

Addressing Online Harassment in the Philippine Digital Landscape: Regulatory Framework for Social Media Influencers and Content Creators

ABSTRACT

The growing problem of online harassment in the Philippine digital environment is examined in this policy proposal, which also makes the case for the creation of a regulatory framework which regulates the moral behavior of social media influencers and content producers. Influencers have enormous influence over public discourse as digital involvement grows, but current laws—like the Cybercrime Prevention Act of 2012 and the Safe Spaces Act of 2019—do not offer proactive rules specific to their duties. The proposal highlights continuing problems like harassment, false information, and uneven platform enforcement based on local, national, and international studies. The policy provides extensive guidelines focused on professional integrity, community protection, ethical communication, and content accountability. Registration, orientation, training, certification, monitoring, auditing, enforcement of sanctions, and ongoing review are all part of its progressive implementation. The idea seeks to improve influencer professionalism, strengthen user protection, and develop a better digital environment by integrating legal issues with ethical frameworks including deontology, utilitarianism, and virtue ethics. All things considered, this policy emphasizes the necessity of structured governance to address changing online harms and promote responsible digital citizenship in the Philippines.

INTRODUCTION

The rapid growth of social media in the Philippines has made digital platforms into influential spaces for community engagement, identity creation, and public conversation. Online spaces are now essential to daily communication and cultural expression, as the nation routinely ranks among the world's most active social media users, with over 80 million and 50 million users, respectively (DataReportal, 2025). In this context, social media influencers and content producers have become influential individuals with the ability to mold attitudes, actions, and social standards. However, their prominence and ability to persuade also put them and their viewers at greater danger of being harassed online. Body shaming, appearance-based criticism, derogatory labeling, and the deliberate spread of false information are examples of behaviors that have become common on major platforms (Sanchez-Danday, 2023; Patacsil, 2019). These behaviors undermine people's dignity and contribute to a larger culture of digital toxicity.

Certain types of digital abuse are punishable by existing laws, such as the Safe Spaces Act (Republic Act No. 11313, 2019) and the Cybercrime Prevention Act of 2012 (Republic Act No. 10175, 2012). Nevertheless, these rules lack proactive regulatory measures that are specific to the function and impact of content providers, as well as explicit ethical requirements. Similar to this, platform community guidelines provide rudimentary safeguards but frequently neglect to take into account local context, consistently enforce rules, or acknowledge the unique obligations involved in fostering sizable online communities. Because of this, users are nonetheless susceptible to unreported wrongdoing and damaging online interactions, and influencers frequently handle difficult ethical dilemmas without official supervision (Tadena, Kim, & Lee, 2021; Maranan et al., 2025).

In order to close these gaps, this policy proposal promotes a thorough regulatory framework that sets moral guidelines, outlines roles, and fortifies accountability systems for Filipino influencers and content producers. The proposed framework emphasizes professional responsibility, ethical communication, content accountability, and the protection of vulnerable audiences. It also outlines mechanisms for training, monitoring, and enforcement that complement existing legal and platform-based structures. By implementing these guidelines into reality, the policy aims to reduce the frequency of online harassment, promote more positive online communities, and motivate content producers to set an example of responsible and polite online conduct.

In the end, this plan aims to improve the digital environment in the Philippines by encouraging moral digital citizenship and making sure that those with substantial public influence have a positive impact on online culture.

RESEARCH AND LITERARY REVIEW

Review of Existing Research

Online harassment is a global problem, with China having particularly high rates, according to international study. According to studies, the anonymity, algorithmic amplification, and engagement-driven dynamics of Chinese platforms like Weibo and WeChat encourage negative interactions. For instance, 64.32% of Chinese college students surveyed said they had experienced cyberbullying, which includes orchestrated shaming, insults, and rumors; these behaviors were closely linked to anxiety and internet addiction (Huang, Zhong, Zhang, & Li, 2021). Increased stress, rumination, and other mental health issues are associated with cybervictimization, according to another study (Zhong, Zheng, Huang, Mo, Gong, Li, & Huang, 2021). Although there is evidence of gendered harassment, such as body-shaming and moral condemnation of female users in anonymous settings (Zhou, Wang, & Zimmer, 2022), China's regulatory response is still in its infancy, with difficulties observed in the implementation of the 2024 Regulations on the Governance of Online Violence Information (Zhang, 2024).

