International Journal of Management and Data Analytics
IJMADA

Vol 6 (1), 132-145
ISSN: 2816-9395
DOI: https://doi.org/10.5281/zenodo.18828429
Journal website: http://ijmada.com

Navigating Online Threats: Understanding, Evaluating, and Mitigating Online Harassment

Authors: Jann Angela Ubana, Khosro Salmani

Mount Royal University, Calgary, Canada.

Corresponding Author: ksalmani@mtroyal.ca

Received: January, 2026 Published: March, 2026

Article Info

Keywords:
Awareness; Digital platforms; Internet security; Online harassment; Online safety; Preventative measures; Social media

Abstract

The Online harassment, abusive, harmful, or threatening behaviour enacted through digital communication technologies, continues to rise across social media and other online platforms. This study investigates the nature and prevalence of online harassment, evaluates the effectiveness of current technological and policy-based protective measures, and identifies gaps in user awareness and platform-level responses. Using a mixed-methods approach, we conducted a comprehensive review of existing literature and case studies and gathered over 100 survey responses from diverse participants about their experiences with online threats, their understanding of online safety, and their perceptions of available preventative measures. Findings show that younger individuals, particularly women, experience higher levels of online harassment, while many users remain uncertain about the effectiveness of reporting processes, platform policies, and law-enforcement responses. Most respondents rely on personal actions such as blocking offenders, yet express low confidence in both platform-moderation tools and legal protections. These results highlight the need for stronger and more consistent platform policies, improved detection systems capable of identifying diverse forms of harassment, and broader education initiatives to enhance user awareness and safety practices. Theoretically, the study contributes to a deeper understanding of how demographic factors and varying user experiences shape perceptions of harassment and influence the effectiveness of digital safety measures. Practically, the findings support the development of more adaptive platform interventions, clearer reporting mechanisms, and enhanced user-education programs aimed at creating a safer and more inclusive online environment.

1.0 Introduction

As the digital landscape continues to evolve, online harassment has become increasingly prevalent as it provides individuals with a platform to target others with harmful, threatening, or abusive behavior [1-5]. Often referred to as digital abuse, online harassment is defined as the use of information and communication technologies by an individual or group to repeatedly cause harm to another person, including threats, embarrassment, or humiliation within a digital setting [1]. It can take various forms such as cyberbullying, hate speech, trolling, doxing, catfishing, cyberstalking, phishing emails, hacking, malware attacks, and more [2].

While such negative encounters can happen anywhere and at any time, most online harassment incidents take place on social media platforms like Instagram, Facebook, X, TikTok, YouTube, including gaming platforms [3]. Furthermore, the accessibility and anonymity promoted by these digital platforms often influence instances of online harassment, making it difficult for online platforms to regulate and mitigate the risks and occurrences of harmful interactions [1]. As the consequences of such incidents go beyond the digital environment, it often results in individuals' psychological distress, physical safety concerns, and reputational damage [1].

For instance, in February 2024, a woman from Ontario, Canada named Angie Sweeney, was shot and killed by her ex-partner, Bobbie Hallaert [16]. Following their breakup, Sweeney blocked Hallaert on social media but continued to harass her by sending one-cent electronic transfers with abusive messages and threats, ultimately leading to Hallaert taking her life [16]. This tragic incident highlights the importance of recognizing the different forms of online harassment, and the need to address systemic gaps in protective measures including the implementation of robust policies and awareness around online harassment [12,18].

As online platforms continually face new challenges in monitoring and regulating online interactions to ensure user safety [7,12,18], online harassment policies and detection strategies must also adapt to address emerging forms of threats and ever-changing online behavior. In addition to implementing protective solutions, it is crucial to evaluate their effectiveness, accessibility, and inclusivity, to adequately address such behaviors. Accommodating new technologies and considering diverse user experiences is also essential to understanding the intricacies of present and future incidents of online harassment [9,12].

To close these gaps, researchers emphasize the value of clear, consistently enforced platform rules aligned with legal frameworks and public facing guidance. Cross platform analyses reveal that definitions, reporting pathways, and enforcement practices still vary widely, creating confusion for users and limiting accountability when harm occurs [12]. Complementing platform policy, national guidance and legal remedies help set baselines for prohibited conduct, deterrence, and victim support, while public reports highlight where harassment most commonly occurs, informing targeted prevention and governance [6, 18]. Together, this points to the need for policy standardization, transparent enforcement metrics, and accessible appeals that are communicated in plain language to different user communities [12, 18].

Monitoring and moderating online interactions remains both technically difficult and resource intensive. Automated systems often struggle to understand context, recognize adversarial or evolving language, and keep up with new forms of abusive behavior. As a result, studies consistently show that detection tools face limitations in accuracy and require continual updates to remain effective [2, 3, 4, 13]. Research on moderation tools and user created blocklists also shows that platforms depend on hybrid, human in the loop approaches to interpret nuance, sort incoming reports, and adjust moderation thresholds as behaviors change over time [7]. Overall, these findings suggest that protecting users requires an ongoing, iterative process that blends machine learning, human judgment, and continuous feedback to address emerging forms of online harm across different communities and contexts [2, 7, 13].

Finally, awareness, user empowerment, and supportive design are critical complements to policy and detection. Community and industry initiatives demonstrate how reporting aids, safety prompts, and onboarding that teach recognition of harassment patterns can reduce harm and improve help seeking efficacy [9, 12]. Studies centered on those most targeted by abuse underscore the importance of inclusive safety features and guidance tailored to victims' lived strategies, such as documentation, escalation pathways, and social support, so interventions are both usable and equitable [17]. Prevention oriented tools and education (including mobile safety apps and public information on rights and resources) further expand reach beyond the platform boundary, helping users prepare for, respond to, and recover from incidents in ways that are practical, timely, and accessible [6, 14].

