

Minoritization And Vulnerability: Essence Of Rule Utilitarianism To Online Speech And Expression

Jethan Divinaflor¹, Maria Desiree Paran², Hans Christian Arce³, Keith Joshua Redoña⁴, Shania Alliah Dichoson⁵, Vincent Andrew Salanga⁶

^{1,2,3,4,5,6} Polytechnic University of the Philippines – Parañaque City Campus

Abstract	Article Info
<p>In response to the escalating challenges posed by online communication, our research aims to thoroughly explore and understand the complexities involved in how rule utilitarianism applies to regulation of online speech and expression, specifically focusing on the vulnerabilities experienced by children and youth. Through an in-depth content analysis of Twitter comments, our study exposes concerning rates of dehumanization, notably in forms such as slut-shaming and sexualization, targeting these vulnerable groups. This exploration underscores the urgency to address these issues, recognizing their potential for harm and perpetuation of discrimination. The persistent tension between upholding freedom of online speech and mitigating harm emphasizes the crucial need for responsible digital tools, inclusive policies, and collaborative initiatives. Our research advocates for enhanced digital literacy, legal frameworks, and collective action, aiming to establish a safer online environment, particularly for marginalized communities. Importantly, the findings highlight the prevalence of dehumanization, urging comprehensive approaches to protect vulnerable communities online.</p>	<p><i>Keywords:</i> Minoritization Vulnerability, Online Speech, Rule Utilitarianism</p>

INTRODUCTION

Within the complex realm of online discussions, the idea of minoritization emerges as a noteworthy social and political occurrence, relegating specific groups to the fringes of society and exposing them to vulnerability (Armstrong, 2019). Online communication is of great importance in the context of liberal democracy, law, and citizen speakers (Cram, 2022). The advent of the digital era has brought forth unparalleled chances for individuals to disseminate their viewpoints on a worldwide scale via platforms such as social media, blogs, podcasts, and online forums (Momen, 2020). Although digital platforms have a role in promoting worldwide movements for justice, equality, and human rights, they also present a dilemma in managing the tension between freedom of speech and the need to prevent harm (Kozyreva, 2023).

Minoritization refers to the social process through which individuals or groups are placed in a subordinate position within society, often due to factors such as race, ethnicity, or gender. On the other hand, vulnerability encompasses the susceptibility to harm or mistreatment, which can arise from various factors, including social, economic, and individual circumstances. The concept of 'Minoritization' and 'Vulnerability' is clearly apparent in this vast digital environment, as demonstrated by the terrible episode of the Christchurch mosque shooting in 2019 (Gorwa et al., 2020). The issue of minority rights and online speech has received growing recognition, highlighting the possible dangers and injustices experienced by marginalized communities (Mackenzie et al., 2019). This study explores the intricacies of vulnerability, differentiating between vulnerabilities that occur in specific situations and vulnerabilities that are inherent to an individual. It highlights the need of addressing vulnerability that arises from the junction of race, ethnicity, and gender, as discussed by Crenshaw in 1989.

Outside the realm of digital technology, there is a connection between social vulnerability in local communities and health disparities, which have a negative impact on the well-being of minority populations (Kawachi et al., 2005). This exploration seeks to examine the challenges involved in regulating online communication using rule utilitarianism, with a focus on children and youth as vulnerable communities. Children and youth are particularly vulnerable to online harms such as cyberbullying, online grooming, and exposure to inappropriate content (Vallejos et al., 2019). The rise of digital technologies and platforms like social media has made online speech and expression crucial for the functioning of liberal democracy, law, and the citizen speaker. However, the conflict between protecting freedom of expression and preventing harm challenges online spaces to provide a clearer set of rules and align with public preferences in addressing moral dilemmas. The 2019 Christchurch Mosque shooting incident underscores the crucial role of algorithmic moderation systems on major platforms. Efforts to build social cohesion, inclusion, and diversity can contribute to preventing or countering extremism. In addition, having a society that is cohesive, inclusive, and embraces diversity is a good in itself.

Hence, this study aims to explore how the principles of rule utilitarianism can be applied to the regulation of online speech and expression, particularly in the context of protecting the rights and safety of children and youth as vulnerable and minoritized communities. It aims to understand the intricacies of balancing freedom of expression with the safeguarding of vulnerable communities, emphasizing the importance of using digital tools responsibly and proportionately. Engage in a conversation or exchange ideas and opinions.