Research conducted at the national level in the Philippines also demonstrates how excessive use of social media encourages online harassment. According to UNICEF Philippines reports, a large number of Filipino teenagers regularly come across abusive or dangerous information online, yet they are hesitant to report these events because they have little faith in response systems (Inao, 2025). Common kinds of online harassment, such as insulting remarks, rumors, harassment through anonymous accounts, and character assaults, have also been documented by empirical research conducted among Filipino college students (Quero, 2022; Cocal, 2019). Within Filipino online forums, text-mining study on unconsented Facebook "pitik" posts uncovers recurrent patterns of digital humiliation, pejorative naming, and shame (Formoso et al., 2023). Scholars note that although national legislation, particularly the Safe Spaces Act (Republic Act No. 11313, 2019) and the Cybercrime Prevention Act of 2012 (Republic Act No. 10175, 2012), establishes legal procedures to penalize digital misuse, these laws are still primarily punitive. There is an absence in preventive guidelines for appropriate online conduct since they define what activity is criminal but do not specify proactive ethical obligations for influencers.

Local research from Mindanao shows how online harassment and cyberbullying can occur even in smaller yet digitally active communities, especially in Davao del Norte and among University of Mindanao campuses. According to a University of Mindanao,

Panabo College survey of senior high school students in Panabo City, 64% of participants said they had experienced cyberbullying, and there was a substantial correlation between these occurrences and daily internet usage (Mangarin & Montaño, 2021). Mild cyberbullying experiences, including psychological and gendered effects, are also reported by Mindanao university criminology students (Deocampo & Yanson, 2024). Additionally, a factor-analysis study conducted among University of Mindanao first-year students revealed a multi-dimensional framework for cybercrime victimization that includes risky online behavior, impersonation, and harassment (Reyes, Orong, Padao, & Pendang, 2025). These results imply that complex types of online violence are experienced by students in more localized Mindanao communities, indicating that digital aggression is not limited to urban or national networks. Research specifically on Panabo-based content creators or influencers is still lacking, though, as little is known about how Davao del Norte's local digital creators deal with harassment, content ethics, or reporting procedures. This gap highlights the need for interventions that promote digital literacy, offer ethical guidelines, and facilitate reporting in this local context, as well as for localized research that looks at students as well as up-and-coming digital influencers in community-based platforms.

Legal and Ethical Frameworks

Although the Philippines' current legal system offers fundamental protection against online harm, it falls short in addressing the moral obligations of social media influencers and content producers. The Cybercrime Prevention Act of 2012 (Republic Act No. 10175, 2012) provides a punitive mechanism against harmful digital behavior by making identity theft, cyberbullying, cyber libel, and other online offenses illegal. Despite its importance, the law offers no guidance on preventing harmful online cultures or defining professional standards for influencers. Complementing this is the Safe Spaces Act (Republic Act No. 11313, 2019), which penalizes gender-based online harassment such as misogynistic comments, body shaming, and stalking. There is a policy divide regarding proactive ethical obligations for public-facing digital personas, even though both statutes cover important forms of misuse.

Social media platforms also provide additional regulatory frameworks through its community guidelines, which prohibit harmful content, hate speech, harassment, and disinformation. However, because of algorithmic constraints, contextual misinterpretations, and different platform goals, enforcement is still uneven. Therefore, without constant accountability, influencers may participate in or unintentionally promote hazardous behaviors. The necessity for a national policy that supports and enhances current platform mechanisms is highlighted by the lack of localized, harmonized standards.

