

Digital Community Stewards:

BUILDING SOCIAL COHESION IN ONLINE COMMUNITIES



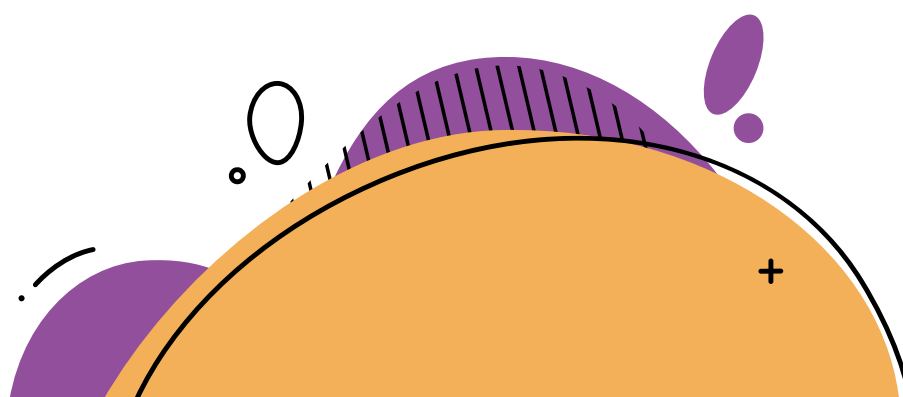


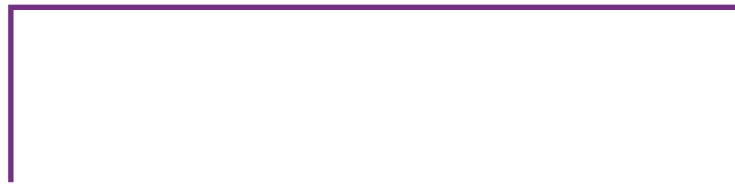
Contents

INTRODUCTUORY MODULE: DIGITAL COMMUNITY STEWARDSHIP	1
Terminology	2
What makes a digital community steward?	3
How do digital community stewards emerge?	4
MODULE ONE: MEMBER ENGAGEMENT	5
What is member engagement?	6
Reinforcers and Challenges in Group Participation	6
Increasing member engagement & utilising Facebook badges	7
Member engagement & Facebook insights	11
MODULE TWO: TRUST AND CONNECTION	13
Building trust	14
Understanding Trust in Digital Communities	14
Building Connection	17
Setting up community guidelines	18
MODULE THREE: UNDERSTANDING INFORMATION DISORDER	21
Misinformation, disinformation & malinformation	22
Types of information disorder and its impact	24
Practising healthy scepticism	28
MODULE FOUR: NAVIGATING INFORMATION DISORDER	31
Fact checking	32
Verification	36
Flagging	40
MODULE FIVE: NON-VIOLENT COMMUNICATION	41
Understanding Online Hate Speech	42
Understanding Non-violent Communication	45
Managing Conflict in Digital Groups	49

MODULE SIX: DIGITAL SAFETY	53
Why is Digital Safety Important for Online Communities?	54
What are the digital safety risks and challenges faced by online ?	55
Responding to Digital Security Risks and Challenges faced by Online Communities	57
MODULE SEVEN: DIGITAL LEADERSHIP	59
Who is a Digital Community Steward - An Exploration	60
How to be an Efficient and Effective Digital Community Steward - Key Strategies	61
How to be an efficient and effective digital community steward	62
How to be an Efficient and Effective Digital Community Steward - Rely on your Team	63
MODULE EIGHT: GROWTH & INCLUSIVITY	65
Expanding your group - Finding new members	66
Expanding your group - diversity and inclusion	70
MODULE NINE: DIGITAL RIGHTS	75
Community Standards - What do we need to know?	76
Digital Rights - What do we need to know?	79
Decolonising Digital Rights	81

This work by Search for Common Ground is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](https://creativecommons.org/licenses/by-nc-sa/3.0/).





INTRODUCTORY MODULE:

Digital Community Stewardship

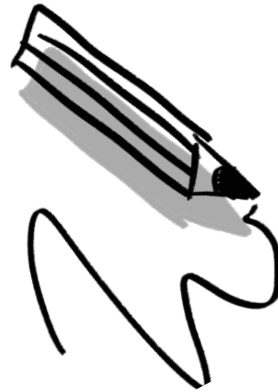
Module Objective: To explore what digital community stewardship means and entails.



Terminology

The descriptions below will describe the difference between three sets of concepts which are mentioned throughout the training.

This training was developed for both moderators and administrators of groups (vs pages) for digital community stewards (vs influencers).



1. MODERATOR VS ADMINISTRATOR

Both Moderators and Administrators are tasked with managing groups. These exact terminology and roles vary from platform to platform. As most social media groups are hosted on Facebook, we use the description by Facebook below.

Moderators: Moderators have the ability to approve or deny posts in the group, remove posts and comments in the group, remove or ban people from the group, and pin or unpin a post.

Administrators: Administrators have the ability to do all the same roles as a moderator. In addition, they are able to manage group settings, remove an admin or moderator, or make another member an admin or moderator.

2. DIGITAL COMMUNITY STEWARDS VS INFLUENCERS

Digital Community Steward: We define 'digital community stewards', as individuals who review user-generated content to ensure members adhere to rules, regulations and community standards of social media platforms. They act, in a sense, as first-line responders who build safe digital experiences within social media groups. This includes admins and moderators of both public and private groups on various social media platforms (i.e. Facebook, WhatsApp, Signal, Telegram, Instagram, Twitch, Discord, and Clubhouse).

Influencers: Influencers in social media are people who have built a reputation for their knowledge and expertise on a specific topic. They make regular posts about that topic on their preferred social media channels and generate large followers of enthusiastic, engaged people who pay close attention to their views. Brands love social media influencers because they can create trends and encourage their followers to buy the products they promote. (Source: [Influencer Marketing Hub](#))

3. GROUP VS PAGE

This training focuses on Groups (vs Pages). A distinction between the two is elaborated below (Source: [Facebook](#)).

Page:

Pages are generally used as a public marketing tool. Artists, public figures, businesses, brands, organisations and charities often use pages to share information with customers or members. People can “like” the page and follow along but have limited interaction between members.

Group:

Groups are places to communicate shared interests with certain people. You can create a group for anything (they are often based on– shared interests, shared identity, events, and community). Group settings offer more privacy settings than pages and allow moderators and administrators to determine membership and rules. When you join a group on Facebook, you start seeing content from that group in your Feed and are able to contribute to the content shared in the group.

This distinction is important, as this training is focused on online communities, and the structure of a “Group” more closely resembles an in-person “Community” than a “Page.”

What makes a digital community steward?



In a research study conducted by [Governance Lab in 2020](#), where the responses of over 15,000 social media users were surveyed, the following were identified as the three most important traits of a digital leader:

- Welcoming differences of opinion among members
- Being visible and communicating well
- Acting ethically at all times

A question for further reflection: What traits and qualities do you think are necessary for successful and efficient digital leaders?

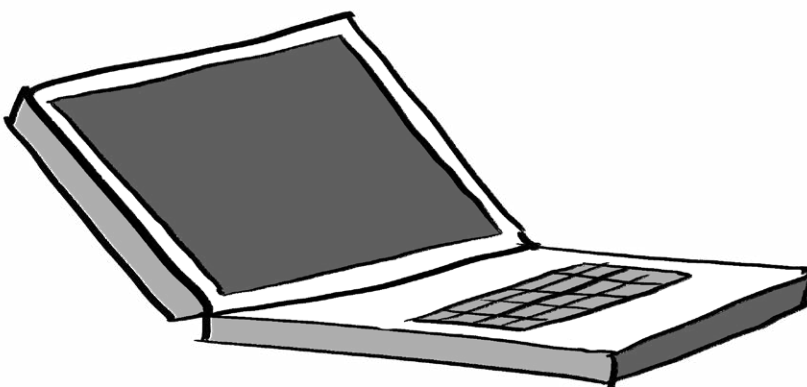
How do digital community stewards emerge?

Digital leaders, and specifically digital community stewards— are a unique group. Most digital community stewards are neither trained nor paid for the work that they do. This is often a labour of love and an act of service. So, it is important to think back on how your journey started as a digital community steward.

More information about digital leaders, based on research from [Governance Lab](#):

A question for further reflection: How did you become a digital community steward? For what reasons did you decide to take on this task? What keeps you interested in continuing this role?

1. Research shows that many group leaders are accidental leaders. They had little idea of how the leadership of their group would unfold when they took on the role. The job is complex and time-consuming and can impose intense pressures for which many leaders are not trained or prepared. For all the focus on leadership in contemporary society, online leadership is an emerging phenomenon that has not yet been well studied or understood.
2. When asked why they lead a group on Facebook, many group leaders said that their main reason for operating on the platform was the extraordinary global reach of Facebook and the fact that the platform and its tools were accessible, easy to use, and free of charge. Many leaders we interviewed also said that they derived significant personal satisfaction and a sense of achievement from their work.



MODULE ONE:

Member Engagement

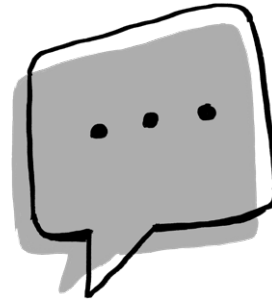
Module Objective: To obtain technical skills that will help increase member engagement in social media groups.

Module Dilemma: My group members are not active and engaged.



What is member engagement?

Member engagement is a measure of how people interact with your social media groups and content. Some standard metrics that are used to gauge engagement include likes, comments, social shares, saved items, click-throughs, or retweets. This can vary from one platform to another. Engagement helps to assess how your content is perceived by your audience based on their interest in and reactions to what you share.



Reinforcers and Challenges in Group Participation

CHALLENGES:

- **Privacy Violations** - There is no guarantee that members will behave and refrain from sharing private information from a group anywhere and everywhere. Screenshots, selecting and copying text, etc., can still be done. Leaks of members-only information can cause members to leave groups or even report unacceptable activity, damaging the group's reputation or evoking action by the platform. Monitoring this can be one of the toughest challenges for admins.
- **Harassment and Bullying** - Bullying and harassment happen in many places and come in many different forms. This can manifest in social media groups in various ways, from trolling to sending threatening messages and making unwanted malicious contact. Admins do not tolerate this kind of behaviour because it prevents people from feeling safe and respected in their online communities.
- **TL;DR** - The "TL;DR" texting abbreviation stands for "Too long; didn't read." Sadly, this is one of the main reasons that group activity is low. Even in groups where members have voluntarily chosen to participate and get information, their level of information overload—often due to too much group chatter—can reduce their activity, sometimes to the point of totally ignoring it.