This research study aims to provide a deeper understanding of the digital risks individuals face, evaluating the effectiveness of current protective strategies across various digital platforms. The contributions of this study are as follows:

1. To begin, we investigate various case studies and current research based on online harassment, along with our preliminary findings to identify patterns, challenges, and impacts.

2. We then explore existing online safety detection software, policies and tools, as well as evaluate the education on preventing online harassment and threats.

3. We analyze a survey, gathering participants' demographic information, their views on online safety, and awareness of online harassment, both before and after presenting our findings.

4. Lastly, we perform a comprehensive analysis of the collected data and offer recommendations on potential solutions to prevent or improve the management of future incidents of online harassment.

The remainder of the paper is structured as follows: In Section 2, we present the literature review, and, in Section 3, we cover where we currently are in addressing online harassment in the digital environment. Section 4 outlines our methodology, detailing the survey design and data collection process. In Section 5, we then present our survey results, and in Section 6, we conduct a comprehensive analysis presenting recommendations on future prevention measures. Lastly, Section 7 provides a summary of our conclusions.

2.0 Literature Review

In this section, we review several case studies and articles that explore different user experiences related to gender identities and racial stereotypes. We discuss current technology solutions used across online platforms and the implementation of policies and legal protections to combat online harassment.

2.1 Experiences in Online Harassment

Implementing online safety solutions that meet the needs of every individual can be challenging, as user experiences vary widely. In [9], studies show that men, compared to women, experience online harassment more often, primarily because men receive more comments on their opinions and attitudes. When it comes to expressing opinions publicly, men tend to be more vocal, whereas women are typically more cautious and selective in their online interactions. In contrast, [5] examines how online harassment directed specifically at women of color, often reflects racial and gender-related stereotypes. For instance, Black women typically experience targeted messages regarding promiscuity, while Hispanic/Latinx women often experience xenophobic and political remarks.

Additionally, online harassment can take many different forms, including but not limited to cyberbullying, catfishing, trolling, and so on. It is not limited to direct text-based or verbal communication but can also occur through visual content such as images, memes, and emojis. Images may contain offensive language or depict incidents of online harassment, and memes can mock or humiliate individuals in ways that are harmful or threatening. In some cases, individuals may strategically use a sequence of emojis that communicate a story, imply threats or mockery without explicit text. Due to the high accessibility and anonymity of the internet across digital platforms, individuals of all ages, genders, and backgrounds are vulnerable to becoming targets or perpetrators of online harassment.

2.2 Detection Software/Models for Online Harassment

The complexities and failures of online platforms to effectively address its various forms highlight the need for more sophisticated detection software. As noted in [2], the authors concluded that many online platforms fail to consider diverse individual experiences, instead design solutions for a homogeneous user base, making these technologies ineffective and inconsistent.

Although diverse user experiences are important to consider when designing solutions that can better detect various online harassment incidents, these solutions are not always the most effective. Furthermore, the way technological solutions detect offensive or harmful content also plays a significant role. In [4], Chen et. al. emphasizes the need for an approach that helps identify users who are likely to post harmful or offensive content. As current message-level detection methods lack accuracy, due to messy, informal, and misspelled language [4], the authors introduce a model called the Lexical Syntactic Feature (LSF) framework that filters bad words, insults, and analyzes sentence patterns to detect content related to online harassment and users. Beyond simple keyword detection, the LSF approach essentially differentiates the types of harmful language, outperforming existing strategies by considering writing-styles and sentence structure to enhance its predictive ability.

Similarly, Bretschneider et. al. [3] explores a pattern-based model that helps classify online harassment by linking harmful phrases or behavior to a user, aiming to block or flag harmful messages within social networks. They propose a sequence-based model which preserves word order within a message improving better detection of offensive and harmful behavior, classification performance, and person identification related to har-assment [3]. While studies show that this solution also outperforms existing solutions, it does not effectively capture online harassment as this approach relies on predefined patterns.

Overall, to address these challenges, recent initiatives to better detect online harassment focus on artificial intelligence or machine learning to automate harassment detection in online platforms. As Kennedy et. al. [9] examines human-reliant moderation such as user-reporting detection measures as ineffective, this is due to the fact that it often results in slow response times and inconsistencies to identify and address harmful behavior. Additionally, the process for users to fill out reports with detailed context is time-consuming and unreliable. To overcome these limitations, machine learning based models have proposed to automate the identification of such behavior to reduce reliance on human-identification solutions. However, due to the limited availability of online harassment datasets and classification, current AI-powered models face challenges in developing and implementing a universal detection framework that can accurately detect harassment across various platforms.

2.3 Policies and Legal Protections Against Online Harassment

The implementation of suitable platform policies and legal protections play a significant role in mitigating online harassment, yet inconsistencies and unreliable enforcement across online platforms, specifically social media present major challenges. Additionally, how they define and classify such incidents can lead to complications in developing effective prevention solutions.

In [12], a review of harassment-related content across fifteen social media platform policies revealed inconsistencies in how they classify and address harassment. Collecting policy documents from those platforms, both informal and formal, their findings presented that not one platform contains a policy that strictly defines online harassment. Though, platforms including Instagram and Twitter/X, describe responses to specific behaviors that are related to harassment such as “repeated unwanted contact” [12]. Moreover, they present additional responses to address harassment as stated in their policy documents, where deleting accounts and involving law enforcement seem to be common amongst the platforms.