The moral obligation of influencers to refrain from harming others and to maintain integrity and justice is emphasized by deontological ethics. For those whose content has a significant impact on wide audiences, utilitarian ethics emphasizes the need to minimize harm and increase societal gain. Virtue ethics emphasizes moral character, accountability, and integrity—qualities crucial for public persons who influence digital culture. When taken as a whole, these ethical theories support the importance of precise rules for behavior.

POLICY PROPOSAL

Policy Guidelines

The proposed policy presents a systematic set of rules intended to encourage moral behavior, professional openness, and community safety among Filipino social media influencers and content producers. These guidelines, which aim to mitigate online harassment while promoting a better digital environment, explicitly address the gaps found in existing legislation and research.

The first set of rules emphasizes professional and ethical responsibility. Influencers need to maintain integrity, truthfulness, and openness, particularly when sharing information or forming alliances. They must make sure that given information is verified, refrain from making false claims, and declare sponsored content. Influencers are prohibited from sharing or promoting content that promotes discrimination, harassment, or false information. Influencers will be obliged to take part in basic ethics training that enhances their comprehension of appropriate online behavior in order to promote this.

The second guideline focuses on ethical communication and audience engagement. Influencers must abstain from using disparaging language, making personal attacks, or engaging in actions that promote fan-driven harassment or online mobbing. They are supposed to set an example of polite conversation, especially when discussing delicate subjects. In order to ensure that everyone participating maintains the same standards of behavior, collaborations must be carried out with knowledge of shared ethical duties. Additionally, influencers need to show cultural awareness and refrain from using disadvantaged populations as a source of content.

Third, the policy provides clear standards for content accountability and community moderation. In order to lessen harassment and false information, influencers must use moderation techniques in their comment areas. They have to keep a close eye on their platforms, eliminate negative interactions, and notify the appropriate

authorities or platform partners of serious instances of abuse. Influencers with a large following will have to release transparency reports detailing their approaches to content moderation and how they deal with online harassment in their communities.

Fourth, the guidelines emphasize a strong emphasis on protection of vulnerable audiences, particularly minors and marginalized individuals. Influencers must refrain from producing or endorsing any content that jeopardizes viewer safety, such as self-harm, unattainable body standards, or dangerous activities. Stricter regulations must be followed when creating content for younger audiences, making sure it is age-appropriate, devoid of explicit content, and considerate of their developmental requirements.

Finally, the strategy promotes protection of vulnerable audiences, particularly minors and marginalized individuals. In these communities, creators are urged to encourage courteous interactions, appropriate online conduct, and digital literacy. By doing this, we can create a more secure online environment that values accountability, empathy, and inclusivity.

All of these regulations are integrated to create a regulatory framework that encourages the production of ethical content, lessens digital harm, and creates safer, healthier online spaces for Filipino social media users.

Implementation Plan

The implementation of this policy will follow a structured, phased approach to ensure systematic adoption, consistent compliance, and long-term sustainability across the influencer ecosystem.

Phase 1: Registration and Orientation

This stage lays the groundwork for the operationalization of the policy. The Official Influencer and Content Creator Registry will be introduced by the Department of Information and Communications Technology (DICT) in collaboration with the National Privacy Commission (NPC) and pertinent partners. Influencers who reach predetermined levels of engagement or followers are required to register. The policy's ethical standards, duties, and compliance processes will be covered in orientation courses. This guarantees that before enforcement starts, influencers are well aware of their responsibilities.

Phase 2: Training and Certification

A required Digital Ethics and Responsible Content Creation Training Program, created in collaboration with academic institutions and digital literacy organizations, will be completed by registered influencers. Cultural sensitivity, content verification, anti-harassment tactics, ethical communication, and moderation techniques will all be included in training modules. A Digital Ethics Certificate, which is necessary to keep registration status, will be awarded to those who successfully complete the course. This stage guarantees that influencers have the information necessary to maintain moral principles.

