

A CRITICAL ANALYSIS ON THE IMPACT OF SOCIAL MEDIA PLATFORMS CONTENT MODERATION POLICIES ON FREEDOM OF SPEECH WITH SPECIAL REFERENCE TO CHENNAI

AUTHORS – LISA.S & HANUSHKA SRINIVASAN .S, STUDENTS AT SAVEETHA SCHOOL OF LAW, SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES (SIMATS) CHENNAI-77

BEST CITATION – LISA.S & HANUSHKA SRINIVASAN .S, A CRITICAL ANALYSIS ON THE IMPACT OF SOCIAL MEDIA PLATFORMS CONTENT MODERATION POLICIES ON FREEDOM OF SPEECH WITH SPECIAL REFERENCE TO CHENNAI, *INDIAN JOURNAL OF LEGAL REVIEW (IJLR)*, 4 (2) OF 2024, PG. 660-675, APIS – 3920 – 0001 & ISSN – 2583-2344.

ABSTRACT

The impact of social media content moderation policies on freedom of speech is a critical area of inquiry in the digital age. This research delves into the complex dynamics between social media platforms' content moderation strategies and individuals' right to free speech, with a focus on the Chennai context. Social media platforms operate as both private enterprises and public squares, leading to a delicate balance between regulation and censorship. Content moderation policies, addressing issues like hate speech and misinformation, often face criticism for potential censorship and inconsistent enforcement. Moreover, they can create a chilling effect, where individuals self-censor to avoid repercussions. The study employs empirical research with 221 participants using convenient sampling, analysing variables such as age, gender, education, occupation, and residential area. Results indicate diverse perspectives, with respondents valuing freedom of speech but also recognizing the need for moderation to combat misinformation and ensure a safe online environment. However, limitations include the small sample size and one-time nature of the study. Recommendations include greater transparency, consistency, and user engagement in content moderation practices to navigate the complexities of balancing regulation and freedom of speech on social media platforms effectively. Overall, while content moderation is necessary, the study suggests the need for nuanced approaches to avoid stifling legitimate expression while curbing harmful content effectively.

KEYWORDS: Content moderation, Free speech, Social Media, Impact, Pros and Cons.

INTRODUCTION:

The advent of social media has revolutionized how individuals communicate, access information, and express their opinions. Platforms such as Facebook, Twitter, Instagram, and YouTube have democratized content creation and dissemination, allowing voices from all walks of life to be heard on a global stage. However, this openness comes with challenges, particularly regarding the moderation of content and its impact on freedom of speech. This research explores the intricate dynamics between social media

platforms' content moderation policies and the freedom of speech of individuals, examining the balance between maintaining a safe online environment and upholding fundamental democratic rights.

The Role of Social Media Platforms

Social media platforms function as both private enterprises and public squares. As private companies, they establish and enforce policies to govern user behavior, protect their business interests, and ensure compliance with legal requirements. As public squares, these

platforms are arenas for public discourse, political debate, and social interaction, integral to the democratic process. This dual role creates a complex environment where content moderation policies must navigate the fine line between regulation and censorship.

Content Moderation Policies

Content moderation involves the processes and mechanisms through which social media platforms review, approve, and remove user-generated content. These policies typically address issues such as hate speech, misinformation, harassment, and graphic content. For instance, Facebook's Community Standards and Twitter's Rules and Policies outline specific behaviors and content that are prohibited on their platforms. These guidelines are designed to foster a safe and respectful environment for users, yet they often involve subjective interpretations of what constitutes harmful or inappropriate content.

Impact on Freedom of Speech

Freedom of speech is a cornerstone of democratic societies, enshrined in various international human rights frameworks, including Article 19 of the Universal Declaration of Human Rights. It encompasses the right to express opinions, seek information, and communicate ideas without fear of censorship or retaliation. In the context of social media, this freedom is tested by the need to balance open expression with the responsibility to prevent harm.

1. Censorship and Overreach—One of the primary concerns regarding content moderation is the potential for censorship and overreach. Critics argue that social media companies, driven by political, social, or economic pressures, may over-censor content, thereby stifling legitimate expression. For example, the removal of political posts or the de-platforming of controversial figures can be perceived as a suppression of dissenting voices. This concern is amplified by the opacity of moderation processes, where decisions are

often made by algorithms or moderators without clear or consistent standards.

2. Arbitrary and Inconsistent Enforcement—Another issue is the arbitrary and inconsistent enforcement of content policies. Users often report experiences of unequal treatment, where similar content is treated differently depending on who posts it. This inconsistency can lead to perceptions of bias, where certain viewpoints are unfairly targeted while others are ignored. The lack of transparency in decision-making processes further complicates matters, as users are frequently left without explanations for why their content was removed or their accounts suspended.

3. Chilling Effect—The concept of a chilling effect refers to the phenomenon where individuals self-censor due to fear of potential repercussions. In the context of social media, this can occur when users refrain from posting controversial or critical content to avoid being penalized. The chilling effect undermines the vibrancy of public discourse, as people may choose to withhold their opinions or engage in less meaningful conversations, diluting the diversity of perspectives that social media platforms are meant to promote.

Legal and Ethical Considerations

The legal landscape governing content moderation and freedom of speech is complex and varies significantly across jurisdictions. In the United States, for example, the First Amendment protects freedom of speech from government interference, but private companies like social media platforms are not bound by these constraints. In contrast, the European Union's General Data Protection Regulation (GDPR) and the Digital Services Act (DSA) impose stricter requirements on platform accountability and user rights. Ethically, social media platforms face the challenge of balancing the protection of free expression with the responsibility to prevent harm. This involves ethical considerations such as the duty to protect vulnerable populations from hate speech and misinformation, while also

respecting the diversity of thought and opinion. The ethical dilemma is further compounded by the global reach of these platforms, which must navigate cultural and legal differences in their moderation practices.