As these inconsistencies highlight a broader issue such as the lack of clear and definitive legal frameworks to address and classify online harassment effectively, they underscore the need for stronger enforcement of consistent and protective measures. While policies and legal protections continue to struggle with modern technological advancements and evolving online behaviors, this often results in insufficient resources and inadequate support systems for individuals who experience online harassment. Ahmed suggests that

to bridge this gap, it is essential to implement comprehensive national policies in addressing harassment while prioritizing the protection of victims' rights. Additionally, [1] advocates for investing in research and data collection to better understand the extent of online harassment to effectively implement robust policies.

3.0 Where We Currently Are

With digital platforms playing an integral part of daily communication and interaction, ensuring online safety has become a crucial challenge to overcome. While many online platforms have implemented software tools and policies to help combat online harassment and threats, the effectiveness of these measures remains a topic of debate. Various companies continue to lack adequate policies and software solutions for detecting and mitigating harmful behavior, making this an ongoing concern in the digital landscape. Although studies show that online harassment incidents occur similarly across social media platforms including Instagram, X, Facebook, TikTok, etc., online safety detection solutions differ between platforms, addressing harassment in different ways. However, because the effectiveness, consistency, and enforcement of these solutions vary widely, they must also be adaptable and inclusive to the needs of individuals across all ages and backgrounds.

3.1 Current Online Safety Detection Tools

One of the few available tools designed to combat such harmful behaviors is BullStop, a mobile application that detects and prevents cyberbullying and online abuse in social media. According to [13], BullStop "uses deep learning models to identify instances of cyberbullying and can automatically initiate actions such as deleting offensive messages and blocking bullies on behalf of the user" (p. 70). Currently available in the Google Play Store, it appears to be one of the only applications specifically developed for addressing cyberbullying across all age groups.

Apart from BullStop, most online platforms now seem to heavily rely on AI-driven mechanisms to detect online harassment and threats, while others continue to rely on human assessment and judgment instead. Although human assessment remains a practical approach in managing incidents of online harassment, it is often an inefficient and ineffective method due to the time required for moderators to analyze and respond to incidents [9]. Typically, users must submit a report detailing their experience, wait for a response, and then receive guidance on how to proceed. While AI-driven detection can proactively identify online threats and harassment, its effectiveness and accuracy remain an ongoing challenge for many online platforms. As previously discussed, because online harassment can occur through the use of images, memes, and emojis, current detection tools struggle to identify such non-textual forms of abuse.

Despite the fact that online platforms have made significant progress in moderating and addressing online harassment throughout the years, continuous improvement is crucial due to the ever-growing digital landscape. The rise of data analytic tools has played a key role in online safety detection including platforms such as Google Analytics, Brandwatch, Hootsuite, Meltwater, etc. Each serves different functionalities in online safety detection: Google Analytics not only analyzes web traffic but also analyzes each visitor and their behavior across different social media platforms [8]. Brandwatch has the ability to filter and duplicate content, being one of the most trusted tools due to its accuracy and effectiveness [8]. Hootsuite is an open-source tool, tracking standard user interactions, categorizing and marking their performance [8]. Meltwater is a social media monitoring platform that tracks user-generated content and monitors relevant conversations within social media platforms [15].

3.2 Current Policies and Legal Frameworks for Addressing Online Harassment

Although the implementation of online safety tools and AI-powered detection solutions are critical in mitigating online harassment, policies and legal protections remain essential to

addressing online harassment. Nonetheless, enforcing such policies and legal protections can establish clear guidelines for effective enforcement. For instance, Canada has implemented strict legal consequences for cyberbullying including the possibility of jail time. In addition, offenders may have their devices taken away or compensate their victims, as well as face additional consequences based on the severity of their actions [6].

While social media platforms have policies aimed at addressing and promoting online safety, they often fail to explicitly reference online harassment and its many forms. Instead, their policies and guidelines refer to broad behaviors that are commonly associated with harmful behaviors and online interactions.

4.0 Methodology

To gain a deeper understanding of common online harassment experiences, how different groups respond, and how online safety and privacy can be improved through enhanced software tools and policies, we conducted a structured 15-minute survey composed of both multiple-choice and short-answer questions. The survey was administered digitally using an online data-collection platform approved by the Human Research Ethics Board (HREB) at Mount Royal University. Prior to beginning the survey, participants were presented with information outlining the purpose of the study, expected time commitment, voluntary nature of participation, and procedures for data handling. Participants provided informed consent electronically. All responses were collected anonymously and stored securely in encrypted, password-protected institutional storage.

4.1 Recruitment

Recruitment involved circulating an online announcement containing the study description and survey link through publicly accessible channels. We also promoted the study on campus by sharing the invitation in common student areas and through classroom announcements, which helped diversify our respondent pool. No incentives were offered for participation, and individuals were eligible to complete the survey if they were at least 18 years old and used the internet regularly. Through these combined recruitment efforts, we received over 100 valid survey submissions.

A total sample size of just over 100 participants was sufficient for the goals of this exploratory study, as our intent was not to make population-wide generalizations but to identify common patterns, experiences, and perceptions related to online harassment. For survey-based exploratory research, samples of 100 or more are generally considered adequate to reveal clear trends, compare responses across categories, and identify recurring issues within a defined community or demographic group. Because the study was primarily advertised on campus, most respondents fell within the 18–24 age range. This age group is also one of the heaviest users of social media and digital communication platforms, and prior research identifies young adults as a population that experiences and witnesses online harassment at notably high rates. As a result, receiving most responses from this demographic aligns well with the study's focus and provides meaningful insight into the online risks faced by a group that is highly active in digital spaces.