Phase 3: Monitoring and Auditing

To check influencer compliance, a centralized monitoring system will be put in place. DICT will regularly evaluate platform data, user reports, and transparency reports submitted by influencers. Monitoring will concentrate on content behavior, how harassment incidents are handled, and compliance with ethical standards. Influencers with sizable followings are required to submit regular reports outlining their moderation techniques and steps taken to prevent negative interactions.

Phase 4: Enforcement of Sanctions

Sanctions for violations found during monitoring will be proportionate. Written warnings, required retraining, registry suspension, or permanent removal for serious or persistent infractions are some examples of these. Enforcement guarantees responsibility and shields online communities from ongoing harm.

Phase 5: Continuous Review

The last stage guarantees adaptation and long-term usefulness. The effectiveness of the policy will be reviewed on a regular basis by DICT and partner institutions, and guidelines will be updated in response to new developments and stakeholder input. This guarantees that the regulatory framework will continue to adapt to the changing digital environment in the Philippines.

Expected Outcomes

Influencer responsibility, user safety, and the general caliber of digital interactions in the Philippines are expected to significantly improve with the introduction of this policy. A formal method for identifying influencers will be established during the Registration and Orientation phase, guaranteeing that everyone covered by the policy receives fundamental instruction on ethical standards and legal responsibilities.

Influencers and regulatory organizations will coordinate more effectively thanks to this organized entrance process, which will also improve initial compliance.

It is anticipated that the Training and Certification phase would improve content creators' digital literacy and ethical competency. Influencers will be better prepared to stop harassment, discourage negative behavior, and encourage polite participation in their online communities with focused training on responsible communication, content verification, and community management. It is anticipated that this will greatly lessen actions like body shaming, disparaging labels, and the dissemination of false information.

The instruments for monitoring and auditing will improve policy enforcement's uniformity and openness. Frequent assessments and reports will establish a data-driven method for detecting infractions, enabling regulatory agencies to react to detrimental actions and new digital threats more skillfully.

The policy will create powerful barriers against unethical behavior through the enforcement of sanctions. Penalties that are appropriate will safeguard viewers, preserve community norms, and emphasize how important influencer obligations are. Lastly, the Continuous Review component makes sure that the policy is still applicable and flexible. Guidelines, training materials, and sanctions can be updated in response to changing platform dynamics and new types of online harassment thanks to ongoing assessments.

In conclusion, these outcomes are expected to promote a more responsible online culture in the Philippines, increase influencer professionalism, and create a safer digital environment.

JUSTIFICATION POLICY

Ethical and Legal Justification

The proposed policy is based on the simultaneous need to create ethical standards for Filipino social media influencers and content creators. Legally speaking, both the Safe Spaces Act of 2019 and the Cybercrime Prevention Act of 2012 offer crucial tools for punishing online abuse, but they focus more on specific offenses than the more general ethical obligations of digital influencers who have a big impact on public opinion. These current laws do not require professional behavior, community management, or responsible information distribution, nor do they provide preventive instructions specific to influencers. As a result, despite the existence of penal

regulations, negative behaviors—such as harassment, false information, and exploitative content—can spread. Therefore, rather than depending only on post-violation sanctions, a regulatory framework designed for influencers is required to enhance current legislation by setting proactive requirements.

The policy is in line with generally acknowledged standards of digital responsibility from an ethical standpoint. Given that their content has actual social repercussions, deontological ethics emphasizes influencers' obligation to refrain from harming others and to communicate honestly. The policy's rules on polite communication, content verification, and harassment avoidance directly promote this goal, which is further reinforced by utilitarian ethics, which emphasizes the need to increase collective well-being and minimize harm. By portraying influencers as moral role models who should exemplify integrity, responsibility, and respect—values crucial in creating healthier online communities—virtue ethics offers an extra basis. Influencers have a disproportionate amount of power in the digital ecosystem, therefore their ethical responsibilities go beyond self-expression to include protecting their followers' welfare and encouraging good digital citizenship.