To address these challenges, there is a growing call for greater transparency and accountability in content moderation practices. This includes clearer guidelines, more consistent enforcement, and improved communication with users regarding moderation decisions. Some platforms have started to implement independent oversight boards, such as Facebook's Oversight Board, which reviews contentious moderation decisions and provides recommendations for policy changes. Moreover, there is a push for more user-centric approaches, where users have greater control over their content and the ability to appeal moderation decisions. Enhancing user education on the rules and the rationale behind them can also foster a more informed and engaged user base. The intersection of content moderation policies and freedom of speech on social media platforms is a critical area of inquiry in the digital age. While these platforms have enabled unprecedented levels of communication and expression, their policies and practices significantly impact the democratic principle of free speech. Striking a balance between regulation and freedom requires ongoing dialogue, legal scrutiny, and ethical reflection. By enhancing transparency, consistency, and user engagement, social media platforms can better navigate this complex terrain, ensuring that they remain spaces for robust and diverse public discourse.

OBJECTIVES:

- To understand the mechanisms and processes involved in social media content moderation policies.
- To analyze the impact of social media content moderation policies on an individual's fundamental right to free speech.

- To investigate potential solutions to mitigate the negative effects of social media content moderation policies while preserving an individual's freedom of speech.

REVIEW OF LITERATURE:

Rita Rexhepi (2013) proposed a paper where it was observed that Content moderation is crucial in today's political era, as social media is used to achieve political goals. Regulation is necessary to preserve democratic standards and promote a healthy online environment. This article compares content sharing regulations in the EU and the U.S., focusing on the benefits and shortcomings of each method. The EU adopts a centralized approach, prioritizing user protection and public interest, while the U.S. adopts a decentralized approach, focusing on free speech protection. However, both regions face challenges such as lack of transparency, over-removal, under-removal, and vague social media standards. The article suggests regulating platform transparency, increasing accountability, and establishing oversight bodies. Platforms should invest in content moderation policies by using higher-level methods to find and remove harmful content.

Shagun Jhaveri et al., (2018) proposed an article, where the researchers present a rich description of Twitter blocklists – why they are needed, how they work, and their strengths and weaknesses in practice. Next, we use blocklists to interrogate online harassment – the forms it takes, as well as tactics used by harassers. Moreover, we find that not all users agree on what constitutes harassment. Based on our findings, we propose design interventions for social network sites with the aim of protecting people from harassment, while preserving freedom of speech.

Sarah Myers West (2018) proposed an article that examines how users understand content moderation systems used by social media platforms to curate discourse. A survey of 519 users found significant impacts of content moderation beyond freedom of speech. The

study suggests that users take steps to assert their agency by seeking redress. The impact of content moderation goes beyond the debate on freedom of speech and suggests an educational model for content moderation systems. The article concludes by presenting an educational approach to understanding the role of social media in shaping discourse.

Robert S. T. (2018) proposed a paper that examines the Facebook content moderation controversy over the Vietnam-era photo "The Terror of War" in late 2016. It highlights the politics and economics of content moderation on social media platforms, where obfuscation and secrecy create an operating logic of opacity. The paper connects the functions of commercial content moderation on platforms like Facebook to their output, which is either the content that appears on a site or content that is rescinded, or digital detritus. The paper argues that the value of user-generated content as a revenue-generating commodity is the key criterion for moderation decisions, leading to commercialized online spaces that have less political and democratic challenge to the status quo.

Thiago Dias Olivia (2020) proposed a study where it was observed that the rise in online content circulation has led to new challenges such as defamatory content, hate speech, fake news, and copyright violations. Artificial intelligence is being developed to automate content removal, but these technologies still lack context understanding, posing risks to users' freedom of expression, access to information, and equality. The article suggests that these technologies should not be the sole basis for decisions affecting user expression, focusing on international human rights law standards.

Giovanni De Gregorio (2020) proposed a research where it was observed that freedom of expression is a fundamental aspect of democracy, but the algorithmic society has led to artificial intelligence technologies governing online information flow. These platforms are

often not accountable for user content, and their content moderation affects users' rights and democratic values. Despite their crucial role, social media platforms are not required to ensure transparency and explanation of their decision-making processes. The liberal paradigm of protecting free speech is no longer enough to protect democratic values in the digital environment, as information flow is driven by profit-maximization rather than democracy, transparency, or accountability.

Mackenzie F. Common (2020) proposed a paper that examines the enforcement stage of content moderation on social media platforms, arguing that the current approach is underdeveloped and poses serious human rights issues. It explains the role of moderators and the enforcement process, identifying problems such as bias in decision-making, over-reliance on efficiency, and inconsistent enforcement of terms and conditions. The paper suggests reforms, such as moving away from efficiency narratives to consider larger issues like human rights and rule of law, and adopting precedents to empower users and provide accountability. The paper concludes by suggesting reforms to improve the enforcement process on social media platforms.

Ivar A. Hartman (2020) proposed an article that discusses the evolution of speech moderation institutions and tools before and after the internet, focusing on legal and institutional aspects of identifying and countering excesses in speech. It contrasts the two scenarios to highlight the challenges faced by lawmakers, lawyers, and judges in a networked society. The article aims to uncover aspects often overlooked or underestimated in the legal debate about content moderation. The third part proposes a new procedural legal framework for online speech moderation without focusing too much on substantive legal standards for balancing speech.

Paolo Petricca (2020) proposed a study that examines Content Moderation, focusing on ethical concerns and cognitive effects. It

highlights issues such as quality of moderation, transparency, and working conditions for human moderators. The study uses Facebook leaked slides to illustrate moderation rules and principles, highlighting a lack of coherence and transparency. This lack of coherence can affect users' cognitive attitudes, perception of reality, and freedom of speech. The study compares this to other online cognitive phenomena and recent EU legislation. The current Content Moderation scheme risks users experiencing cognitive distortions, emphasizing the need for greater transparency and improved working conditions for moderators.