4.2 Survey Framework

The survey was designed to gather both quantitative and qualitative insights. It followed a structured sequence of themed question blocks to ensure consistency across participants and to support organized analysis of trends and patterns.

Classification and Demographics. Participants first completed demographic questions, including age range, gender identity, highest level of education, and perceived digital literacy or comfort with online safety practices. These variables allowed us to examine whether certain demographic groups reported different levels of exposure or responses to online harassment.

Online Activity and Exposure. In the second section, participants were asked about their general internet use. This included frequency of use, primary online activities (e.g., social networking, gaming, streaming), and which social media platforms they used regularly. Participants also rated how safe they felt online using a Likert scale. These questions provided contextual information about individual exposure levels and perceptions of online safety across different platform environments.

Experience with Online Harassment and Threats. Next, participants were asked about their understanding and experiences with online harassment. We used a five-point Likert scale to measure awareness, followed by questions asking whether they had ever experienced or witnessed online harassment. This section also assessed recognition of common harassment forms, including cyberbullying, hate speech, impersonation, and threats. Participants identified the platforms where such incidents occurred and indicated whether they knew the identity of the perpetrator(s). These responses allowed us to gauge broader awareness gaps and the contexts in which harassment is most commonly encountered.

Reporting Online Harassment and Threats. This section focused on individual reporting behavior. Participants were asked whether they reported incidents, the reasons for reporting or not reporting, the platform or authority to which they reported, and the estimated time it took to receive a response. They also evaluated whether the reporting process was effective in improving or resolving the situation. These responses helped identify procedural barriers, perceptions of platform responsiveness, and levels of trust in existing reporting systems.

Addressing Online Harassment and Threats. The final section examined participants' attitudes and perceived responsibilities related to preventing or addressing online harassment. Using Likert-scale items, participants rated statements about the role of bystanders, victims, and online communities in preventing future incidents. They were also asked open-ended questions about their opinions on the effectiveness of current prevention measures, platform policies, and broader awareness efforts. These responses offered insights into perceived system shortcomings and potential areas for improvement.

5.0 Analysis and Results

The following section presents the results of our survey, conducted over the span of 2.5 weeks. In May 2024, we began our preliminary study, where we collected information from various case studies and articles to present to participants. During the initial phase of our study, we requested and were granted approval from The Human Research Ethics Board of Mount Royal University, as human involvement was required to officially conduct our study. In February 2025, we began promoting our survey. We created a LinkedIn post inviting individuals to participate. Using Google Forms to conduct our survey, we aimed to ensure that the questions being asked were relevant and structured to gather meaningful insights into each participants' experiences and perspectives related to online safety.

Noting the limitations of our study, our participation pool consists of individuals from diverse age and gender identities. As participants could select multiple options, the provided percentages indicated the frequency of each response rather than a cumulative total. The results from the survey highlighted various user experiences based on certain demographic factors, articulating how diverse groups navigate online threats and harassment. It revealed that the younger demographic report higher exposure to online harassment in social media, while awareness of online safety measures varies significantly among different age and gender groups. These insights provided us with a deeper understanding of the digital risks individuals face and the effectiveness of current protective strategies across various digital platforms.

5.1 Classification and Demographics

The results exhibited a diverse range of user backgrounds including different age groups, gender identities, and levels of online exposure. Among the 104 participants, 49% identified as "Male" and 49% identified as "Female", while the remaining 1.9% identified as "Other". The majority of participants belonged to the age group 18 to 25, either completing or have obtained a Diploma or Bachelor's Degree, which reflects a population that is highly engaged with digital platforms like social media. This age group represents one of the most active users of social media and digital communication platforms. Prior research consistently identifies young adults as a population that both experiences and witnesses online harassment at disproportionately high rates. Consequently, the predominance of responses from this demographic is well aligned with the study's objectives and offers valuable insight into the online risks encountered by individuals who are highly engaged in digital environments.

Considering these demographic factors, the perceptions of online safety differed as most participants expressed their confidence in their online safety, while a smaller percentage reported concerns or uncertainty when it came to the question "I feel safe using the internet...". Our findings presented key trends and differences in how certain groups experience and navigate online harassment, ultimately offering valuable insights into the general landscape of online exposure and safety.

5.2 Online Activity and Exposure

As digital platforms provide various avenues for social media, education, communication, entertainment, etc., they have also become primary spaces for online threats and harassment. As expected, while conducting our survey, 99% of participants use the internet daily, primarily for social media, communication, and entertainment (see Fig. 1). With 86.5% using Instagram regularly, it is the most popular social media platform among participants as it essentially allows for communication, content sharing, and user interaction. Our findings indicate that 49% of participants had no strong opinion when asked about their feelings towards online safety. Surprisingly, a higher percentage of females within the 18 to 25 age group chose this response, which is interesting given that females are typically more susceptible to online harassment and threats compared to males. This suggests that the majority remain uncertain or neutral regarding their internet security, which emphasizes the lack of confidence in existing online safety measures across different digital platforms.

