conceptualization, and intellectual revisions. Data collection, interpretation, and drafting of manuscript**
- The authors agree to take responsibility for every facet of the work, making sure that any concerns about its integrity or veracity are thoroughly examined and addressed

• **Conflict of Interest:** NIL

• **Funding Sources:** NIL

Correspondence:

Ghulam Abbas Balti

ghulam.abbas@uobs.edu.pk