REINFORCERS:

- **Consistency** - Being consistent in terms of topics discussed and frequency of posting activity can go a long way toward keeping groups engaged by setting standards and expectations. A disciplined approach helps build trust
- **Patience** - Groups can take a long time to gain traction. Some admins give up on the group because of low numbers and lack of participation. Unless the purpose of the group dictates a shorter term, it is important to have at least a year's worth of data to review which will provide more valuable insight to decide the group's lifespan.
- **Leadership** - An admin who establishes a protocol and culture, and manages it effectively and equitably for members, can do a lot to keep a group on purpose and on task.

Increasing member engagement & utilising Facebook badges

ENCOURAGING MEMBERS TO PARTICIPATE



Ever feel like you're the only one posting to your group? You're not alone, this happens in every group, especially at the beginning. There can be a number of reasons why people may not actively participate in your group. Just as there are introverts and extroverts in any real-world gathering, there are active and less vocal folks online.

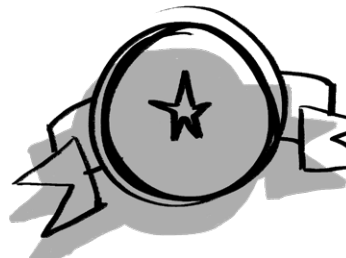
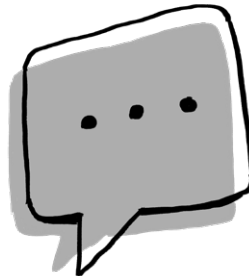
Some members want to comment on everything. Others may read or observe an ongoing discussion without participating, and only sometimes comment. Experienced group admins remind us that many quiet members are still getting value from your group. They're a part of your community, too.

Put yourself in a new member's shoes: What are the top few things you'd want to know about a group and its members before contributing? Use the group description and welcome post to explain some of these group basics. This helps members feel comfortable participating confidently.

Some members might need ideas about what to post. Asking easy, open-ended questions is a great way to increase group participation. If a less active member responds, interacting with them in the comments can let them know their voice matters. Many admins use polls to encourage less vocal members to add their opinion with a single tap. Going Live in your group allows members to comment and connect with you in real time. Admins tell us that even members who are reluctant to post will sit in on Facebook Live. Once people participate in these ways, they might feel more comfortable posting on their own.

TOOLS AND STRATEGIES TO IMPROVE ENGAGEMENT

1. **Welcome Post for New Members** – This is an effective way to make your new members feel welcome and valued. Stewards can develop a common welcome post for their group – and simply customise it by tagging the new members when they post it when necessary. Facebook automatically tags all the new joiners when you hit the ‘Write Welcome Post’ option under the ‘Members’ tab.
2. **Repost Best Engagement** – If there are 100 posts in a month in a Facebook group and 5 out of those received high engagement, stewards can repost any or all of those five posts that performed well in the group. Facebook Insights is useful in identifying posts with high engagement.
3. **Recognize Top Contributors** – This is something stewards can do on a weekly or monthly basis. Facebook Insights will help admins to identify members who are top contributors to the group. The simple act of making a ‘Thank You’ post to acknowledge and recognise the participation and engagement of these members can encourage others to contribute as well.
4. **Themed Days or Weeks** - This is a simple yet very effective way to increase engagement in Facebook groups. For example, ideas such as TipsTuesday, which motivates members to share tips and advice on a relevant question, or SelfieSunday, which encourages willing members to post a selfie, keeps the group active and alive – and gives them something to look forward to from the group.



USE OF BADGE FEATURES ON FACEBOOK

WHAT ARE BADGE FEATURES?

Facebook group badges or Facebook group member badges are the visual tags that you see beside group members' names. These icons next to names on Facebook appear in all group posts, comments, and member profiles if the badges are enabled in the group. The members can earn most of the badges based on how they contribute to the group. Members can request some badges like an expert badge on Facebook, based on their skills.



WHY ARE BADGES IMPORTANT?

Badges help you learn about community leaders and single them out from the crowd of group members, making it easier for new members to reach out whenever needed. This is also a way for admins to recognise group members for their significant contributions to the group.



HOW DO BADGES INCREASE MEMBER ENGAGEMENT?

- **They help you recognise your most engaged members** - Whether a community drops out or thrives fully depends entirely on how active its members are. Most communities have a few main members who keep the conversations going and ensure others get what they need from the group. These members are your community's greatest asset, and it's important to give them recognition in the community. Badges are a great tool for providing this kind of recognition. They make it easier for other users to spot the people who are most helpful, knowledgeable, and consistently responsive.
- **They promote healthy competition and set achievable goals** - For your more committed users, having a goal to work toward can serve as a powerful incentive. When community members start to see that some of their peers have badges, it alerts them to the possibility of earning one (or more) themselves. For some, it will become an aspirational milestone, something that drives them to participate more—either out of competition with others or their own internal motivation to be a star helper.
- **They provide a way to reward community members** - By bestowing badges onto members who participate the most often and in the most valuable ways, you provide positive reinforcement for their actions and incentivise them to do more. People awarded with badges get to feel the pride of being recognised for their contributions. They include gaining valuable skills badge-holders can add to their resumes, particularly those who want to expand their careers as confident Digital Community Stewards.
- **They help scale support even further** - One of the greatest benefits of a community is that your community members can help each other. Badges can help scale these efforts even further. The Verified Problem Solver—or whatever label you choose—helps take work off the community steward's plate.

THE BADGES

The following are badges that are relevant to community-based groups on Facebook.

Admin	For members who manage the membership, moderators, settings, and posts for the group.
Moderator	For members who manage the membership and posts for the group.
Group expert	For members who answer questions and share their expertise with the group.
New Member	For new members. (Visible for two weeks).
Founding Member	For members who helped build this group by sharing it, inviting people, or creating posts when the group was new.
Valued responder	For people who share valuable answers to questions asked in the group. (Shown for one month).
Visual Storyteller	Recognize those who consistently share valued photos and videos.
Rising Star	Recognize those who created engaging posts within their first month of joining.
Conversation Booster	Recognize those who created the most engaging conversations in the past month.
Greeter	Rewards users who make new members feel welcome.
Top Fan	Someone who can become a brand ambassador and increase the reach of your messaging. Top fans can serve as a catalyst, helping you gain the trust of new members quickly.

HOW TO MANAGE FACEBOOK BADGES

The following are some tutorials that will help you understand how to manage your badge features on Facebook.

[How to use badges in Facebook group 2020](#)

[How To Add Badges in Your Facebook Group](#)

Member engagement & Facebook insights

It is important for digital community stewards to understand the advantage of being a leader in a digital community. Group leaders on Facebook and other online platforms also have an important advantage over offline leaders: They can access analytics to see who is joining and leaving the group and how many members are engaging, even at what time of day it might be good to post to get the attention of more members. Most offline

can't see aggregate trends in their communities at any given time, so they don't always know when issues are bubbling up. The key difference in online communities is the ability to measure and see the breadth and depth of member behaviours - and how to influence it.



WHAT ARE FACEBOOK INSIGHTS?

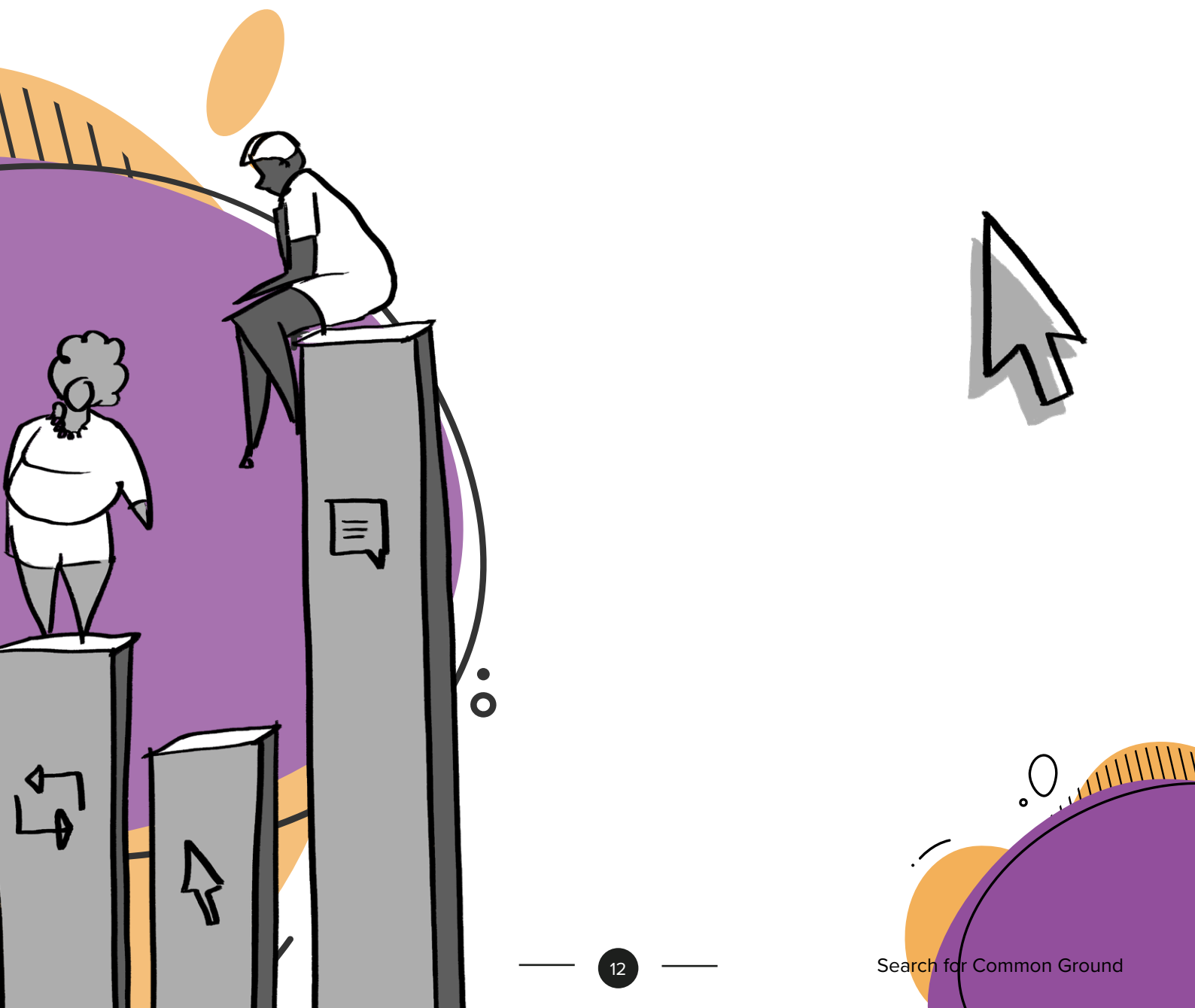
A Facebook group is so much like a human body. The minute we stop taking care of either of them, they will stop functioning and ultimately become dead. Hence, it is necessary to keep checking on our health and that of our Facebook group. For the latter, we obviously need the right set of engagement metrics.

