

Digital neighbourhood moderation: online community safety and the role of Facebook group moderators in Queensland, Australia

Maram Samiee, Keith Heggart, Philip Birch and Hazel Wallace

Abstract

Purpose – The convergence of digital technologies and local public safety practices has led to the emergence of Facebook-based community groups as informal but influential platforms for neighbourhood governance. Group administrators and moderators are increasingly central to managing online safety, yet their roles remain under-examined. This paper presents findings from Phase 1 of a four-phase study that investigates the roles, responsibilities and challenges faced by Facebook group moderators in Queensland, Australia. Being a preliminary examination, this study aims to understand how these individuals contribute to digital community safety, manage misinformation and interact with formal policing structures.

Design/methodology/approach – A mixed-methods survey was conducted with 65 moderators overseeing 61 neighbourhood Facebook groups across Queensland, Australia. The survey gathered quantitative and qualitative data on moderator practices, challenges and perceptions of their role in community safety.

Findings – Findings indicate that moderators function as digital first responders, juggling content moderation, community expectations and unpaid workloads. Key challenges include the spread of misinformation, low digital literacy among group members and the absence of formal guidance from authorities. In spite of these obstacles, this study reveals opportunities for collaboration between moderators and the Queensland Police Service to foster trusted, resilient digital spaces.

Practical implications – This paper offers recommendations for police forces, digital platform stakeholders and local policymakers to strengthen collaborative frameworks and improve digital community safety infrastructure. Moderators are emerging as a new civic layer in the governance of community safety. Structural support, targeted training and formal partnerships are essential to sustain their effectiveness in digital neighbourhood governance.

Originality/value – This paper presents findings from Phase 1 of a four-phase study that investigates the roles, responsibilities and challenges faced by Facebook group moderators in Queensland, Australia. Being a preliminary examination, this paper aims to understand how these individuals contribute to digital community safety, manage misinformation and interact with formal policing structures.

Keywords Police, Social media, Community safety, Misinformation, Facebook moderators, Neighbourhood groups

Paper type Research paper

(Information about the authors can be found at the end of this article.)

Received 2 May 2025
Revised 24 June 2025
29 July 2025
Accepted 29 July 2025

© Maram Samiee, Keith Heggart, Philip Birch and Hazel Wallace. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <https://creativecommons.org/licenses/by/4.0/>

Funding: This study was supported by Queensland Police Service (No. DICE Prevention Project Grant).

Introduction and background

The proliferation of social media platforms has transformed the way communities share information, raise concerns and engage with local institutions, particularly in the area of public safety (Bullock *et al.*, 2021; Williams *et al.*, 2018; Wukich, 2021). Facebook neighbourhood and crime-focused community groups are now a primary source of local information for many Australians. Moderators (and administrators) of these groups play a pivotal role in shaping

online discourse, managing misinformation and fostering constructive engagement among residents. However, arguably, the complexities and pressures of digital community leadership remain under-researched in both criminological and communication studies literature.

The Queensland Police Service (QPS) Digital Intelligence and Community Engagement (DICE) Prevention Project in Australia was initiated to investigate how neighbourhood Facebook groups function as mechanisms of community safety and social cohesion. Phase 1 of the project, conducted in collaboration with the University of Technology Sydney (UTS), used a mixed-methods survey targeting Facebook community group administrators and moderators across Queensland [1]. The study sought to understand their motivations, strategies, challenges and support needs in managing online spaces related to crime, safety and community issues.

Acting as a preliminary examination, the article reports on the findings of that survey and situates them within the broader context of digital citizenship and public safety. It explores how voluntary moderators, often acting without formal training or guidance, are navigating complex terrains of misinformation, privacy, legal uncertainty and community concerns or distrust. Importantly, the study surfaces opportunities for structured collaboration between moderators and formal authorities, particularly the police, to enhance messaging and public safety outcomes in digital environments. The article commences by reviewing existing literature on online moderation, digital misinformation and community safety, before outlining the methodology of the study. The key findings are then thematically presented, before considering the implications of the findings.

Framing the issue

The convergence of digital communication technologies and local safety practices has given rise to new forms of informal governance within online communities. Social media groups, particularly on platforms such as Facebook, have increasingly assumed roles traditionally associated with neighbourhood watch (NHW) programs, local councils and policing authorities (Pepper and Bullock, 2025; Walby and Joshua, 2021). This shift has introduced new challenges and opportunities in the governance of community safety, especially in digital spaces where misinformation can spread rapidly and regulatory frameworks are often ill-defined.

Online moderators perform a range of functions that blend elements of community management, content curation and social enforcement (Gillespie, 2018). Their responsibilities often include setting and enforcing rules, adjudicating disputes, and making decisions about the visibility and credibility of information, and therefore they perform the role of gatekeeper. In community safety contexts, these functions are especially sensitive, as moderators may be required to respond to posts involving victimisation, allegations of criminal behaviour, personal safety concerns, or the sharing of potentially harmful content (Bloch-Wehba, 2021; Gillespie, 2018). Moderators' decisions can shape public perception, influence trust in institutions, and either mitigate or amplify community tensions.

A growing body of literature explores the role of misinformation in undermining public trust and complicating online discourse (Firdaus *et al.*, 2024; Marwick and Lewis, 2017; Pérez-Escobar *et al.*, 2023; Shahbazi and Bunker, 2024). In neighbourhood Facebook groups, misinformation may take the form of false crime reports, unverified CCTV footage, or scam-related posts. The digital literacy of both moderators and group members plays a critical role in how such information is interpreted and managed. Yet, many volunteer moderators lack formal training in digital and media literacy or investigative protocols, and platform tools for managing misinformation remain limited.

In addition to studies on misinformation and content moderation, recent work has examined the broader dynamics of platform governance and the civic role of digital communities. For instance, Rahman *et al.* (2019) explored the ethical implications of crowdsourced

surveillance in neighbourhood apps, while [Alnemr \(2021\)](#) and [Crosset and Dupont \(2022\)](#) analysed the algorithmic shaping of civic discourse in Facebook community spaces. These studies underscore how platform design and moderation tools shape user agency, civic participation, and information credibility. Comparative analyses across platforms suggest that community moderation is not only a social process but one deeply structured by the technological affordances of the platform itself ([Kiene et al., 2019](#); [Seering, 2020](#)). Integrating this perspective highlights the importance of platform-specific safety protocols and collaborative infrastructure for community-led digital governance.

Contemporary policing strategies increasingly emphasise community engagement and partnership-based models ([McKee and Lewis, 2016](#); [O'Neill and McCarthy, 2014](#)). However, the integration of these strategies into online spaces has been uneven. Studies show that while police departments may maintain official social media presences, their involvement in community-led online groups is rare and often informal ([Meijer and Torenvlied, 2016](#)). This absence creates both a gap and an opportunity: moderators are often left to fill the role of informal communicators of crime-related information, but they do so without the authority, training, or resources afforded to police. Bridging this gap requires rethinking how law enforcement agencies support and collaborate with digital community leaders.

Therefore, in response to these developments, this article presents findings from Phase 1 of a four-phase study examining the roles, responsibilities, and challenges faced by Facebook group moderators in managing digital community safety across Queensland. By investigating how these moderators navigate misinformation, mediate public discourse, and liaise with police, the study positions them as emerging civic actors and community champions in local safety governance. In doing so, it contributes new empirical insights into the informal infrastructures of digital neighbourhood watch and highlights the need for structured support, training, and collaboration between moderators, policing authorities, and digital platforms.

Finally, it is important to note that while this study focuses specifically on Facebook because of its widespread use among online neighbourhood communities in Queensland, it is important to acknowledge that other platforms such as Reddit and WhatsApp also serve as venues for digital neighbourhood engagement and moderation. These platforms differ in affordances, ranging from anonymity and decentralised moderation (as seen in Reddit) to encrypted, closed-group communication (as with WhatsApp). Facebook's structure, with visible membership and layered administrative controls, uniquely facilitates community visibility, content approval, and partnerships with formal institutions such as police. This platform specificity is a key factor in how community safety is practised digitally. These platform-specific distinctions are supported by prior research into community moderation affordances. For example, [Zhang et al. \(2024\)](#) highlight how Facebook's group architecture, such as administrative hierarchies, post approval settings, and visible membership, supports layered governance and information curation, which is less easily achieved on platforms with flat or anonymous structures such as Reddit. Similarly, [Seering \(2020\)](#) notes that Facebook's persistent group histories and structured norms allow for forms of self-regulation that more closely resemble offline community safety mechanisms. These features made Facebook a suitable and strategic focus for this phase of the DICE Prevention Project. Whilst specifically, from a QPS context, DICE Prevention's own crime and community surveys have repeatedly revealed Facebook as a preferred source for residents to receive crime and safety news and alerts.

Methodology

This study adopted a mixed-methods approach, combining quantitative survey data with qualitative insights from open-ended responses. The survey was designed by researchers from the UTS in collaboration with the QPS DICE Prevention team. Participants were

administrators/moderators of Facebook-based community groups in Queensland, identified through targeted social media outreach and UTS promotion channels. Groups and moderators were selected through purposive and convenience sampling, with recruitment promoted via the QPS official channels and to stakeholder contacts, including NHWQ and Crime Stoppers Qld; UTS communication channels; and as a paid Facebook promotion targeted to Queensland community and crime group administrators/moderators applying keywords that aligned with the study focus. While participation was voluntary, efforts were made to ensure variation in group size, group purpose, and geographic representation. However, no quotas were imposed, and representation of official Neighbourhood Watch groups versus informal community forums was not confirmed as part of the survey. Eligibility criteria included being over 18 years old, residing in Queensland, and actively serving as a relevant community or crime group administrator/moderator. The lack of publicly available data on the total number or demographic profile of Facebook neighbourhood groups and their moderators in Queensland restricts the ability to assess the representativeness of those sampled. Nevertheless, the findings provide valuable early insights into a largely undocumented but increasingly influential form of digital civic engagement.

Out of 209 responses, 144 were excluded because of incomplete data or failure to meet inclusion criteria, leaving 65 valid responses from individuals representing 61 unique Facebook groups. The survey included both closed-ended and open-ended questions covering topics such as group membership and management, moderation practices, exposure to misinformation and support needs.

