



ADVANCE SOCIAL SCIENCE ARCHIVE JOURNAL

Available Online: <https://assajournal.com>

Vol. 02 No. 03. July-Sep 2024. Page# 86-106

Print ISSN: [3006-2497](#) Online ISSN: [3006-2500](#)Platform & Workflow by: [Open Journal Systems](#)**Digital Hate Speech as a Non-Traditional Security Threat in South Asia****Dr. Saeed Ahmed Butt**

Assistant Professor, Government College University, Lahore.

saeedbutt@gcu.edu.pk**ABSTRACT**

The adoption of digital communication technologies has radically changed political social interactions, communication and information sharing in South Asia. Social media has grown possibilities for democratic engagement and freedom of expression, but has also helped the widespread, rapid spread of online hate speech, misinformation and extremist narratives. Digital hate speech is increasingly no longer just a human rights issue, but a major non-traditional security threat that can impact on social cohesion, political stability, communal harmony and regional security. Digital hate speech is not like the traditional form of military hate speech and spreads across international boundaries, rapidly through communication channels online and frequently brings about real-life violence against religious, ethnic, linguistic and cultural minorities. These challenges are especially acute in South Asia, where societies are diverse and there are historical and social grievances, and where the Internet is widely used. In this article, the authors consider a new phenomenon, digital hate speech, to be a non-traditional security issue, and discuss its evolution, motivations and consequences in South Asia, focusing on Pakistan, India, Bangladesh and Sri Lanka. The study uses Human Security Theory, Securitization Theory and framework of Social Identity Theory and Network Society to examine link between online hate campaigns and radicalization, political polarization, communal violence, electoral manipulation, and cyber-enabled extremism. It also assesses current state of the legal and policy context, including the protection provided under constitution and existing cybercrime laws, as well as the international human rights instruments, including Universal Declaration of Human Rights (UDHR), the International Covenant on Civil and Political Rights (ICCPR) and United Nations Strategy and Plan of Action on Hate Speech. The fight against digital hate speech is vital not just for the protection of minorities' rights, but also for ensuring peace, the resilience of democracies and enduring human security in South Asia.

Keywords: Digital Hate Speech, Non-Traditional Security, Human Security, South Asia, Social Media, Religious Extremism, Minority Rights, Cybersecurity, Online Radicalization, Human Rights.

1. Introduction:

Digital communication technologies have revolutionized political, social and security dynamics of twenty-first century. Social media platforms like Facebook, X (formerly Twitter), YouTube, TikTok, Instagram and WhatsApp have transformed how information is disseminated, allowing millions of people in the South Asian region to engage in public discourse, political action and civic engagement. The Internet penetration has increased significantly in South Asia over the last 10 years, and hundreds of millions of people now turn to digital platforms to consume news, communicate, learn, and engage in politics. This digital revolution has resulted in a freedom of expression and democratisation of information, but has also opened up unprecedented opportunities for hate speech, misinformation, disinformation and extremist ideology. (Nations, United Nations Strategy and Plan of Action on Hate Speech, 2019). Online

hate speech is becoming one of most serious issues facing today's societies. Digital hate speech is characterized by its speed, often spreading via algorithm-driven platforms, encrypted messaging and transnational online network, as opposed to more traditional hate speech printed or broadcasted. In the anonymity of cyberspace and the speed of viral digital communication, hateful stories can travel at breakneck speed, reaching millions of people in just a few hours. This means that more and more, online hate speech manifests as offline discrimination, communal violence, political instability and human rights violations. (Benesch, 2014). However, South Asia is a special context for the study of digital hate speech, owing to its religious, ethnic, linguistic and cultural diversity. This area is home to almost a quarter of world's population, societies based on pluralism and old political conflicts, communal tensions, historical grievances and socio-economic inequalities. But in some countries like Pakistan, India, Bangladesh, Sri Lanka and Nepal, sectarian conflict, ethnic violence and political polarization are occurring and getting increasingly shaped by digital communication technologies. The social media has become a space for political actors, extremist groups, ideological and even ordinary users to challenge narratives in terms of religion, nationalism, ethnicity and identity. The intersection of cyberspace and physical security has been observed in many cases where online hate campaigns have preceded or exacerbated in-person violence targeting minority groups (Howard, 2020). The concept of non-traditional security (NTS) has broadened the definition of security beyond the military sphere to incorporate threats that impact on the security of individuals, communities and states via non-military means. Some scholars, including Barry Buzan, Ole Wæver and Jaap de Wilde (Buzan, 1998), have suggested that security can be considered on a variety of levels, from societal, political, economic, environmental to cyber dimensions. Likewise, the United Nations Development Programme (UNDP, 1994) presented the concept of human security, meaning security against chronic threats like violence, discrimination, fear and deprivation. Digital hate speech is a form of non-traditional security threat in this broader context as it undermines social cohesion, weakens democratic institutions, contributes to radicalization and poses risks to the safety and dignity of vulnerable groups. Digital hate speech is not the same as an honest criticism or a robust political debate, as it is directed against a person or a group of people on the basis of a protected characteristic, including religion, ethnicity, race, nationality, language, gender or other characteristics. According to United Nations (2019), hate speech is defined as communication that attacks or is communicated in a discriminatory manner based on characteristics of identity. The right to freedom of expression is guaranteed under International Covenant on Civil and Political Rights (ICCPR) (United Nations, 1966) and Article 20 of ICCPR (United Nations, 1966) allows for legitimate restriction of this right, for instance when there is a call for national, racial or religious hatred that is inciting discrimination, hostility or violence. Digital hate speech in South Asia takes many forms, such as anti-religious propaganda, sectarian incitement, ethnic vilification, gender-based harassment, misinformation targeting minority groups, coordinated online abuse and inflammatory political messaging. These stories often reinforce existing divisions in society and promote stereotypes and legitimise discrimination against marginal groups. Minority communities, such as Muslims, Hindus, Christians, Sikhs, Ahmadis, Buddhists, and other ethnic and religious communities (depending on context), are often targeted as a primary group in coordinated online attack. These kinds of campaigns can help lead to psychological trauma, social exclusion, forced displacement, and even communal violence (International, 2023).

Another key worry is the role of algorithm-driven social media platforms in promoting polarising content. Platform algorithms can be shown to be more likely to prioritise emotionally

charged and sensational content, as this is more likely to engage users. This can often result in hate speech, conspiracy theories, and misinformation being more likely to get attention than balanced or evidence-based information. (Vaidhyanathan, 2018) In addition, fake narratives and inflammatory messages can be easily shared on private, hard-to-trace messaging platforms like WhatsApp, which makes it difficult for authorities and platforms to moderate content. The strategy of using digital platforms to shape public opinion has been increasingly realized by political actors. Online hate speech is frequently used to rally people and to discredit their opponents during election campaigns, as well as in situations of political unrest and communal crises to build exclusionary narratives based on religion or ethnicity. Foreign influence operations, orchestrated disinformation campaigns and trans-national digital propaganda further muddy waters of South Asian security landscape, further emphasising transnational nature of digital threats in South Asia (Bradshaw, 2019) Although awareness is increasing, there are still a number of legal and institutional issues. While most South Asian countries have laws against cybercrime, the implementation of these laws is not uniform, and fears exist that law could be weaponised to stifle dissent, rather than respond to genuine hate speech. Moreover, free expression/incitement balancing issue remains a complex one of constitutional and ethical questions. Gaps in content moderation policies and practices, such as inconsistencies, are also a source of criticism for tech firms, as are their limited ability to have knowledge of local language expectations in South Asian languages and lack of transparency about how algorithmic decisions are made (Meta Oversight Board, 2023). In this context, the present study focuses on issue of digital hate speech as a non-traditional security threat in the context of South Asian region. Online hate is not viewed as a legal or technical matter alone, but rather from a more interdisciplinary perspective, combining security studies, human rights, political science, media studies, and cyber governance. It claims that digital hate speech is a multi-layered security threat, which poses a threat to human security, democratic governance, regional security, and social cohesion.