Evelyn Douek (2021) proposed an article where it was observed that Social media platforms are increasingly adopting international human rights law (IHRL) as a basis for their content moderation rules. This trend has been criticized for potentially imposing constraints on their operations. Some argue that these companies genuinely care about human rights, while others believe they are embracing the terminology because it may not be as constraining. This article critiques the contributions IHRL can make to content moderation and highlights its limitations as a practical guide for platforms in difficult cases. It argues that failing to acknowledge the limitations of IHRL will only serve the interests of platforms rather than their users, as it allows them to wrap themselves in the language of IHRL while the requirements of the body of norms remain indeterminate and contested.

Hannah Bloch-Weha (2021) proposed a paper where it was observed that Technology platforms are becoming the new governments, and content moderation is the new law. As platforms adopt automated mechanisms to enforce their rules, the power of the private sector seems to grow. However, complex relationships between public and private authorities challenge this power. Law enforcement and police exert influence over platform content rules, giving governments a voice in supposedly "private" decisions. Social media also aids law enforcement in detecting,

investigating, and preventing crime. This article examines the relationship between content moderation and surveillance, tracing how it intersects with law enforcement through formal demands for information, informal relationships between platforms and law enforcement agencies, and the impact of end-to-end encryption.

Tannef Mirrlees (2021) proposed an article that examines the power of Google, Apple, Facebook, Amazon, and Microsoft (GAFAM) over social platforms, highlighting their role in promoting hate speech and using hate content moderation apparatuses to de-platform alt-right users. Drawing on a political economy of communications approach, the chapter examines GAFAM's terms of service agreements and community guidelines to identify these hate content moderation apparatuses.

Anastasia Kozyreva et al., (2022) proposed a study that examining online content moderation found that citizens prefer quashing harmful misinformation over protecting free speech. Respondents were more reluctant to suspend accounts than to remove posts, and were more likely to do so if the consequences of the misinformation were severe or if sharing it was a repeated offense. The study also found that content moderation of harmful misinformation was a partisan issue, with Republicans consistently less willing than Democrats or independents to remove posts or penalize the accounts that posted them. The results can inform the design of transparent rules for content moderation of harmful misinformation.

Greyson K. Young (2022) proposed an article that discussed about the darker side of social media which has been brought to light by recent events like the 2020 presidential election and the U.S. Capitol Insurrection, when hate speech, fake news, and violent content have become the norm. Platforms have had to handle this problem by choosing whether to delete or withdraw offensive content. This article addresses the shortcomings of the two main

platforms' approaches to content moderation, as well as outside legal considerations such as FOSTA-SESTA and Section 230.

Nicole Buckley and Joseph S. Schafer (2022) proposed a research that explores and evaluates these policies compared to mainstream platforms and analyzes how moderation policies interact with the ideological framework asserted at an alternative platform's nascence. The U.S. ideological right has been criticizing mainstream social media platforms' election-related speech policies, leading to right-leaning influencers and followers migrating to alternative platforms like Parler, Bitchute, Gab, and Gettr.

Yi Wu (2022) proposed an article where it was observed that one of the keystones of Internet platform firms' operations is content control. Innovation coexists with concealment. When it comes to protecting human rights through content moderation, the United States and the European Union are at the top of the global rankings. The Constitution's wording and legal precedents have combined to create a largely comprehensive system of protection. Converging content moderation, privacy, and free speech are the public and private sectors, online and offline communities, and the ex-ante and ex-post balance. We must conduct further research on the online moderation model's design in order to reconcile the concerns about human rights legislation, policy, and the need to make online environments safer for a diverse global populace.

Elkin-Korean Niva eat al., (2022) The debate over social media platforms' discretion to suspend users and remove content has primarily focused on free speech implications and constitutional restraints. Platforms exercise discretionary powers under contracts defining their Terms of Service, which have been misinterpreted as dyadic contracts. The authors argue that platforms' contracts should be interpreted as contractual networks, focusing on interrelated obligations among independent agents who share a common goal. This

approach may facilitate a bottom-up check on content moderation via private ordering, increasing platforms' accountability.

Sahana Udupa et al., (2023) proposed an article that highlights the limitations of existing machine learning content moderation methods in dealing with hateful language online. It identifies the complexities of annotating problematic text in AI-assisted moderation systems and diagnoses the shortcomings as arising from the liberal-modern idea of "the human." The authors propose "ethical scaling" as a decolonial critique of the "human vs machine" conundrum and highlight the structuring effects of coloniality on extreme speech. Ethical scaling calls for a transparent, reflexive, and replicable process of iteration for content moderation, with community participation and global parity. This should evolve alongside addressing algorithmic amplification of divisive content and resource allocation.

Sai wang and Ki Joon Kim (2023) proposed a study involving 478 participants that found that AI is increasingly used in content moderation to detect and remove hate speech on social media. The study found that users' perceptions and acceptance of removal decisions for hate speech targeting certain social groups, such as religion or sexual orientation, were consistent regardless of the moderation agent type. When explanations for the removal were provided, removal decisions made jointly by humans and AI were perceived as more trustworthy, increasing users' willingness to accept the verdict. However, this moderated mediation effect was only significant when Muslims were the target of hate speech.

Shagun Jhaver and Amy X. Zhang (2023) proposed a study that compares platform-wide moderation and user-controlled, personalized moderation tools for regulating hate speech, sexually explicit content, and violent content. A survey of 984 US adults found that perceived negative effects on others negatively predict preference for personal

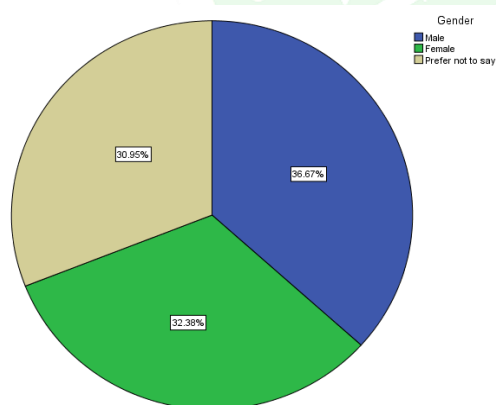
moderation settings, while free speech support positively predicts preference for platform-directed moderation. The findings suggest that platform governance initiatives should consider both actual and perceived media effects to increase user satisfaction. Users view personal moderation tools as a means to assert greater control over their social media feeds.