Descriptive statistics were used to analyse the quantitative data, while thematic coding outlined by [Braun and Clarke \(2006\)](#) was applied to the qualitative responses to identify recurring challenges, motivations and priorities among participants. The application of thematic analysis, and [Braun and Clarke's \(2006\)](#) six-step process, including familiarisation with the data, initial code generation and identification of patterns, allowed for the responses to be independently coded by two of the researchers before collaboratively reviewing and defining the themes through iterative discussion. This process ensured the themes were grounded in participant voices while reflecting shared meanings relevant to digital community safety. Given the exploratory and policy-focused nature of this phase of the research, analysis was limited to descriptive statistics. This approach prioritised accessibility for stakeholders and ensured clarity in yielding community moderator needs. While the study drew on [Braun and Clarke's \(2006\)](#) foundational six-step method, we acknowledge the evolving nature of reflexive thematic analysis and the shift away from rigid coding reliability towards researcher reflexivity and interpretive depth. Future stages of this project will consider the guidance of [Braun and Clarke \(2023\)](#) in enhancing reflexive praxis across analysis stages. Furthermore, future research could allow for more advanced quantitative analyses to explore relationships between moderator confidence, group structure, and community outcomes. Given the exploratory and foundational nature of this research, descriptive statistics were used to provide a clear overview of patterns in the data. This decision was informed by both the small sample size ($n=65$), which limits statistical power, and the need to ensure accessibility and transparency for non-academic stakeholders, including community moderators and law enforcement personnel. With that said, findings are reported thematically to reflect the structure and focus areas of the study.

The research team included both academic researchers and practitioners from the QPS. While this collaboration facilitated access to key stakeholders and enriched understanding of community safety contexts, measures were taken to maintain analytic independence. Data analysis was conducted by academic researchers not involved in day-to-day police operations. The team adopted a reflexive approach, remaining aware of potential positional biases stemming from professional affiliations and seeking to foreground participant voices throughout the interpretation of findings. With the intent of the survey from a QPS perspective being to apply an evidence-based process and ensure the overall project

responded directly to identified community needs. The study received ethics approval from the UTS HREC (ETH24-10240).

Results

The following findings yielded from the survey analysis contain three themes, those being group information; group guidelines, content and moderation; and moderator knowledge and resources.

Results 1: group information

Out of the 209 surveys collected, 64 were removed because of a failure to meet the inclusion criteria, which required respondents to be over 18, based in Queensland, and actively serving as a group administrator/moderator. A further 65 cases were removed because of a significant proportion of missing data, and 15 more were excluded as their community group name did not align with the study objectives. This resulted in a final sample of 65 cases for analysis, representing 61 unique community groups. As seen in Figures 1 and 2, most respondents were female and over 54 years of age. A geographical distribution of survey respondents state-wide can be seen in Figure 3, with Figure 4 showcasing a closer look at the distribution within the Brisbane area.

It is important to note that throughout the charts and tables seen in this report, a small percentage of respondents refrained from answering certain questions; this is included or excluded based on their relevance to the analysis.

Responses to “What is the purpose of the group?” range from focusing on support and community interaction to promoting crime awareness strategies. The data in Figure 5 resonates with these responses, with a sizeable proportion aligning their group with engagement in local issues, education and awareness, and support and assistance.

A particular response states how “The group was an extremely important resource during the north Queensland floods from Cyclone Jasper. I fact checked and shared news to help people during this crisis”, conveying the importance of community groups during emergencies for relaying important, accurate information. This conveys that while the primary purpose of a majority of the community groups is not related to emergency response and preparedness, it is valuable for moderators to be aware of reliable avenues for fact-checking and news reporting in the case of emergencies.

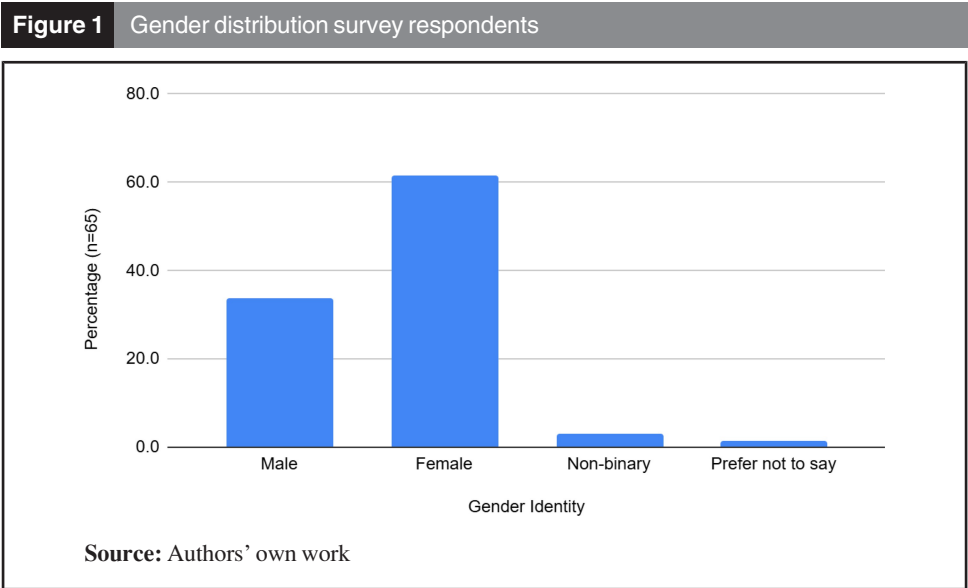
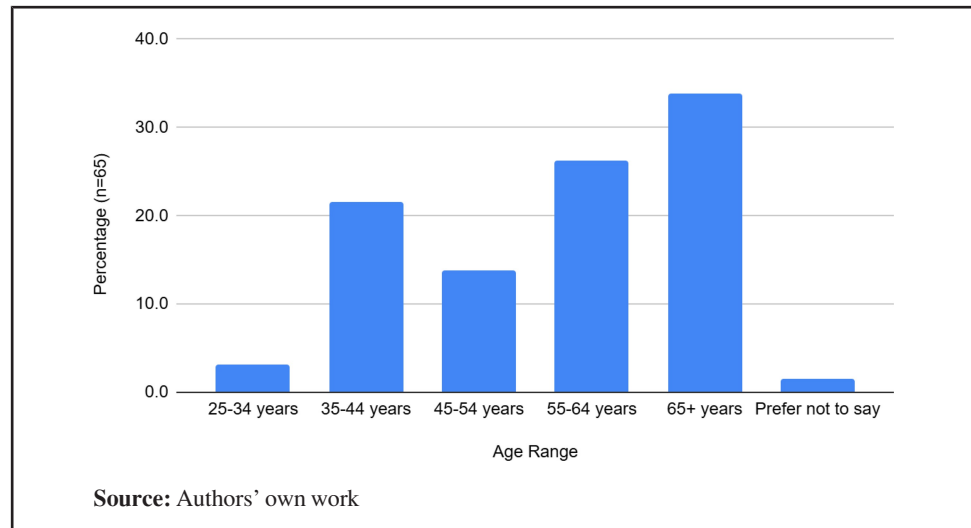


Figure 2 Age distribution of survey respondents



Figures 6 and 7 show the estimated number of group members and the number of moderators managing them. The most frequently reported group size falls between 1,001 and 10,000 members, with the majority of groups being managed by three to ten moderators. However, 6% of respondents refrained from answering this question, which is likely because of their role as a singular or co-moderator with fewer than three moderators.

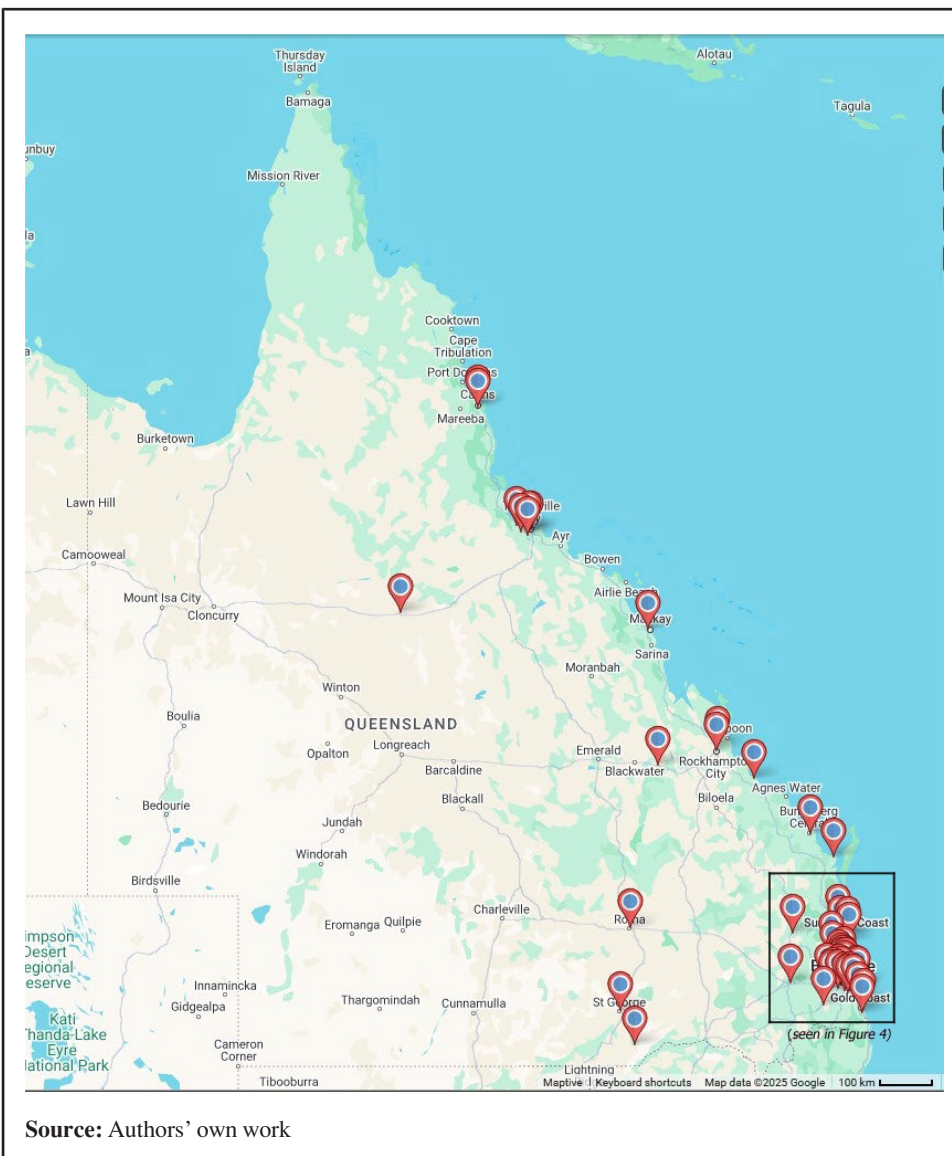
Figures 8 and 9 depict the average number of posts published per week in a group and the time moderators will typically spend managing them. The most frequent response is that groups generate 21–50 posts per week, which aligns with the most reported administrative workload of 6–10h per week. As shown in Figure 10, many respondents have been moderators for their respective groups for over three years, demonstrating a long-term commitment to their roles in spite of their voluntary nature. To note, a small portion of respondents identify as solo moderators and spend only 2–3h a week moderating, which was not available as an option on the multiple-choice survey question. This highlights how most moderators of community groups do so in their spare time rather than as a full-time commitment, which limits their ability to effectively moderate online activity.

