

Why Instagram's Anti-Bullying Policies Fail Pakistani Teens: A 2026 Analysis of Enforcement Gaps and Proposed Reforms

EMAAN MUSTAQEEM

emaanmustaqeem569@gmail.com

Abstract

Instagram has Community Standards that ban bullying and harassment, including rules against fake accounts, repeated unwanted messages, shaming, and threats. The platform also provides tools such as AI comment filters, reporting options, and "Teen Accounts" launched with Pakistan's PTA in October 2025 (PTA, 2025). Yet, enforcement is still lacking for Pakistani teens. The main reasons include language differences, delayed responses, and limited local support. This paper reviews the policies, examines relevant studies, explains the research approach, analyzes enforcement challenges, and proposes practical reforms.

Introduction

Instagram is widely used by Pakistani teenagers for staying connected and sharing experiences. However, it has also become a common space for cyberbullying that affects students at home and in school. Official data from the Ministry of IT shows that Pakistan recorded 1,285 cybercrime cases in 2025, with online harassment forming the largest share at 545 cases (42 percent). Of the 545 harassment cases, 360 victims were women, 110 were men, and 15 were children (Ministry of IT, 2025).

However, these registered cases represent only a fraction of the actual total complaints. In reality, the Pakistani people filed over 142,000 cybercrime complaints in 2025 (Daily Aaj, 2025). According to the Digital Rights Foundation's Digital Security Helpline Annual Report 2025, major social media platforms including Instagram accounted for 53% of the reported cyber harassment cases, highlighting Instagram's central role in the crisis (Digital Rights Foundation, 2025). Despite the large scale, many more incidents go unreported due to social stigma and fear of family. Although Meta's anti-bullying policies are detailed, their effectiveness in Pakistan remains limited. This paper explores the main reasons for these problems and suggests steps to improve protection for youth.

Literature Review

Cyberbullying is a growing problem everywhere in the world, especially among teenagers who spend most of their time on social media. While Instagram has created community guidelines and AI moderation tools, studies show these systems are built mainly around English language content, meaning harassment in other languages frequently goes undetected. Research confirms that automated content detection methods for harmful language remain significantly underdeveloped for Urdu (Raza et al., 2025).

In Pakistan, the cyberbullying problem is well documented in various studies. Saleem et al. (2021) found cyberbullying victimization to be widespread among Pakistani youth with serious psychological effects. A recent study by Aftab and Iqbal (2026) reports that over 55% of adolescents surveyed in Lahore had experienced online victimization. According to the Digital Rights Foundation (2023–2025), Urdu-language harassment frequently slips through moderation filters, while cultural stigma discourages victims from reporting. Yet no existing research has examined why Instagram's policies struggle specifically in Pakistan's linguistic and cultural setting, nor offered practical reforms. This study addresses that absence by showing how Urdu harassment and stigma bypass enforcement, and by proposing actionable solutions for Meta, the PTA, and Pakistani schools.

Research Question

Why do Instagram's anti-bullying policies fail to provide enough protection for Pakistani teens,

even after the launch of Teen Accounts in October 2025, and what practical reforms can help fix the main enforcement gaps?

Methodology

This paper uses a qualitative approach to identify gaps in Instagram's anti-bullying policies and suggest practical solutions. The analysis is mainly based on a careful review of official documents and existing studies. These include Meta's Community Standards and transparency reports (2025–2026), PTA announcements about Teen Accounts (October 2025), the Ministry of IT Cybercrime Report 2025, and peer-reviewed studies such as Saleem et al. (2021) and Aftab and Iqbal (2026). Reports from the Digital Rights Foundation (2023–2025) are thoroughly examined. All sources were carefully read and cross-checked as of April 2026.

To gather direct perspectives, an anonymous short survey will be conducted. The survey will consist of 10-12 questions targeting around 50-60 Pakistani teens (aged 13–17) about their experiences with bullying on Instagram, whether they use Teen Accounts, why many do not report problems, and how well the safety tools work. This data will be collected over a period of 2-3 weeks. The responses will be analyzed thematically and this will help compare official policies with the real experiences of students in Pakistan.

Conclusion

Instagram's anti-bullying policies are thorough in design but provide insufficient protection for Pakistani teens. Poor detection of Urdu-language harassment, slower response times, low reporting due to stigma, and limited reach of Teen Accounts are the major enforcement gaps. These weaknesses lead to increased anxiety, depression, and reduced online participation, especially among girls. To improve safety, Meta should strengthen AI moderation for local languages. They should hire more local reviewers and publish Pakistan-specific enforcement data. Stronger cooperation between the PTA and Meta is also needed to speed up content removal and expand protections for users under 18. Meanwhile, Pakistani authorities and schools must enforce cyber laws more effectively and introduce digital safety education. Coordinated action can help prevent future harm to the country's youth.

References

Aftab, M., & Iqbal, S. (2026). Socio-demographic factors, cyberbullying victimization, and digital resilience among adolescents in Pakistan. *Humanities and Social Sciences Communications*. <https://doi.org/10.1057/s41599-026-07012-6>

Daily Aaj. (2025, December 1). *Pakistanis file over 142,000 cybercrime complaints in 2025*. <https://en.dailyaaj.com.pk/Pakistan/details/15947/Pakistanis-File-Over-Cybercrime-Complaints-in>

Digital Rights Foundation. (2023–2025). Reports on online harassment and digital rights in Pakistan. <https://digitalrightsfoundation.pk/>

Digital Rights Foundation. (2025). *Digital Security Helpline Annual Report 2025: Navigating evolving threats, strengthening digital resilience*. <https://digitalrightsfoundation.pk/digital-security-helpline-annual-report-2025/>

Meta Transparency Center. (2026). Community Standards: Bullying and Harassment. <https://transparency.meta.com/policies/community-standards/bullying-harassment/>

Ministry of IT. (2025). *Pakistan Cybercrime Report 2025*. National Cyber Crime Investigation Agency (NCCIA), Government of Pakistan.

PTA. (2025, October 29). PTA and Meta launch Instagram's "Teen Accounts" to strengthen online safety for youth in Pakistan. Pakistan Telecommunication Authority. <https://www.pta.gov.pk/>

Raza, G. M., Naseem, A., Iqbal, S., Masood, T., Kim, B. S., & Nguyen, T. N. (2025). Intelligent model for target identification from Urdu threatening tweets in low-resource language using deep learning and fine-tuning of language models. *Journal of Information and Telecommunication*. <https://doi.org/10.1080/24751839.2025.2576951>

Saleem, S., Khan, N. F., & Zafar, S. (2021). Prevalence of cyberbullying victimization among Pakistani university students. *Technology in Society*, 65, Article 101577. <https://doi.org/10.1016/j.techsoc.2021.101577>