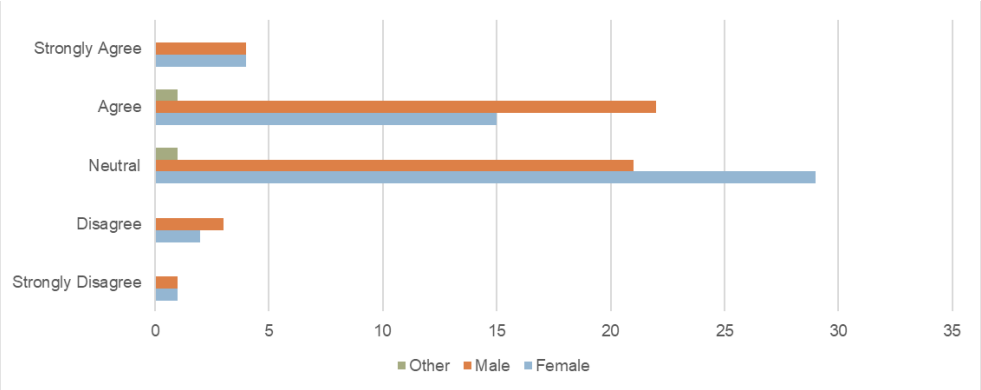

Fig 1. Perceived Online Safety Levels by Gender.

5.3 Individual Experiences with Online Harassment and Threats

When it comes to participants' individual experiences, 53 participants showed confidence in their understanding of what online harassment is, while 34 participants showed some level of awareness but possibly lacked complete understanding (see Fig. 2). Although online harassment is a prevalent issue in modern society, this lack of understanding may be due

to potential gaps in education or clarity as to what constitutes online harassment and threats. While only 86.5% of participants claimed to have experienced or witnessed online harassment, we learned it is possible that some incidents of online harassment may go unrecognized or aren't deemed as "harassment" due to different user experiences and interpretations of what defines harassment or desensitization to behaviors within the digital environment.

This finding suggests the need for more awareness and implementation of preventative measures to help users recognize and respond to online threats and harassment effectively. Significantly, our findings also revealed that female participants reported higher numbers compared to male participants when it comes to experiencing online harassment and threats, highlighting the need for gender-specific approaches to addressing these issues. It is also likely that individuals who have experienced or witnessed online harassment possess a stronger understanding of the issues, while those who have not may be less informed about their different forms and classifications.

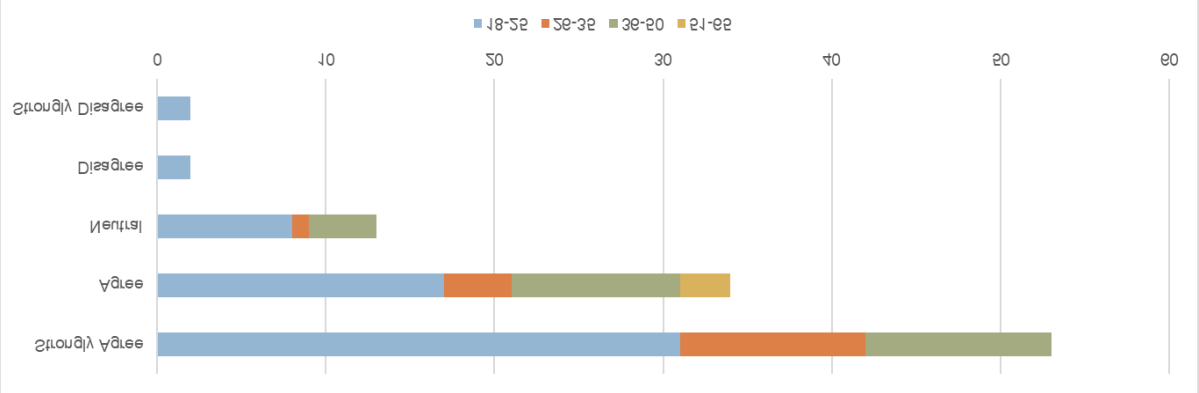

Fig. 2. Awareness of Online Harassment by Age Group

Additional key assumptions from participants' experiences, present that the most common form of online harassment and well-known or understood amongst the majority of participants is "Cyberbullying", "Trolling", and "Catfishing". Although "Doxing" was recognized as a form of online harassment, the lack of participants who selected this option might suggest that other existing forms may not be as widely acknowledged or understood due to the absence of coverage regarding these issues.

5.4 Reporting Online Harassment and Threats

As individuals interpret online harassment and threats differently, efforts to identify and respond to such issues effectively can be difficult due to the varying levels of awareness. While a majority of participants have experienced or witnessed online harassment, only 47.11% claimed to have reported such incidents, whereas 33.65% chose not to. When asked in the survey "If you chose not to report an incident of online harassment, what were your reasons?", many indicated that they did not care or feel as if the incident was "severe enough" or believed that harassers have greater freedom to express harmful behaviors compared to in-person interactions, thus encouraging them to ignore it. Some participants also expressed concerns with involving law enforcement as it could pose a greater risk, or believed authorities would only act if the incident was more serious.

Our findings revealed that even when users took action to address incidents of online harassment, they were unsure whether their efforts were effective at resolving the situation. When asked "Do you feel that the steps you took were effective at resolving the situation/making the situation better?" 22.8% of participants reported that the steps they took were effective, while 17.4% reported that their efforts did not resolve or improve the situation. Particularly, 38% expressed uncertainty, indicating that they were unsure whether their efforts made a difference. This finding suggests that although users take action to help

mitigate online harassment, there is no guarantee that such efforts will lead to an effective or positive resolution.