2. Literature Review:

Digital hate speech has become an international issue and sparked significant research on the topic in fields such as security studies, political science, communication studies, sociology, international law, and human rights. Previous research tended to focus on the legal and ethical aspects of hate speech in relation to freedom of expression and discrimination. As social media, AI, and digital communication technologies have rapidly evolved, so too is scope and effect of hate speech. Increasingly, digital hate speech is understood as a multi-layered phenomenon that affects human security, democratic governance, social cohesion and cyber security and international peace (United Nations, 2019).

The term "hate speech" is a contested one, since laws on subject differ among jurisdictions. United Nations Strategy and Plan of Action on Hate Speech (2019), defines Hate speech as communication that attacks or uses discriminatory language against persons or groups by reason of their religion, ethnicity or nationality, race, gender, or other identity characteristics. Likewise, 'hate speech' is further referred to as incitement to discrimination, hostility or violence, as it becomes particularly dangerous (Rabbat Plan of Action, 2012). Hate speech is differentiated from offence or controversial expression by scholars. Waldron states that hate speech can be harmful to dignity and equal social position of communities targeted by such speech as it suggests they do not deserve equal citizenship. (Waldron, 2012) Whereas hate speech does not just offend people, hate speech harms social conditions that are essential for democratic living together. Not all hateful speech is dangerous, according to Benesch (2014). Rather, speech is dangerous when certain contextual conditions such as political unrest, print

and audience susceptibility, influential speakers and historical grievances heighten the possibility of mass violence. This contextual approach is especially pertinent for South Asia, where communal tensions have a long history of intermingling with digital communication. Social media has made hate speech a transnational phenomenon from local to global level. Castells (Castells, 2010) discusses how digital networks have fundamentally transformed the nature of power in world, as they provide a platform for millions of individuals to communicate with each other without central control. Information moves virtually instantaneously across geographical lines these days, and extremist narratives have never had more exposure. Vaidhyanathan (Vaidhyanathan, 2018) found that social media platforms are not neutral communication tools. Their algorithmic systems value emotionally engaging content as it brings more engagement and ad income. As a result, there is a higher likelihood that divisive political messages and conspiracy theories, misinformation and hate speech will be amplified through algorithm rather than balanced or evidence-based information. Bradshaw and Howard (Bradshaw, 2019) also state that governments, political groups, and ideological factions are increasingly using computational propaganda, automated bots and orchestrated online campaigns to influence public opinion. Digital hate speech thus not only exists as spontaneous user-generated content but also as a political campaign. The developments indicate that digital technologies have changed hate speech in ways that impact its speed, scale, anonymity and persistence in relation to traditional media. The traditional focus of security studies is on threat from outside to states or military threat. Since publication of report on human development by United Nations Development Programme (UNDP, 1994), however, academic interest has been turned to concept of human security, which focuses on protection against fear, violence, discrimination and social exclusion. Human security scholars believe that focus of security needs to be on individual and not on state. In this context, digital hate speech poses risks to various aspects of security including Personal security (harassment and incitement to violence online), Community security (increasing hostilities towards ethnic and religious groups), Political security (polarisation of democracy and electoral manipulation) and Psychological security (fear, trauma and intimidation). Newman contends that nature of today's security issues is far more likely to come from non-state sources, digital means, and societal fragmentation than from traditional interstate warfare. (Newman, 2016) Thus, Internet hate speech is a new and non-conventional security threat. The concept of 'securitization' was introduced in Copenhagen School, notably in work of Buzan, Wæver and de Wilde (Buzan, 1998), which greatly expanded security studies. From this point of view, when political actors are able to convincingly state that issue is an existential threat that necessitates use of extraordinary policy measures, it turns into a security threat. This is particularly true of digital hate speech, as it now has become a threat to security, not just a communication problem and is increasingly recognized as such by governments, international bodies and technology giants. The Copenhagen School also coined term 'social security' that described security of collective identities, cultural values, languages and social cohesion.

Digital Hate Speech poses a threat to society's security by fostering intolerance, polarization of identities and communal hatred. Such risks are especially high in pluralistic societies like those of South Asia. But academics also warn of over securitisation which can lead to censorship, surveillance and legitimate political dissent. It is then a core issue of balancing security and civil liberties (Buzan et al., 1998). Social Identity Theory (SIT) by (Tajfel, 1979) Tajfel and Turner is one of crucial psychological theories to explain digital hate speech. The theory states that people get a part of their identity from their social groups. Digital platforms can exacerbate In-group favouritism, Out-group hostility, Identity-based political mobilization, Religious

nationalism and Ethnic polarization. These psychological tendencies are reinforced by online echo chambers, which show people primarily content that aligns with what they already believe. Sunstein states that personalization of algorithms leads to decreased exposure to conflicting ideologies and to the deepening of ideological extremism. (Sunstein, 2017) The research shows that identity politics is now increasingly digital in nature, in which hashtags, memes, short videos, and viral misinformation are able to reinforce communal narratives more quickly than traditional political communication. There are a growing number of articles discussing the role of tech giants in hate speech regulation. Content moderation is now one of the key governance issues in digital age (Gillespie, (2018)). At the same time social media platforms are criticised for taking away too much content, not enough, being inconsistent, biased against language and lacking transparency. Similarly, UNESCO (2023) highlights that automated moderation systems are less effective for many languages spoken in South Asia, including low-resource languages, because of the limited linguistic data sets and cultural context. While AI has enhanced ability to identify hate speech, scholars recognize key challenges in subject. Algorithms often fail to detect common terms in political discourse that are used sarcastically, in code, as a meme, in a historical context, or within a particular slur or colloquialism. In fact, as the internet penetrates more and more into South Asia, and as political polarization increases, the region is drawing more and more scholars' attention. Studies regarding India reveal a relationship between digital nationalism, electoral politics, religious polarization and online misinformation. Scholars have observed that narratives in online space have been inflammatory, coinciding with communal violence and mob attacks at times. Research on Pakistan is related to hate speech, extremist recruitment, cyber-encouraged radicalization, blasphemy accusations on internet, political polarization, and disinformation campaigns. Hate speech targeting religious minorities, secular activists, bloggers and journalists has occurred in Bangladesh online.

In Sri Lanka, digital misinformation has been one of factors that has fueled communal tensions between Buddhist and Muslim communities. Although each of these countries are different, there are some commonalities that can be found in the region, including rising smartphone usage, growing reliance on social media, and rising disposable income. Political identity, low digital literacy, platform accountability and changing cybercrime laws. The common characteristics indicate that issue of digital hate speech is not just a domestic concern but rather a challenge to regional governance. There is a strong focus by human rights scholars on need to strike a balance between rights in addressing digital hate speech. Freedom of expression is guaranteed by Article 19 of Universal Declaration of Human Rights, 1948 and International Covenant on Civil and Political Rights, 1966. But Article 20 of ICCPR calls on states to forbid incitement to discrimination, hostility or violence through advocacy of national, racial or religious hatred. Rabat Plan of Action (2012) suggests a six-part threshold test based on context, speaker, intent, content, extent of dissemination and likelihood of harm. This framework aims to separate protected expression from unlawful incitement. In general, scholars have found that a democratic society should not punish offensive opinions, but should do more to punish speech that is directly connected to violence or systematic discrimination. Despite growing body of literature on digital hate speech, there are some knowledge gaps. First, most of current research focuses on legal, communication or technological aspects of hate speech, and even fewer focus on it from the broader security perspective. Second, there are limited comparative studies across different countries of South Asia. Most studies on topic of Pakistan-India relations are analytical and look at India or Pakistan individually, ignoring regional dynamics. Third, there is a lack of comparative studies that focus on regulation of

platforms and simultaneously address the relationship between digital hate speech, human security and minority rights and regional stability. Fourth, unprecedented pace of technological change, whether in form of generative artificial intelligence, deep fakes, recommendation algorithms or encrypted messaging apps, has also posed novel challenges and challenges are not sufficiently examined in field of South Asian security studies. Last but not least, little interdisciplinary work that brings together the fields of security studies, human rights law, political science, communication studies and digital governance is undertaken within a coherent analytical framework.