The motivations for becoming a group moderator can be summarised into the following three themes: crime awareness, community support and group management. These themes have considerable overlap, as those motivated to become moderators for the purpose of improving community support do so for reasons related to safety and security. Similarly, those that fall under the group management theme are either motivated by inadequate existing moderation or do so as a part of their volunteering for local NHW programs, which are directly linked to crime awareness strategies.

Crime awareness. A key motivation for becoming a group moderator was to promote crime awareness within the community. One respondent explained that they: “Set the group up to make the community aware of the juvenile thugs committing burglary, vehicle break ins, and vandalism”. Crime awareness emerges as a common driving force among respondents, intertwined with their motivations to foster community outreach and improve group management and administration, as reflected in another moderator’s statement about their desire to “[. . .] help support our community in many ways and also advising about scams”.

These insights reflect how crime awareness is not simply a concern about specific incidents but often serves as a gateway into broader forms of civic vigilance. Moderators act as

Figure 3 Geographical distribution of respondents (state-wide)



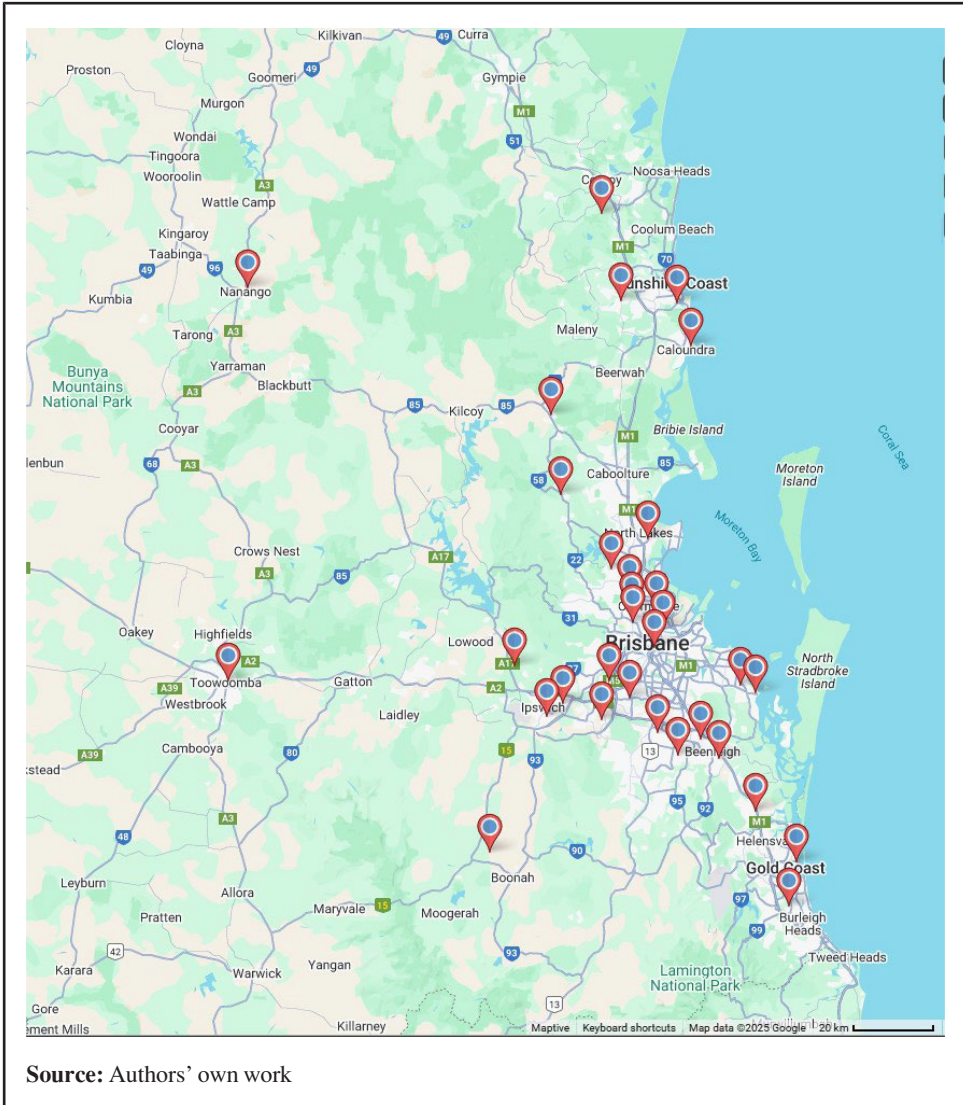
Source: Authors' own work

informal guardians of local order, stepping into spaces where residents feel underserved or vulnerable.

Community support. The strongest motivation appears to be related to fostering community outreach and supporting neighbours. One respondent explained how they aim to “Support our local region by compiling and sharing relevant information in one place”. This perspective overlaps with motivations related to crime awareness strategies, as some respondents share that they wish “To live in a safe and informed neighbourhood that looks out for each other supported by Qld Police”, which motivates them to contribute to community groups.

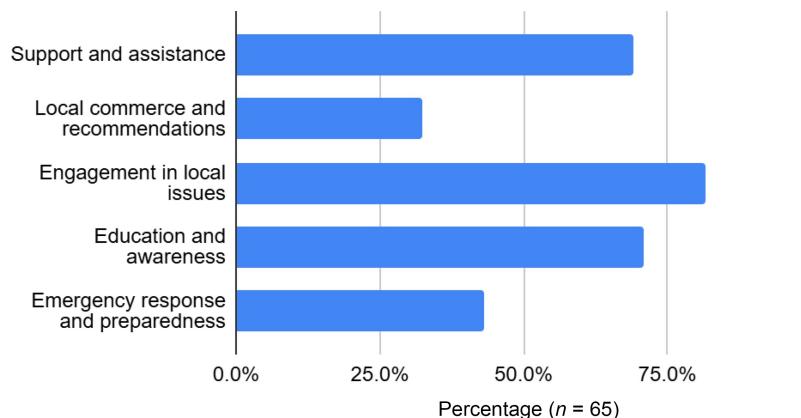
These responses illustrate the importance of a strong social network in maintaining safety and security within communities, especially through sharing accurate and reliable neighbourhood crime information. As one respondent stated: “An informed community is a resilient community” and emphasises their goal as moderator to “provide the community

Figure 4 Geographical distribution of respondents (Brisbane)

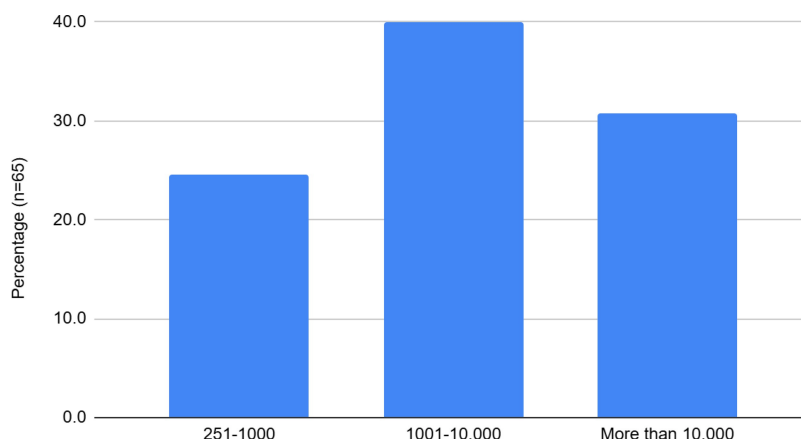


with timely accurate information in detail that is relevant, critical and non-invasive to the persons involved with respect to the circumstances and the duties of officers”. This further reinforces the commitment to responsible information sharing and community resilience. The emphasis on community support underscores the social cohesion function of these groups. For many moderators, digital participation is an extension of neighbourhood care and mutual aid, particularly in contexts where physical forms of community engagement may be limited.

Group management. Similar to the previous theme, the motivations here stem from a desire to be beneficial to the community. One respondent shared that: “The previous admin wasn’t able to manage the workload, I’d recently moved to the area and wanted to help the local community”, highlighting their willingness to step in to foster a safer, more connected environment. Other respondents shared similar sentiments, describing how their groups were: “A great resource but people were too scared to post because abuse was rampant” and that “It was embarrassing to see how the page could easily get out of control and I didn’t want people to think that was what our town was like”. These statements emphasise the importance of moderators in maintaining a safe and supportive online space.

Figure 5 Purpose of the group

Source: Authors' own work

Figure 6 Approximate member count

Source: Authors' own work

Additionally, some moderators are motivated by their responsibilities as volunteers for their local NHW groups; as one respondent explained, it is a “Part of Roles and Responsibilities of NHW volunteer role & District volunteer role”. Others shared similar situations, with one explaining that their goal was to “establish a readily accessible site for Neighbourhood Watch and residents at [suburb name] [2]”. These responses demonstrate how community engagement and structured volunteer roles often drive individuals to take on administrative responsibilities. To be clear, no confirmation was done as a part of the survey to identify groups as official NHW Australasia or NHW Queensland groups.