METHODOLOGY:

The research method followed here is empirical research . A total of 221 samples have been collected out of which all samples have been collected through convenient sampling methods. The independent variables utilised in this study are age, gender, educational qualification, occupation and residential area. The dependent variables are related to the impact of Social Media Content Moderation Policy on an individual's freedom of speech. The tools used for analysis are pie charts and clustered bar graphs through SPSS.

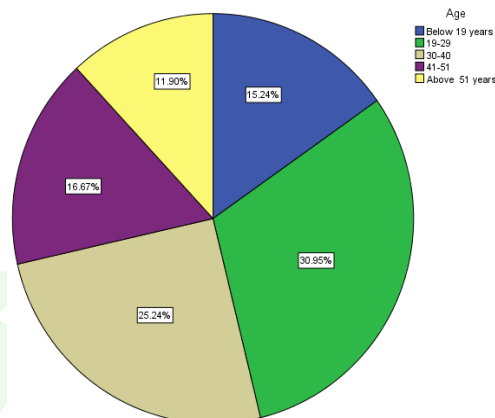
ANALYSIS:

Figure 1



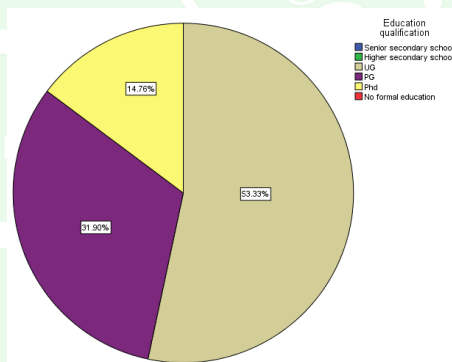
Legend: The above given figure shows the gender of the respondents.

Figure 2



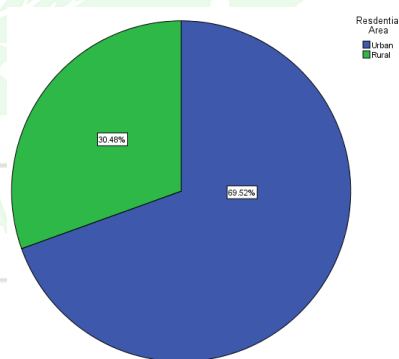
Legend: The above given figure shows the age of the respondents

Figure 3



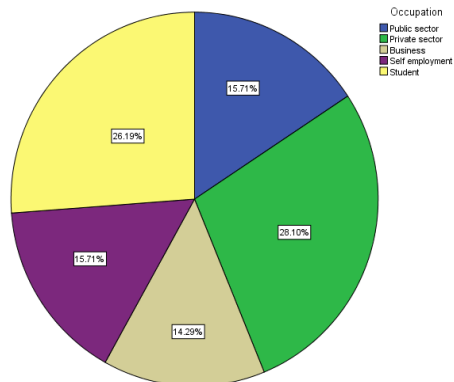
Legend: The above given figure shows the education qualification of the respondents.

Figure 4



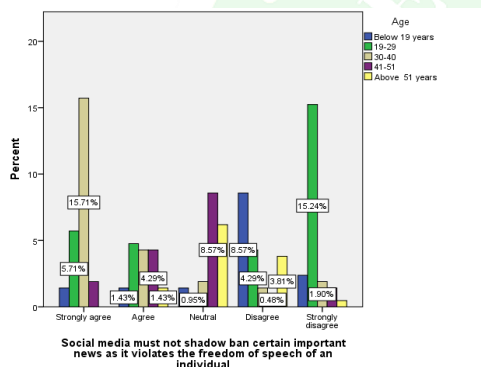
Legend: The above given figure shows the residential area of the respondents.

Figure 5



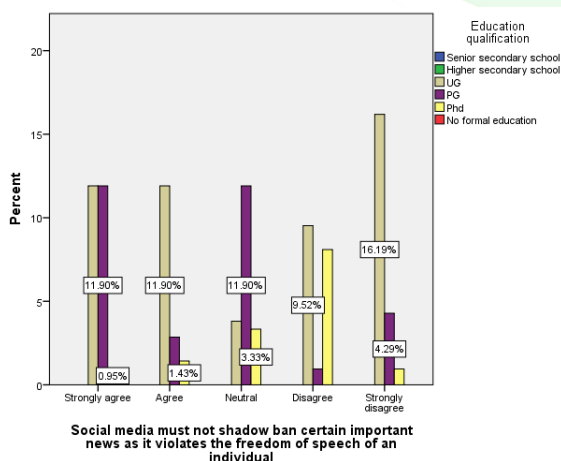
Legend: The above given figure shows the occupation of the respondents.

Figure 6



Legend: The above given figure depicts a cluster bar graph being combined and compared with the question, Social media must not shadow ban certain important news as it violates the freedom of speech of an individual and respondents age.