LITERATURE REVIEW

Online speech and Vulnerable communities

The targets of online harassment are more likely to be those who belong to a minority or can be defined as different. These people experience harassment in various forms ranging from less severe forms such as Offensive name and Purposeful embarrassment, to more severe forms of online harassment ranging from Sexual Harassment to Physical threats and Stalking (Vogels, 2021). Individuals belonging to minority groups, particularly children and youth, are disproportionately targeted by online harassment. This harassment takes various forms, ranging from offensive name-calling and purposeful embarrassment to more severe forms such as sexual harassment, physical threats, and stalking (Antonio et al., 2023). For instance, young women and individuals from minority backgrounds are often targeted based on their appearance and sexual orientation, leading to severe instances of online harassment (Bussu et al. 2023).

Increased online time among youth corresponds to a higher likelihood of encountering hate in the online realm. Predictably, engaging in online communication with strangers is linked to an elevated risk of exposure to hateful content (Harriman et al., 2020). Additionally, youth using the internet are impacted by online hate

speech, particularly those who are part of social minorities or actively involved in online political engagement, expressing opinions, and advocating for vulnerable social groups (Obermaier and Schmuck, 2022).

Youth and Children in the Context of Online Speech and Expression

Young people are the fastest growing adopters of new online technology, this may bring many benefits, it also exposes them to various risks and dangers (Way, A., & Redden, S., 2017). Specifically, youth and children are particularly vulnerable to harassment and exploitation. They are disproportionately targeted by online harassment, which can take various forms, ranging from less severe forms such as offensive name-calling and purposeful embarrassment to more severe forms such as sexual harassment, physical threats, and stalking (Towhid et al., 2023). This vulnerability is further exacerbated by the increased time spent online by young people, which increases their likelihood of encountering hate speech (Prakash et al., 2021). Furthermore, according to the study by Bozzola et al. (2022), direct interpersonal communication and reducing the use of media devices, even inside the house can be increasingly beneficial in minimizing the negative adverse effect of social media.

Balancing between freedom of Online speech and Expression and Harm mitigation

The tension between the promotion of free expression and the need to mitigate harm in the online space has been a subject of study. One such study, "Resolving content moderation dilemmas between free speech and harmful misinformation," discusses the conflict between upholding freedom of expression and preventing harm caused by misinformation. The researchers found that most US citizens preferred quashing harmful misinformation over protecting free speech, and that people were more likely to remove posts and suspend accounts if the consequences of the misinformation were severe or if it was a repeated offense (Kozyreva et al., 2023). In a related study, "The New Governors: The People, Rules, and Processes Governing Online Speech," an analysis is provided on how major online platforms moderate content and the evolution of free expression values in the private sphere. This study outlines how the online platform Twitter moderates' content, taking into account free speech norms, corporate responsibility, and the economic necessity of creating an environment that aligns with the expectations of its users (Klonick, K., 2017).

Accountability and Responsibility

The internet is a tool that uses being global communication platform, it plays a crucial role specifically in shaping online discourse and navigate the freedom of the minoritized communities in terms of speech and expression. The duties and obligations of individuals in a contemporary democratic society extend beyond the mere act of electing a government. In the present era, citizens possess the capacity to actively observe and react even to public policies. Moreover, if they perceive it as essential, they have the right to oppose in response to unfavourable state affairs (Momen, 2019). On the contrary, La Rue cautions against entrusting censorship measures to private entities, arguing for accountability only for original authors of internet content, while acknowledging a multi-stakeholder governance strategy for internet regulation involving a mix of private, public, domestic, and international actors (Joyce, 2015).