One respondent shared that they “have retired and have the time to do what is necessary. Keeping the community aware of what is happening. Tips on safety etc”. This highlights how valuable becoming a moderator can be, not only for the benefit of the community but also to provide a sense of purpose to retirees. Similarly, another moderator shares that they “wanted to volunteer in our local community. I am not always at home so this is a way that I can contribute to our community”, which highlights how

Figure 7 Number of admins/moderators managing the group

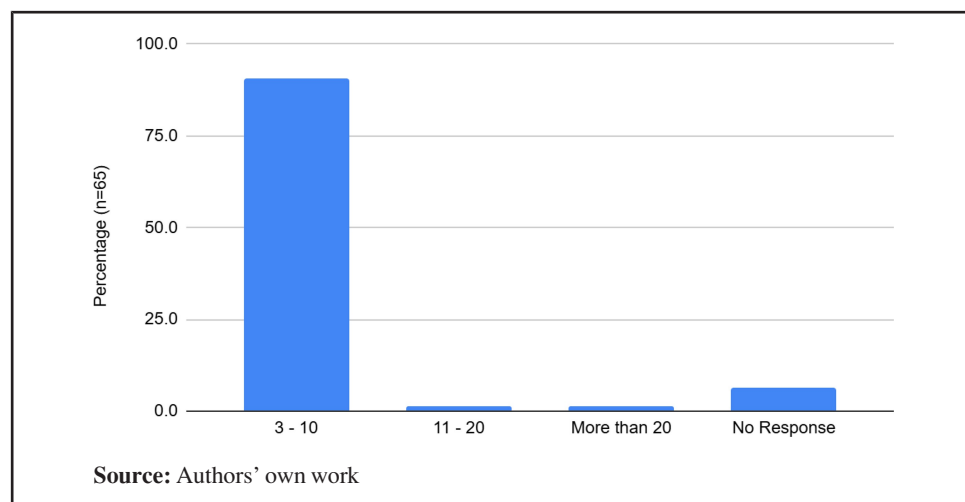
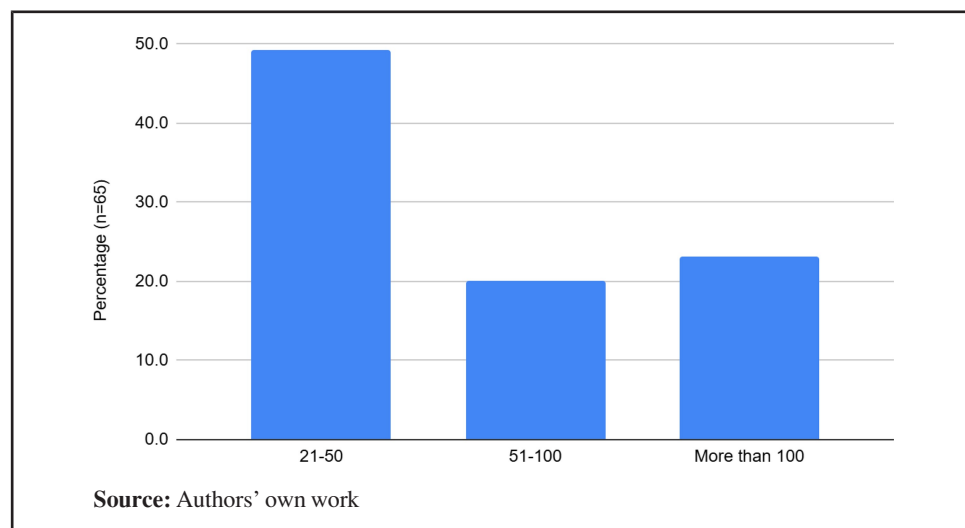


Figure 8 Posts published in a week



these online communities allow an opportunity for community participation to those that may struggle with clashing commitments in their personal lives.

Taken together, these motivations reveal that moderation is driven not just by concern for public safety, but also by a desire for belonging, identity, and recognition. The moderator role becomes a means of asserting positive influence in a digital civic space.

Results 2: group guidelines, content and moderation

As shown in [Figure 11](#), most groups have established guidelines that are communicated to their members. A significant number of rules are closely aligned with the guidelines of their NHW group. Common rules include: Being kind and courteous, no hate speech, no promotions or spam, reporting criminal activity before posting, no abusive language, respecting privacy and no doxxing. Additionally, multiple groups restrict entry by location to preserve a local community focus.

Figure 9 Hours spent managing the group in a week

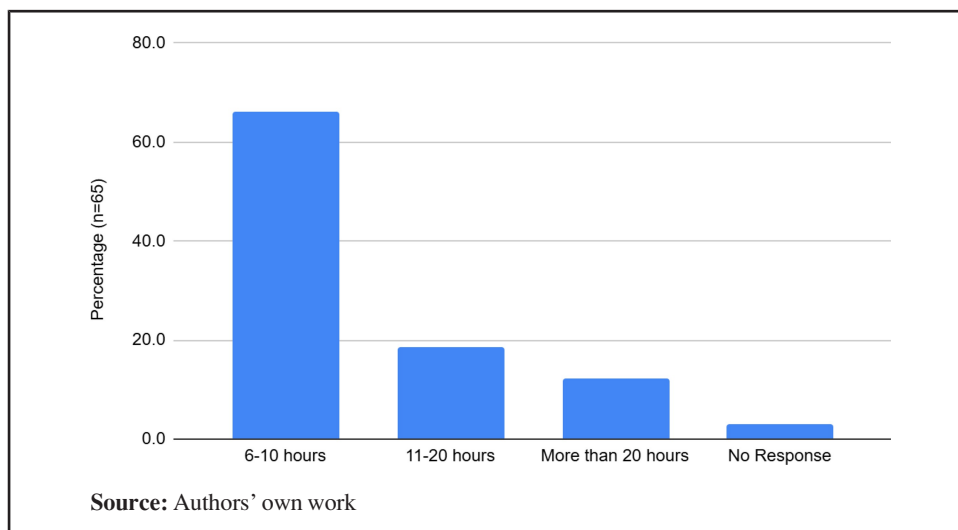
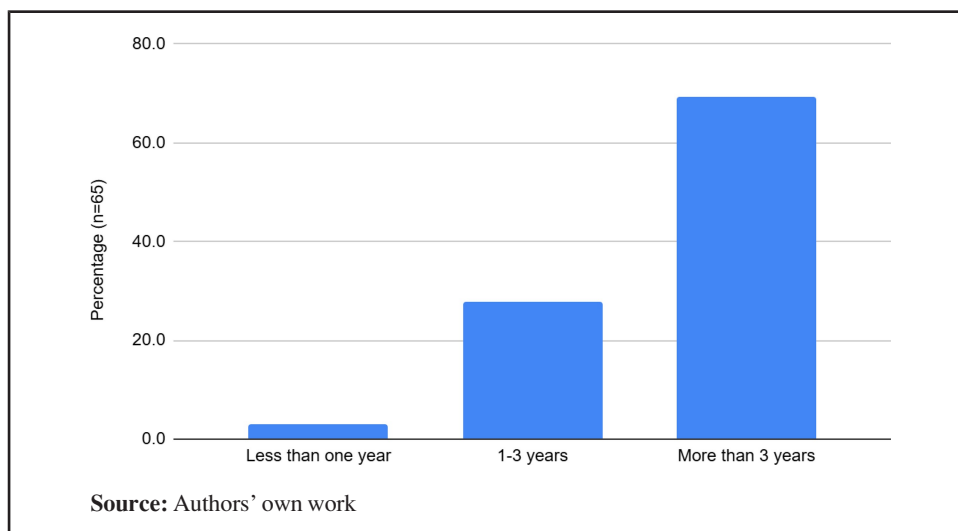


Figure 10 Length of time as admin/moderator



The data in [Figure 12\(a\)](#) and [\(b\)](#), shed an insight into the type of content that is shared in community groups in a typical week. Participants were requested to select from a range of options that aligned with the research goals, and thus this data does not represent the overall group's content. [Figure 12\(a\)](#) shows that common content that appears in the groups is related to CCTV footage, potentially stolen property, suspicious activity, and members seeking help and sharing crime prevention tips. Some respondents shared that their community: "Have direct contact with QPS & DCPC" and that "QPS are already members of the group and they do a great job of posting/commenting/replying to members". However, another respondent explained that they "Share official information from local & state Police FB pages & associated organisations" in their group.

Others share that their community group is sometimes comprised of shared experiences by victims of crimes and road incidents; however, they require: "Accurate and up to the minute details of stolen vehicles confirmed by QPS. Our source is currently from vehicle owners", to be able to better help their community. Additionally, confusion about the reporting process