5.5 Addressing Online Harassment and Threats

In efforts to address incidents of online harassment, 80.2% of participants blocked the harasser as a means of protecting themselves, while 53.5% chose to change their passwords and update their privacy settings. This finding may suggest that blocking harasser is often seen as a more immediate and controlled approach to addressing online harassment compared to alternative protective measures like reporting the user.

Additionally, 11.9% of participants addressed incidents of online harassment by deactivating their accounts or adopting other methods to protect themselves online. While individuals are inclined to take proactive steps to protect their online presence, other protective measures may be more effective, accessible, or convenient. For instance, blocking a user is a quick and direct method that immediately allows for restricted interaction, whereas changing passwords and updating privacy settings require additional steps for the user to perform. Thus, making these methods less immediate. In addition, account deactivation appears to be perceived as a last resort likely due to the idea of losing entire access to an account and its associated content.

Moreover, blocking the harasser offers instant relief and control over the situation, in comparison to reporting a user which involves a more tedious and complex process. Reporting a user typically requires filling out a form or report with additional context and specified reasons for their actions which can delay the resolution process. This extra effort may discourage users from reporting the situation, especially if it's a significant threat. Users often choose to block a user simply because they find them annoying and not dangerous, blocking the preferred method for managing online harassment in less severe situations.

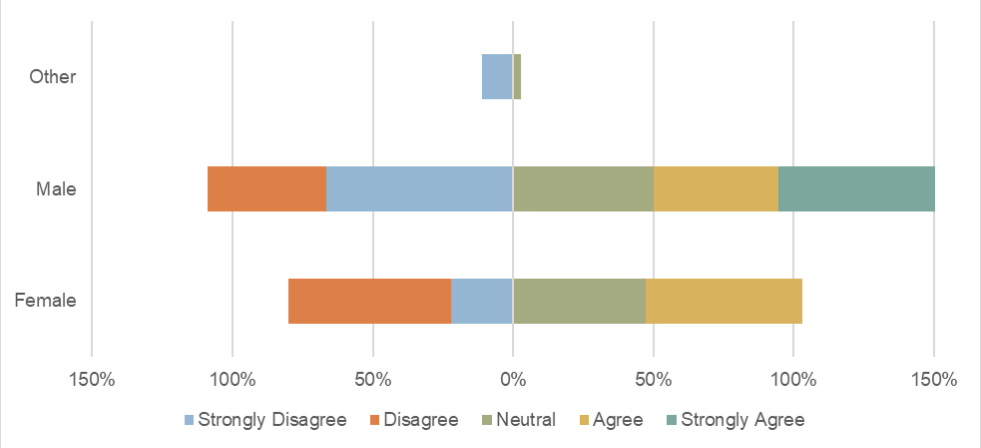

Fig. 3. Perceptions of Law Enforcement's Response to Online Harassment by Gender.

Further reflected in our findings, participants expressed skepticism towards external intervention. The majority of participants claimed that they are unaware of software tools available to detect online harassment, suggesting that many lack access to or knowledge of alternative protective measures other than blocking and updating their account's privacy settings. Participants also expressed low confidence in law enforcement and social media platforms when it comes to addressing incidents of online harassment (see Fig. 3). In response to the question "Law enforcement takes incidents of online harassment seriously..." 45% of participants disagreed and 38% were indifferent. Similarly, when asked "Social media platforms are doing enough to combat online harassment..." a significant percentage strongly disagreed, disagreed, or remained neutral, highlighting that digital platforms are not effectively addressing incidents of online harassment (see Fig. 4.).

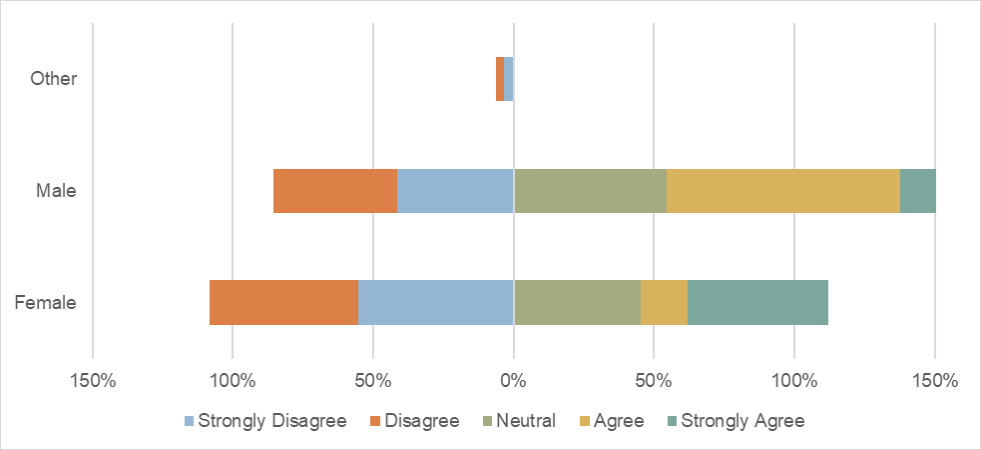

Fig. 4. Perceptions of Social Media Platforms' Efforts to Combat Online Harassment by Gender

Finally, while users take personal actions to protect themselves from online, most participants believe that systemic solutions are necessary for effectively preventing online harassment and threats. When asked "What do you think is the MOST effective way to prevent online harassment?" 77.9% of participants pointed to the need for stronger policies and tools from social media platforms to better enforce blocking, reporting, and content filtering, while 75% emphasized the importance of increasing public awareness and education on online harassment. This suggests that though users rely on blocking and other personal safety measures, long-term solutions require greater efforts from both digital platforms and legal institutions to effectively combat incidents of online harassment.