Ethical Consideration

Another research which conducted a study about the Online Sexual Exploitation of Children (OSEC) focused on finding the evidence of the activity. It shed the light on the limitations of facilitative technology, social media and financial transactions as the key vulnerabilities that enabled the exploitation of children online. The research suggested developing these key vulnerabilities into a evidence-based practice or approach as a way to prevent OSEC and suggested stronger policy responses that enforce child safety measures (Roche et al., 2023). According to a 2021 policy brief by the International Telecommunication Union (ITU), the rapid growth of Information and Communication Technology (ICT) offers new opportunities for children to learn about their rights. However, it also introduces challenges such as privacy threats, cyberbullying, and online exploitation. To effectively address online risks for children, a national child online protection strategy is essential. This strategy, integrated with existing policies, provides a comprehensive framework to tackle challenges such as digital safety, inclusivity, and empowerment. Stakeholders, including government ministries, law enforcement, social services, and the ICT industry, must coordinate efforts. The strategy should prioritize child rights, align

with existing laws, involve children in decision-making, and be adaptable to technological advancements. Policymakers are urged to standardize the definition of a child, review and align legal frameworks, and address specific issues like child sexual abuse material.

DATA AND METHODOLOGY

In this research endeavour, we delve into the complexities of online speech and expression, exploring the impact of minoritization and vulnerability, with a specific focus on children and youth communities. By examining posts and comments from Twitter, we aim to identify prevalent themes and patterns related to dehumanization and demonization within the context of minoritization and vulnerability.

To gather insights, we conducted content analysis on posts and comments on Twitter. This method allows for a comprehensive understanding of the challenges faced by children and youth in online spaces. Content analysis is a systematic and objective method for identifying and interpreting patterns in data, such as text, images, or videos. It involves several steps, including defining the scope, selecting a sample, developing a coding scheme, coding the data, analysing the data, and interpreting the results.

By using content analysis, we can gain a deeper understanding of the prevalent themes and patterns, and by categorizing type of hate speech comments within the context of minoritization and vulnerability, specifically focusing on children and youth communities. This approach allows us to identify the specific challenges faced by these groups in online spaces and inform educational initiatives that aim to prevent exposure to hateful content and promote online safety.

In addition to the content analysis, we have collected bad comments that indicate hate speech related to children and youth. We have categorized these comments into dehumanization and demonization which can be breakdown into slut-shaming and sexualization, body shaming, racial discrimination and derogatory remarks based on sexual orientation next category is the violence and incitement lastly is the early warning, this segments to understand the origin and scope of the hate speech content. This categorization will help in visualizing the data using a table, which will provide a clear overview of the distribution of hate speech content about children and youth on a global scale.

RESULTS AND DISCUSSION

Perspective	Frequency	Results
Dehumanization and Demonization: Slut-shaming and Sexualization Body Shaming Racial Discrimination Derogatory Remarks Based on Sexual Orientation	39	93.2%
Violence and incitement (<i>An act of explicitly calling for violence</i>)	3	6.8%

Figure 01. Table

The table represents comments about youths and children reveal that the majority of the comments fall under the category of dehumanization and demonization, which accounts for 93.2% comments. s

Slut-shaming and sexualization represent one of the most common forms of dehumanization and demonization, with 22 comments. Slut-shaming is defined as the stigmatization of an individual based on their appearance, sexual availability, and actual or perceived sexual behaviour. These comments often involve sexualization, objectification, body-shaming, as well as degrading comments and rape threats. Examples of these can be found in the comments below:

Perspective	Frequency	Results
Dehumanization and Demonization: Slut-shaming and Sexualization Body Shaming Racial Discrimination Derogatory Remarks Based on Sexual Orientation	39	93.2%
Violence and incitement (<i>An act of explicitly calling for violence</i>)	3	6.8%

Body shaming accounts for 8 comments under the dehumanization and demonization. This targets individual based on their physical appearance, often leading to negative self-perception and mental health issues. Examples of these can be found in the comments below:

Comment 1: *"What a disgrace, my teenage gender neutral they/them looks up to Jenna. Its time to find them a new role model, smoking isnt allowed in my household!"*

Comment 2: *"Panget mo promise. HAHAAHHAHA. Pag nagsama kayong magtotropa*

mukha ka lang utusan, pagkapareho nyo lang ng besties mo, pareho kayong malandi 😏"

[I swear you are so ugly. HAHAAHHAHA. When you're with your friends you look like a servant, you're the same as your besties, sluts]

Comment 3: *"NGAUN KA MAGSALITA, NOON NG TAPANG MO KASI KINAKAYA NYO SI RICCI NGAUN SIGW MGSALITA KA TINGNAN NATEN ANG TAPANG MONG PANGET KA MUKHANG ASONG ULOL MANANG MANA SA*