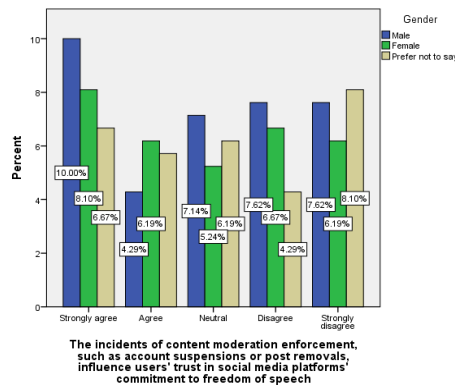
Figure 7



Legend: The above given figure depicts a cluster bar graph being combined and

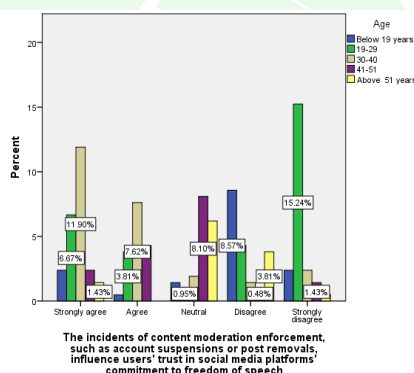
compared with the question, Social media must not shadow ban certain important news as it violates the freedom of speech of an individual and respondents education qualifications.

Figure 8



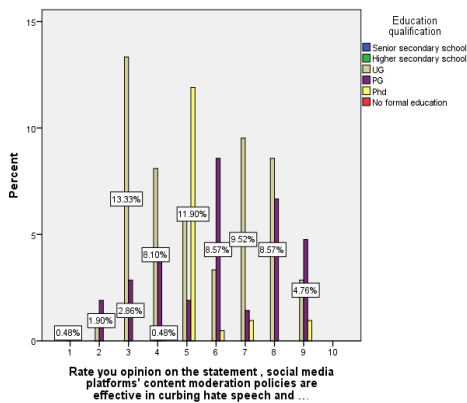
Legend: The above given figure depicts a cluster bar graph being combined and compared with the question, The incidents of content moderation enforcement, such as account suspensions or post removals, influence users' trust in social media platforms' commitment to freedom of speech and respondents gender.

Figure 9



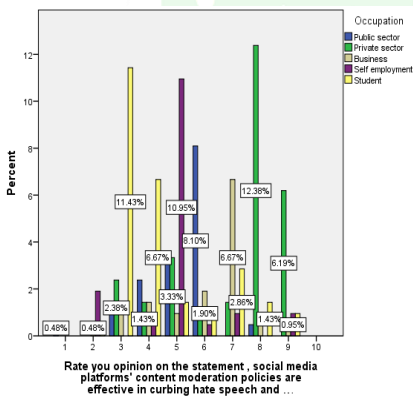
Legend: The above given figure depicts a cluster bar graph being combined and compared with the question, The incidents of content moderation enforcement, such as account suspensions or post removals, influence users' trust in social media platforms' commitment to freedom of speech and respondents age.

Figure 10



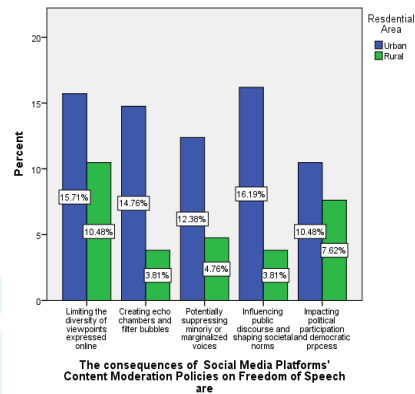
Legend: The above given figure depicts a cluster bar graph being combined and compared with the question, Rate your opinion on the statement , social media platforms' content moderation policies are effective in curbing hate speech and misinformation and respondents educational qualification.

Figure 11



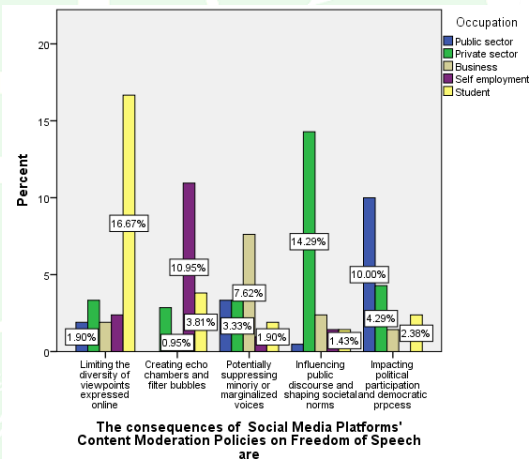
Legend: The above given figure depicts a cluster bar graph being combined and compared with the question, Rate your opinion on the statement , social media platforms' content moderation policies are effective in curbing hate speech and misinformation and respondents occupation.

Figure 12



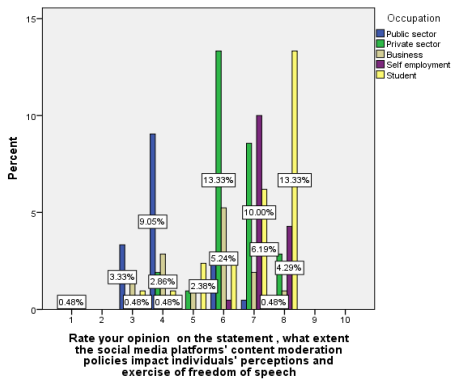
Legend: The above given figure depicts a cluster bar graph being combined and compared with the question, The consequences of Social Media Platforms' Content Moderation Policies on Freedom of Speech are and respondents residential area.

Figure 13



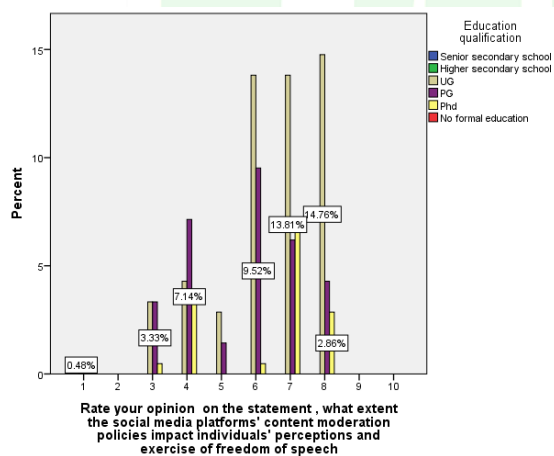
Legend: The above given figure depicts a cluster bar graph being combined and compared with the question, The consequences of Social Media Platforms' Content Moderation Policies on Freedom of Speech are and respondents occupation.