That's where Facebook group insights help. They provide a holistic understanding of how our group is growing, how our members are engaging with the content, who is engaging the most when they are the most active, and so on.

Couple these insights with our group health data from Convosight and we will be able to upskill our community management game forever.

The following are some resources will help you understand how to use Facebook Insights to improve your group engagement:

[New Facebook Group Insights](#)
[How to Improve Your Facebook Group Engagement](#)



MODULE TWO:

Trust and Connection

Module Objective: Identify strategies to build trust and connection within the group to encourage better interaction among members

Module Dilemma: There is no real trust or connection in the way my group members interact with each other



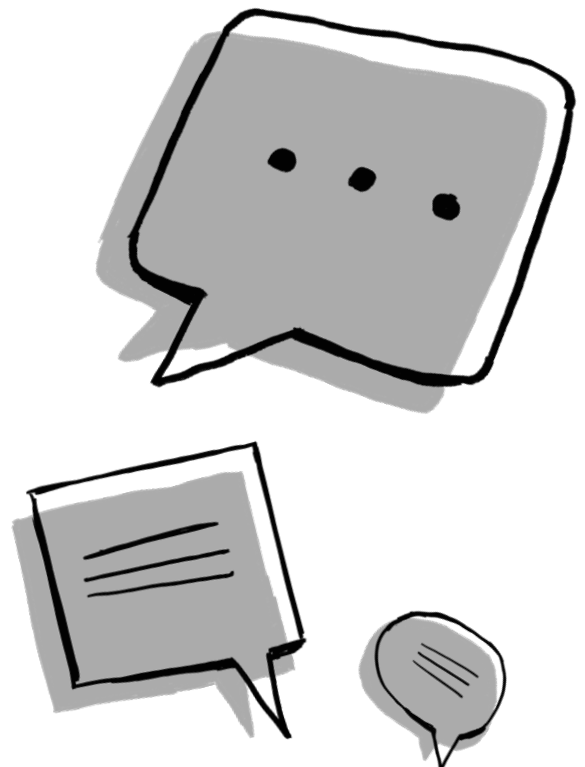
Building trust

UNDERSTANDING TRUST IN DIGITAL COMMUNITIES

It is not enough that the members share information, post regularly, and connect with the objective of the group. They must connect with each other. There need to be interpersonal relationships within the group.

- People can experience a strong sense of community from membership in such groups despite the lack of physical proximity.
- Online groups are a still fluid form of human organisation that in many cases attract members and leaders who are marginalised in the physical societies they inhabit, and who use the platform to build new kinds of community they could not form in real space.
- Many of these groups have counter-cultural norms. These groups cut across traditional social groupings and bring together people normally divided by geography around a shared trait or interest.
- The flexible affordances of online platforms have enabled new kinds of leaders to emerge in these groups with unique skills in moderating often divisive dialogues, sometimes among millions of members.
- The leaders of many of these groups run them as a labour of love; they are neither trained nor paid, the rules that govern their internal operations are often uncoded, and the hosting platform - in this case, Facebook - holds significant power over their operations and future.

Note: The above findings were taken from a research [report by The Governance Lab](#) (2020), which interviewed leaders of 50 Facebook groups and 26 global academic and industry experts for this study of the governance and leadership of Facebook groups.



BUILDING TRUST IN DIGITAL COMMUNITIES

PILLAR 1: TRANSPARENCY

Transparency and plain talk are imperative to build trust within digital communities. The maximisation of transparency translates to authentic community stewards who constantly keep the members updated about the purpose, objectives, and health of the group. The following are a few strategies through which we can achieve the first pillar of trust:

- Being clear and transparent about the objectives and boundaries of the group
- Reiterating the community standards of the platform and repercussions for breaking them

PILLAR 2: RESPONSIBILITY

This refers to making ethical decisions and safeguarding the welfare and dignity of the members. Any online platform is only as ethical as its creators and users. This means as leaders of these online groups, digital community stewards have the responsibility to promote fairness and well-being among interactions within the group.

- The following are a few strategies through which we can achieve the second pillar of trust:
- Ironing out complaints in a sensitive and timely manner
- Stopping misinformation in its tracks





PILLAR 3: PRIVACY

All members of the group have access to the information shared within the group. So, the group must acknowledge the preferences of the users regarding what information can be collected, shared, and saved. Consent must be valued and respected at all times.

The following are a few strategies through which we can achieve the third pillar of trust:

- Inform members about boundaries for sharing private information, highlighting the need to practise filtering personally identifiable information
- Setting clear standards and regulations regarding access, control, and sharing of information within the group

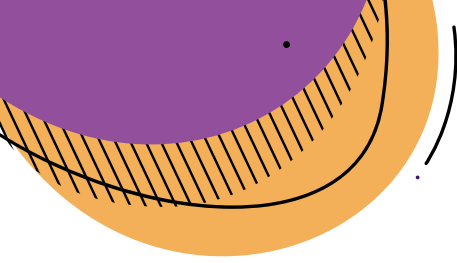
PILLAR 4: SECURITY

This requires digital community stewards to inform the members about the need for digital hygiene – both within the group and outside of the group. Heightened awareness of cyber risks and violations of digital rights will contribute to developing trust and reliability within an online community.

The following are a few strategies through which we can achieve the fourth pillar of trust:

- Be mindful of common cyber risks and digital threats to digital communities.
- Promote and practise basic digital safety practices within the group.

Note: The above information on digital trust was obtained and adopted through resources from [Deloitte Insights](#).



Building Connection

UNDERSTANDING INCLUSIVITY AND EXCLUSION



WHAT IS INCLUSIVITY?

Inclusivity is about welcoming, developing, and advancing a diverse mix of individuals. It's about making all people feel valued, including changing practices that might unfairly benefit any one group and making sure that everyone feels they have the same opportunity to advance and make an impact.

WHAT IS DIGITAL INCLUSION?

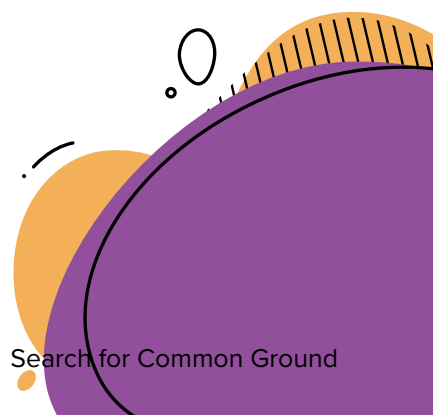
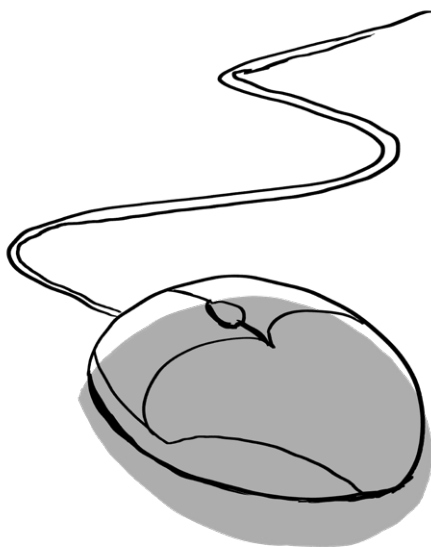
The following are some key components of digital inclusion:

1. Affordable, robust broadband internet service
2. Internet-enabled devices that meet the needs of the user
3. Access to digital literacy training
4. Applications and online content are designed to enable and encourage self-sufficiency, participation, and collaboration.

BUILDING EMPATHY AND INCLUSIVITY IN DIGITAL COMMUNITIES

The following are some ways digital communities can take specific measures to make marginalised and vulnerable groups feel welcome and included in online communities.

- **Remember community standards** – Most digital platforms have community standards that are designed and practised to protect the safety and welfare of its users. Since your group is hosted on such a platform, the members must be reminded that discrimination and harassment within the group would violate the community standards. Therefore, all members are expected to practise empathy and inclusivity at all times.
- **Reaffirm through your group rules** – Most groups, as we will see soon, will follow the rules. It is important for community stewards to reiterate the importance of inclusivity and empathy through these rules. This shows members that these values are not simply practised due to the necessity to follow the community standards established by the organisation. But also because the group values diversity, inclusion, and equity for all.
- **Celebrate diversity** – Many social media groups make celebratory posts for big holidays. This practice should not be restricted to majority communities. For instance, if a group has members of diverse religions, posting a celebratory message only on Christmas is not enough. Similarly, depending on the group and context, key celebrations of diverse communities, such as International Day for People with Disabilities and the International Day Against Homophobia, Biphobia, and Transphobia, should also be acknowledged and celebrated.
- **Promote image descriptions** – Alternative text can be attached to images so that when a screen reader encounters an image, it reads out the text to the end-user rather than just skipping over it. If you are using text in images, say as a header or banner, the end-user is missing out on that key information. Images on social media are becoming more accessible too, and the ability for end-users to add alt-text is now available on most platforms (like Facebook, Twitter, and Instagram). Where this function is not available, the best practice is to provide an Image description at the bottom of your post.



Setting up community guidelines

A core strategy for building trust and promoting connections is to practise empathy and inclusivity in our digital communities is to have clear, well-defined, and uncompromisable guidelines that will protect the members of the groups – including those who are marginalised and made vulnerable online.



DEVELOPING REGULATIONS FOR DIGITAL COMMUNITIES

The following are some focus areas you can focus on to develop the community regulations:

- Transparency
- Privacy
- Accessibility
- Safety & Safeguarding
- Non-discrimination & Inclusivity

The following are some guiding notes and recommendations you can utilise to develop the community regulations:

- Think of the best way to present these rules: Bullet points, Dos and Don'ts
- Outline and reiterate the purpose of the group
- Be clear about core group values
- Be clear about your moderation policy
- In addition to stating what is not allowed, explain what is welcomed and encouraged.
- Reflect on the overall community standards of the platform
- Consider if there are any legal restrictions or regulations preventing members from commenting on certain topics.

The following is one such sample set of community guidelines for reference. Note that these need to be further fleshed out and developed according to context and requirements.

- Be respectful of others and their opinions.
- Do not make personal attacks in your comments.
- Do not post content that is discriminatory, obscene, inflammatory, harassing, hateful, threatening, profane or personally abusive.
- Do not post anything that could be libellous or defamatory: if an accusation is made against a named person, for example, a member of the group, we will hide it.
- Do not post content with explicit or technical detail about self-harm or suicide.
- Do not post adverts for commercial products or services.
- Do not repeatedly post about issues that are off topic, i.e., spam.

