MODULE SIX:

Digital Safety

MODULE DURATION
3
HOURS

CORE MODULE INFORMATION:

Module Type: Phase 3 – Safeguarding digital communities

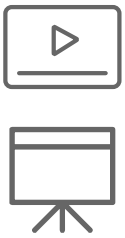
Module Objective: Improve digital safety skills and practices to mitigate the risk of violence in online communities.

Module Dilemma: I want to keep myself and my members safe.

Module Delivery: This module was developed to be delivered physically but may be converted to a remote module with some customisation.

WHAT IS NEEDED IN PREPARATION FOR THIS MODULE:

- Facilitators should review this Module in detail and customise the content to suit their participants, as needed (including adding case studies/examples relevant to your region or country).
- Facilitators should prepare notes for each activity. While this guide provides some discussion points and explanation as a base, further explanation at times will be needed (and participants may ask clarifying questions, so the facilitator should be well prepared).
- Review Content for Training Activities for a list of general training materials and module-specific activities (this link includes sample questions for Menti questions and Kahoot quizzes and information about how to make them). Note: Before the training, be sure to have these activities prepared.



MATERIALS

- Powerpoint slides (linked to sample PPT slides)
- Links to videos and MP4 files downloaded for backup (links are embedded in the above PPT and linked below, per session).
- Module 6 Specific Resources: Dilemma activity papers printed out or written out if the module is taking place in person (Sample Dilemma Activity Printout).

Session 1: Why is Digital Safety Important for Online Communities?

Session Objective: Understand the importance of and need for digital safety in online communities

The trainer will introduce Module 6 objective: to improve digital safety skills and practices to mitigate the risk of violence in online communities.

Session 1 will focus on the importance and need for digital safety in online communities.
Session 2 will focus on the identification and understanding of digital safety risks faced by online communities.
Session 3 will focus on responding to digital safety risks and challenges faced by online communities.

20 MINS

THE DILEMMA – A QUICK RATING AND DISCUSSION

Note for Facilitator: This activity was developed to take place in person. However, it can be altered to include two Menti questions if this is taking place online. If this activity is taking place in person, the trainer should prepare two sets of paper, numbered 1-5 (Sample Dilemma Activity Printout)

THE DILEMMA- A QUICK RATING

The trainers will begin the module by dissecting the module dilemma through a sharing exercise.

The trainers will ask participants to stand up and spread out across the space as they will be moving around the room for this activity.

There will be 5 pieces of paper (numbered 1-5) on the ground horizontally (for an X-Axis). The trainers will ask the following question: ***How safe is the internet for your respective online community?*** (on a scale of 1-5, with 1 being "least safe/secure" and 5 being "most safe/secure". Participants will be asked to stand close to their respective numbers on the ground.

Next, the trainers will set down five pieces of paper (numbered 1-5) on the ground vertically (for a Y-Axis). The trainers will ask participants to rate themselves on a scale of 1-5: ***How confident/equipped are you to utilise digital safety skills and practices?*** , with 1 being 'I'm not very confident or equipped' to 5 being 'I'm very confident and equipped.' Participants will then move vertically and be situated somewhere on this continuum.

Based on the results from this activity, the trainers will get a sense of where the group of community stewards stands on their understanding of digital safety and their perceived ability to safeguard digital communities. The trainers can also ask a couple of participants why they picked these numbers. ***What makes your group safe or unsafe? Why would you say you are equipped/ill-equipped and confident/unconfident to address safety concerns? Would you consider yourself sceptical or optimistic?*** This activity will help us understand how relatable and relevant this dilemma is for the participants.

20 MINS

155

SAFETY AND ONLINE COMMUNITIES – A CASE STUDY

Before the trainers get into the digital safety risks and concerns about online communities, they will first dissect the importance of safety for online communities. As highlighted by the Governance Lab report, the following are some perspectives on why digital safety is important for groups:

- Research shows that safety, or the sense of being able to exist and express oneself without fear of judgement, is a prerequisite for participants in groups.
- Creating this sense of safety demands active management. This applies to social media groups as well.

The trainers will help the participants further dissect this by analysing a case study.

The following case study shows that leaders of online groups face particular challenges in balancing the maintenance of group identity, civil discourse, protection of members' safety, and the need to respond to contemporary political and social events.