6.0 Discussions and Recommendations

While a significant percentage of participants expressed confidence in their online safety and presence, many also continue to remain uncertain or impartial, implying the lack of awareness surrounding online harassment and trust in existing protective measures. The survey highlighted that the younger demographic, particularly within the female population, experience higher levels of online harassment and threats, emphasizing the need for gender-specific approaches to addressing these issues. Despite the prevalence and risks of online harassment across digital platforms, protective measures remain inadequate and ineffective, leaving users unsure about the effectiveness of their actions or attention given by digital platforms and authorities to address incidents of online harassment.

Nonetheless, with continued efforts to take proactive steps such as blocking the harasser or updating their account settings, the responsibility to mitigate online harassment cannot depend on users. Implementing effective long-term solutions encourages systemic changes like better policies, more effective tools for detecting and blocking online harassment, and increased public awareness and education. To say the least, the survey emphasizes the need for both individual action and institutional responsibility in creating a safer digital environment for everyone.

Our findings that younger participants, especially those in the 18–25 group, report higher exposure on social media align with prior work showing that harassment is concentrated on mainstream platforms and disproportionately affects certain demographics [18]. The higher self-reported exposure among female participants in our sample also echoes research documenting gendered patterns in harassment and the distinct ways in which women, and particularly women of color, are targeted online [5]. Consistent with these accounts, our respondents' uncertainty about online safety and mixed confidence in their own protections mirrors the uneven risk landscape observed in earlier studies of user experiences and community harms [5, 18]. Together, these parallels suggest that the

patterns we observe locally are not idiosyncratic but instead reflect broader, well-documented dynamics in the literature.

At the same time, participants' reliance on blocking, their ambivalence toward reporting, and their low confidence in platform and law-enforcement responses are consistent with prior critiques of moderation workflows and policy frameworks. Comparative policy analyses show that platforms vary widely in definitions, reporting pathways, and enforcement, contributing to user confusion and skepticism about outcomes [12]. Likewise, studies of moderation tools and community blocklists emphasize that platforms continue to face technical and operational limits in detecting and managing evolving abusive behaviors at scale, which helps explain users' preference for immediate, self-help actions over formal reports [7]. Technical work on detection further clarifies why users experience inconsistent remediation: message-level models struggle with noisy, informal, and adversarial language, requiring continual updates and hybrid (human-in-the-loop) approaches to maintain effectiveness [3, 4]. In this context, our respondents' call for stronger policies, clearer processes, and better tools directly resonates with recommendations across the literature to standardize platform rules, improve transparency and enforcement, and invest in socio-technical moderation that can keep pace with emerging harms [7, 12, 18].

In the following section, we offer potential solutions and recommendations to help reduce incidents of online harassment and enhance existing protective strategies. It is important to acknowledge that our recommendations are based solely on our survey results and findings. Through this study, we intend to promote a safer online environment for individuals of all ages and gender identities, regardless of their level of awareness of digital platforms.

6.1 Recommendations

Based on our research findings and survey results, we propose several recommendations aimed at reducing incidents of online harassment and improving current protective strategies. Though there is no single solution to entirely eliminate online harassment, prioritizing the development of robust detection and prevention tools, strengthening platform policies, and enhancing user awareness can significantly contribute to a safer digital environment for everyone.

Enhanced Reporting Mechanisms. One recommendation is to improve reporting mechanisms in online platforms, particularly social media. To help users feel more comfortable with the reporting process, increasing transparency around how reports are handled and reducing response times could promote quicker and more effective resolutions. Additionally, providing clear, accessible, and user-friendly interfaces may influence a more proactive approach to addressing online harassment.

Robust Platform Policies. Similar to the idea of blocklists [7], online platforms, specifically social media, should enforce stronger content moderation policies. By blocking certain interactions or implementing automatic bans for repeat offenders, implementing stronger policies in general can help prevent harmful behaviors and control negative interactions. Furthermore, setting clear guidelines and encouraging effective enforcement to ensure individuals are held accountable for their actions can allow users to feel empowered to report incidents of online harassment without fear of retaliation.

Enhanced Detection Systems. Integrating advanced technological detection systems such as automated threat detection algorithms or community driven moderation can help identify harmful behavior and flag incidents effectively to reduce incidents of online harassment and threats. By implementing prevention solutions that are tailored to diverse user experiences, rather than adopting a one-size-fits-all approach, may also improve the effectiveness and accuracy of detecting online harassment and threats in online platforms.

Enhanced Education Initiatives. Although many online platforms advocate for online safety and provide resources for harassment prevention, schools and workplaces play a significant role in educating individuals on the importance of recognizing and addressing

online harassment. Introducing online safety education can empower users across all demographics to better identify various forms of online harassment and effectively mitigate associated risks by implementing tutorials or support guides for users on how to monitor, report, and address harassment on online platforms. Nonetheless, enhanced education initiatives can help raise awareness of the prevalence of online harassment and threats.

7.0 Conclusions

This study offers exploratory evidence about how a primarily campus-recruited sample (N≈104), largely aged 18–25, perceives and responds to online harassment. Within these bounds, the data indicate that many respondents are uncertain about their online safety, that younger participants report higher exposure on social media, and that blocking is the most common immediate response, while formal reporting remains inconsistent and is often viewed with skepticism.