Some other resources and templates for social media guidelines of various organisations:

[Social Media Guidelines – Community First Development](#)

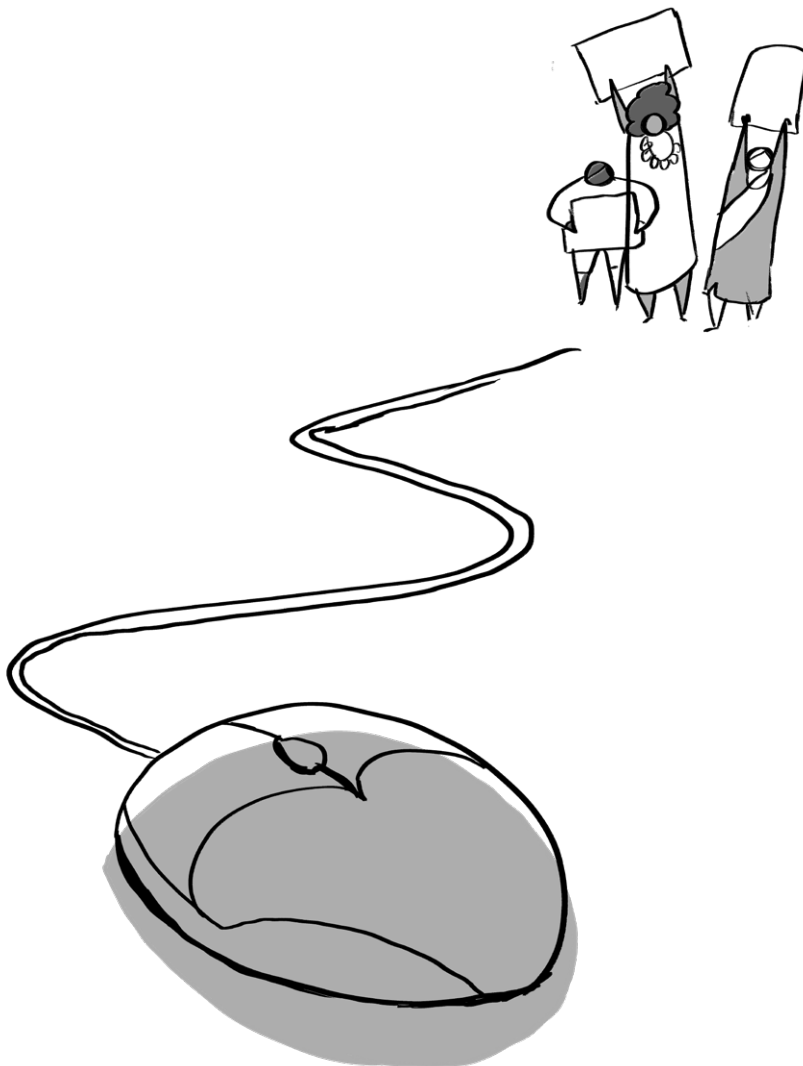
[Social Media Community Guidelines – Fresh Visions](#)

[Social Media Guidelines and House Rules Template](#)

FINALISING REGULATIONS FOR DIGITAL COMMUNITIES

The following are some suggestions for the digital community stewards to be mindful of when reviewing and further developing the set of guidelines:

- Contextualize and prioritise the guidelines according to the experiences and requirements of your group
- Include group members in the finalisation process. Listen to their suggestions and opinions on the draft guidelines. This is important to promote inclusivity.
- Think about how you are going to hold breaches and violations of these guidelines. Reflect on accountability.
- Decide how often these guidelines need to be reviewed and updated and decide who will be responsible for this process.
- Ensure these guidelines are available in the local languages and/or commonly spoken languages within the group.
- Remember to practice – not just preach. Lead by example.
- Consider accessibility and where these regulations will be hosted – about section, pinned post, external link, etc.



MODULE THREE:

Understanding information disorder

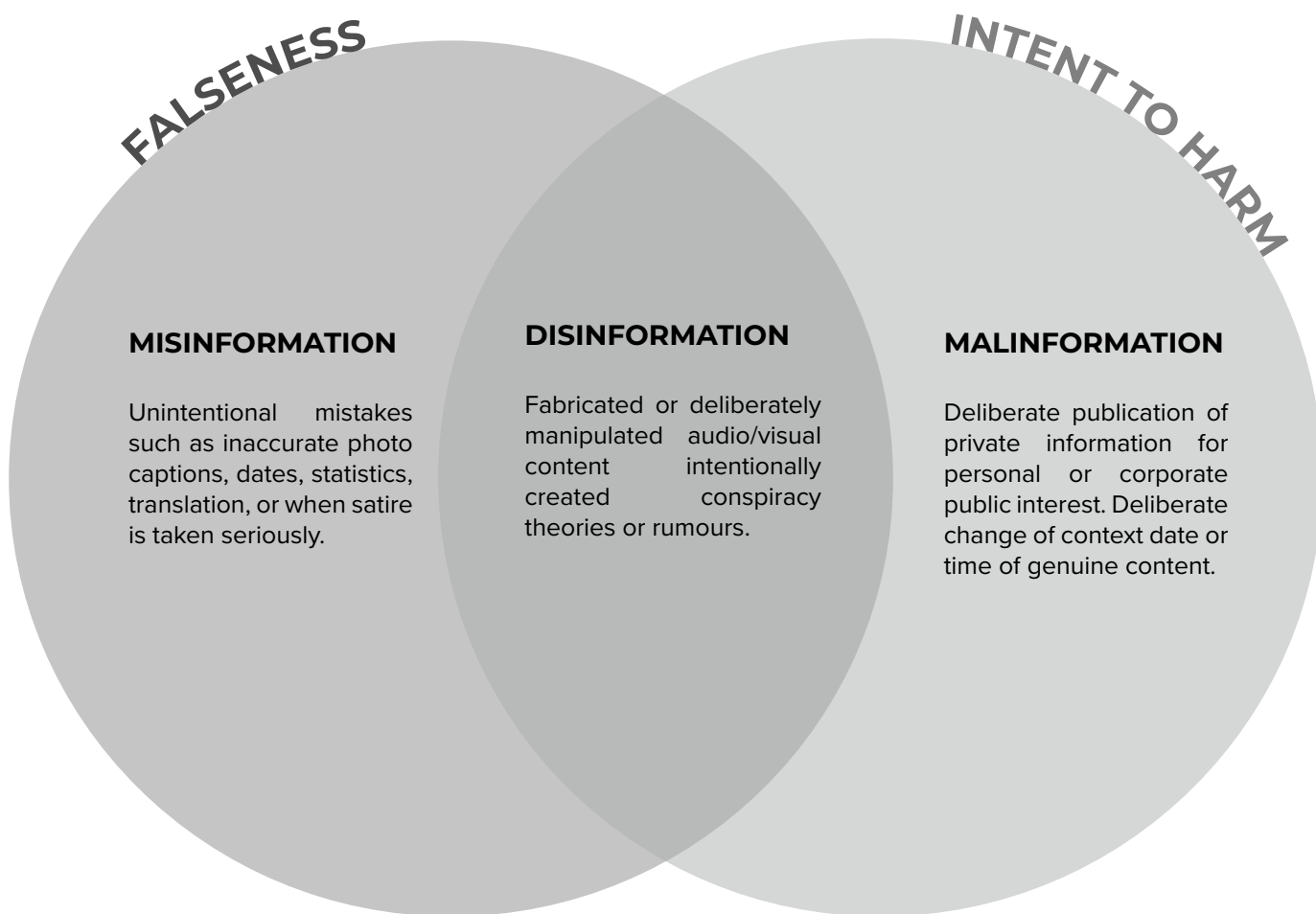
Module Objective: Understand information disorder and its impact on digital communities

Module Dilemma: My group members are promoting misinformation and disinformation



Misinformation, disinformation & malinformation

The following information on Information Disorder was obtained and adopted through resources from [First Draft News](#).



MISINFORMATION


Definition - Misinformation is false information shared by people - but they don't realise it's false or misleading, often because they're trying to help.

Example - A terror attack on the Champs Elysees in Paris on 20 April 2017 inspired a great deal of misinformation as is the case in almost all breaking news situations. Individuals on social media unwittingly published a number of rumours, including the news that a second policeman had been killed, for example. The people sharing this type of content are rarely doing so to cause harm. Rather, they are caught up in the moment, trying to be helpful, but fail to adequately inspect and verify the information they are sharing. One example was that Muslims in the UK celebrated the attack. This was debunked by the CrossCheck project on 22 April 2017.

FALSE

Did London Muslims "celebrate" a terrorist attack on the Champs-Elysees?

Posted on April 22, 2017



Did Muslims in London celebrate the Champs-Elysees terrorist attack?

MANIPULATED

They're celebrating a cricket victory

The video dates back to June 2009

CrossCheck

India lynchings: WhatsApp sets new rules after mob killings

© 20 July 2018



DISINFORMATION

Definition – Disinformation is false or misleading information intentionally created to make money, have political influence, or maliciously cause trouble or harm.

Example – In India, between 2017 and 2018, rumours of child kidnapping spread through WhatsApp, inciting violence against certain population segments and resulting in at least 33 murders and more than 99 attacks. After this incident, WhatsApp had to limit the number of times a message can be forwarded after it was seen that a spate of mob lynchings was linked to messages that circulated on WhatsApp groups in India.

MALINFORMATION

Definition – Genuine information that is shared with the intent to cause harm. This could be personal details, sexual images published without consent, or leaked emails to damage someone's reputation.

Example – In the late 90s and early 2000s, anti-abortion activist Neal Horsley collected names, pictures, and home addresses of abortion providers and published them on a website called the Nuremberg Files. He labelled that list as a “hit list.” Eight doctors from Nuremberg's listings have been killed so far. The website celebrated the death of such murders and encouraged pro-life activists to continue killing other doctors from the hit list.

Publishing Doctors' Names is a Threat

The Nuremberg Files was 1990s anti-abortion doxxing



Types of information disorder and its impact

Within the three overarching types of information disorder (mis, dis and mal information), we also refer to seven main categories. These help us understand the complexity of this ecosystem and the shades of grey

that exist between true and false. They live along a spectrum, and more than one category can apply to a specific type of content.