3. Theoretical Framework:

Digital hate speech is not a conventional security issue, and a theoretical approach that comes from an interdisciplinary perspective is necessary to understand it. The purpose of scholarship was initially to examine issue of freedom of expression or criminal law, but current research increasingly takes into account implications for human security. Nowadays, governments, international institutions and civil society have termed online hate campaigns as a danger to national unity, democratic stability, public order, social harmony, protection of minorities, cyber resilience etc. So many states have introduced cybercrime laws, internet content laws, and digital governance policies that seek to regulate out-of-control online communication. Securitization Theory also has a useful warning, though. As Buzan (et al.) state, posing an issue as a security threat can be sufficient to justify extraordinary powers on part of the state, such as increased surveillance of individuals, closure of the internet, censorship and restrictions on civil liberties. (Buzan, 1998) However, digital hate speech should be taken seriously but not be used to justify measures that limit legitimate political criticism or peaceful dissent. Security-Liberties balance is still one of the critical policy issues in South Asia. Social Identity Theory (Tajfel and Turner) suggests that people create their identities via their social group membership. (Tajfel, 1979) Religion, ethnicity, religion, nationality, language, caste, and political party are natural categories. These identities influence attitudes and behavior. Social categorization, social identification and social comparison are three psychological processes identified by theory. People divide up society into "us" and "them. Individuals form an attachment to their own group. Groups want to be differentiated positively from the other groups, and they generally form a stereotype of out-groups. These processes are greatly enhanced by the use of digital media. Online spaces enable people to form communities of ideas and learn to reinforce common values and ideals over and over again. Similar content is recommended by algorithms and limiting the view of opposing points of view. This results in higher polarization levels, normalization of stereotypes, dissemination of conspiracy theories, and greater social acceptability of hostility toward minorities. In South Asia, identity politics frequently relates to ethnicity, religion, language, nationalism and beyond. Online hate speech targets such identities by suggesting a threat to national culture, religion or security against minorities. According to SIT, this is because emotional stories are disseminated more quickly online than factual stories.

The digital revolution has greatly impacted communication patterns. Castells developed Network Society Theory, which states that power is shifting from traditional power authorities to a global digital information network. Castells says that communication is no longer centralised, information is consumed and produced at the same time, information disseminates immediately across borders and digital networks are changing political power. (Castells, 2010) In network society, hate speech has some properties not found in traditional media like Speed, Scale, Permanence, Anonymity and Transnational Reach. Millions of people can be reached in minutes with information. One single post can be seen around the world by sharing, reposting

and algorithms for recommendation. Digital content is frequently available for an extended period of time. Users can remain anonymous or only under a pseudo name, which decreases accountability. Internet hate campaigns are often international and affect the politics of a region. Network Society Theory is a theory that can be used to account for increasing power of digital hate speech over traditional propaganda. While traditional newspapers or television fare have their own ecosystems, online communication enables misinformation and hate narratives to proliferate via millions of interconnected users in a decentralized ecosystem. Theoretical perspectives account for different aspects of digital hate speech. Human Security Theory is a theory that addresses harm that digital hate speech does to individual security, dignity, minority rights, and social cohesion. Securitization Theory (ST) provides an account of the way in which governments are increasingly defining online hate speech as a problem of national security and one that justifies policy intervention. Social Identity Theory provides an understanding of how these types of stories evolve quickly on internet, how religious polarization and ethnic hostility contributes to this spread and why. The Network Society Theory has been theorized as an explanation of how digital technologies, algorithms and transnational communication amplify hate speech. These theories do not compete with each other, but rather supplement one another. As a group, they illustrate that the issue of digital hate speech is a communication problem, a psychological phenomenon, a governance challenge, a human rights issue and a regional security issue.

4. Conceptualizing Digital Hate Speech:

Hate speech is one of the most important challenges today in the information age. Internet and social media have helped to facilitate communication among people, allowing them to share ideas across geographical and cultural barriers, but they have also helped to spread hatred, prejudice, misinformation and extremist ideologies quickly. While newspaper, radio or TV-based hate speech has certain characteristics that make it particularly harmful to society, digital hate speech has unique traits, such as speed, anonymity, permanency, algorithmic amplification and transnational reach, that enhance its impact on society. (Gillespie, (2018)) Religious, ethnic, linguistic and cultural diversity has created digital hate speech a grave threat to social cohesion and democratic governance in South Asian societies. Growing polarization, historical grievances, identity politics and the ubiquity of social media have made it easy for hostility on the internet to quickly turn into discrimination and violence in real life. It is therefore necessary to adopt an interdisciplinary perspective to define 'digital hate speech', combining legal, technological, political and human rights approaches. The definition of Digital Hate Speech: There is no legal consensus about hate speech. National traditions, political cultures, and legal systems have led to different national policies for regulation of harmful speech. There are a number of international bodies that have created influential working definitions. United Nations Strategy and Plan of Action on Hate Speech defines hate speech as 'Any form of communication (speech, writing or behavior) that attacks or denigrates or uses pejorative or discriminatory language in reference to a person or a group because of who they are.' (Nations, 2019) This definition emphasises three key points: discrimination and communication based on identity; hostility or prejudice targeting individuals; and individuals or groups. Like-wise, Council of Europe (1997) defines hate speech as words, gestures, songs, symbols and actions spread with intent to incite, promote or justify racial hatred, xenophobia, anti-Semitism, or other forms of intolerance. The Rabat Plan of Action refrains from using a strict nomenclature. (OHCHR., 2012) Rather, it suggests a contextual perspective that considers whether speech is likely to lead to discrimination, hostility or violence. Online hate speech is thus speech that dehumanises specific communities, stereotypes them, fuels discrimination,

glorifies violence, fuels hostility and normalises exclusion. It is important to note that hate speech is not just criticism that is legal against religious beliefs, public institutions, political parties, or governments. Coherent and vigorous debate, even if it is controversial and offensive, is protected in democratic societies so long as it does not veer too far over line into unlawful incitement. (Waldron, 2012).

4.1 Evolution from Traditional to Digital Hate Speech:

In past, newspapers, pamphlets, political rallies, radio broadcasts and television broadcasts were main vehicles for spreading hate speech. Traditional media were typically controlled and regulated at national level with editorial oversight. This dramatically changed with the digital revolution. Today, billions of people are able to produce and propagate content without needing to go through gatekeepers on social media platforms. Castells states that digital communication has moved away from a hierarchical broadcasting system to networked communication. (Castells, 2010) Digital hate speech differs in several aspects from traditional hate speech. Time is a factor that helps to make messages go around the world within minutes via internet. A post can be seen by millions of people over a period of time via reposts, hashtags, recommendation systems, and algorithmic promotion. In the case of digital content, it is possible to take screenshots, archives and reposts, even after it has been deleted, this is called permanence. Anonymity means that anonymous accounts should have less accountability, and more aggressive behaviour. Automation can help bots and organized networks to artificially magnify hateful narratives. The cross-border influence enables digital platforms to carry extremist messages across borders and impact regional politics and relations between nations. The characteristics above constitute the reasons digital hate speech is a non-traditional security threat.