NANAY NYANG NAKAGAT NG ASONG ULOL"

YOU CAN PUSH RICCI AROUND NOW TRY SPEAKING AGAIN SHOW YOUR BRAVERY YOU UGLY LOOKING DOG IN HEAT JUST LIKE YOUR MOTHER THAT WAS BITTEN BY A DOG IN HEAT]

Comment 4: *"eh kung putang ina mo di ka naman maganda"*

[you motherf*cker you're not even pretty]

Racial discrimination accounts for 5 comments under the dehumanization and demonization. It involves targeting individuals based on their race or ethnicity, which can lead to feelings of exclusion, isolation, and self-stigma. Examples of these can be found in the comments below:

Comment 1: *"These Americans think they gods"*

Comment 2: *"You are now hypocritical racism, I feel sick listening to your song now"* Comment 3: *"Fix up Nigga!!!"*

Derogatory remarks based on sexual orientation accounts for 4 comments under the dehumanization and demonization. It involves targeting individuals based on their sexual orientation, which can lead to feelings of exclusion, isolation, and self-stigma.

Comment 1: *"Why are you gay?"*

Comment 2: *"So u gay and a psycho too?? It is finished..."*

Comment 3: *"A FCKING GROOMER! U THOUGHT U DID SOMETHING BEA ALIPORES LMAOO"*

These findings highlight the prevalence of dehumanization and demonization comments about youths and children, with slut-shaming and sexualization being the most common form. It is crucial to address these issues and raise awareness about the negative impacts of such comments on the mental health and well-being of young people.

Another category is violence and incitement this one accounts for 6.8% comments. It involves incitement to violence or hatred, often through online platforms and social media.

Hence, this category often involves the dissemination of content that actively encourages or supports violent actions, fostering a climate of hostility and animosity. As a result, it becomes imperative to address and mitigate such instances to ensure a safer and more inclusive online environment.

Comment 1: *“seriously, i just wish somehow she could get burned so bad that she had to stop all this nonsense. like children could see this and think it’s ariana. it’s disgusting”*

Comment 2: *“Look at this shitty dad having cum”*

Comment 3: *“young mother sitang! yellow teeth kadiri and putangina! mamatay kana!
Dwende!”*

CONCLUSION

The analysis of comments on Twitter uncovered a disturbing prevalence of dehumanization, particularly in slut-shaming and sexualization. The findings highlight the urgent need to address dehumanization, recognizing its potential to incite harm and perpetuate discrimination. Our research emphasizes the importance of raising awareness and implementing interventions to mitigate the adverse impact of dehumanization and demonization on the mental well-being of young individuals. Notably, within this context, there exists a subset of comments categorized as violence and incitement, which, despite its minority representation, explicitly calls for violent actions.

Despite the aim to balance freedom of expression with safeguarding vulnerable communities, our results indicate that children and youth remain at significant risk in online spaces, particularly in the form of dehumanizing comments. The tension between freedom of online speech and harm mitigation persists, advocating for responsible digital tools and inclusive policies to create a safer online environment.

This research advocates for a clear and inclusive framework, enhanced digital literacy, legal measures, and collaborative efforts. Our research urges a comprehensive approach to address dehumanization and demonization comments, contributing to the broader discourse on fostering a digital environment that prioritizes the equilibrium between freedom of expression and harm prevention while safeguarding vulnerable communities.