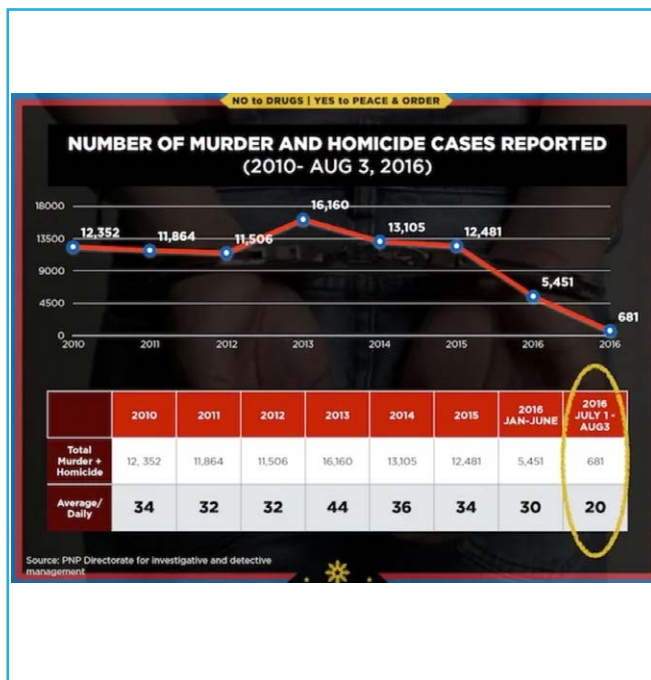
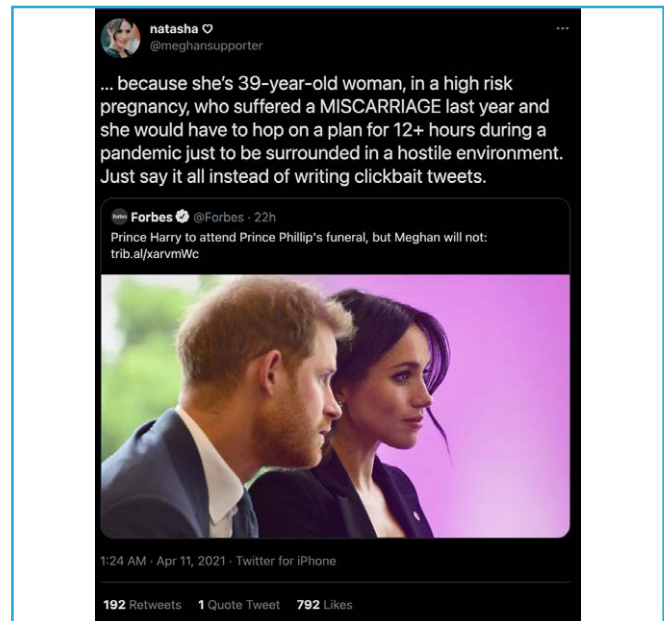


SATIRE

Satire is a literary technique that employs humour, irony, or exaggeration to expose flaws and criticise individuals, governments, or society itself. Although satirical pieces are meant to be humorous, their greater purpose is often constructive social criticism. For example, you might be aware of The Onion, a very popular satirical site in the United States. El Deforma, Mexico's version of The Onion, News Curry from Sri Lanka, and Revista Barcelona from Argentina are similar publications. The problem is when satire is used to strategically spread rumours and conspiracies. When challenged, it can be simply shrugged off “as a joke”, something not meant to be taken seriously. Furthermore, satire can also be dangerous when from its original source, it gets spread online and turned into screenshots or memes, losing its original context in the process

FALSE CONNECTION

When headlines, visuals, or captions do not support the content, this is an example of a false connection. The most common example of this type of content is clickbait headlines. With the increased competition for audience attention, editors increasingly have to write headlines to attract clicks, even if when people read the article, they feel that they have been deceived. This can also happen when visuals or captions are used, particularly on sites like Facebook, to give a certain impression, which is not backed up by the text. For example, the satirical news website The Science Post published an article titled 'Study: 70% of Facebook users only read the headline of science stories before commenting' in 2018. The body of the article didn't have any actual text, just paragraphs of "lorem ipsum" as a placeholder. But you'd only know that if you clicked through to read it. It was shared more than 125,000 times and proved the point of the headline.



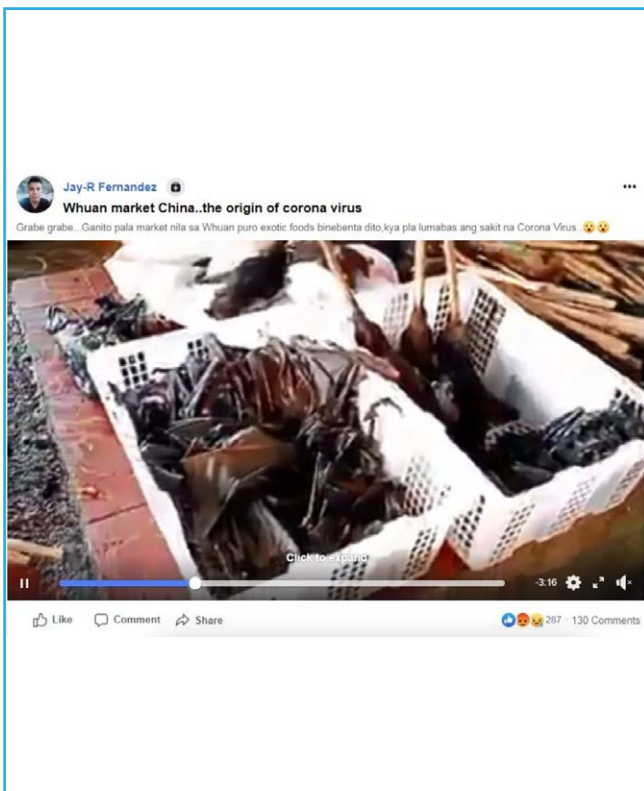
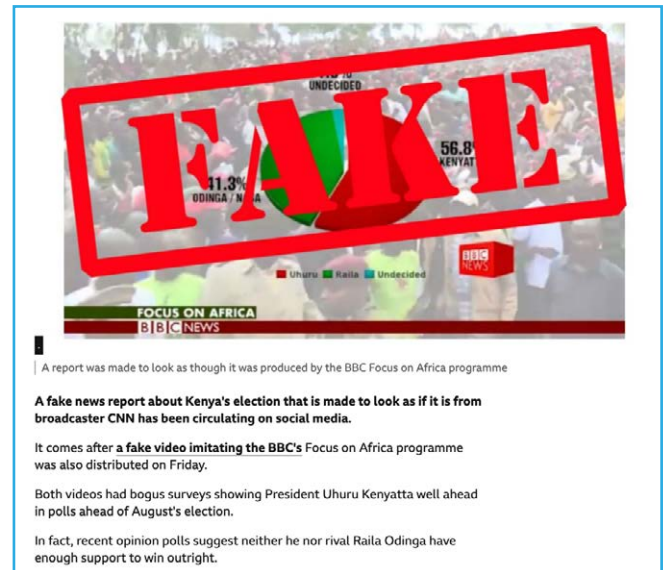
MISLEADING CONTENT

What counts as 'misleading' can be varied and hard to define, but it usually involves omitting pieces of information to tell a story in a certain way (i.e. cropping photos to change its message, choosing statistics selectively). This is also called 'framing'. Even the most advanced technology cannot easily detect misleading use of information because it involves contextualisation and nuance. This means it requires our brains to analyse the whole story or the bigger picture to judge whether the content intentionally misleads or not.

On August 22, 2016, during the first Senate hearing on extrajudicial killings, then-senator Alan Peter Cayetano showed a line graph (Figure 1) that purports to show the declining number of murder and homicide cases reported since President Duterte assumed office. But the line graph dipped at the end mainly because the data for 2016 was split into two periods: January to June, and July 1 to August 3.

IMPOSTER CONTENT

We always like to employ mental shortcuts to help us understand information. One very powerful shortcut is seeing a brand or person we already know and trust. When we get information coming from trusted brands or people, we are not as doubtful. But the problem is, it is very easy to make fake accounts and pretend to be someone else online. Imposter content is false or misleading content that claims to be from established brands, organisations, or personalities. For example, ahead of the Kenyan elections in 2017, BBC Africa found out that someone had created a video with a photoshopped BBC logo and strapline, and it was circulating on WhatsApp. They, therefore, had to make a video that they shared on social media, warning people not to be fooled by the fabricated video.



FALSE CONTEXT

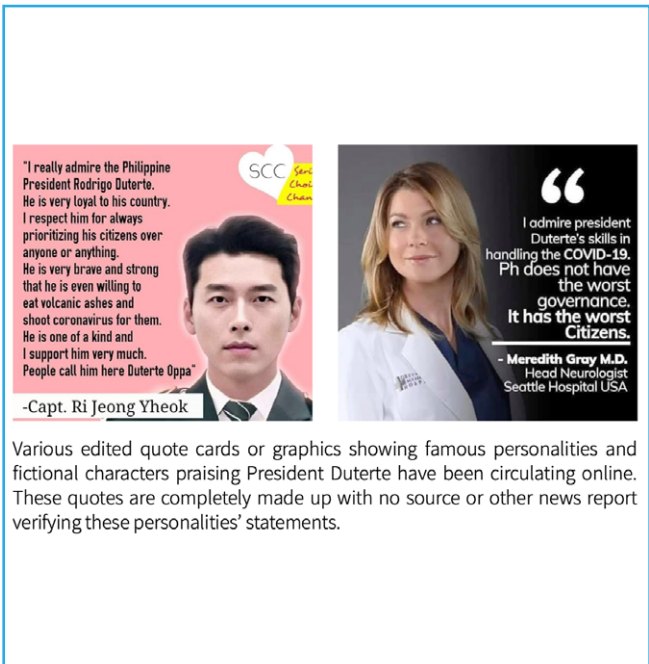
When genuine information is shared out of its original context, such as when old news stories are re-shared in the present time, it can be very dangerous. Sharing information in its proper context is very important because the context (i.e., the time, place, and situation) within which an event or news story existed helps explain the event. Sometimes, it is only a plain case of misinformation where a person mistakenly re-shares an old story. Other times, the purpose is more deliberate: to mislead people by sharing information in a different context.

One of the first viral videos after the Coronavirus outbreak in January 2020 showed a market selling bats, rats, snakes, and other animal meat products. Different versions of the video were shared online, claiming to be from the Chinese city of Wuhan, where the new virus was first reported. However, the video was originally uploaded in July 2019, and it was shot in Langowan Market in Indonesia. It was shared widely online because it played on people's anti-Chinese sentiments and preconceptions.

MANIPULATED CONTENT

Manipulated content is genuine content that is altered or edited to change the message. It is not completely made up or fabricated. This is most often done with photographs and images. This kind of manipulation relies on the fact that most of us look at images while quickly scrolling through content on small phone screens.