Participants also expressed limited confidence in platform and law-enforcement responses and identified a need for stronger policies, clearer processes, and better awareness resources. A key finding from our research is that diverse user experiences significantly influence individuals' definition and perception of online harassment, which in turn can strongly impact the effectiveness of online safety measures. Through our research, we deepened our understanding on the nature of online harassment and how it has evolved alongside technological advancements. The convergence of high exposure among young adults, variable awareness, and preference for self-help actions highlights actionable gaps for platforms and institutions, particularly around reporting usability, transparency of enforcement, and accessible safety education.

In this study we also outlined four key recommendations: enhanced reporting mechanisms, more consistent platform policies, improved detection support, and targeted education. These recommendations should be viewed as practice-oriented priorities suggested by respondents' experiences, rather than prescriptive solutions validated across all demographics.

Future work should broaden sampling beyond a campus context, incorporate richer behavioral measures (e.g., platform logs or longitudinal follow-ups where feasible and ethical), and examine subgroup differences more systematically to understand how needs vary across age, identity, and platform use. Such extensions would allow for stronger external validity and a clearer link between specific platform features, policy enforcement practices, and user outcomes over time.

References

[1] R. Ahmed (2024). The cyber harassment in the digital age: Trends, challenges, and countermeasures. Radinka Journal of Science & Systems Literature Review, 2, 442–450.

[2] L. Blackwell, J. Dimond, S. Schoenebeck, & C. Lampe (2017). Classification and its consequences for online harassment: Design insights from HeartMob. Proceedings of the ACM on Human-Computer Interaction, 1, 1–19. https://doi.org/10.1145/3134659

[3] U. Bretschneider, T. Wöhner, & R. Peters (2014). Detecting online harassment in social networks. In Proceedings of the International Conference on Information Systems – Building a Better World through Information Systems (Auckland, New Zealand, December 14–17, 2014).

[4] Y. Chen, Y. Zhou, S. Zhu, & H. Xu (2012). Detecting offensive language in social media to protect adolescent online safety. In Proceedings of the 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust (Amsterdam, Netherlands, September 3–5, 2012).

[5] S. C. Francisco & D. H. Felmlee (2022). What did you call me? An analysis of online harassment towards Black and Latinx women. Race and Social Problems, 14, 1–13. https://doi.org/10.1007/s12552-021-09330-7

[6] Government of Canada (n.d.). Legal consequences of cyberbullying. https://www.canada.ca/en/public-safety-canada/campaigns/cyberbullying/cyberbullying-against-law.html (accessed March 24, 2025).

[7] S. Jhaver, S. Ghoshal, A. Bruckman, & E. Gilbert (2018). Online harassment and content moderation: The case of blocklists. ACM Transactions on Computer-Human Interaction, 25, 1–33. https://doi.org/10.1145/3185593

[8] S. Kaur & T. Pal Singh Brar (2021). Social media analytics: A review of social media tools and techniques. In Online National Conference on Wireless Communication, Computing and Informatics (pp. 91–93). Mata Sundri University Girls College, Mansa, Punjab.

[9] G. Kennedy, A. McCollough, E. Dixon, A. Bastidas, J. Ryan, C. Loo, & S. Sahay (2017). Hack Harassment: Technology solutions to combat online harassment. In Proceedings of the First Workshop on Abusive Language Online (pp. 73–77). Association for Computational Linguistics, Vancouver, Canada.

[10] M. Lindsay, J. M. Booth, J. T. Messing, & J. Thaller (2015). Experiences of online harassment among emerging adults: Emotional reactions and the mediating role of fear. Journal of Interpersonal Violence, 31, 3174–3195. https://doi.org/10.1177/0886260515584344

[11] M. Nadim & A. Fladmoe (2021). Silencing women? Gender and online harassment. Social Science Computer Review, 39, 245–258.

[12] J. A. Pater, M. K. Kim, E. D. Mynatt, & C. Fiesler (2016). Characteristics of online harassment: Comparing policies across social media platforms. In Proceedings of the 2016 ACM International Conference on Supporting Group Work (pp. 369–374). Association for Computing Machinery, New York, NY.

[13] A. Perera & P. A. Fernando (2021). Accurate cyberbullying detection and prevention on social media. Procedia Computer Science, 181, 605–611. https://doi.org/10.1016/j.procs.2021.01.207

[14] S. Salawu, Y. He, & J. Lumsden (2020). BullStop: A mobile app for cyberbullying prevention. In Proceedings of the 27th International Conference on Computational Linguistics (pp. 70–74). International Committee on Computational Linguistics (ICCL), Barcelona, Spain.

[15] I. Stavrakantonakis, A. E. Gagiu, H. Kasper, I. Toma, & A. Thalhammer (2012). An approach for evaluation of social media monitoring tools. In Common Value Management (CVM2012), 1st International Workshop on Common Value Management, Proceedings of the Extended Semantic Web Conference (Heraklion, Greece, May 27–31, 2012). Fraunhofer Verlag, Stuttgart.

[16] CBC News (n.d.). https://www.cbc.ca/news/canada/e-transfer-abuse-1.7125623 (accessed February 24, 2025).

[17] J. Vitak, K. Chadha, L. Steiner, & Z. Ashktorab (2017). Identifying women's experiences with and strategies for mitigating negative effects of online harassment. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (pp. 1231–1245). Association for Computing Machinery, New York, NY.

[18] E. A. Vogels (2021). Online harassment occurs most often on social media, but strikes in other places, too. https://www.pewresearch.org/short-reads/2021/02/16/online-harassment-occurs-most-often-on-social-media-but-strikes-in-other-places-too/ (accessed March 24, 2025).