Figure 11 Group guidelines have been established and communicated to all members

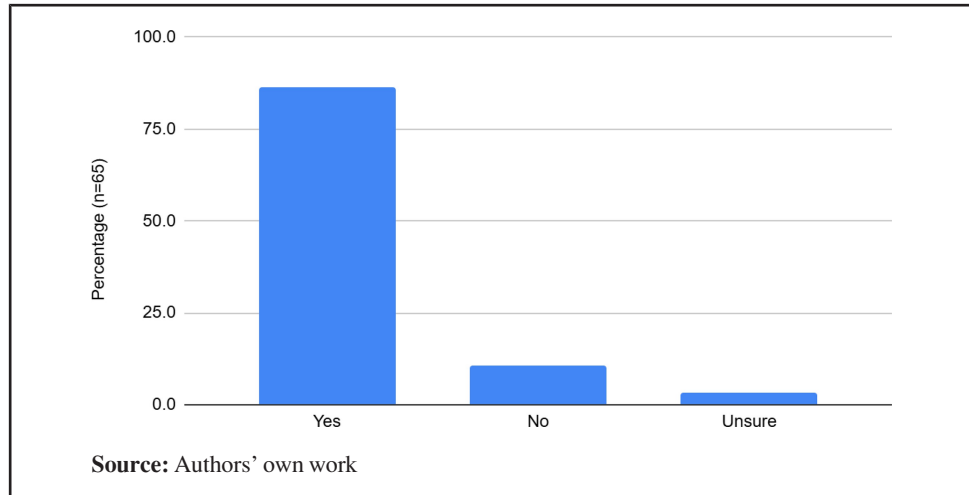
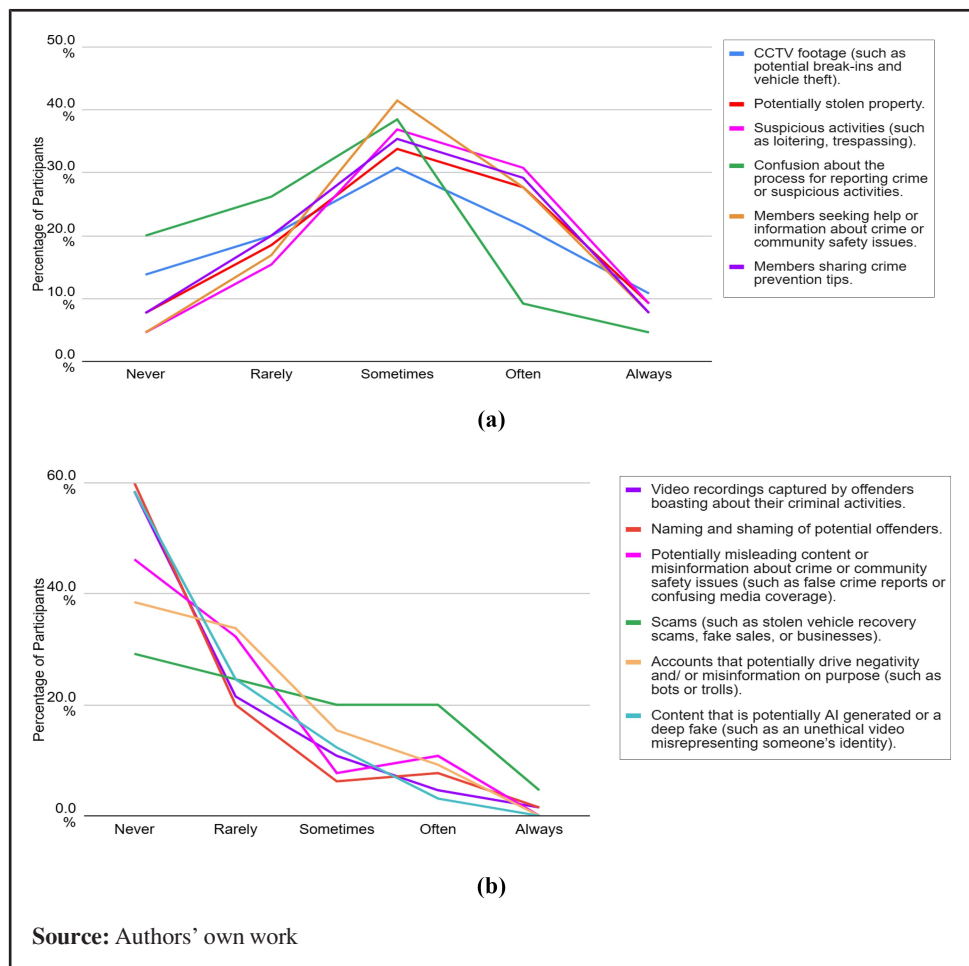


Figure 12 Group content in a typical week



for crime and suspicious activities appears often within the community groups, which suggests a lack of knowledge and an information gap related to crime reporting.

Comparatively, the data seen in [Figure 12\(b\)](#) is positively skewed, which illustrates that these content types are rarely seen within the community groups surveyed. Notably, over 50% of respondents reported that “name and shame” content, offender-generated content, misinformation and AI-generated content are never posted within community groups, which is indicative of the moderation efforts of moderators in maintaining a secure and reliable environment for the community. Alternatively, this could demonstrate a lack of ability in identifying such posts or an unwillingness to disclose such information. One respondent highlighted that: “All posts require admin approval, so this prevents scams being posted” aligning with the data.

These findings suggest that while problematic content is reportedly rare, its perceived presence is shaped by moderators’ confidence in recognising and classifying such material. The relatively low self-reported frequency may reflect successful prevention or, conversely, unawareness or underreporting because of digital literacy gaps.

While [Figures 13](#) and [14](#) show that harmful or problematic content is rarely blocked, deleted or reported, there are still a significant number of respondents that express that problematic content does need to be blocked on occasion. A respondent explained how their group has “post approvals set, so rarely do we encounter this issue”, others noted that they “usually just block or ban. Rarely have we reported to Facebook”. This dynamic highlights moderators’ preference for localised, self-managed responses (such as blocking or banning), rather than escalation to platform-level or formal authorities. This reflects a pragmatic, community-first approach to conflict resolution. Together with the information from [Figure 12\(a\)](#) and [\(b\)](#), this is indicative of the significant role of moderators as information gatekeepers to filter content before it is shared to the main group forum.

As seen in [Figure 15](#), of the posts that are reported, they are most reported to Facebook, with the second most frequent being the QPS. Some respondents explain that they will report it to all services available to them, including local police contacts and other group moderators.

Results 3: moderator knowledge and resources

As seen in [Figure 16](#), the majority of the respondents agree on the importance of moderators being able to identify content and discussions that can impact crime and community safety

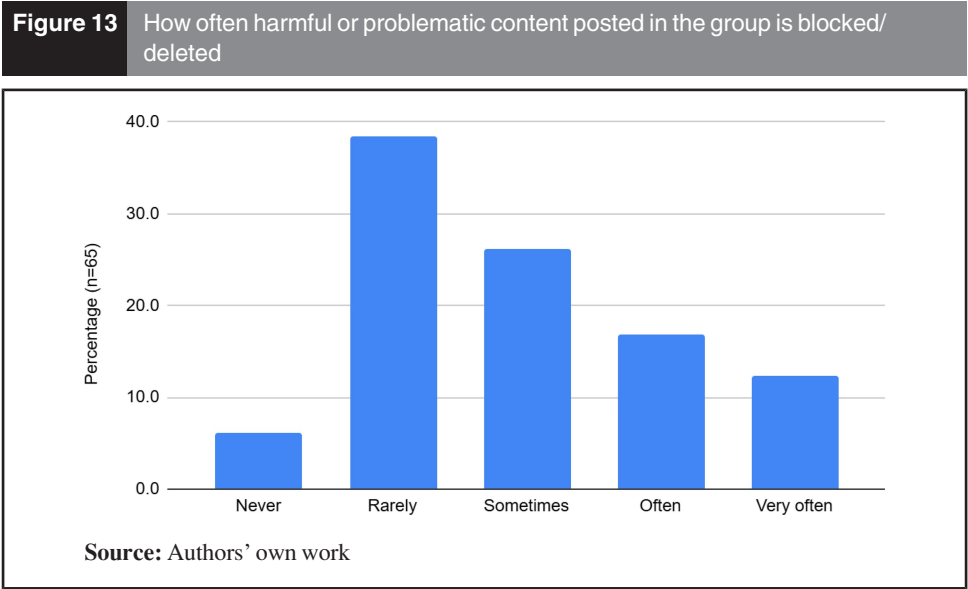


Figure 14 How often harmful or problematic content posted in the group is reported

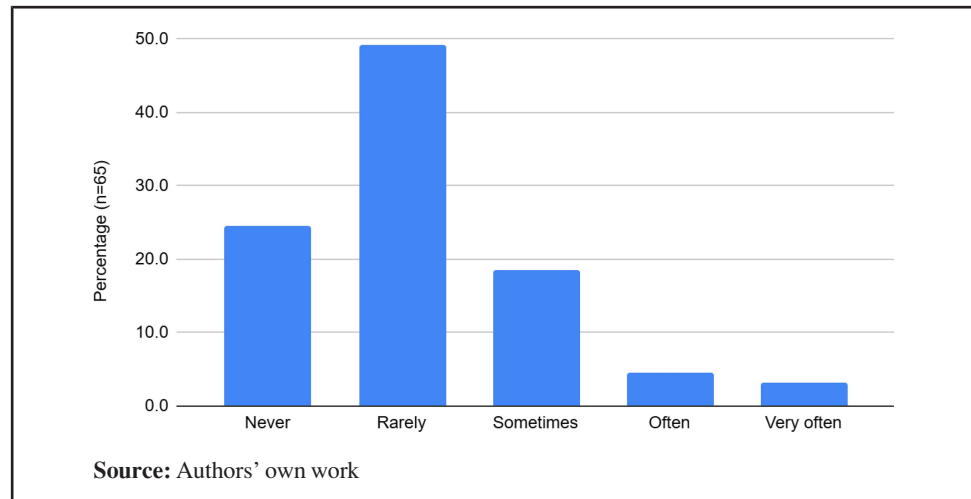
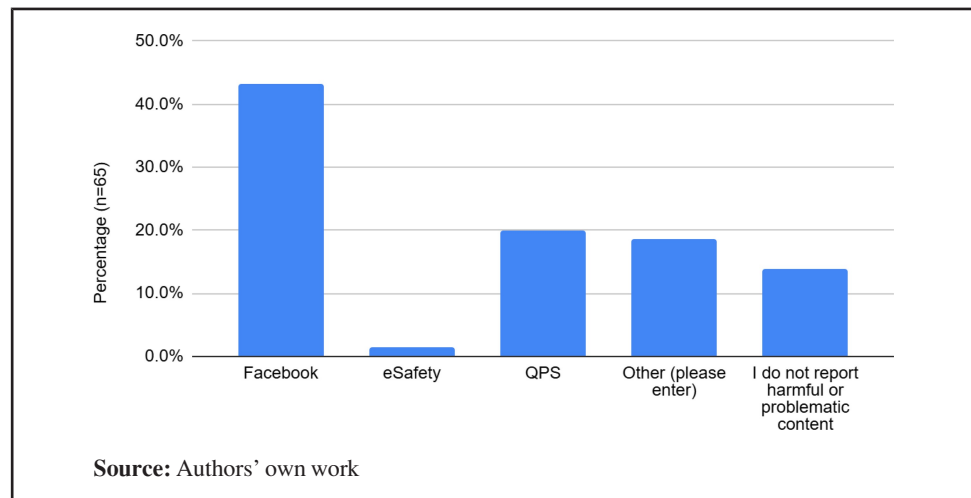


Figure 15 To whom is potentially harmful or problematic content reported to?