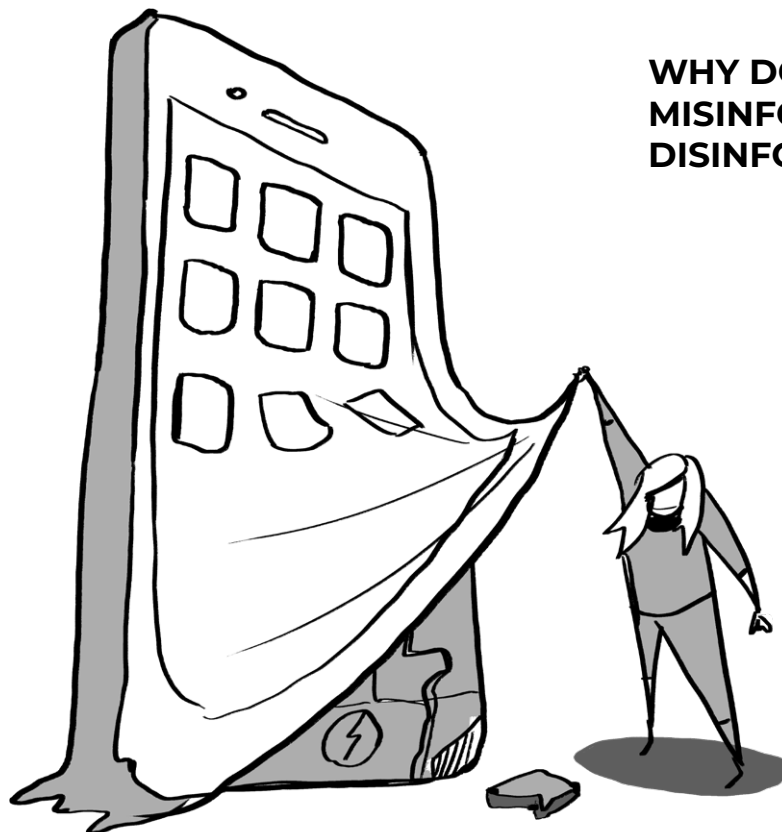
On February 3, 2020, the Sudanese Minister for Foreign Affairs and the Chinese Ambassador to Sudan met to discuss the ongoing Coronavirus outbreak. In the next couple of weeks, the photographs of that meeting were photoshopped to show the Sudanese Minister wearing a face mask. The images were shared widely on social media, including comments like “Africans don’t want to take chances with the Chinese”.



FABRICATED CONTENT

Fabricated content is anything that is 100% false. This is the only type of content that we can really consider purely ‘fake’. Staged videos, made-up quotes, and fake websites fall under this category. ‘Deepfakes’ or ‘synthetic media’ are fabricated media produced using Artificial Intelligence (AI), which usually combines different elements of video and audio to create ‘new’ content that never actually happened.

Practising healthy scepticism



WHY DO PEOPLE BELIEVE MISINFORMATION AND DISINFORMATION?

INFORMATION APPEALS TO OUR EMOTIONS.

Studies show that people remember information better when they appeal to their emotions. These are stories that make people angry, scared, anxious or make them jump for joy. One perfect example of misinformation that banked on people's fear is when Philippines social media personality DJ Loonyo hinted about the alleged 'dangers' of coronavirus mass testing. Through a

Facebook Livestream, he expressed fears over what one might be asked to drink or ingest in a 'trial-and-error' process for mass testing. His statement went viral and drew flak for spreading fear and misinformation about COVID-19 testing, which does not require ingestion nor is a trial-and-error process.

Another example is this clickbait story from the tabloid Abante Tonite about a bill making religious mementoes in hospitals optional. Its misleading headline states, “Hindi lahat Katoliko! Krus sa mga ospital pinapatanggal”. The story was shared on Facebook in multiple pages and groups, garnering “angry” reactions from many Facebook users. The headline purposefully misled the readers to think that the bill intends to ban religious mementoes instead of simply making them optional.



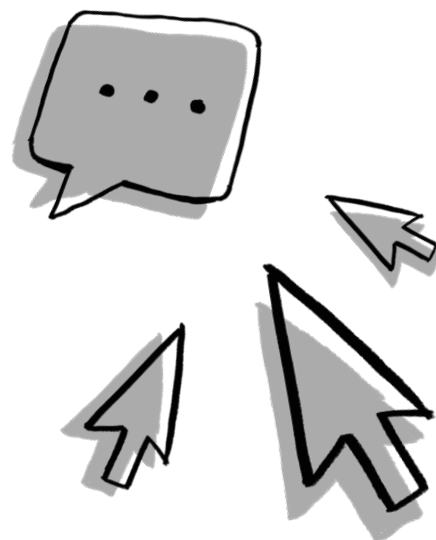
WE CARRY MANY BIASES WITHIN US.

Aside from our emotions, we also tend to accept information faster and easier when they confirm our existing views. This is called ‘confirmation bias’. The danger here is when we think something is true when we feel that it must be true. This is most applicable to misleading content -- information that has some amount of truth to it rather than being entirely made up. If an online post is partly true, and you are already convinced by half of it, you may disregard that that post is also ‘partly false’ or, at the very least, incomplete.

Aside from confirmation bias, there are many other hidden biases that influence one’s way of thinking. It is useful to be aware of these, too.

1. Implicit bias: we associate two different things, which in our minds, are usually linked
2. Sunk-cost fallacy: the more time or emotions we invest into something, the more we want to keep investing in it
3. Anchoring bias: the first piece of information we hear tends to have more influence on us
4. Bandwagon effect: if a lot of people act or think in a certain way, we tend to act or think the same

If we encounter a post online that feels right to us and triggers our emotions, our tendency is to share it with others. So, the very first step is crucial: pause, calm down, and recognise your emotional response. This is called ‘emotional scepticism’ or questioning your own emotional reactions to the messages around you.



THE BALANCE BETWEEN TRUSTING AND DOUBTING

This is a crucial task for digital community stewards. One of the main challenges in dealing with information disorder is finding trustworthy sources amidst the

information overload about the coronavirus pandemic. As a rule of thumb, stewards must act with caution.

HEALTHY SCEPTICISM VS CYNICISM

To be sceptical means to have an attitude of doubt, to be always questioning. This is a really important skill for dealing with information disorder, but too much of it can be unhealthy too. It can quickly slide into cynicism which is an attitude of scorn, negativity, and general distrust in people's motives and integrity. When you see too much disinformation everywhere, it is easy to be disheartened and develop hatred, and this is what we must strive to avoid. Remember that not all information is designed to deceive or manipulate. Our goal is to maintain the right amount of scepticism of the news we consume without sliding into the idea that good journalism does not exist.

An important strategy so that you can avoid cynicism is to learn to ask questions about ALL media messages, not just those with which you may disagree. We must be aware of and open to questioning not only the biases of media producers but also our own biases. This way, we find the right balance between trusting and doubting.

SKEPTICS	CYNICS
<ul style="list-style-type: none"> • are open-minded • challenge negative factors • can be convinced by presenting evidence 	<ul style="list-style-type: none"> • are close-minded • focus on the negative factors • cannot be convinced by presenting evidence

MODULE FOUR:

Navigating Information Disorder

Module Objective: Navigate information disorder to reduce its impact on digital communities

Module Dilemma: I want to flag certain content to the platform I'm moderating my group on and be safe doing it (e.g., remain anonymous)



Fact checking



An important thing all digital community stewards should ask themselves while trying to fact-check a claim is whether it is verifiable, or what is referred to as “fact-checkable”, or not.

To understand this, let’s look at the following examples.

Opinion - An opinion is based on a belief or point of view. It is not based on evidence that can be checked and often displays a bias as it is based on someone’s own experiences, world views and belief systems. (E.g. “Pink is a better colour for girls than blue”).

Prediction - A prediction is a statement predicting or promising something that will happen in the future. E.g., “It will snow in March.”

Fact - A fact is backed up with evidence and can be checked or verified. E.g., “Spiders are arachnids, not insects.”

Thinking something is true because we agree with it, does not make it a fact.

Two people might be in the same room – one saying it is hot and the other that it is cold. These opinions cannot be fact-checked. What can be fact-checked, however, is the exact temperature in the room.

Digital community stewards should only fact-check facts – not opinions or predictions.

It can be a difficult task to figure out what posts and what claims need to be fact-checked. This filtering and prioritisation depend entirely on the admin of the group – who sometimes might be guided and supported by the group members.

In this case, it is important for digital community stewards to focus on the following. Asking these questions can help you decide whether a post should be fact-checked or not.

- Is the topic important?
- Is the claim framed as a statement of fact or opinion?
- Does the claim matter? What is the potential harm?
- Have I flagged content by this group member before?

LET'S TALK ABOUT BOTS

Bots are basically social media accounts that are operated entirely by computer programs and are designed to generate posts and/or engage with content on a particular platform. In disinformation campaigns, bots can be used to draw attention to misleading narratives, to hijack platforms' trending lists and to create the illusion of public discussion and support. Researchers and technologists take different approaches to identify bots, using algorithms or simpler rules based on a number of posts per day.

Below is an infographic from FirstDraftNews, sharing some indicators to "Spot a Bot".

HOW TO SPOT A BOT

Just because it acts like a doesn't mean it is a bot. These indicators of automated or co-ordinated online activity can help, but look for a combination of signs, not just one.

ACCOUNT

- Recent creation date
- Lack of personal information
- Profile photo is ambiguous, stolen or non-existent
- Divisive words, hashtags, URLs or emojis in bio
- Suspicious handle e.g. lots of numbers

CONTENT

- Tweeting in more than one language
- Engaging in multiple international narratives
- Signs of automation or account management software like buff.ly
- Posting inflammatory memes or GIFs
- Hashtag spamming
- Occasional off-brand retweets
- Very few reliable news sources
- Awkward turns of phrase





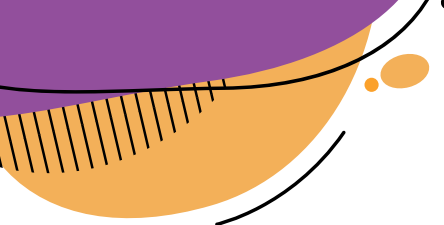
NETWORK

- Followers and following is high and almost identical
- High number of following and no followers
- Following a suspicious mix of sources
- Connected to other suspicious accounts
- Duplicated account
- Previously circulating suspicious content
- Previously identified by other organisations

More information on the above indicators can be found [here](#).

ACTIVITY

- High volume of tweets (more than 100/day)
- High percentage of retweets (more than 80s%)
- Posting persistently day and night
- Posting only at specific times of day
- Sudden spike in activity or change in interests



Several infographics have been developed, explaining to people how to fact-check fake news, mis and disinformation.

We have adapted an infographic from the International Federation of Library Associations ([IFLA](#)) and [Simon Fraser University](#) to share some general and key guidelines for fact-checking:

GUIDELINES FOR FACT CHECKING

Consider the source:

Click away from the story to find the site and contact info

Read beyond:

headlines can be developed to be outrageous, creating “click bait” for more sharing. Read the entire article.

Check the author:

Quickly search the author, and ask: Are they credible or controversial? Are they real?

Supporting sources:

Click on the sources in the article, and ask: Does the information accurately support the story?



Check the date:

Is this old news? Reposting old news may not mean it is relevant to more current events.

Is it a joke?

If it is extremely difficult to believe, it might be satire. Look into the site and article to be sure.