Figure 14



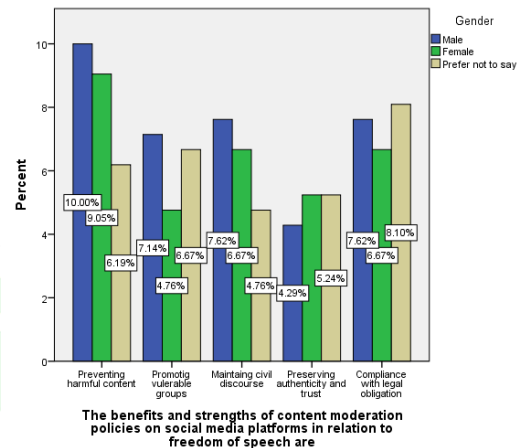
Legend: The above given figure depicts a cluster bar graph being combined and compared with the question, Rate your opinion on the statement, what extent the social media platforms' content moderation policies impact individuals' perceptions and exercise of freedom of speech and respondents occupation.

Figure 15



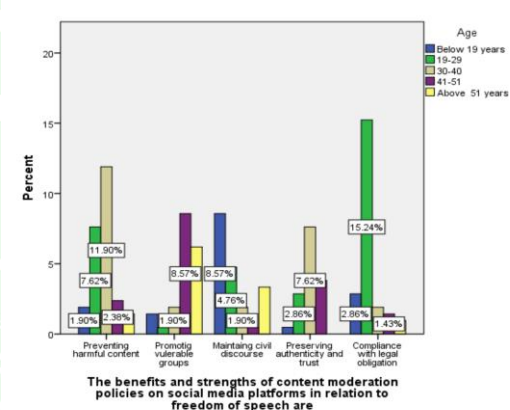
Legend: The above given figure depicts a cluster bar graph being combined and compared with the question, Rate your opinion on the statement, what extent the social media platforms' content moderation policies impact individuals' perceptions and exercise of freedom of speech and respondents education.

Figure 16



Legend: The above given figure depicts a cluster bar graph being combined and compared with the question, The benefits and strengths of content moderation policies on social media platforms in relation to freedom of speech are and respondents gender.

Figure 17



Legend: The above given figure depicts a cluster bar graph being combined and compared with the question, The benefits and strengths of content moderation policies on social media platforms in relation to freedom of speech are and respondents age.

Table 1

	Case Processing Summary					
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Occupation * The benefits and strengths of content moderation policies on social media platforms in relation to freedom of speech are	210	100.0%	0	0.0%	210	100.0%

to 19-29 years, 25.24% belongs to 30-40 years, 16.67% belongs to 41-51 years and 11.90% belongs are above 51 years. **(Fig:2)** It is revealed that 53.33% of the respondents have UG degree, 31.90% have PG degree and 14.76% have PhD degree. **(Fig:3)** It is revealed that 69.52% of the respondents are from urban areas and 30.48% are from rural areas. **(Fig:4)** It is revealed that 15.71% are working in public sector, 28.10% are working in private sector, 14.29% are Business, 15.71% are self employed and 26.19% are student. **(Fig:5)** It is revealed that majority of the respondents i.e 15.71% are from age group 30-40 years and they strongly agree to the statement Social media must not shadow ban certain important news as it violates the freedom of speech of an individual. **(Fig:6)** It is revealed that majority of the respondents i.e 16.19% have UG degree and they strongly disagree to the statement Social media must not shadow ban certain important news as it violates the freedom of speech of an individual. **(Fig:7)** It is revealed that majority of the respondents i.e 10% are male And strongly agree to the statement The incidents of content moderation enforcement, such as account suspensions or post removals, influence users' trust in social media platforms' commitment to freedom of speech. **(Fig:8)** It is revealed that majority of the respondents i.e 15.24% belongs to age group 19-29 And strongly disagree to the statement The incidents of content moderation enforcement, such as account suspensions or post removals, influence users' trust in social media platforms' commitment to freedom of speech. **(Fig:9)** It is revealed that majority of the respondents i.e 13.33% have UG degree and rated 3 on the rating scale for the question Rate your opinion on the statement , social media platforms' content moderation policies are effective in curbing hate speech and misinformation. **(Fig:10)** It is revealed that the majority of the respondents i.e 12.38% are working in private sector and rated 8 on the rating scale for the question Rate your opinion on the statement , social media platforms' content moderation policies are effective in

curbing hate speech and misinformation. **(Fig:11)** It is revealed that the majority of the respondents i.e 16.19% are from urban area and choose the option of influencing public discourse and shaping societal norms for the question The consequences of Social Media Platforms' Content Moderation Policies on Freedom of Speech are. **(Fig: 12)** It is revealed that the majority of the respondents i.e 16.67% are students and choose the option of limiting the diversity of viewpoints expressed online for the question The consequences of Social Media Platforms' Content Moderation Policies on Freedom of Speech are. **(Fig: 13)** It is revealed that majority of the respondents i.e 13.33% are students and have rated 8 in the rating scale for the question Rate your opinion on the statement , what extent the social media platforms' content moderation policies impact individuals' perceptions and exercise of freedom of speech. **(Fig: 14)** It is revealed that majority of the respondents i.e 14.76% have UG degree and have rated 8 in the rating scale for the question Rate your opinion on the statement , what extent the social media platforms' content moderation policies impact individuals' perceptions and exercise of freedom of speech. **(Fig: 15)** It is revealed that the majority of the respondents i.e 10.00% are male and have chosen the option of preventing harmful content for the question The benefits and strengths of content moderation policies on social media platforms in relation to freedom of speech are. **(Fig: 16)** It is revealed that the majority of the respondents i.e 15.24% belong to the age group 19-29 and have chosen the option of compliance with legal obligations for the question The benefits and strengths of content moderation policies on social media platforms in relation to freedom of speech are. **(Fig: 17)**

DISCUSSION:

Fig 6–The majority of respondents aged 30-40 strongly agree that social media should not shadow ban important news because this age group values freedom of speech and likely relies

heavily on social media for news consumption. They may perceive shadow banning as a form of censorship that restricts access to diverse viewpoints and essential information, which they see as crucial for a healthy democracy.