4.4 Distinguishing Hate Speech from Misinformation, Disinformation & Extremist Propaganda:

There are multiple similar ways in which harmful content can be transmitted digitally. Hate Speech is communication that is based on identity and that encourages discrimination, hostility or violence towards protected groups. Misinformation is information that is false or inaccurate, but is shared by accident. For example, a user reporting incorrect data about a religion or religious community. Disinformation is false information that is intentionally created or spread for the purpose of manipulating public opinion. For instance, during election times, fabricated stories are promoted against minority communities. Malinformation is a term used for information that is intentionally shared with the intent to cause suffering to a person or group, typically with a carefully selected context or information that is deliberately tweaked to provide false information. Extremist Propaganda is content meant to recruit, justify violence or extremist ideological objectives. These forms can be conceptually separate but often overlap. For instance, disinformation or rumors can foster hatred or hate speech that can lead to mobilisation of mobs and potentially culminate in communal violence. This interaction has been more and more evident in the political crises of South Asian countries.

4.5 Drivers of Digital Hate Speech

The proliferation of digital hate speech is influenced by multiple interacting factors such as political polarization, religious extremism, ethnic Nationalism, economic incentives, weak Digital Literacy and algorithmic amplification. Political competition increasingly relies upon identity-based narratives that portray opponents or minority communities as threats to national identity. Election campaigns often witness increased circulation of inflammatory digital content. Extremist organizations exploit digital platforms to spread sectarian hatred and recruit supporters. Religious identity remains one of the most frequently exploited fault lines across

South Asia. Digital platforms facilitate mobilization around ethnic identity, historical grievances, and territorial disputes. Social media companies generate revenue through user engagement. Highly emotional or controversial content often attracts greater attention than balanced reporting. Vaidhyanathan (2018) argues that engagement-based algorithms unintentionally reward divisive communication. Limited ability to evaluate online information enables misinformation and hate speech to spread more easily. Recommendation systems prioritize content likely to maximize user interaction rather than social responsibility. Political polarization, religious extremism, ethnic nationalism, economic incentives, weak digital literacy and algorithmic amplification all influence the proliferation of digital hate speech. The competition for political power is increasingly based on identity narratives, which speak of threats to national identity from opponents or minority communities. During election campaigns, there is a rise in spread of digital inflammatory content. Sectarian hatred and recruitment of members by extremist organizations is a major concern on digital platforms. Religious identity is one of the most exploited lines of cleavage throughout South Asia. Digital platforms enable mobilization on ethnic issues, historical grievances and territorial issues. The money that social media companies make is based on engagement by users. Balanced reporting is not as likely to garner attention as highly emotional or contentious material. According to Vaidhyanathan, engagement-based algorithms inadvertently incentivize divisive communication. (Vaidhyanathan, 2018) Limited ability to evaluate online information enables misinformation and hate speech to spread more easily. There is no social responsibility priority in the recommendation system, it is content that is likely to get user interacting more. You can also use AI and algorithmic amplification to enhance your business.

4.6 Artificial Intelligence and Algorithmic Amplification:

AI and algorithmic amplification can also power up your business. AI has emerged as a key player in modern-day social media policy. Algorithms are used to suggest videos, trending topics, news feeds, suggested accounts, ads, and more. Studies are becoming more abundant that recommendation systems can reinforce polarizing content, precisely since sentiment and engagement are linked. As more and more research indicates, recommendation systems can inadvertently exacerbate divisive content due to its emotional content. Individuals visiting political or religious material are more likely to be shown ever more extreme suggestions. This is what helps to create echo chambers, polarization, confirmation bias and online radicalization. Meanwhile, AI presents significant opportunities as well. Artificial intelligence helps platforms identify hate speech, eliminate extremist material, spot coordinated manipulation and recognise abusive language. But automated systems are still far from flawless. They often have difficulty understanding sarcasm, satire, coded language, memes, multilingual communication and culture. Challenges are even greater in South Asia where hundreds of languages and dialects are spoken.

4.7 International Human Rights Framework:

International law deals with hate speech in a manner that strikes a balance between freedom of expression and protection from hate speech, one of most complex issues in the field. The right to free expression is guaranteed in Universal Declaration of Human Rights (1948) Article 19: "Everyone has right to freedom of opinion and expression." This is bedrock of democratic societies. The International Covenant on Civil and Political Rights (1966) also guarantees freedom of expression, in Article 19. But Article 20 states that states must prevent the following: "Any advocacy of national, racial or religious hatred which is incitement to discrimination, hostility or violence." So, international law affirms that freedom of expression is not unconditional. The Rabat Plan of Action (2012) suggests six test points to guide decision-

making in the realms of law and politics on issues of legal restriction of speech. There are six factors: context, speaker, intent, content, extent of dissemination, and potential for harmful effects. This framework aims to ensure that there is no arbitrary censorship and that vulnerable communities are protected. UN Strategy and Plan of Action on Hate Speech, 2019, highlights importance of prevention through education, digital literacy, counter-speech, responsible platform governance and international cooperation. (Nations, International Covenant on Civil and Political Rights., 1966) The strategy will not over-reliance on criminal sanctions.

4.8 Digital Hate Speech in South Asian Context:

South Asia has a set of features which make them even more susceptible to effects of digital hate speech. These include differences in religion and ethnicity, lingering historical conflicts, rising internet usage and smartphone penetration, rising polarization and use of encrypted messaging applications. Digital transformation of region has opened up new avenues for democratic engagement, but also raised risk of being exposed to communal misinformation, sectarian propaganda, disinformation, information warfare across borders and extremism using digital tools. Sensitive issues like religion, language, ethnic, migration, national identity and territorial issues are often targeted in online hate campaigns. They frequently cross borders; digital hate speech is a domestic governance problem as well as a regional security issue.

4.9 Digital Hate Speech as a Non-Traditional Security Threat:

Traditional security focuses upon military aggression and territorial defense. Digital hate speech challenges this perspective because its consequences emerge through societal rather than military mechanisms. Its security implications include weakening democratic institutions, increasing communal violence, undermining minority rights, facilitating radicalization, disrupting elections, damaging regional stability, and threatening human security. Unlike conventional warfare, these threats evolve gradually through digital communication, psychological influence, and identity mobilization. Consequently, many scholars now classify digital hate speech as an emerging non-traditional security threat requiring coordinated responses from governments, technology companies, educational institutions, civil society organizations, and international organizations.

5. Digital Hate Speech as a Non-Traditional Security Threat in South Asia:

Digital technologies have changed South Asia's security landscape forever. The region has been traditionally concerned with interstate conflicts, territorial disputes, terrorism and nuclear deterrence, but in this century, novel security threats, namely non-traditional security (NTS), have taken the form of threats that transcend national boundaries and impact individuals, societies, and states. Of these, digital hate speech is one of the greatest threats at present as it shapes public opinion, exacerbates communal conflicts, weakens democratic institutions and leads to acts of violence. Unlike traditional military threats, digital hate speech is transmitted via digital communication networks, social media platforms and through cyberspace. It is very fast to cross borders and can take advantage of existing divisions in the society, religion and politics. All of its impacts are multiple: on human security, harmony, political stability, economic activity, and regional peace. Security in present-day is not only a military issue but also one of society, politics, economy and information, as Buzan, Wæver, and de Wilde (1998) state. Digital hate speech is thus a paradigmatic case of the non-traditional security threat since it undermines society without resorting to traditional military means. This section explores some of the key security dimensions in which digital hate speech impacts South Asia.

5.1. Human Security and Individual Vulnerability:

The Human Security paradigm focuses on the human over the national state in the security analysis (UNDP, 1994). In this context, digital hate speech poses a risk to personal dignity,

personal safety, psychological wellbeing and human rights. Online abuse often targets religious minorities, ethnic communities, women, journalists, academic personnel, human rights activists and political activists. Constant harassment in online sphere can lead to psychological trauma, anxiety, depression, reputational damage, self-censorship and withdrawing from public participation. Online hostility can create sense of insecurity in minority groups, regardless of whether there has been any physical violence. Digital intimidation that cannot stop may lead to a reduction in freedom of religion, of expression and political participation, and thus democratic citizenship. Studies reveal that online hate campaigns can often follow or be a part of offline discrimination, social exclusion and targeted violence (Benesch, 2014). Thus, digital hate speech is more than just a virtual reality; it is a real threat to human security.