Note to Facilitator: The following cases/examples should be altered to best suit your country context and participant profiles. The examples below are samples but should be customised to suit your audience.

THE CHALLENGE OF GLOBAL SOCIAL ISSUES AND RISKS TO MEMBERS – HOW LEADERS RESPOND

In recent years, Facebook Group leaders have found themselves having to take positions on issues entirely unrelated to the subject of the group. For groups in many countries, not just the United States, this issue came to a head in mid-2020 after the police killing of unarmed African American George Floyd.

- The killing caused an outpouring of debate, and sometimes conflict, in Facebook Groups of all kinds. In the United States, leaders of Boss-Moms Facebook and a Billion Vegans, among others, deleted posts protesting the killing on the ground that the posts were irrelevant to the original purpose of the group. The leaders' position outraged many of the groups' members, who also pointed out that the groups had no moderators of colour.
- Facebook published guidance on how to talk about racial justice and urged groups to ensure that their admin and moderator teams were appropriately diverse and inclusive. As new groups, such as White People DOING Something., were formed to fight racism, older groups had to confront this issue. In Canterbury, Edd Withers and his fellow admins expressed solidarity with Black Lives Matter and personally contacted group members who made "White Lives Matter" or "All Lives Matter" posts, explaining why they thought these opinions were "problematic" at this time. The approach won some support but prompted others to leave the group. Many joined a sister group that discouraged political threads and promoted discussion on "positive topics."
- Subtle Asian Traits had always blocked overtly political content, instead favouring humorous memes about hyphenated Asian identity. But in June 2020, a group of 45 moderators held an emergency three-hour global call to discuss how to respond to the Floyd killing and whether it should allow Black Lives Matter content.

The above will be shared with the participants, after which the participants will discuss and answer the following questions.

What was the key issue these groups faced? Can you think of a social issue or particular incident that triggered such a reaction within your own group?
How did the leaders respond to the situation? How would you have responded?
What are some steps leaders can follow to avoid or better handle such situations in the future?

20 MINS

WHY IS DIGITAL SAFETY IMPORTANT FOR ONLINE COMMUNITIES?

Trainers can encourage the participants to add more points to the following list.

1. **Communities – including online communities – seek safety.**
   The various communities we belong to, including our family, friend groups, workplace, etc., need a sense of safety. We find it difficult to belong to and exist in communities that do not feel safe. Safety is not conditional. It is a right. Therefore, it must be available and accessible for all individuals regardless of whether the community is hosted on an online or offline platform.

It is also worth noting that Facebook defines 'community' as follows: "A collection of people in which they receive a sense of belonging, connection, and a feeling of safety.

2. **Online Communities are an alternative to individuals and groups that cannot host offline communities.**
While social media groups are often created due to the benefit of proximity and convenience, many community-based groups host themselves online as it is a safer alternative when compared to offline communities.

   This preference is often influenced by local legislation. For example, in countries where abortions are banned, or homosexuality is criminalised, individuals might feel safer in a digital community than in a physical one.

   As these communities, and the individuals who belong to these communities, are already vulnerable and at risk, it is crucial that these online communities are safe and well-protected.

3. **Online Communities are easier to attack than offline communities.**
Community-based online communities, especially ones such as the above, which discuss issues that attract scrutiny and judgement, can be at risk of attacks and smear campaigns. This could be through an individual or even another group that opposes the views, beliefs, and perspectives of the group.

   While physical attacks might be rare and require coordination and effort, cyber-attacks are, unfortunately, easier to facilitate. Groups on social media platforms, including Facebook, are vulnerable to such attacks – especially when necessary digital safety mechanisms are not in place. This is why it is important for online communities to be aware of and follow the necessary digital safety mechanisms.

   It is important to note that these attacks and retaliation can sometimes even be arbitrary and state-led violence. A good example of this is when three social media admins, including a journalist, were wrongfully arrested in Sri Lanka in May 2022 during the police protests and outrage (Note: trainers should seek to find an example from their context).

4. **Online Communities value privacy and anonymity.**
There are certain online communities, such as ones on Discord, that operate entirely anonymously. There are other online communities, such as ones on Facebook, that remain private. These decisions regarding privacy and anonymity are often made to safeguard the members as well as the overall interest of the online community.