REFERENCES

- Bozzola, E., et al. (2022). The Use of Social Media in Children and Adolescents: Scoping Review on the Potential Risks. *Int. J. Environ. Res. Public Health* 19(16):9960. <https://doi.org/10.3390/ijerph19169960A>
- Chen, X., Huang, C., & Cheng, Y. (2020). Identifiability, Risk, and Information Credibility in Discussions on Moral/Ethical Violation Topics on Chinese Social Networking Sites. *Frontiers in Psychology*, 11, Article 535605. <https://doi.org/10.3389/fpsyg.2020.535605>
- Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanit Soc Sci Commun* 10, 567. <https://doi.org/10.1057/s41599-023-02079-x>
- Cram, I. (2022). *Liberal Democracy, Law and the Citizen Speaker: Regulating Online Speech* (First Edition). Oxford: Hart Publishing. <https://doi.org/10.5040/9781509945856>
- Goedhart, N.S., Pittens, C.A.C.M., Tončinić, S. et al. Engaging citizens living in vulnerable circumstances in research: a narrative review using a systematic search. *Res Involv Engagem* 7, 59 (2021). <https://doi.org/10.1186/s40900-021-00306-w>
- Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 1–15. <https://doi.org/10.1177/2053951719897945>
- Harriman, N. W., Shortland, N., Su, M., Cote, T., Testa, M. A., & Savoia, E. (2020). Youth Exposure to Hate in the Online Space: An Exploratory analysis. *International Journal of Environmental Research and Public Health*, 17(22), 8531.

- International Telecommunication Union (ITU). (2021). Keeping children safe in the digital environment: The importance of protection and empowerment. https://www.itu.int/en/ITU-D/Cybersecurity/Documents/Policy_Brief_Keeping_Children_Safe_in_the_Digital_Environment.pdf.
- Joyce, D. (2015). Internet freedom and human rights. *European Journal of International Law*, 26(2), 493–514. Doi: <https://doi.org/10.1093/ejil/chv021>
- Kozyreva, A., Herzog, S. M., Lewandowsky, S., & Reifler, J. (2023). Resolving content moderation dilemmas between free speech and harmful misinformation. *Proceedings of the National Academy of Sciences*, 120(7), e2210666120. <https://doi.org/10.1073/pnas.2210666120>
- Klonick, K. (2017). *The New Governors: The People, Rules, and Processes Governing Online Speech*. *Legal Perspectives in Information Systems eJournal*.
- Momen, M. N. (2020). Threats to Free Expression in the Digital Age. In *Global Encyclopedia of Public Administration, Public Policy, and Governance* (1st ed.). Springer, Cham. https://doi.org/10.1007/978-3-319-31816-5_3897-1
- Momen, M. N. (2019). Myth and Reality of Freedom of Expression On the Internet, *International Journal of Public Administration*, 43:3, 277-281. <https://doi.org/10.1080/01900692.2019.1628055>
- Mostert, F. (2019). To regulate or not to regulate online speech, that is the question. *Journal of Intellectual Property Law & Practice*, 14(10), 741–742. <https://doi.org/10.1093/jiplp/jpz098>
- Obermaier, M., & Schmuck, D. (2022). Youths as targets: factors of online hate speech victimization among adolescents and young adults. *Journal of Computer-Mediated Communication*, 27(4). <https://doi.org/10.1093/jcmc/zmac012>
- Roche, S., Otarra, C., Fell, I., Belle Torres, C., & Rees, S. (2023). Online sexual exploitation of children in the Philippines: A scoping review. *Children and Youth Services Review*, 148, 106861. <https://doi.org/10.1016/j.childyouth.2023.106861>
- Rogers, w., Lange, M. (2013). Rethinking the Vulnerability of Minority Populations in Research. *American Journal of Public Health*. 103(12), 2141-2146. <https://doi.org/10.2105/AJPH.2012.301200>
- Perez Vallejos, E., Douthwaite, L., Creswich, H., Portillo, V., Koene, A.R., Jirotko, M., McCarthy, A., & McAuley, D. (2019). The impact of algorithmic decision-making processes on young people's well-being (Preprint).
- Vashishth, A. (2022). India's online freedom of speech and expression. School of Legal Studies, ApeejayStya University, Gurugram. DOI: <https://doi.org/10.52458/9789391842673.nsp2022.eb.asu.ch.19>
- Vallor, Shannon, "Social Networking and Ethics", *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/fall2022/entries/ethics-social-networking/>.
- Vogels, E. A. (2021, January 13). *The State of Online Harassment*. Pew Research Center | Nonpartisan, nonadvocacy, public opinion polling and data-driven social science research | Pew Research Center. https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2021/01/PI_2021.01.13_Online-Harassment_FINAL-1.pdf
- Way, K. & Redden, M. (2017). The study of youth online: a critical review and agenda, *Review of Communication*, 17:2, 119-136, DOI:10.1080/15358593.2017.1293838