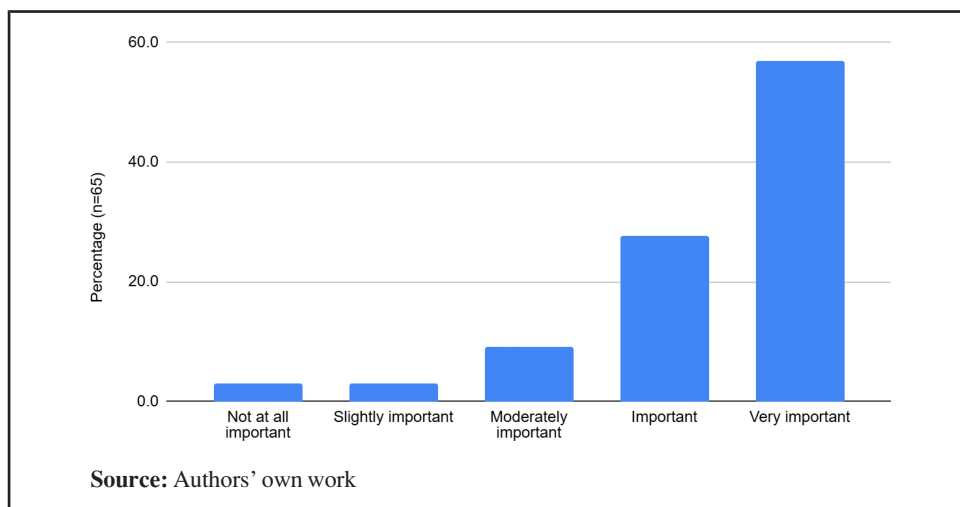


both online and offline. Furthermore, in [Figure 17\(a\)](#) and [\(b\)](#), it can be seen that a majority of respondents are highly confident in all their abilities regarding various administrative and online moderating skills. Of all the options available, respondents show that they are most unsure about their ability to identify and de-escalate conflict, apply digital literacy skills, and develop membership/participation questions.

What such findings reveal is that confidence levels vary most in areas such as digital literacy and de-escalation, skills that are critical in managing misinformation and conflict. This suggests targeted support in these areas could significantly enhance overall group resilience and safety.

Additional support. The most requested support when questioned on the setup and general running of the group was about legal issues related to running a Facebook community group, developing and applying group rules and membership questions, having direct contact with the QPS for assistance managing crime, and identifying and managing content about crime and community safety (see [Table 1](#)).

Figure 16 Importance for group admins/moderators to identify and manage crime and community safety-related content



Similarly, when asked to rank the topics they would like support with to improve their knowledge and confidence in identifying content related to crime and community safety, the most requested was for assistance with CCTV footage, potentially stolen property, reporting crime or suspicious activities, and members seeking help or information about crime and community safety issues (see [Table 2](#)).

This supports the themes that have emerged from the survey thus far, as there appears to be a significant gap in valid and reliable sources of information available for community groups to refer to, or a lack of awareness of what does exist. Community members would benefit from easy access to reliable QPS-endorsed resources that clear up confusion.

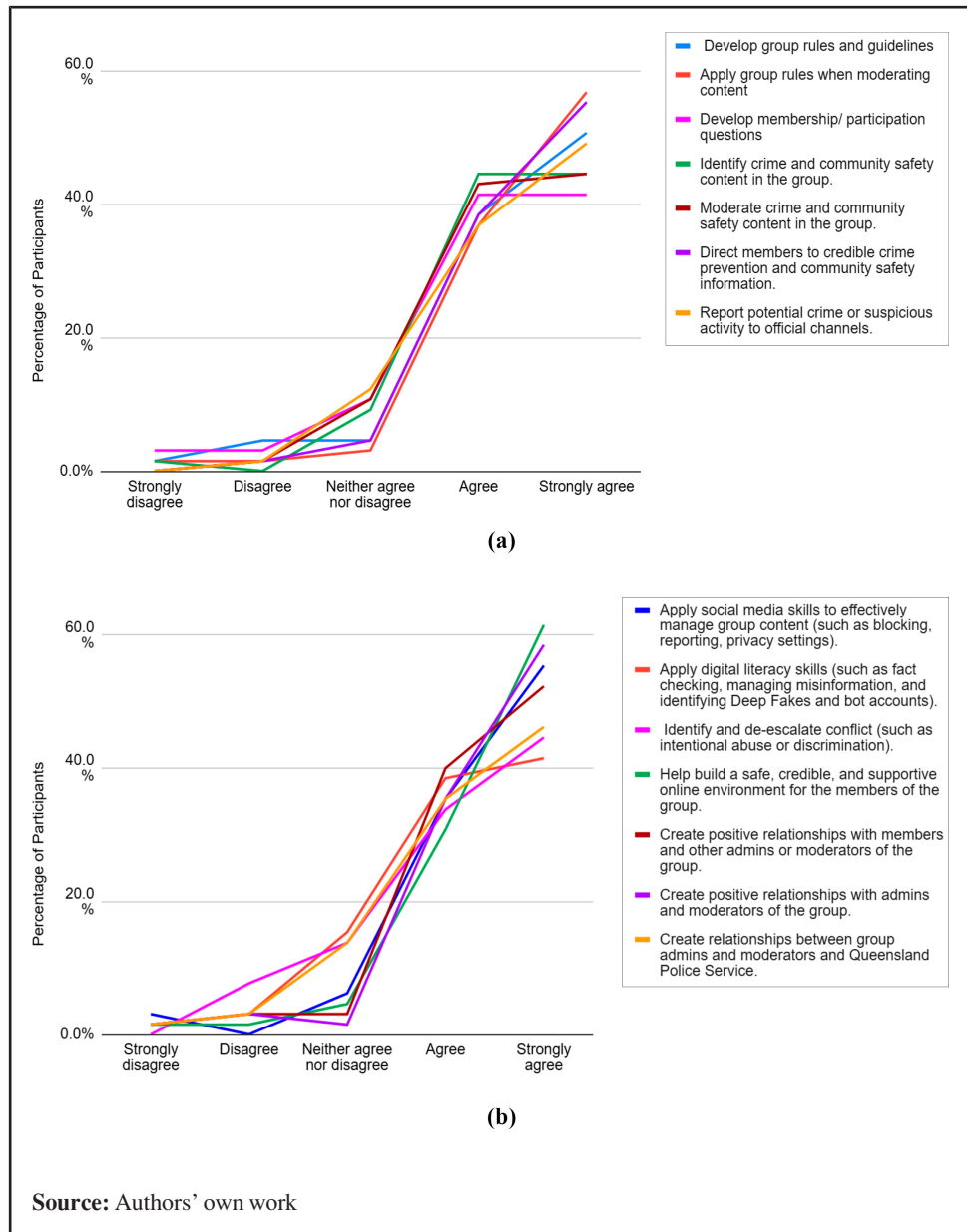
As shown in [Figure 18](#), the most available support resources to share in response to crime or suspicious activity posted in the groups are links to existing QPS resources, information on how and where to report suspicious or potentially criminal activities, and information related to scams. Furthermore, [Figure 19](#) shows that the most preferred learning resource to receive would be direct contact with QPS officials, which supports the theme of respondents wanting clearer and more effective access to QPS-endorsed information and communications to improve information security. One moderator noted that they would benefit from: "A 5 min explanation of what we can and cannot publish when it comes to CCTV or crime reports - in any format".

[Figure 20](#) shows that respondents are highly receptive to receiving help and resources from QPS to help build a safer community environment.

Challenges faced by moderators. The challenges that moderators face have revealed four overarching themes that impact them: Community distrust and Scammers, lack of reliable information, and a heavy workload.

Community distrust and scammers. Respondents shared that some members of their community question the reliability of the information provided by their groups. One explained that: "Because of the ages of our seniors it is sometimes difficult to get them to believe in genuine safety topics that might be presented or help them understand how to recognize scams this maybe because of age, stubbornness or a steadfast belief they know better, it seems even with so much information available some senior people don't access safety information also regarding them as not true", this highlights the challenge of effectively communicating reliable information to members of the community that struggle with digital literacy.

Figure 17 Confidence in abilities



Another shared that there is a “reluctance by residents in our area to report to Police or accept advice”, which suggests a lack of trust in local law enforcement, and an unrest within the community. This is supported by another respond who stated that they struggle with: “Getting a commitment from the community to report crime to Police”, emphasising the distrust in local law enforcement. Additionally, this suggests a low community expectation for efficient results when reporting issues to authorities, and dissatisfaction with the reporting process.

Furthermore, moderators face pushback when enforcing group rules. One moderator notes frequent criticism stating, “I always get lots of backlash how I’m stopping people’s ability of free speech”. Another acknowledged that: “Well meaning people wanting to keep the community safe but going about it the wrong way, including

Table 1 Priority of topics for additional support related to the setup and general running of the group

Rank	Item	%
1	Legal issues related to running a Facebook neighbourhood community group	15.3
2	Developing and applying group rules and guidelines	10.3
3	Developing membership/participation questions	9.4
4	Having direct contact with Queensland police service for assistance managing crime and community safety-related content	9.1
5	Identifying and managing content about crime and community safety	8.6
6	Accessing credible support information to respond to crime and community safety concerns	8.5
7	Applying safety features (such as blocking and reporting of potentially harmful or illegal content)	7.7
8	Applying digital literacy skills (such as fact-checking, managing misinformation and identifying deep fakes)	7.3
9	Applying social media privacy features	6.6
10	Identifying and de-escalating conflict (such as intentional abuse or discrimination)	4.6
11	Identifying and managing bots and/or trolls	4.4
12	Building supporting relationships with members of the group	3.0
13	Building effective relationships with other administrators/moderators of the group	2.1
14	None of the above	2.0
15	Something else (please enter)	1.1

Source(s): Authors' own work

Table 2 Priority of topics for additional support related to crime and community safety content

Rank	Item	%
1	CCTV footage (such as potential break-ins and vehicle theft)	15.8
2	Potentially stolen property	12.8
3	Suspicious activities (such as loitering and trespassing)	11.9
4	Reporting crime or suspicious activities	8.6
5	Members seeking help or information about crime or community safety issues	8.5
6	Naming and shaming of potential offenders	8.0
7	Video recordings captured by offenders boasting about their criminal activities	7.6
8	Potentially misleading content or misinformation about crime or community safety issues (such as false crime reports or confusing media coverage)	7.2
9	Scams (such as stolen vehicle recovery scams, fake sales or businesses, romance scams)	6.0
10	Accounts that potentially drive negativity and/or misinformation on purpose (such as bots or trolls)	5.1
11	Content that is potentially AI-generated or a deep fake (such as an unethical video misrepresenting someone's identity)	4.7
12	None of the above	2.6
13	Something else (please enter)	1.1

Source(s): Authors' own work

naming and shaming, providing dangerous solutions and borderline vigilantism". This highlights the challenge of maintaining positivity, trust, security and cohesion within the group while preventing harmful behaviours. Similarly, respondents express ongoing struggles with identifying and preventing scammers and fake accounts from joining their groups. This breakdown in trust not only affects the spread of reliable information but also limits moderators' ability to function as credible intermediaries. The challenge is as much relational as it is informational, highlighting the social complexity of digital public safety work.