   If the privacy of the members is violated or even threatened, it can greatly affect their trust in the group and might even influence them to leave the group. This is why online communities must prioritise digital safety mechanisms that ensure the privacy and anonymity of the group members.

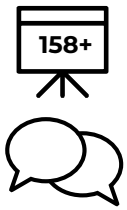
5. **Online Communities help individuals build resilience and confidence.**
When individuals become part of a safe space and trust the members within, they begin to develop a sense of belonging and emotional security. Rather than completely shielding individuals from the adverse effects of society, safe spaces can help individuals build the strength and confidence needed to confront discrimination and isolation in a supportive environment and allow them the opportunity to engage as their authentic selves.

Session 2: What are the Digital Safety Risks and Challenges Faced by Online Communities?

Session Objective: Identify and understand the digital safety risks faced by online communities.

Trainers will introduce Session 2 and its objective to identify and understand the digital safety risks faced by online communities.

10 MINS



DISCUSSION - HOW SAFE IS MY GROUP?

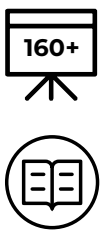
Note for Facilitator: Some of these points may be covered during the dilemma and discussion for Session 1. However, this is a useful point to discuss in more depth and see if participants can identify additional factors that contribute to risks and challenges in online group settings.

The trainers will then ask participants to discuss to better understand the level of safety and level of vulnerability of their groups. It is important for the participants to understand that the risks and challenges faced by social media groups depend on a number of factors. This includes:

- The size of the group
- The objective of the group
- Relevant local legislations
- Privacy settings of the group (including accessibility and visibility)
- The capacity of the admins/moderators.

The trainers will provide examples to elaborate on each of the above. Participants will be encouraged to add to this list. Each participant will be asked to cross-examine the vulnerability of their social media group based on the above factors. This will enable them to identify what areas need more work and what needs to be further strengthened in order to improve the safety and security of their group.

20 MINS



COMMON FORMS OF DIGITAL VIOLENCE - ACTIVITY AND DISCUSSION

Note: It is intended to have a discussion section to be entirely participant-facilitated.

DISCUSSION

Before discussing the specific digital safety issues of social media groups, the trainers will facilitate a more general discussion. Note that there can be an overlap between the general risks faced by digital citizens in the overall platform and specific risks faced by social media groups.

The purpose of this section is to be aware of common risks and challenges faced by all digital citizens, not just those who exist in online groups and communities. The trainers will help participants understand various ways in which digital violence can manifest itself on social media platforms.

Some common forms of online violence include: cyber harassment, cyber stalking, image-based sexual violence, cyber grooming, impersonation, hacking, Doxxing. hate speech and disinformation can also be categorised as online violence.

Participants will be divided into pairs. Each pair must identify one form or manifestation of online violence. Once that is done, each pair must do a short presentation about their topic. The participants will be encouraged to make the presentations as creative as they would like.

The presentation must include the following information:

- ***What is this form of online violence?***
- ***How and why does it happen?***
- ***Who does it affect the most?***
- ***What are its consequences?***
- ***How can people prevent it or be safe from it?***

Once all presentations are completed, the trainers will add more forms of digital violence to the existing list.

30 MINS

161+

WHAT RISKS AND CHALLENGES DO I NEED TO KNOW ABOUT SOCIAL MEDIA GROUPS?

Moving on to risks and challenges faced by admins and groups, it is important to note that researchers have also studied how certain safety factors can undermine incentives or discourage people from participating in online groups. In many cases, the risk of harm or loss of privacy can result in participation being disincentivised.

The following are six specific challenges the trainers will share with the group of admins.

Note: These challenges will be further dissected and elaborated, as time permits, through case studies, video materials, and examples from your local context.

1. **Polarisation**

Michela Del Vicario et al. (2016) studied how Facebook users formed echo chambers or groups isolated from the outside environment, which enforce members' existing behaviours and beliefs. The authors found that highly active users in those groups exhibited more negative emotions and that higher activity in these communities correlated with greater polarisation. Likewise, a study of political discussions on Twitter found that tweets that contained moral and emotional language spread more widely than those that did not, but that this diffusion was largely contained within each ideological group, which the authors argue "may partly explain increasing polarisation between liberals and conservatives (Brady et al. 2017)."