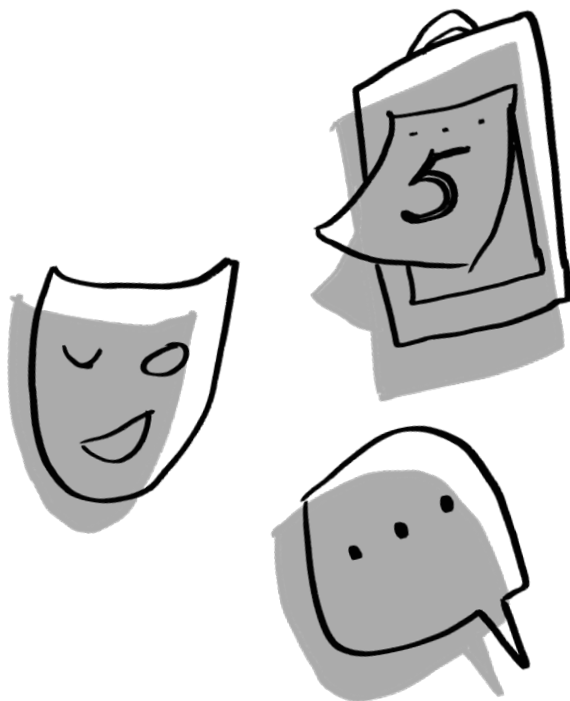
Check your biases:

Could your own position on the subject or beliefs be affecting your judgement? Remember to take a step back and think critically about this before making judgements.

Misinformation and Disinformation is created to appeal to our emotions and instincts. Think about why this news is making you react in a certain way.

Ask the experts:

Ask a topic expert, or cross-check it with other trusted or official sources or a fact-checking site.



You can watch [this video](#) to learn more about these key factors of spotting mis and disinformation

An interesting resource: [why do we prefer information disorder over fake news?](#)

Verification

WRITTEN INFORMATION

BOOLEAN SEARCH QUERY

Smart searches cut through social media chatter by finding precise snippets of information based on keywords. When searching for newsworthy content online, you've got to know exactly what you're looking for and have the skills to find it. Using the right keywords to search in the right places is key. This is where Boolean search queries help. These strings of words allow you to cut through the usual social media chatter by upgrading a default search to a multifaceted, specific search to find more precise snippets of information.

Boolean search is used when you require stricter search criteria to eliminate generic results that are loosely connected to your brand or other keywords.

Boolean searches help you to specify exactly what you are looking for or not looking for. For example, let's say you're searching for posts during a breaking news event, such as the Notre Dame fire. You want to search for Notre Dame, but you won't want posts about the Disney film. A boolean search will allow you to include posts that mention "Notre Dame" but exclude ones about the Disney film to refine your search results and find the information you're after.

This is possible with 'operators', which allow you to combine multiple keywords. There are three operators for basic searches: AND, OR, and NOT.

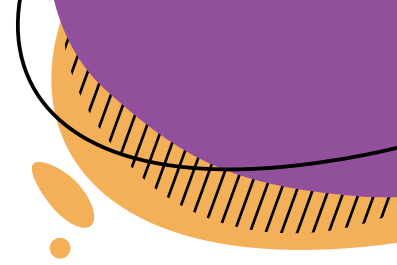
A quick note: Twitter, Reddit and Google currently support Boolean searches, while YouTube supports them to some extent on top of its advanced search function. Facebook used to support Boolean searches, though it appears it no longer does.



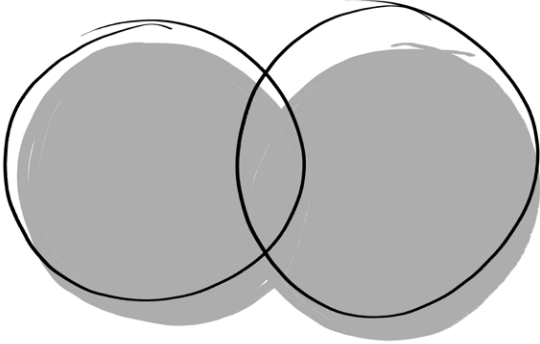
AND – This allows you to narrow your search to only retrieve results that combine two or more terms. For example, you might want to search for "Notre Dame" and fire.

OR – This allows you to broaden your search to retrieve results connecting two or more similar terms. This can be good for misspellings and typos. In the case of Notre Dame, you could search for "Notre Dame" OR "Notre Dam". This will retrieve all results containing either phrase.

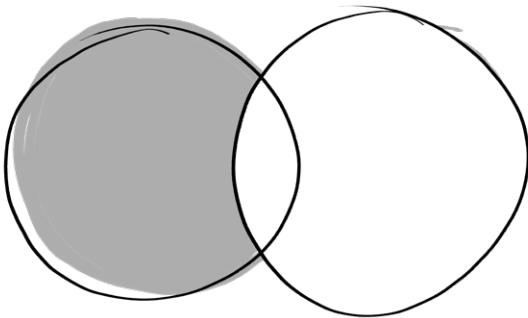
NOT – This allows you to exclude terms you don't want to appear in your search results. For example, you might want to exclude mentions of the Disney film The Hunchback of Notre Dame, and so you could write "Notre Dame" NOT "hunchback".



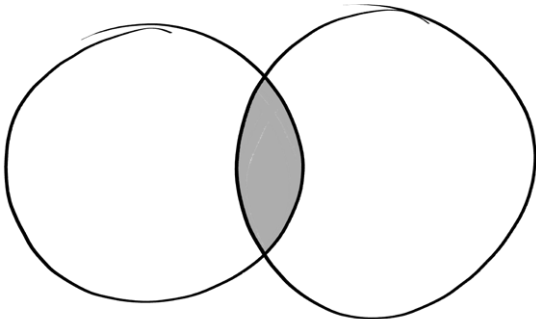
“Notre Dame” OR fire



“Notre Dame” NOT fire



“Notre Dame” AND fire



Key points:

- Operators (AND, OR, NOT) must be written in capitals, or they won't work
- If you're searching for phrases (terms made up of multiple words) then you have to put them in quotation marks (eg “Notre Dame”)
- You won't be able to find information that has been made private by a user

You can find more ‘operators’ for Boolean searches [here](#).

Traditionally, to search for content within a certain timeframe, you can search using the “since:” and “until:” operators (on Google it's “before:” and “after:”).

For example, if looking for examples of people turned away from a polling booth on the day of the European Parliament elections, you could search: “polling booth” AND “turned away” near:London since:2019-05-23 (year-month-day).

You can also look for content posted up to a certain date. For historical examples of people being turned away from polling booths, search: “polling booth” AND “turned away” until:2019-05-22.

Twitter, Google and YouTube also all provide dates in their advanced search functions which allow you to select a custom date frame.

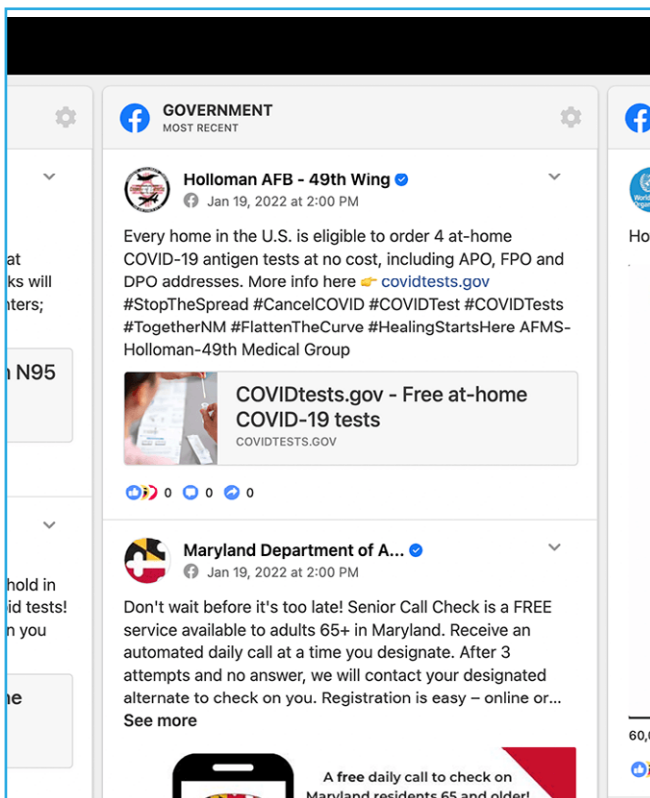
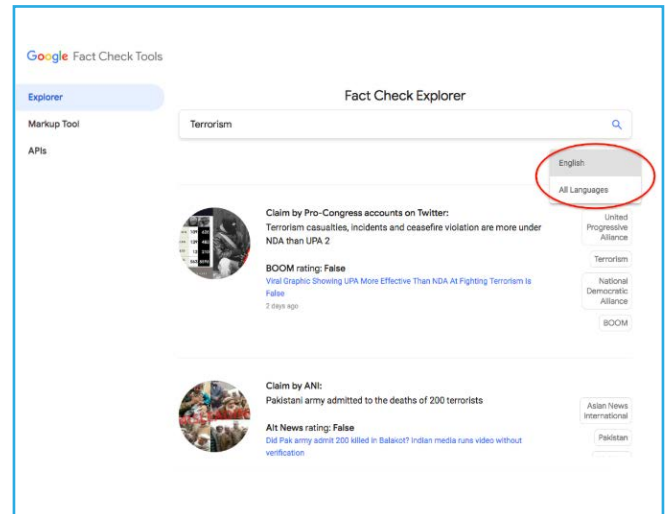
To search for social posts within a geographic area to find a source, you can search for content that was published in a specific location.

While very few social media users geotag their posts, it's possible to search for the specific locations of the ones that do.

For example, if there was a lack of booths at a London polling station, you might search: “polling booth” near:London within:15mi. You could even search: “polling booth” AND “turned away” near:London within:15mi.

FACT CHECK EXPLORER – GOOGLE

This tool allows you to easily browse and search for fact checks. For example, you can search for a politician's statement or for a topic. You can also restrict results to a specific publisher. You can search by keywords and see a list of matching claims and the corresponding fact checks. Keywords can be anything from specific topics to full politician quotes. Basically, anything you would also put into the regular Google Search bar.



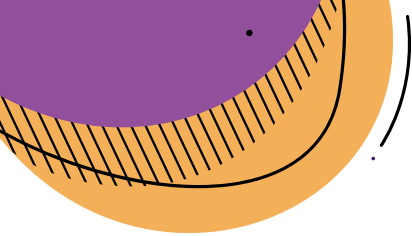
CROWDTANGLE LIVE DISPLAYS ON COVID-19

This is a very specific and [very useful resource](#) to navigate information disorder regarding the COVID-19 pandemic.

CrowdTangle is a social analytics tool owned by Facebook. Their main portals require sign-up, but everyone can access their Public Live Displays. It's a quick, visual way to see how information on coronavirus is being spread on social media.

Public Live Displays are organised by region and country and show content from local media, regional World Health Organization pages, government agencies, and local politicians, as well as social media discussion from Facebook, Instagram, and Reddit.

Each Public Live Display shows Covid-19 related posts in real-time, sorted by keyword, with public pages and accounts for each region.



VISUAL INFORMATION



A picture is worth a thousand words, and when it comes to disinformation it can also be worth a thousand lies. One of the most common types of misinformation we see at First Draft looks like this: genuine photographs or videos that have not been edited at all but get reshared to fit a new narrative. Let's look at such an incident.