The public welfare justifies the policy as well. It has been demonstrated that social exclusion, mental health problems, a decline in digital trust, and the normalization of negative behaviors are all impacted by online harassment. Influencers may unintentionally intensify toxic relationships or facilitate fan-driven harassment in the absence of clear ethical standards. The policy clearly addresses these concerns and aids in safeguarding vulnerable users, including women, children, and marginalized populations, by establishing defined standards, training requirements, and accountability systems.

Challenges and Considerations

While the proposed regulatory framework has significant advantages, its implementation and long-term efficacy may be limited by a number of issues. Differing degrees of influencer compliance are a major obstacle. Because of perceived limitations on their creative freedom or worries about the administrative load, influencers may oppose required registration, training, or monitoring. Influencers vary greatly in background, platform presence, and professional experience. Clear communication, easily accessible procedures, and incentives for compliance are necessary to ensure widespread involvement.

Coordination between platforms and institutions is another challenge. Government organizations, social media businesses, marketing firms, and educational

institutions must work together to implement the policy. The development of consistent standards may be hindered by variations in platform policies, enforcement strategies, and data-sharing procedures. To ensure consistent enforcement, formal agreements with platforms and the development of strong interagency relationships will be crucial.

Difficulties may also arise from resource constraints. Reliable personnel, technical infrastructure, and data management systems are necessary for efficient monitoring and auditing. Processing reports and carrying out routine compliance checks may be difficult for smaller organizations or regional offices. It will be necessary to strategically allocate resources and provide capacity-building support in order to ensure fair implementation across regions, especially rural or under-resourced areas.

Another factor to take into account is public approval. Users may have different expectations about the conduct of influencers, and excessive penalties could be seen as a danger to the right to free speech. Therefore, to preserve public confidence and prevent perceptions of overregulation, it is essential to strike a balance between accountability and freedom of speech.

Finally, a constant problem is the quickly changing nature of digital platforms. Emerging technologies, changing platform algorithms, and new types of harassment could surpass current regulations. In order to maintain the policy's relevance and efficacy, ongoing review procedures are crucial for updating guidelines, updating training materials, and improving sanctions. By addressing these issues, the policy's execution will be strengthened and the Philippines' digital environment will become more moral and responsible.

CONCLUSION

The urgent necessity to create a formal regulatory framework that directs the moral behavior of social media influencers and content producers in the Philippines is highlighted by this policy proposal. The absence of clear guidelines makes both producers and audiences susceptible to online harassment, false information, and detrimental digital behaviors as digital platforms continue to influence public conversation and cultural expression. The policy offers a comprehensive strategy for developing safer and more moral online environments by introducing guidelines on responsible communication, content accountability, and community protection. These guidelines are supported by phased implementation through registration, training, monitoring, enforcement, and ongoing review.

By emphasizing preventive and educational methods rather than depending only on punitive action, the proposed framework enhances current legal protections. Additionally, it gives influencers the skills and information they need to promote healthier community engagement and set an example of positive digital behavior. Adopting this policy will ultimately encourage stakeholders to promote and sustain ethical norms across Philippine social media landscapes, resulting in a more respected, accountable, and resilient digital environment.