5.3. Threats to Societal Security and Social Cohesion:

According to Buzan and Ole Wæver, one of the meanings of 'security' in a society is the protection of collective identities, cultural values and social cohesion (Buzan et al., 1998). Peace in South Asian multicultural societies lies in harmonious relationship between people of diverse religions, languages and ethnicities. Digital hate speech is threatening this coexistence with the promotion of narratives of exclusion and "othering". Often online propaganda represents certain communities as outsiders, security threats, enemies of the nation, culturally inferior, or politically disloyal. Exposure to such stories normalizes prejudice, eroding intergroup trust. Digital platforms also help dissemination of rumors when it is needed the most during communal tension. Rumours of religious activities, alleged offences or political plots can become rampant before confirmation of facts is possible. Such rumours often help to perpetuate existing stereotypes and contribute to polarization. Putnam suggests that democratic societies, to function, need social capital in terms of trust, reciprocity and civic engagement. (Putnam, 2000) Digital hate speech undermines these pillars with promotion of suspicion and hostility towards communities.

5.4. Political Security and Democratic Governance:

The use of digital hate speech is a growing menace to democracy across South Asia. However, identity-based rhetoric is often used to rally electoral support by political actors, by ideological organizations and by coordinated online networks. During election campaigns, misinformation and disinformation are common, along with inflammatory speeches and communal propaganda and harassment of political opponents online. These practices undermine deliberation in a democracy because they take place of evidence-based discussion with emotionally driven narratives. According to Howard (2020), social media has become the main battlegrounds of political contestation, and algorithms are used to boost information based on emotional appeal rather than its factual veracity. This means digital hate speech plays a role in political polarization, loss of trust in public institutions, delegitimization of democratic systems, diminished electoral integrity and democratic accountability. Online abuse is often directed at a minority of candidates, women politicians and civil society groups, decreasing their participation in democratic processes.

5.5 Radicalization and Violent Extremism:

A key threat to security posed by digital hate speech is its connection with radicalisation. Digital communication is being increasingly used by extremist groups for ideological indoctrination, recruitment, fundraising, coordination and propaganda operations. Digital recruitment differs from traditional extremist recruitment in that it allows organizations to advertise their message to much wider audiences than they could with traditional methods, at a relatively low cost. Algorithms can also inadvertently radicalize users by feeding them more extreme content, if they are already reading polarized content. The process is termed 'radicalization pathway'

(Conway, 2017) as it involves normalization of violence against perceived outgroups through a series of incremental exposures to increasingly hostile narratives. But it is crucial to understand that hate speech does not necessarily lead to violence. Multiple factors contribute to radicalization, such as political grievances, social exclusion, identity crises, economic factors and ideological influences, all of which combine to create the phenomenon.

5.6 Digital Hate Speech and Communal Violence:

The most obvious impact of digital hate speech is ease with which it can make communal violence possible. Digital platforms speed up violence in a number of ways. Firstly, when tensions are high, misinformation can be fast and easy to share. Secondly, pictures and video images that are emotionally charged are more likely to generate outcry. Thirdly, rumors spread faster than official corrections. Fourth, online mobilization facilitates the organisation of mass mobilisations or violent acts. The false digital narratives about religion and ethnicity have been seen to drive mob violence, assault on religious institutions, displacement, and property destruction in several incidents across South Asia. Violence is not always directly triggered by online material, but can also be a catalyst, by helping to magnify underlying grievances and push collective action forward. For this reason, United Nations (Nations, 2019) stresses the need to prevent and preempt conflicts, with a particular focus on early detection.

5.7 Cybersecurity and Information Security:

There are links between digital hate speech and cyber security. Cybersecurity has long been about the protection of computer systems and digital infrastructure, but today, the manipulation of information is an important cyber threat that today's scholars are more aware of. Hate campaigns on the internet can include fake accounts, automated bots, coordinated inauthentic activity, manipulated images, deepfake videos and AI-generated content. These techniques make it harder to trace back to the original sources and have the perpetrators held accountable. Coordinated influence operations by foreign actors, aimed at heightening political instability, can also be carried out using domestic social divisions. It is therefore not only a social problem but also an emerging segment of information warfare, where digital hate speech acts as a tool. Therefore, digital hate speech is not just a social problem, but rather it is a tool of an emerging part of information warfare.

5.8. Cross-Border Implications and Regional Security:

A unique aspect of digital hate speech is that it is transnational. Historically, South Asian nations have had a common connection, share languages, religions, and media, and have a high level of digital interaction. Thus, the online narratives often cross the national borders. Digital propaganda across borders can aggravate tensions between nations, nationalism, misinformation, mistrust and polarisation in the region. Digital communication is then a challenge to interstate relations because it allows domestic political stories to creep into neighbouring societies. So far, implications of digital hate speech have not received significant institutional discussions in region in terms of threats to regional stability, especially from regional bodies like SAARC. Improved regional coordination and cooperation in field of cyber governance, information sharing, digital literacy and platform regulation can help to lower transnational risks.

5.9 Impact on Minority Communities:

Digital hate speech often has the most impact on minority communities. The narratives about minority religions as threats in the Internet contribute to discrimination and social exclusion. Digital intimidation deters people from voting, participating in political discourse, or participating in civil society. Targeted businesses can face organized boycotts and damage to their reputation. Continued online vitriol creates fear, anxiety and long-term emotional

suffering. Some minority people are afraid of being harassed or retaliated against if they voice their opinion online. These cumulative effects have an impact on equality before the law and on democratic pluralism.

5.10 Platform Governance Challenges:

Technology companies are a key player in combatting digital hate speech. But multilingual communication, cultural diversity, coded language, satire, fast-changing slang and encrypted messaging can pose a challenge for effective moderation. South Asia also has its own set of problems as many regional languages are moderately supported in very few languages. There have also been criticisms about non-uniform implementation between countries and languages. There is a need for increased transparency, more linguistic expertise, external supervision, cooperation with domestic civil society and scheduled human rights checks to improve accountability of platform. South Asian governments have enacted cybercrime laws, digital media regulations, and initiatives to ensure online safety. Although these are steps towards filtering out the bad stuff on the internet, there are notable concerns. International human rights organisations have continually stressed that legislation against hate speech must be clear, conform to international human rights standards, be proportionate, not be arbitrary and be able to defend legitimate political criticism. Excessive restrictions are likely to have a chilling effect on journalists, opponents of government, academics, and civil society groups. Hence, there is need to balance security and constitutional freedoms in the context of good governance.

6. Digital Hate Speech in South Asia: A Comparative Country Analysis

While the nature of digital hate speech is different from one country to another, depending on the political context, legal framework, religious makeup, media landscape, and degree of digital penetration, it has become a pervasive issue in South Asia. Rapid penetration of the internet, increased access to affordable smartphones and use of social media platforms have changed manner by which citizens obtain information, participate in political discourses and share religious or ethnic identities. Concurrently, these advances have enabled hate speech, misinformation and disinformation to be spread with great speed and intensity over social and digital platforms, exacerbating social divides. Even though there are significant national variations, there are a number of structural similarities amongst South Asian countries that facilitate digital hate speech. They range from historical communal grievances to identity politics, varying digital literacy, the lack of content moderation in regional languages, and changing cyber governance. Digital hate speech has thus emerged as an issue both for domestic governance and as a regional non-traditional security challenge to democratic stability, minority rights and interstate relations.