Trainers can show participants this video about how filter bubbles work and how this can lead to polarisation.

2. **Emotional Harm**

The risk of emotional harm affects both a person's decision to join an online group and their behaviour as a member. Some have studied how the emotional distress that some experience within online groups, for instance, due to cyberbullying or online harassment, can negatively impact their experience or even drive them off of social media altogether. Jon-Chao Hong et al. (2015) studied users' motivations to participate in Facebook communities and found that people with higher levels of general anxiety and social anxiety were less likely to continue participating in communities on the platform.

3. **Security**

An individual has to weigh the potential benefit of engaging with their online community against the risk that other community members, or the platform itself, may breach their expectation of privacy. This tradeoff intersects with social traits, such as gender (Fogel and Nehmad 2014). For instance, studies of social media use by young women in urban India have found that while participating in online communities that cross social boundaries can be empowering (Kumar 2014), many women have also experienced breaches of privacy both online (e.g. unsolicited messages from other users) and offline (e.g. stalking and sexual harassment) as a result of using social media (Karusala et al. 2019).

4. **Privacy**

Information collection and sharing may now be so pervasive that privacy concerns are diminishing as a deterrence to participating in online spaces (Al-Ameen et al. 2020). Nusrat Jahan Mim and Syed Ishtiaque Ahmed (2020) studied the conflicts between the image-sharing culture of online social networks and the privacy that is expected within the home in urban Bangladesh. Rather than change their image-sharing practices, many participants actually modified the functions of rooms within homes to accommodate picture-taking, much to the dismay of family members who felt their privacy was being violated (Mim and Ahmed 2020). Another possible explanation is that it is difficult for individuals to make informed decisions about what to share when they are not aware of who is privy to the information they share, as is often the case on social media (Acquisti and Gross 2006).

5. **Lurkers**

Nu Sun et. al (2014) examined a variety of these factors in their study of lurkers or those who are members of online groups but do not actively participate. The authors pointed to a number of reasons for this behaviour, including environmental reasons (i.e. the social environment does not make posting easy), personal reasons (e.g. introversion), relationship reasons (e.g. a lack of intimacy in the group), and also reasons related to security and privacy.

6. **Termination**

While leaders have a great deal of control over their group's activities, it is social media that ultimately decides what is allowed to happen on its platform, including whether groups can exist at all. A fear that the platform could unilaterally remove their community, or require them to start paying, is common among many group leaders.

Session 3: Responding to Digital Safety Risks and Challenges Faced by Online Communities

Session Objective: Identify various tools and practices that can reduce the risk of and mitigate the impact of digital safety challenges faced by online communities

In this session, the trainers will inform the participants about three core digital safety practices that are crucial for social media groups: choosing the right privacy setting, using post approvals correctly, and using membership questions. These tools and practices can help mitigate the risk and impact of digital safety challenges in online communities.

15 MINS

CHOOSING THE RIGHT PRIVACY SETTING

Choosing the right privacy setting for your group can depend on the type of group you have and the sensitivity of what's discussed in your group.

On Facebook, when it comes to visibility and accessibility, there are three key settings for the group:

- Make your group public if you want it to be easily found and if you want posts in the group to be visible to non-members.
- Choose the private but visible in the search setting if you want people to be able to search for and easily discover your group but keep posts visible to members only.
- Make your group private and hidden in search if you want to have the most restrictions over group visibility and want membership to be by invitation only. You may consider this if the group covers a sensitive subject.

The trainers will do a quick tutorial on how to activate and use the above privacy settings via admin tools.

A FEW KEY THINGS TO KEEP IN MIND:

- Private groups with 5,000 or more members can't change their privacy to a public group.
- Groups are limited to one privacy setting change every 28 days.
- Admins who change their group's privacy will have a 24-hour grace period to change it back.
- All members of your group will get a notification that you have changed the group's privacy.
- Invited members are members of a group who can see the group in preview mode.
- When a Page joins a Group, there could be several admins on that Page. All of them can see and interact with posts and members of the group.

- If an admin of the group has added a third-party app, the app has access to posts and comments in the group.
- Third-party apps can't access who wrote posts and comments unless the app has been given permission by the author.