Fig 7- Many respondents who have completed their UG might strongly disagree with the statement about social media shadow banning important news due to the perception that unchecked misinformation can spread rapidly on these platforms, causing real-world harm. For instance, during the COVID-19 pandemic, false information about vaccines proliferated online, potentially endangering public health. Thus, some believe that regulating content is essential to prevent misinformation and protect societal well-being.

Fig 8- Many male respondents strongly agree with the statement because recent high-profile incidents, such as the suspensions of prominent figures like former President Donald Trump, highlight concerns about biased content moderation. These actions often spark debates about censorship and freedom of speech, leading users to question whether social media platforms are genuinely committed to upholding these principles or are selectively enforcing rules based on political or ideological leanings.

Fig 9- Many respondents aged 19–29 might strongly disagree with the statement because they often view content moderation as necessary to curb misinformation and harmful behavior. Growing up in the digital age, they might prioritize safety and accuracy over absolute freedom of speech. They may see moderation not as a trust issue, but as a platform's responsibility to maintain a healthy online environment.

Fig 10- Many respondents with undergraduate degrees may rate the effectiveness of social media content moderation policies as a 3 due to perceived inadequacies in these measures. Current events, such as the spread of misinformation during elections or the proliferation of hate speech during social

movements, highlight the limitations and inconsistencies in enforcement. These issues demonstrate that while efforts are made, they often fall short of significantly curbing harmful content.

Fig 11- The majority of private sector workers might rate social media content moderation highly because they often witness firsthand the efforts made by their companies to tackle hate speech and misinformation. Many private sector companies invest heavily in AI algorithms and human moderators to ensure their platforms maintain a positive environment, as seen in recent collaborations between tech giants and independent fact-checkers. Additionally, employees may also be aware of the legal and reputational risks associated with failing to address such content effectively, driving their endorsement of these policies.

Fig 12- In urban areas, access to social media is often widespread, making it a central platform for information sharing and debate. Consequently, urbanites may prioritize the option of influencing public discourse and shaping societal norms because they recognize the significant impact social media content moderation policies can have on shaping public opinion and cultural attitudes. For instance, recent controversies surrounding platforms like Facebook and Twitter highlight how their moderation decisions can spark widespread discussions and shape societal norms regarding free speech, misinformation, and online behavior.

Fig 13- Students, especially those who are not working, might prioritize a controlled online environment because they perceive it as safer and more conducive to learning. In an increasingly polarized digital space, where misinformation and harmful content abound, they might opt for platforms that limit diversity of viewpoints to ensure a more curated and trustworthy experience. For instance, amidst rising concerns about online radicalization and misinformation, students may prioritize platforms with stricter content moderation to

safeguard against harmful influences and ensure a more constructive online discourse.

Fig 14–The scenario could reflect a growing concern over social media's influence on free speech. Students, often more vocal about social issues, might rate it higher (8), perceiving restrictions on speech as more impactful. On the other hand, private sector workers, possibly cautious due to professional implications, might rate it lower (6), acknowledging the impact but also valuing moderation policies for maintaining order. This split in opinion mirrors broader societal debates on balancing free expression with online regulation, seen in recent discussions surrounding misinformation, hate speech, and platform accountability.

Fig 15–One possible reason could be that those with undergraduate degrees might have a deeper understanding of the complexities surrounding freedom of speech and the influence of social media content moderation policies. They may recognize the significant impact these policies have on shaping public discourse and individual expression, leading them to rate the importance of such policies higher than those with less education. For instance, recent debates over social media censorship of political content could underscore the relevance of this issue.

Fig 16–In many societies, there's a prevailing belief that curbing harmful content safeguards public safety and preserves social harmony. With the rise of online harassment, hate speech, and misinformation, male respondents might prioritize preventing such content to foster a safer online environment. They may see it as a means of protecting marginalized groups, reducing toxicity, and promoting constructive discourse, even if it entails some restrictions on free speech. This perspective aligns with broader societal trends towards accountability and responsible online behavior.

Fig 17–The prevalence of compliance with legal obligations among the 19–29 age group could be influenced by growing awareness of the legal repercussions tied to content moderation.

For instance, recent high-profile cases of platforms facing lawsuits or fines for failing to enforce content policies might emphasize the importance of legal compliance. Additionally, this age group, being more digitally savvy, may prioritize adherence to regulations as a means to safeguard freedom of speech while navigating complex online spaces.

LIMITATION OF THE STUDY:

A Convenient sampling method has been used in this study. The limitation of this study is that it is one time research conducted over a short period of time. Another limitation is that the research data collected was from a small size population of 221 samples. The data collected may not be completely accurate. However the responses were collected from people who have good awareness about this matter.

CONCLUSION:

The advent of social media has revolutionised how individuals communicate, access information, and express their opinions. Platforms such as Facebook, Twitter, Instagram, and YouTube have democratised content creation and dissemination, allowing voices from all walks of life to be heard on a global stage. However, this openness comes with challenges, particularly regarding the moderation of content and its impact on freedom of speech. The major objective of the research study is to Understand how Social Media Content Moderation Policy works, their Moderation Policy interferes and affects an individual's basic fundamental right of free speech. The findings of the research study is, it does affect an individual's freedom to express their opinion. The suggestion is to not to make the violation policy stringent if any speech or post is made with accurate information. It can further enhance their violation policy extending to defamation if it involves false accusations. The conclusion of the research study is content moderation should be restricted. But, a true and impactful statement should not be termed as content moderation and subject to violation policy in accordance to an influence made.