6.1. Pakistan:

Internet connectivity and social media usage have increased significantly in Pakistan in last ten years. Social media platforms like Facebook, YouTube, WhatsApp, TikTok, Instagram and X (Twitter) have emerged as significant channels of political communication, religious discourse and news intake. Digital media have contributed to greater public engagement but also to quick spread of misinformation, sectarian discourses and online harassment. Digital hate speech is often seen in Pakistan in context of religious sectarianism, extremist propaganda, ethnic tensions, political polarization, online harassment of women, accusations of religious sensitivities and collective attacks on journalists and civil society organizations. When elections or politically-charged protest rallies or events happen online, they generate intense campaigns and narratives that can be disseminated quickly via social media and encrypted messaging platforms. Pakistan has taken several steps to regulate cyberspace, including Prevention of

Electronic Crimes Act (PECA) 2016, which criminalizes certain acts of cyber harassment, cyberstalking and illegal online content. The Pakistan Telecommunication Authority (PTA) has the power to regulate and block online information and content that is deemed illegal in its own laws. These actions aim to regulate harmful online content, but they have been criticized by academics and international human rights groups for failing to align with constitutional freedom of expression and due process rights limits and international standards (Human Rights Watch, 2024). The digital presence of hate speech in Pakistan has potential to threaten interfaith harmony, sectarianism, democratic discourse, human security and counter-extremism. The relationship between online misinformation and existing social tensions has shown that digital communication can have a large impact on interactions offline that affect security.

6.2. India:

India has one of the world's biggest digital population with hundreds of millions of users of internet. Social media are integral to political communication, political campaigns, civic action and reporting. Social media are essential to political communication, political campaigns, civic engagement and news reporting. Adopting a language-first approach provides opportunities for content moderation due to the country's linguistic diversity, while there are challenges arising from a digitally active population. There are some common recurring themes in academic research, such as religious polarization, communal misinformation, caste-based abuse, ethnic hostility, political disinformation, gender-based online violence and coordinated trolling. The narratives tend to escalate during times of elections, communal conflict, or key political events online. There has been several empirical studies that have recorded how misinformation through encrypted messaging platforms has led to episodes of mob violence and communal unrest. While there are several socioeconomic and political reasons for such occurrences, digital communication has become an important catalyst. The government of India has laid down regulations for online communication under provisions of the Information Technology Act and the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules. Authorities also have broadened platform duties in the area of removing illegal content and user due diligence. The debate is still on-going whether current laws strike right balance between national security, public order, privacy rights and freedom of expression. Digital hate speech has an impact on India's democratic process, communal peace, protection of minorities, electoral integrity and societal security. The experience of this country provides an example of the potential for digital technologies to reinforce identity politics in vast multicultural democracies.

6.3. Bangladesh:

The government of People's Republic of Bangladesh (PRBPLA) has witnessed fast digitalization process in country through government's policy of "Digital Bangladesh". The use of Internet, mobile and social media has grown considerably over the last decade. Research reveals several recurring trends in online hate speech, such as religious intolerance, misinformation about minorities, harassment of journalists on internet and abuse against women, all of which have been considered some of the major forms of hate speech online. Lack of accurate information spread via Facebook has sometimes been a cause of communal tensions with vulnerable communities. Bangladesh has enacted laws on cybercrime and digital communication such as Digital Security Act (which was revoked and replaced by updated legislation). The laws are aimed at tackling cybercrime, misinformation and online incitement to action, and ignite ongoing discussions about freedom of expression and legal protections. Digital hate speech poses a risk to the coexistence of religions, public order, democratic

participation, security of minorities and trust in digital information. Preventive strategies that integrate education, media literacy and good governance are more and more being stressed.

6.4. Sri Lanka:

The post-conflict context in Sri Lanka is an important factor to consider when understanding digital hate speech. In past, conflicts over ethnicity and religion have occasionally been echoed and exacerbated in online communication. Numerous online narratives have been documented that entail religious intolerance, ethnic polarization, misinformation, nationalist rhetoric, and coordinated social media campaigns. Digital platforms have been used, on occasion, for speedy mobilisation when communities are in tension. Sri Lanka has intermittently banned social media in times of communal violence to curb the dissemination of inflammatory material. Although temporary restrictions could help to slow the spread of misinformation, there is discussion about their effectiveness in the longer term and about how they can be applied in a way that respects democratic freedoms. As the Sri Lankan experience illustrates, online communication can impact on post-conflict reconciliation, ethnic relations, democratic stability and public confidence. A long-term sense of resilience is dependent on two things: institutional reforms and community-based dialogue.

6.5. Nepal:

Nepal has documented relatively lesser instances of digital hate speech, when compared to other South Asian countries. However, the growing penetration of the Internet has created new governance problems. Political polarization, ethnic identity issues, misinformation, and narratives of discrimination against marginalised groups have become part of online discourse. Nepal is still working on legal and institutional frameworks for regulating cyberspace and safeguarding the rights of constitutional freedom. The success of the country's experience demonstrates the need for proactive digital governance before the establishment of harmful online ecosystems.

6.6. Comparative Analysis:

While there are some similarities across the region, there are also some differences. Causes are identical in South Asia, digital hate speech is shaped by identity politics, religious polarization, misinformation, algorithmic amplification, weak digital literacy, lack of multilingual content moderation, and politics competition. Each of five countries has been affected by various levels of social cohesion erosion, growing communal distrust, polarization, minority rights violations, democratic problems and cybersecurity issues. There are significant variations, too. Moderating India's very vast digital ecosystem is a unique challenge. The Pakistan Digital Environment is a reflection of the concerns regarding sectarianism, political polarization and cyber governance. Government of the Bangladesh has given a high priority to the regulation of misinformation in online platforms. The connection between digital communication and reconciliation in post-conflict context is illustrated by Sri Lanka's experience. While Nepal is at a relatively early stage of digital governance development, it is being faced with growing risks related to digital connectivity.

There are no specific regional frameworks in place in South Asia to tackle digital hate speech. Current issues are a lack of cooperation between countries, diverse legal frameworks, definition of hate speech, challenges of jurisdiction, inadequate information sharing, and inadequate digital literacy programs within the region. There can be scope of strengthening cooperation through dialogue, capacity-building, judicial cooperation, digital literacy programmes, best practice exchanges and research exchange at the regional level, such as SAARC level. The comparative analysis shows that digital hate speech is a new form of non-traditional security threat in South Asia. Although the nature and extent of online hate varies in

Pakistan, India, Bangladesh, Sri Lanka and Nepal, there are common structural factors such as identity politics, algorithmic amplification, misinformation and the growing connectivity of digital technology that lead to similar governance and security challenges in all five countries. However, national legal responses have developed in response to these challenges, and the need to reconcile protection of public order with international human rights commitments is continuing to be a concern. To prevent security threats arising from digital hate speech without compromising democratic freedoms, it will be vital to strengthen digital literacy, improve platform accountability, promote independent oversight, and foster regional cooperation.

7. Social Media Platforms, Artificial Intelligence, and the Future of Digital Hate Speech

Global information ecosystem has drastically changed due to the rapid growth of social media platforms and artificial intelligence (AI). From Facebook, X (formerly Twitter), YouTube, Instagram, TikTok, to WhatsApp, social media platforms have become vital components of political communication, news sharing, civic participation, and personal relationships in South Asian countries. They are easily accessible and can link millions of users, thereby enhancing democratic participation and freedom of speech. Simultaneously, these technologies have made it easier to spread hate speech, misinformation, disinformation and extremist narratives. Meanwhile, these technologies have also made it easier to spread hate speech, misinformation, disinformation and extremist narratives, creating governments, technology companies and civil society with complex challenges. Social media platforms differ from traditional media in that they have algorithm-driven recommendation systems that focus on user engagement. These can inadvertently amplify content that is emotionally charged and polarising, leading to exposure of hateful narratives, to the extent that digital hate speech has become a major non-traditional security threat (Vaidhyanathan, 2018). However, efforts to protect human rights and freedom of expression are increasingly hampered by the rise of generative AI, deep fake technology and ever-more complex forms of computational propaganda. It critically explores the role of social media platforms, Artificial Intelligence, Algorithmic governance and new technologies on future of digital hate speech in South Asia.