It is highly recommended to notify members of a privacy change in a post or group announcement several weeks beforehand to make sure members feel comfortable with the new privacy setting. If the privacy level of a group changes, all members of the group receive a notification of the change.

15 MINS

USING POST APPROVALS

All posts from group members appear in your group by default unless you turn on the post-approval option. With post approvals, admins can review all pending posts within a group and must approve them before they appear. Experienced admins use post approvals for the following reasons:

- You're on holiday, and you can't moderate posts effectively.
- An inflammatory topic is taking the group away from its intended purpose.
- People post about the same thing over and over again.

Admins agree that whether or not you use post approvals can depend on the nature of the group. A supportive group, where members need to talk immediately and freely, might not want to use post approvals. A small family or friend group may not want to limit member participation in any way. A group that is about one specific topic may use post approvals as a way to prevent members from posting about something off-topic and unrelated.

On the rare occasions when group conflict occurs, admins recommend temporarily turning on post approvals until tensions have eased as a way to manage member conflict.

If you are going to use post approvals, be sure to review pending posts quickly to let members feel heard.

Trainers will hear an audio statement from a group admin who discusses how and why he uses post approvals.

The trainers will demonstrate how to use post approvals on Facebook. Also, if they are willing, trainers will also ask the present admins in the group who are familiar with post approvals to do the demonstration for the other participants.

15 MINS

CREATING MEMBERSHIP QUESTIONS AND GROUP RULES

Note: The trainers will also inform the participants about digital safety mechanisms they would like to know about (e.g. a quick refresher can be shared on reporting procedures of particular platforms). The information below is relevant mostly to Facebook groups but can be customised as needed.

The following link to the Module 6: Session 3 Annex, contains information the trainers will share on digital safety for admins on WhatsApp and Telegram.

Trainers will first show this video where admins discuss why it's important to have membership questions.

When first setting up your group, you might invite close friends or people who share the same interests. As the group grows, the number of people asking to join increases. Experienced admins cite that prioritising and managing new member requests is an important part of making sure that new members are a good fit for the group.

Membership questions help you learn more about people who want to join your group by asking them up to three free-form questions. Only admins and moderators see the answers, which can be reviewed within the member request queue. Since membership in private groups is limited to people who were personally invited to join, membership questions are only available to groups in a public or private setting.

Membership questions can be as simple as "what do you hope to get out of the group?" to more detailed questions like "how do you feel you might be able to contribute to this group as a member?" Experienced admins recommend mentioning the group's rules in one of the questions, asking potential members to agree to read and abide by them once they join the group.

Once your group has grown, and you're receiving numerous membership requests, you may find using membership request filters helpful. With membership request filtering, you can select and sort predefined filters to apply to pending requests. These can be accepted or declined individually or in bulk. The list of pre-set filters includes location or gender, whether a person was invited by a current member, whether they are in other groups you manage, and more. This information is made available based on what that person shares on their public profile. Member request filtering is one of the time-saving screening techniques recommended by admins.

The trainers will demonstrate how to use 'member request filtering'. Once that is done, all participants will be given a few minutes to either come up with the membership questions for their group or reevaluate and reframe the existing membership questions of the group based on the learnings from the training. Participants, especially those with similar groups and objectives, will be encouraged to support each other.

5 MINS

SHORT DISCUSSION

Trainers can ask participants the following discussion questions: ***Is there anything else that you do to respond to risks or challenges faced in your online community? Can you share an example of a time you responded to safety risks or challenges in your group? Is there anything you would change about the approach you took to address these concerns?***

10 MINS

MODULE 6 CLOSING ACTIVITIES

Next, the Trainer will facilitate a Q&A session.

The module will end with a Pop Quiz on Kahoot (this is an optional activity; however, this is a great way to energise the participants at the end of the module).

Review Content for Training Activities for quiz content and instructions on how to make a Kahoot Quiz.

The PPT slide can be linked to the Kahoot quiz for ease of access and presentability. Trainers can encourage participation by handing out chocolates to the pop quiz winners.

Finally, the Facilitator will ask the training participants to complete a short feedback form. This can be optional and created according to the organiser and facilitator's needs, therefore a sample is not shared.

The content of this module was adopted and inspired by the following resources:

- Navigating disinformation: UN Women
- The power of virtual communities: Governance Lab
- Understanding Privacy Settings: Facebook Community