REFERENCES:

1. I.Reita Rexhepi.(2013).Content Moderation: How the EU and the U.S. Approach Striking a Balance between Protecting Free Speech and Protecting Public Interest.Trento Student Law Review.Volume: 5.Issue: 1. <https://ssrn.com/abstract=4433201>
2. 2.Shagun Jhaveri, Sucheta Ghoshal, Amy Bruckman and Eric E Gilbert.(2018).Online Harassment and Content Moderation: The Case of Blocklists.ACM Transactions on Computer-Human Interaction.Volume: 25.Issue: 2.Article No.: 12.pp: 1–33 <https://doi.org/10.1145/3185593>
3. 3.Sarah Myers West.(2018).Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms.New Media & Society.Volume:20.Issue: 11. <https://doi.org/10.1177/1461444818773059>
4. 4.Roberts S. T.(2018). Digital detritus: 'Error' and the logic of opacity in social media content moderation. First Monday.Volume: 23.Issue: 3. <https://doi.org/10.5210/fm.v23i3.8283>
5. 5.Thiago Dias Olivia.(2020).Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression.Human Rights Law Review.Volume: 20.Issue: 4.Pages 607–640. <https://doi.org/10.1093/hrlr/ngaa032>
6. 6.Giovanni De Gregorio.(2020).Democratising online content moderation: A constitutional framework.Computer Law & Security Review.Volume: 36.105374. <https://doi.org/10.1016/j.clsr.2019.105374>
7. 7.MacKenzie F. Common.(2020).Fear the Reaper: how content moderation rules are enforced on social media.International Review of Law, Computers & Technology.Volume: 34.Issue: 2. <https://doi.org/10.1080/13600869.2020.1733762>
8. 8.Ivar A. Hartman.(2020).A new framework for online content moderation.Journal of Computer law and Security Review.Volume: 36.ISSN: 0267–3649. <https://doi.org/10.1016/j.clsr.2019.105376>
9. 9.Paolo Petricca.(2020).Commercial Content Moderation: An opaque maze for freedom of expression and customers' opinions.Rivista internazionale di Filosofia e Psicologia.ISSN:2039-4667 (print) - E-ISSN: 2239-2629 (online) <https://doi.org/10.4453/rifp.2020.0021>
10. 10.Evelyn Douek.(2021).The Limits of International Law in Content Moderation.6 UCI J. INT'L TRAN'L & COMP. L.Volume: 37. <https://dx.doi.org/10.2139/ssrn.3709566>
11. 11.Bloch-Wehba, Hannah.(2021).Content Moderation as Surveillance. Berkeley Technology Law Journal.Volume: 36.Issue: 3.pp. 1297-1340.Texas A&M University School of Law Legal Studies Research Paper No. 21–37. SSRN: <https://ssrn.com/abstract=3872915>
12. 12.Tanner Mirrlees.(2021).GAFAM and Hate Content Moderation: Deplatforming and Deleting the Alt-right.Media and Law: Between Free Speech and Censorship.ISBN: 978-1-80071-730-5.eISBN: 978-1-80071-729-9.Volume: 26.PP: 81-97. <https://doi.org/10.1108/S1521-613620210000026006>
13. 13.Anastasia Kozyreva, Stefan M. Herzog, Stephan Lewandowsky, Ralph Hertwig, Philipp Lorenz-Spreen, Mark Leiser and Jason Reifler.(2020).Resolving content moderation dilemmas between free speech and harmful misinformation.PNAS.Volume: 120.Issue: 7. <https://doi.org/10.1073/pnas.2210666120>
14. 14.Greyson K. Young.(2022).How much is too much: the difficulties of social media content moderation.Information &

- Communications Technology Law
Journal.Volume: 31.Issue: 1.
<https://doi.org/10.1080/13600834.2021.1905593>
15. 15.Nicole Buckley and Joseph S. Schafer.(2022).Censorship-free' platforms: Evaluating content moderation policies and practices of alternative social media.For(e)Dialogue Journal.Volume: 4.Issue: 1.
<https://doi.org/10.21428/e3990ae6.483f18da>
16. 16.Yi Wu.(2022).Privacy, Free Speech and Content Moderation: A Literature Review and Constitutional Framework Analysis.Journal of Innovation In Science And Technology.Volume: 1.No: 4.ISSN: 2788-7030.
<https://www.paradigmpress.org/ist/article/view/303>
17. 17.Elkin-Koren Niva, De Gregorio Giovanni and Perel Maayan.(2022).Social Media as Contractual Networks: A Bottom Up Check on Content Moderation.Iowa Law Review.Volume: 107.Issue: 3.p987.ISSN: 0021-0552.
<https://openurl.ebsco.com/EPDB%3Agcd%3A16%3A78768/detailv2?sid=ebsco%3Aplink%3Ascholar&id=ebsco%3Agcd%3A156173624&crl=c>
18. 18.Sahana Udupa, Antonis Maronikoulakis and Axel Wisiolek.(2023).Ethical scaling for content moderation: Extreme speech and the (in)significance of artificial intelligence.Sage Journals.Big Data & Society.Volume: 10.Issue: 1.
<https://doi.org/10.1177/20539517231172424>
19. 19.Sai wang and Ki Joon Kim.(2023).Content Moderation on Social Media: Does It Matter Who and Why Moderates Hate Speech?.Cyberpsychology, Behavior, and Social Networking Journal.Volume: 26.Issue: 7.
<https://doi.org/10.1089/cyber.2022.0158>
20. 20.Shagun Jhaver and Amy X. Zhang.(2023).Do users want platform

moderation or individual control? Examining the role of third-person effects and free speech support in shaping moderation preferences.New Media And Society.Volume: 10.
<https://doi.org/10.1177/14614448231217993>