7.1. Social Media Platforms as Information Ecosystems:

Unlike the traditional media such as newspapers, radio, and television, social media is a medium of communication in which user is also a producer of information. The Spanish sociologist *Manuel Castells* maintains that current societies are part of a network society, in which communication takes place through decentralized digital networks instead of hierarchical institutions. (Castells, 2010) These platforms offer many democratic advantages such as promoting political participation, empowering citizen journalism, enhancing information accessibility, boosting social mobilization and bridging divides among vulnerable populations. The same attributes also make platforms susceptible to abuse. This is because digital hate speech can be easily disseminated, shared instantly via social media, reposted inflammatory content, organized online campaigns, coordinated online harassment and targeted transnational audiences. This means that false stories can propagate more quickly than fact checks and official statements.

7.3 Algorithmic Amplification of Hate Speech:

Algorithmic amplification is one of the most crucial technological contributors to digital hate speech. Recommendation algorithms are used to generate news feeds, trends, recommended videos, recommended accounts, ad priorities, etc. Most systems have been developed so that users are primary focus and not social impacts of content. According to research conducted by Vaidhyanathan (Vaidhyanathan, 2018) and Gillespie (Gillespie, (2018)), emotional, sensational and controversial content tends to have more engagement. As a result, algorithms can

inadvertently boost exposure to inflammatory political messaging, conspiracy theories, communal misinformation, identity-based propaganda and hate speech. This process leads to formation of a divisive information ecosystem, as content with a polarizing message gets more attention due to its comments, shares, and reactions. While platforms have taken steps to minimize harmful recommendations, there are still some issues to address, especially in multilingual settings like South Asia.

7.4 Facebook and Digital Hate Speech:

Facebook continues to be a major social media network in South Asia. Because of its large membership, it has become a significant forum for political campaigning, religious talk, community networking and information. However, there are a number of challenges to governance identified by academic research and personal investigations. This includes poorly moderated regional language content, delayed takedown of inappropriate content, spread of misinformation by viruses and coordinated inauthentic activity. Facebook's focus on content sharing via social networks allows misinformation to spread rapidly among trusted individuals, boosting credibility of the spread. Additionally, independent researchers have reported that recommender systems can inadvertently show users more and more polarising content in recommended groups and pages. Meta has increased spending on content moderation, AI, and third-party oversight. However, there is ongoing debate about the extent to which current instruments are effective in addressing linguistic and cultural diversity in South Asia.

7.5 X (Formerly Twitter):

In political communication, journalism, diplomacy, public debate, etc. X is an important tool. X also allows for the real-time sharing of information, with hashtags, reposts and trending topics compared to Facebook. Benefits are many like quick communication when an emergency happens; leaders and citizens can interact directly and government becomes more transparent. The platform, however, is also used to facilitate coordinated trolling, targeted harassment, misinformation campaigns, political polarization, and cross-border propaganda. Individuals like political and media personalities, academics, women politicians and minority activists are often subject to coordinated abuse. The platform's changing requirements for content moderation have sparked a continuing debate about how to balance the rights of free expression with the need to protect against harmful content.

7.6 WhatsApp and Encrypted Communication:

The reason that WhatsApp is unique in South Asia is that it is used extensively for personal communication. WhatsApp uses end-to-end encryption instead of open social media platforms. Encryption enhances privacy, confidential communications, and the prevention of unauthorized surveillance. It also makes the process of locating and eliminating bad content more difficult. False rumours, manipulated videos and inflammatory messages can spread quickly among private groups even before authorities or independent fact-checkers are notified. Misinformation spreads through encrypted messaging apps and has been demonstrated to impact public action in elections, communal tensions, health crises and security crises. The question then is how to manage privacy demands and devise measures to limit the viral dissemination of harmful content.

7.7 YouTube, TikTok, and Short-Video Platforms:

In South Asia, platform for video sharing is increasingly influencing political communications. YouTube merges entertainment with educational information, news and political commentary. TikTok and other short video apps have revolutionized the way digital communication is done with highly personalized recommendation systems. There is a lot of learning and gain of money in these platforms. But it is not only that algorithms can cause ideological echo chambers,

political polarization, conspiracy theories etc. Misinformation with visual imagery can be more effective than text-only misinformation as it can elicit emotional reactions that are associated with video information. As a result, platforms have started to invest more in automated fact-checking and collaborations with independent fact-checking groups.

7.8 Artificial Intelligence and Automated Content Moderation:

Platform governance has taken center stage and is defined by AI. Machine Learning systems are being used to identify hate speech, terrorist propaganda, coordinated manipulation, spam and fake accounts. AI can help platforms to moderate billions of posts every day, which is not possible by human review alone. Important caveats are, however, noted. Sarcasm, satire, historical references, local dialects, multilingualism and culture are often difficult to decipher and are not always understood by algorithms. South Asian region is a region of exceptional challenges as hundreds of languages and dialects are used in digital platforms. In other words, automated systems are less accurate in regional languages than English. Therefore, there is still a need for human intervention.

7.9 Generative Artificial Intelligence and Deepfakes:

The most recent development of generative artificial intelligence is one of the biggest challenges in pipeline. Generative AI can create fake videos, synthetic images, cloned voices, fake text and automated social media posts. Along with education, healthcare, research and innovation benefits, these technologies have potential opportunities for malicious actors in education. There is possibility of fabricated political speeches, manipulated religious messages, fake evidence, impersonation and coordinated disinformation campaigns. Deepfake technology can be especially harmful during election times or communal conflicts where bogus AV information can spread quicker than verification. The increasing use of AI in governance has led many scholars to suggest that transparency, authentication technologies, media literacy and swift fact-checking processes should be an integral part of the digital governance of the future.

7.10 Platform Governance and Corporate Responsibility:

Platform Governance and Corporate Responsibility was designed. More and more technology companies realize that they have quasi-public governance roles. Content moderation comes with several principles that must be considered, such as freedom of expression, user safety, privacy, transparency, accountability and legal compliance. Both UNESCO (UNESCO, 2023) and United Nations (Nations, 2019) refer to a human rights approach to platform governance. Core values are transparency of moderation decisions, independent oversight, appeal procedures, accountability for algorithms, multilingual moderation and regular human rights impact assessments. While corporate governance is essential, it is not enough to end digital hate. Cooperation between governments, technology companies, civil society organisations, universities, journalists and international organisations is needed to get effective responses.

7.11 Future Challenges for South Asia:

The future digital world will be more complicated than ever. New technologies will come into being such as artificial intelligence, virtual and augmented reality, decentralized digital platforms, which will bring new opportunities but also new threats. The key challenges are AI-driven misinformation, the rise of highly advanced deepfakes, trans-border information warfare, algorithmic bias, multilingual moderation, the fragmentation of regulatory standards, and the protection of digital rights. To boost digital literacy initiatives, cyber diplomacy, cyber cooperation, independent regulatory bodies, research capabilities, and awareness among the South Asian governments. Similarly, technology companies need to enhance their moderation functions for South Asian languages and culture.