Lack of reliable information. Misinformation and a lack of strict guidelines for how to run a group is a struggle that multiple respondents relate to. One moderator expressed concerns over: "Not being familiar with what is acceptable & unacceptable for members to post", and another shared that: "It is totally frustrating not having close contact with a DCPC [3] for advice and Stats", which indicates how beneficial it would be for local law enforcement to have a higher involvement and improved communications with community groups, potentially through a reinvigoration of the NHW Queensland framework promoting improved use of digital spaces.

Figure 18 Support resources possibly shared in response to crime or suspicious activity posted in group

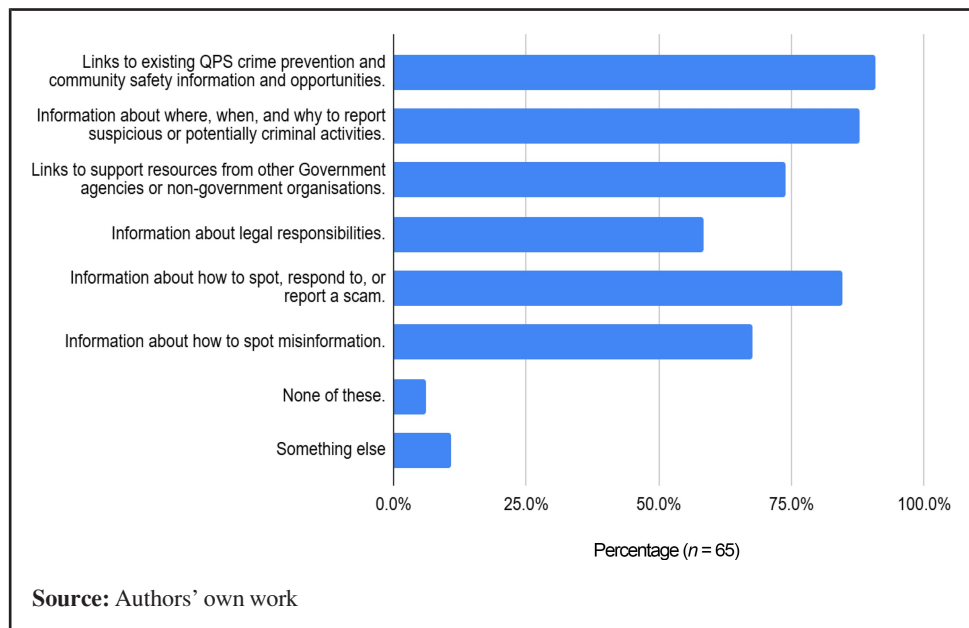
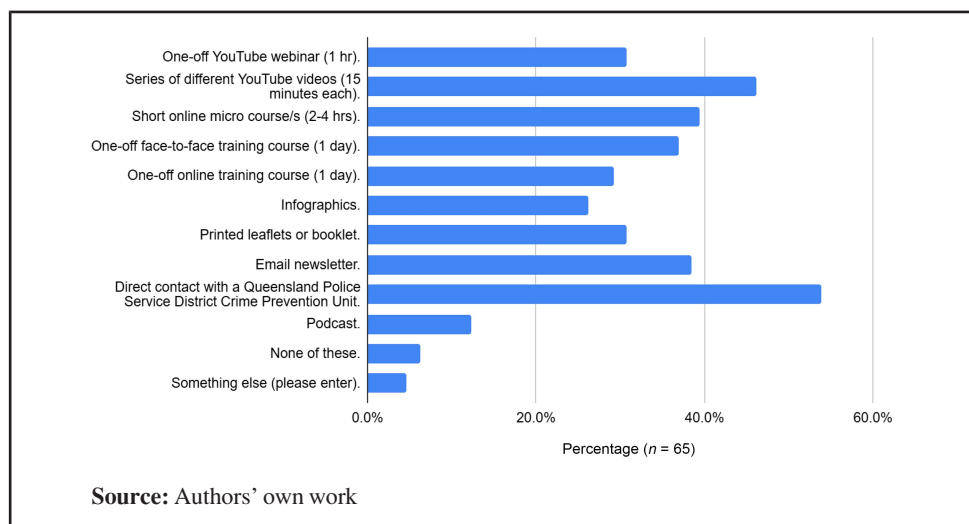


Figure 19 Preferred learning resource to receive

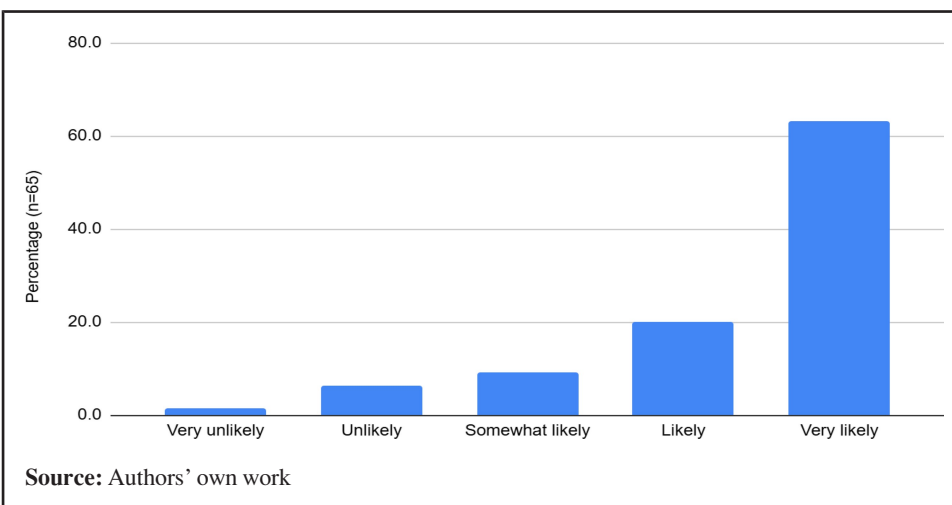


Conversely, this highlights that moderators may be unable to or are unfamiliar with the processes necessary to access information that has been provided. Multiple survey participants also indicate that clearer guidelines from policing authorities would help with streamlining support within the community. Their perspectives are reflected in the following quotes:

But with clear guidelines in place it makes it easier if people start abusing each other, or share content that is misleading or not factually correct

I would like a FB Admin 'handbook' that clearly identifies the rules for administering a Neighbourhood watch group. That would help me to be more confident in dealing with the group.

Figure 20 Likelihood to use resource developed by the QPS to support you to facilitate a safe, credible and supportive neighbourhood community Facebook group

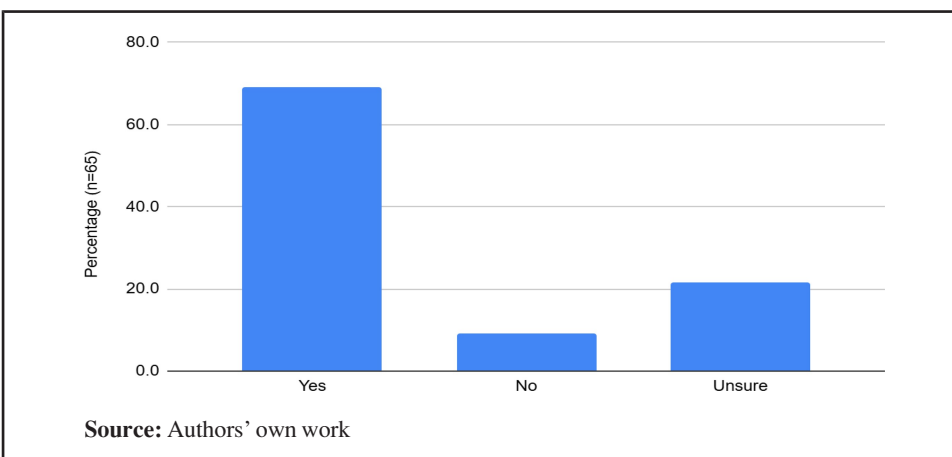


Helping members under the limitations we have as admins in relation to reporting crime. Members don't understand why we can't identify suspects etc.

These responses define the need for structured guidance and stronger connections with law enforcement to enhance moderation efforts and maintain community trust and satisfaction. These views are reflected within [Figure 21](#), as a majority of respondents agree that they would like a QPS media team member to join their group to facilitate smooth communication.

Heavy workload. Because of personal responsibilities and commitments, many respondents find it difficult to balance moderating their respective groups and express difficulties in recruiting volunteers to assist. One moderator highlighted the difficulty of: "Finding other suitable Admins that I would have confidence in allowing them to post", reflecting concerns about maintaining group integrity and trust. Another noted that: "Keeping interest and getting volunteers to assist when crime is low" can be particularly challenging, suggesting that engagement fluctuates based on perceived community safety. These insights emphasise the need for sustainable support systems to ensure the longevity and long-term effectiveness of community group moderation.

Figure 21 Would you like the QPS media to join and interact in this group?



The sustainability of volunteer moderation hinges not just on workload but on trust, digital capacity and community recognition. Without structural support, the risk of community backlash and moderator burnout could arguably weaken this valuable layer of informal governance.

Additional feedback captured in the data. Survey feedback included a mix of recommendations and positive remarks. A key suggestion was for general email accounts to be set up for each NHW group to streamline communications. Additionally, concerns were raised about the effectiveness of online safety advice for seniors, given their limited digital literacy. As one respondent noted, “Advice for Senior people on internet safety is out there but because it is actually online where in fact a lot of misinformation is they don’t seem to regard it as safe information either”.

While another participant emphasised the need for more timely police updates, stating that they “Would like local Police community announcements on crime and persons of interest posted on their webpage to be more frequent”, which supports all previous findings.

Many respondents also expressed appreciation for law enforcement engagement, sharing that they: “Really appreciate the support and engagement with our SW Qld police teams. Happy to address any issues when notified and also to share positive pathways with the communities”. Others acknowledged the broader effort to maintain safe online space, stating, “Thank you for recognising the importance of keeping online community groups safe spaces for us all and taking steps to help those of us passionate about connecting genuine community-minded members effectively”.

These findings collectively illustrate how moderators are navigating complex intersections between care, control and credibility. The work of moderators reveals itself to be neither fully private nor fully public but rather situated in a hybrid governance space that bridges community service and digital stewardship.