REFERENCES

- Cocal, C. J. (2019). *Cyberbullying practices and experiences of Filipino college students' social media users*. **Journal of Education, Management and Social Sciences**, 2(1).
- Cyberbullying in social media and online games among Chinese college students and its associated factors. (2021). *International Journal of Environmental Research and Public Health*, 18(9), 4819. <https://doi.org/10.3390/ijerph18094819>
- Cybercrime Prevention Act of 2012, Republic Act No. 10175. (2012). https://lawphil.net/statutes/repacts/ra2012/ra_10175_2012.html
- DataReportal. (2025). *Digital 2025: The Philippines*. We Are Social & Meltwater. <https://datareportal.com/reports/digital-2025-philippines>
- Deocampo, C. B. C., & Yanson, J. B. (2024). Cyberbullying experiences of University of Mindanao criminology students. *American Journal of Humanities and Social Sciences Research*, 8(4), 222–228. <https://www.ajhssr.com/wp-content/uploads/2024/04/Y24804222228.pdf>
- Formoso, J. X. V., Lumacad, M. E. S., Lumacad, G. S., Dacer, D. R. D., Taldo, K. K. D., & Gaylo, A. G. L. (2023). Uncovering cyberbullying themes from unconsented Facebook Pitik posts through text-mining techniques. *European Journal of Humanities and Social Sciences*, 3(3), 459. <https://doi.org/10.24018/ejsocial.2023.3.3.459>
- Inao, C. (2025). *Young people in the Philippines speak out on online safety*. UNICEF Philippines. <https://www.unicef.org/philippines/stories/young-people-philippines-speak-out-online-safety>
- Li, Y., Gong, X., & Zhang, Y. (2023). Cybervictimization and cyberbullying among college students: The chain mediating effects of stress and rumination. *Frontiers in Psychology*, 14, 1067165. <https://doi.org/10.3389/fpsyg.2023.1067165>

Mangarin, R. A., & Montaño, M. D. (2021). The influence of internet usage to cyberbullying of SHS students. *International Journal of Research and Innovation in Social Science*, 5(4), 85–88.
<https://rsisinternational.org/journals/ijriss/Digital-Library/volume-5-issue-4/85-88.pdf>

Maranan, M. H., Kalalo, A., Lumbis, A., & Escueta, J. (2025). Addressing cyberbullying among junior high school students: An in-depth examination of prevalence, impacts, and preventive strategies. *JPAIR Multidisciplinary Research*, 59(1), 114–127.
<https://doi.org/10.7719/jpair.v59i1.922>

Patacsil, F. F. (2019). Analysis of cyberbullying incidence among Filipina victims: A pattern recognition using association rule extraction. *International Journal of Intelligent Systems and Applications*, 11(11), 48–57. <https://doi.org/10.5815/ijisa.2019.11.05>

Qin, Y., Duan, W., Wu, Q., & Lu, Z. (2024). *Dismantling gender blindness in online discussion of a crime/gender dichotomy on Weibo*. arXiv.
<https://arxiv.org/abs/2403.01697>

Quero, S. M. C. (2022). *Coping with FB cyberbullies: A survey among college students in the Philippines*.
https://www.researchgate.net/publication/364830920_Coping_with_FB_Cyberbullies_A_Survey_among_College_Students_in_the_Philippines

Reyes, C. C., Orong, M. J. T., Padao, M. P. Jr., & Pendang, C. E. (2025). Unveiling the dimensions of cybercrime victimization: An exploratory factor analysis. *International Journal of Research and Innovation in Social Science*, 9(7), 315–327.
<https://dx.doi.org/10.47772/IJRISS.2025.90700023>

Safe Spaces Act, Republic Act No. 11313. (2019).
https://lawphil.net/statutes/repacts/ra2019/ra_11313_2019.html

Sanchez-Danday, A. S. Q. (2023). Cyberbullying among college students: A case study. *European Modern Studies Journal*, 7(3), 19–31.
[https://doi.org/10.59573/emsj.7\(3\).2023.19](https://doi.org/10.59573/emsj.7(3).2023.19)

Tadena, S., Kim, S.-J., & Lee, J. (2021). Empathy, cyberbullying, and cybervictimization among Filipino adolescents. *Child Health Nursing Research*, 27(1), 65–74.
<https://doi.org/10.4094/chnr.2021.27.1.65>

Zhang, R. (2024). The impact of the “Regulations on the Governance of Online Violence Information” on user behavior in Chinese social media platforms. *Law and Economy*, 3(11), 23–27. <https://doi.org/10.56397/LE.2024.11.04>

Zhou, Z., Wang, Z., & Zimmer, F. (2022). *Anonymous expression in an online community for women in China*. arXiv. <https://arxiv.org/abs/2206.07923>