8. Policy Recommendations:

Hate speech has become a big non-traditional security threat in South Asia via the internet. Digital hate speech differs from traditional security threats, such as armed conflict or land disputes, in that it attacks people, communities, democratic processes, and regional peace and security via the Internet. The intertwining of algorithmic social media, artificial intelligence (AI), misinformation, and identity-based politics have further contributed to the proliferation of harmful narratives threatening to undermine social cohesion, polarize, and exacerbate the vulnerability of minority communities. This challenge must be met more than with punitive legislation. Human rights, democratic governance, cyber security, digital literacy and technological innovation and regional cooperation are essential to sustainable solutions. The policy recommendations for governments, technology companies, civil society organizations, and educational institutions and regional organization are given in this final part.

8.1. Recommendations for National Governments:

The governments of South Asia should develop holistic national policies to address digital hate speech, taking into account its multidimensional nature as a security threat, without infringing on constitutional protections and best international practices on digital human rights. There should be clear legal demarcation between protected freedom of expression, offensive but lawful speech and incitement to discrimination, hostility and violence which is unlawful. Domestic legislation ought to be in line with Article 19 of the Universal Declaration of Human Rights (UDHR), Articles 19 and 20 of the International Covenant on Civil and Political Rights (ICCPR) and the Rabat Plan of Action (OHCHR, 2012). Any legislation must be written in a way that does not allow for arbitrary restrictions on legitimate political criticism, journalism, or academic discussion. Independent regulators need to track the trends of digital hate speech, to coordinate with technology companies, to be transparent about their activities through annual reports, to enable research and to ensure the protection of digital rights. They should be politically independent, to build trust with the people. Police need to be trained in cyber investigation, cyber evidence collection, cyber extremist networks, and international cyber cooperation. Judicial officers also need on-going training on digital rights and cyber-governance.

8.2. Recommendations for Social Media Companies:

Today's information ecosystem is shaped by technology companies to a great extent. Platforms should publish greater information regarding recommendation systems, ranking algorithms, content moderation criteria, and enforcement procedures. There needs to be increased access to anonymised data for independent research for public interest research. There are hundreds of languages and dialects in South Asia. Platforms should provide local language experts, enhance the AI for regional language as well as understand the local political and religious text. Improvements in linguistic ability would greatly enhance accuracy of moderation. Platform policies should be subjected to regular reviews to consider whether they impact negatively on religious minorities, ethnic communities, women, journalists and political activists. The UN Guiding Principles on Business and Human Rights offer a suitable framework for these. The universities, schools and institutes of research have a key role in long-term prevention. Critical media literacy, identification of misinformation, responsible online communication, digital citizenship and ethical use of Artificial Intelligence should be part of the education curriculum. Pupils should be taught to check before they share information online. Intercultural exchanges, debates, peace education and civic involvement should be promoted in educational institutions. These programs help to counteract prejudice and enhance democratic values. Civil society will play a crucial role in the fight against digital hate speech. Organizations should follow closely the developments in the online world, offer legal support to victims, organize

awareness campaigns, support counter-speech projects and encourage fact checking organizations. Community-based responses are likely to be more effective than punishment alone and build social resilience.

8.3. Recommendations for Regional Cooperation:

Often, digital hate speech crosses national borders. Cooperation between the regions is thus necessary. While the South Asian Association for Regional Cooperation (SAARC) has always focused on economic and political cooperation, it could be expanded to include cyber governance, digital literacy, online extremism, AI governance and digital human rights. Regional dialogue would help to ensure sharing of information and policy harmonization. The collaboration of South Asian universities in research centers on digital security, artificial intelligence, misinformation, online extremism and human rights should be established. Comparative research would support evidence-based policy-making. Technical assistance, judicial training, capacity-building programmes, AI governance, digital literacy campaigns and human rights monitoring should stem from the support of global, intergovernmental bodies like the United Nations, UNESCO, the International Telecommunication Union (ITU), and the Office of the High Commissioner for Human Rights (OHCHR). International cooperation will be vital, especially as digital communication crosses national borders.

9. Conclusion:

This study has explored digital hate speech as a new manifestation of non-traditional security threats in South Asia, which has an impact that transcends cyberspace and reach the other areas of human security, democratic governance, social cohesion and regional stability. The analysis shows that digital communication technologies are now so fast and extensive that they have revolutionized process of producing, spreading and consuming hateful narratives. Digital platforms can differ from other hate speech activities because the content can be transmitted instantly to other countries through algorithmic amplification, network communication and increasingly advanced Artificial Intelligence systems. The article has been based on the three theories: Human Security Theory, Securitization Theory and Social Identity Theory and Network Society Theory, in order to highlight that digital hate speech should not be perceived merely as a legal and technical issue. Instead, it's a multi-faceted problem, which relates to identity issues, polarization, misinformation, cyber government and international human rights. The comparative examples of Pakistan, India, Bangladesh, Sri Lanka and Nepal show that negative online ecosystems have developed in each of these nations due to a number of common structural factors, such as increased internet penetration, identity politics, lack of digital skills, and platform governance issues. The study also emphasizes the significance of social media and AI in the current information landscape. Recommendation algorithms, automated amplification and generative AI technologies pose both opportunities and risks. These innovations have improved communication and democratic engagement but they have also made digital hate speech faster, more widespread and more sophisticated. Therefore, for good governance, not only technological innovation but transparency, accountability, multilingual moderation and respect for internationally recognised human rights are needed. An important point the article makes is that responses to hate speech on the internet should not be simplistic. Overly tight restrictions can have a negative impact on freedom of expression, on legitimate political opposition and on the effectiveness of democratic institutions. However, inadequate regulation has the potential to lead vulnerable groups to be victims of discrimination, intimidation, and violence. Therefore, a balanced approach based on legal, necessary, proportional, transparent and judicial principles is crucial. The results indicate that successful sustainable solutions require a coordinated effort of the governments, technology providers, education, civil society,

regional institutions, and international institutions. Long-term investments in digital literacy, ethical AI, independent oversight, community resilience, and regional cooperation are more likely to be effective than punitive measures. In conclusion, protecting the digital future of South Asia demands the idea that human security and freedom of expression are not mutually exclusive goals. Balancing protection and the promotion of an open democratic discussion is key to ensuring the preservation of individuals from online hate and to helping digital technologies to foster pluralism, peaceful coexistence and inclusive and sustainable development, instead of social fragmentation. The importance of rights-based governance, evidence-based policymaking and regional cooperation in combating one of the most important non-traditional security challenges of the twenty-first century will not abate as digital ecosystems grow.

References

- (1997), C. o. (1997). *Recommendation No. R (97) 20 of the Committee of Ministers to Member States on hate speech*. Council of Europe.
- Benesch, S. (2014). *Defining and diminishing hate speech*. Minority Rights Group International.
- Bradshaw, S. &. (2019). *The global disinformation order*. Oxford Internet Institute.
- Buzan, B. W. (1998). *Security: A new framework for analysis*. Lynne Rienner.
- Castells, M. (2010). *The rise of the network society (2nd ed.)*. Wiley-Blackwell.
- Gillespie, T. ((2018)). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*.
- International, A. (2023). *The state of the world's human rights 2022/23*. Amnesty International. Amnesty International.
- Nations, U. (1966). *International Covenant on Civil and Political Rights*.
- Nations, U. (2019). *United Nations Strategy and Plan of Action on Hate Speech*. United Nations.
- Newman, E. (2016). *Human security*. Routledge.
- OHCHR. (2012). *Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred*.
- Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon & Schuster.
- Sunstein, C. R. (2017). *Republic: Divided democracy in the age of social media*. Princeton University Press.
- Tajfel, H. &. (1979). *An integrative theory of intergroup conflict*. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Brooks/Cole.
- UNDP. (1994). *Human Development Report 1994*. Oxford University Press.
- UNESCO. (2023). *Guidelines for the governance of digital platforms*. UNESCO.
- Vaidhyanathan, S. (2018). *Antisocial media: How Facebook disconnects us and undermines democracy*. Oxford University Press. Oxford University Press.
- Waldron, J. (2012). *The harm in hate speech*. Harvard University Press.