Discussion and conclusion

This article has presented the findings from surveys conducted with online community and crime groups across Queensland, providing valuable insights into the roles and responsibilities of group moderators. These individuals play a crucial part in fostering community support, raising crime prevention awareness, championing protective behaviours and maintaining safe and credible online spaces. The data shows that, in spite of the voluntary nature of these roles, most moderators demonstrate long-term dedication and a strong commitment to civic participation and community well-being.

Moderators are driven by overlapping motivations centred on crime awareness, community support, and the effective management of online groups. Many began moderating to fill perceived gaps in safety communication or to foster a more connected local community. Their actions frequently involve curating information, supporting distressed members, and de-escalating online tensions – all functions traditionally seen as part of formal policing or social service roles. This highlights the emergence of a new kind of digital civic leadership and the need for law enforcement to recognise the importance of two-way engagement in online spaces. Whilst most groups have well-established rules aligned with values such as kindness, courtesy and safety, with clear prohibitions against misinformation, hate speech, or naming and shaming. These groups typically use post-approval mechanisms, ensuring that moderators act as critical information gatekeepers. This proactive approach limits harmful content, but it also places considerable pressure on moderators to make rapid and accurate decisions – often without formal training or guidance.

In spite of these efforts, moderators frequently encounter challenges related to the spread of misinformation, the posting of unverified crime-related content, and difficulties in managing controversial or emotionally charged discussions. Some also face backlash from

members who misunderstand or contest group rules, suggesting a need for clearer community expectations and external validation of moderation practices.

Key challenges

The survey results identify three key challenges. First, *Misinformation and Lack of Authoritative Resources*: Moderators report difficulties in identifying false or misleading content, particularly around crime and safety topics. Many feel underprepared to assess or respond to such content without the support of reliable and accessible guidelines. This gap increases the risk of misinformation spreading and places undue burden on individual moderators to make legal or ethical judgements. Second, *Community Distrust and Digital Literacy*: A recurring issue is the lack of trust in both group moderators and official police information. This is compounded by digital literacy barriers, especially among older community members. As a result, even when accurate safety information is available, it may be met with scepticism or ignored. Respondents highlighted the need for stronger collaboration with police to reinforce the credibility of the information shared in their groups. Finally, third, *Workload and Volunteer Sustainability*: The time commitment required for effective moderation is substantial, with many moderators managing groups with thousands of members. Volunteer burnout is a real risk, exacerbated by difficulties in recruiting co-moderators and sustaining engagement during periods of low activity. These findings point to the need for sustainable support structures to maintain moderator effectiveness and well-being over time.

Furthermore, such findings build upon earlier research into online community moderation (e.g. Gillespie, 2018; Walby and Joshua, 2021) by demonstrating how Facebook moderators are assuming hybrid roles that blend digital caretaking with local public safety facilitation. While the pressures and responsibilities outlined by respondents align with concerns raised in broader literature on civic digital labour and informal policing (Seering, 2020; Crosset and Dupont, 2022). However, this study contributes new insights by foregrounding the Queensland context and revealing moderators' appetite for formalised collaboration with policing agencies. Future research, arguably, should examine the longitudinal impact of co-designed moderator–police partnerships, explore platform differences in moderation models, and incorporate the perspectives of group members to complement the moderator-centred focus.

Opportunities for collaboration and capacity building

A strong theme emerging from the survey is the desire for closer and more structured collaboration with law enforcement representatives. Respondents expressed interest in direct lines of communication with local police personnel, access to verified resources on crime reporting and community safety, and co-developed training modules on moderation, legal risks, and online safety. The most requested supports include guidance on legal issues, tools for dealing with potentially criminal content, and advice on applying privacy and digital literacy strategies.

The data also reveal that moderators want clarity on how to handle content such as CCTV footage, stolen property posts, and scam warnings – areas where misinformation and community anxiety can be particularly high. Providing streamlined, easy-to-use QPS-endorsed resources would significantly strengthen the confidence and capacity of group moderators.

Building digital resilience through partnerships

The findings suggest that Facebook-based community groups are vital platforms for digital community engagement and grassroots public safety. However, without appropriate

support, the challenges of misinformation, digital distrust, and overwork threaten to undermine their impact. Moderators are acting as informal digital first responders and community communicators, but they are often operating without the tools, authority, or backing needed to do this work effectively. Ongoing collaboration between law enforcement agencies and community moderators – grounded in co-design and mutual trust – will be essential to addressing these challenges. Structured partnerships can enhance the legitimacy of group moderators, improve the quality of online information, and build digital resilience at the neighbourhood level. The next phases of the QPS DICE Prevention pilot project present a novel and important opportunity to trial and refine these approaches, ensuring that Queensland's digital communities are safe, positive, informed, and well-supported.

In conclusion, the survey findings affirm that online neighbourhood groups are a valuable asset in fostering community safety and cohesion. However, to fully harness their potential, ongoing support, structured collaboration with law enforcement, and capacity-building initiatives are essential. Empowering moderators through tailored resources and stronger partnerships will ensure these digital spaces remain safe, trusted and resilient cornerstones of the community safety landscape.

Acknowledgment

The authors wish to acknowledge the support and assistance from the Queensland Police Service in undertaking this research. The views expressed in this publication are not necessarily those of the Queensland Police Service, and any errors of omission or commission are the responsibility of the authors.

Notes

- [1.] While the roles of Facebook administrator and moderator are distinct, from herein we refer to both roles as 'moderator' for the purpose of the article.
- [2.] Suburb name removed for anonymity purposes.
- [3.] District Crime Prevention Co-ordinators (DCPC) are the first point of contact for NHWQ volunteers.

References

- Alnemr, N. (2021), "Citizens and intelligent machines: algorithms in deliberative democracy", Doctoral dissertation, University of Canberra.
- Braun, V. and Clarke, V. (2006), "Using thematic analysis in psychology", *Qualitative Research in Psychology*, Vol. 3 No. 2, pp. 77-101.
- Braun, V. and Clarke, V. (2023), "Toward good practice in thematic analysis: avoiding common problems and be (com) ing a knowing researcher", *International Journal of Transgender Health*, Vol. 24 No. 1, pp. 1-6.
- Bloch-Wehba, H. (2021), "Content moderation as surveillance", *Berkeley Technology Law Journal*, Vol. 36 No. 3, pp. 1297-1340.
- Bullock, K., Garland, J. and Coupar, F. (2021), "Police-community engagement and the affordances and constraints of social media", *Policing and Society*, Vol. 31 No. 4, pp. 373-385.
- Crosset, V. and Dupont, B. (2022), "Cognitive assemblages: the entangled nature of algorithmic content moderation", *Big Data & Society*, Vol. 9 No. 2, pp. 1-13.
- Firdaus, N., Jumroni, J., Aziz, A., Sumartono, E. and Purwanti, A. (2024), "The influence of social media, misinformation, and digital communication strategies on public perception and trust", *The Journal of Academic Science*, Vol. 1 No. 3, pp. 131-138.
- Gillespie, T. (2018), *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*, Yale University Press.

- Kiene, C., Jiang, J.A. and Hill, B.M. (2019), "Technological frames and user innovation: exploring technological change in community moderation teams", *Proceedings of the ACM on Human-Computer Interaction*, Vol. 3 No. CSCW, pp. 1-23.
- McKee, A.J. and Lewis, A.L. (2016), "The new community policing: developing a partnership-based theoretical foundation", *Interdisciplinary Journal of Partnership Studies*, Vol. 3 No. 3.
- Marwick, A. and Lewis, R. (2017), *Media Manipulation and Disinformation Online*, Data & Society Research Institute, New York, NY, Vol. 359, pp. 1146-1151.
- Meijer, A.J. and Torenvlied, R. (2016), "Social media and the new organization of government communications: an empirical analysis of twitter usage by the dutch police", *The American Review of Public Administration*, Vol. 46 No. 2, pp. 143-161.
- O'Neill, M. and McCarthy, D.J. (2014), "(Re) negotiating police culture through partnership working: trust, compromise and the 'new'pragmatism", *Criminology & Criminal Justice*, Vol. 14 No. 2, pp. 143-159.
- Pepper, M. and Bullock, K. (2025), "The community supporting the community... that's where its strength is": exploring the contemporary position of neighbourhood watch in the United Kingdom", *Safer Communities*, Vol. 24 No. 2, pp. 116-129.
- Pérez-Escolar, M., Lilleker, D. and Tapia-Frade, A. (2023), "A systematic literature review of the phenomenon of disinformation and misinformation", *Media and Communication*, Vol. 11 No. 2, pp. 76-87.
- Rahman, Z., Mattingly, S.P., Kawadgave, R., Nostikasari, D., Roeglin, N., Casey, C. and Johnson, T. (2019), "Using crowd sourcing to locate and characterize conflicts for vulnerable modes", *Accident Analysis & Prevention*, Vol. 128, pp. 32-39.
- Seering, J. (2020), "Reconsidering self-moderation: the role of research in supporting community-based models for online content moderation", *Proceedings of the ACM on Human-Computer Interaction*, Vol. 4 No. CSCW2, pp. 1-28.
- Shahbazi, M. and Bunker, D. (2024), "Social media trust: fighting misinformation in the time of crisis", *International Journal of Information Management*, Vol. 77, p. 102780.
- Walby, K. and Joshua, C. (2021), "Community crime prevention and crime watch groups as online private policing", *Safer Communities*, Vol. 20 No. 4, pp. 237-250.
- Williams, C.B., Fedorowicz, J., Kavanaugh, A., Mentzer, K., Thatcher, J.B. and Xu, J. (2018), "Leveraging social media to achieve a community policing agenda", *Government Information Quarterly*, Vol. 35 No. 2, pp. 210-222.
- Wukich, C. (2021), "Government social media engagement strategies and public roles", *Public Performance & Management Review*, Vol. 44 No. 1, pp. 187-215.
- Zhang, A.X., Bernstein, M.S., Karger, D.R. and Ackerman, M.S. (2024), "Form-from: a design space of social media systems", *Proceedings of the ACM on Human-Computer Interaction*, Vol. 8 No. CSCW1, pp. 1-47.

Author affiliations

Maram Samiee, Keith Heggart and Philip Birch are all based at Faculty of Design and Society, University of Technology Sydney, Sydney, Australia.

Hazel Wallace is based at Queensland Police Service, Brisbane, Australia.

Corresponding author

Philip Birch can be contacted at: Philip.Birch@uts.edu.au

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com