In January, Facebook posts receiving thousands of shares featured the photograph (embedded above) and claimed the people in the photo were coronavirus victims in China. A quick look at the architecture shows that it looks very European, which might raise suspicion. Then if we take the image, run it through a reverse image search engine, and look for previous places it has been published, we find the original from 2014. It was an image, originally published by Reuters, of an art project in Frankfurt, which saw people lying in the street in remembrance of the victims of a Nazi concentration camp.

But with a few clicks, we can verify these types of images when they are shared online and in messaging groups.

Just like we can “Google” facts and claims, you can ask a search engine to look for similar photos and even maps on the internet to check if they've been used before. This is called a ‘reverse image search’ and can be done with search engines like Google or Bing.

REVEYE REVERSE IMAGE SEARCH

[This \(chrome\) extension](#) allows performing an inverse image search by right-clicking on any image in a website. You can configure the context menu to contain either a single button with your default search engine or a cascaded menu with all included search engines.

This extension allows us to perform a search by image. It even allows us to choose between the image search engines Google or Bing.

A sample tutorial can be found [here](#).

THUMBNAILS AND VIDEOS

Whenever we upload a video to the internet, it creates a thumbnail or screenshot to show as a preview. We can manually change it, but most people don't. Just like we can use a reverse image search to find out if a photograph has been published on the internet before, we can use thumbnails to see if a video has been previously posted online. Using reverse image search, you can take several thumbnails from any video and check whether it's been posted on the Internet before.

INVID

The InVid project develops a video verification tool called WeVerify to detect emerging stories and assess the reliability of video content spread via social media. It is a plug-in toolkit designed to assist in fact-checking through video verification. The tool provides users with contextual information on videos, reverse image searching, video metadata, and video copyright information, along with other features to assist in verifying content.

A sample tutorial can be found [here](#).

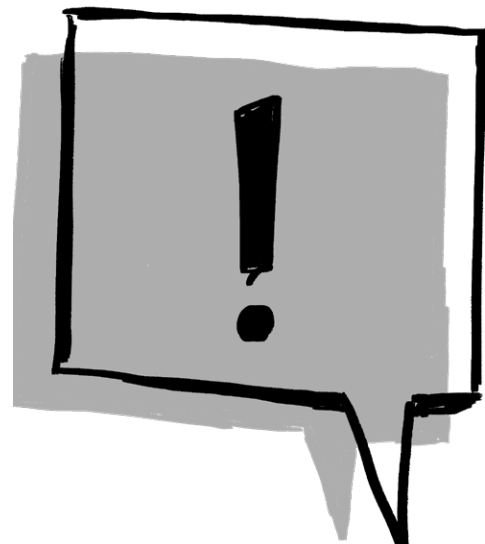
Watch [this video](#), which looks at a case study on verifying a social media post.

Flagging

HOW DO WE FLAG CONTENT?

What comes after is reporting such information to the relevant social media platforms. For this purpose, take a look at [this resource](#) from WHO, and additional resources, to inform yourself about the reporting practices relevant to your platforms. This resource will help you understand how to report and flag content on social media platforms such as: Facebook, Instagram, Twitter, WhatsApp, YouTube, TikTok, Discord, LinkedIn, Viber, etc.

Note: An important aspect of flagging such content in social media groups is to ensure the members do not repeatedly post or spread such content. Therefore, an essential part of this process is to address this issue with the member who posted the misinformation or disinformation in the group.



MODULE FIVE:

Non-violent Communication

Module Objective: Understand and encourage the need for and practice of non-violent communication in digital communities

Module Dilemma: My group members are insulting each other, encouraging hate speech and bullying



Understanding Online Hate Speech

WHAT IS HATE SPEECH?

The thing about hate speech is that it does not have a uniform definition in human rights law.

Hate speech regulations vary significantly by jurisdiction, particularly in how they define what constitutes hate speech and to what extent they differ by speech that is offline versus online.

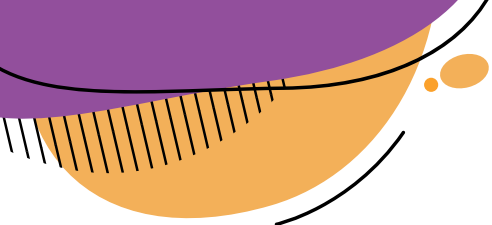
There is a need for clear and narrowly circumscribed definitions of what is meant by the term “hate speech”, or objective criteria that can be applied. Over-regulation of hate speech can violate the right to freedom of expression, while under-regulation may lead to intimidation, harassment, or violence against minorities and protected groups.

Here are a couple of international definitions by various organisations.

The International Covenant on Civil and Political Rights (ICCPR) Article 20 (2): Any advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence shall be prohibited by law.

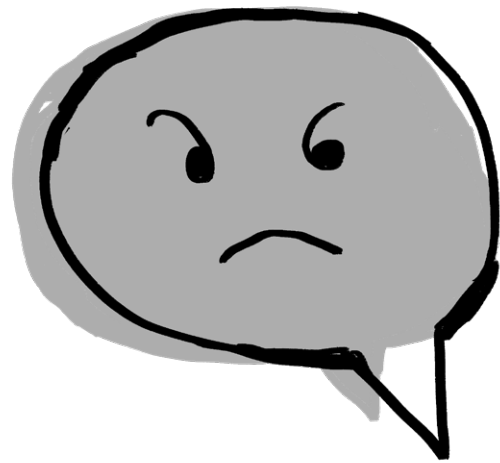
Article 4(a) of the International Convention on the Elimination of All Forms of Racial Discrimination: Dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts against any race or group of persons of another colour or ethnic origin, must be declared an offence that is punishable by law.

United Nations Strategy and Plan of Action on Hate Speech: Any kind of communication in speech, writing, or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factors.



While the third definition is not a legal definition and is broader than the notion of “incitement to discrimination, hostility or violence” - prohibited under international human rights law - it highlights three important attributes: Hate speech can be conveyed through any form of expression, including images, cartoons, memes, objects, gestures, and symbols and it can be disseminated offline or online.

- Hate speech is “discriminatory” - biased, bigoted, intolerant - or “pejorative” - in other words, prejudiced, contemptuous, or demeaning - of an individual or group.
- Hate speech makes reference to real, purported, or imputed “identity factors” of an individual or a group in a broad sense: “religion, ethnicity, nationality, race, colour, descent, gender,” but also any other characteristics conveying identity, such as language, economic or social origin, disability, health status, or sexual orientation, among many others.



You can access a list of targeted groups and relevant resources for each group [here](#). This list includes national, ethnic, religious, and linguistic minorities; migrants and refugees; women and girls; LGBTQI+; vocational targets such as journalists and activists.

HATE SPEECH VS FREEDOM OF EXPRESSION

Free speech refers to the right to seek, receive and share information and ideas with others. But this freedom must be used responsibly and can be restricted when considered as threatening or encouraging hateful activity.

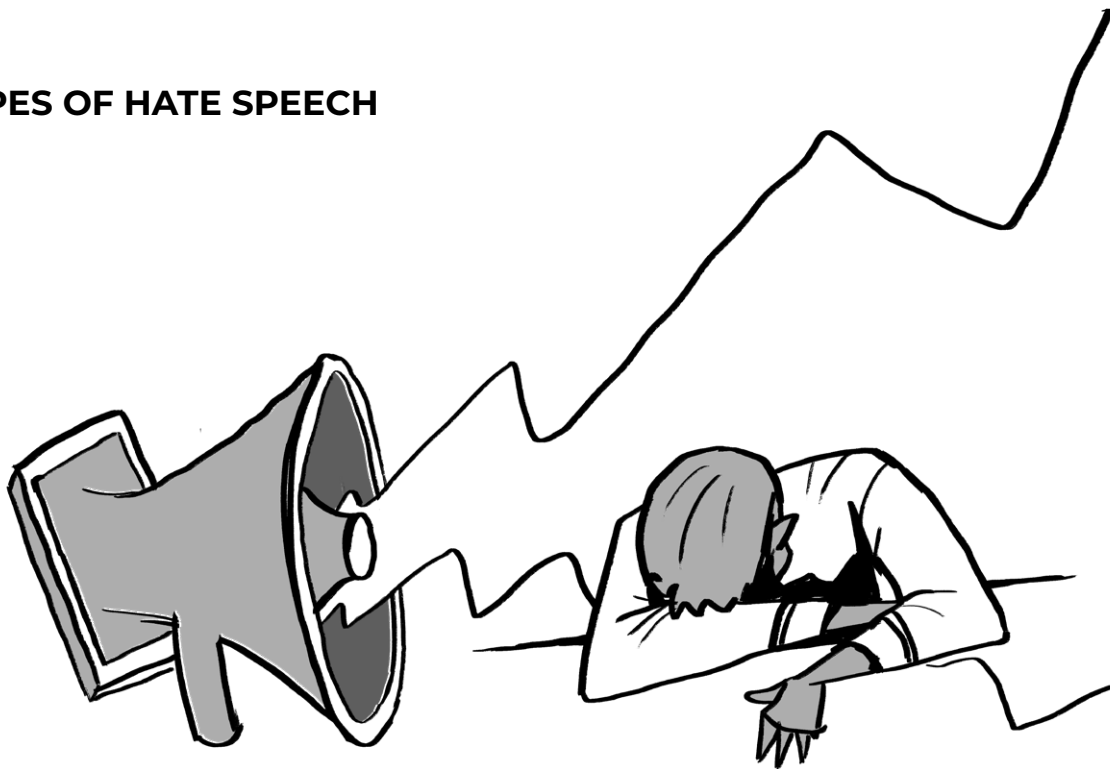
Hate speech, particularly online hate speech, targets particular groups of people – often minorities and dehumanises them. Hate speech perpetrators often see “the other” as enemies and have a tendency to connect all issues in society to these targeted communities.

Remember the key difference: Free Speech is a way to exchange, teach, learn and challenge each other’s perspectives,, whereas Hate Speech is targeting particular groups with malicious intentions and insulting individual identities. All internet platforms can be places where people post hateful content, whether as words, videos, photos, or memes, and cause great harm. It is up to all of us to consider our own online content and make sure we are not crossing the line from free speech to hate speech.

A good way to distinguish this difference between free speech vs hate speech is to use the constitution. For example, this is the case in Sri Lanka:

Some people might believe that they are within their right to say what they want – both on online and offline spaces – due to the freedom of speech and expression guaranteed by our constitution. If this is the case, then it must be also pointed out that the same constitution also guarantees that all persons are equal before the law and entitled to equal protection of the law and that no citizen will be discriminated against on the grounds of race, religion, language, caste, sex, political opinion, place of birth or any such grounds. Explain the relevance of both Article 14(1)(a) (Freedom of Expression) and article 12 (right to equality) of our constitution so that our participants understand that hate speech is not only immoral and unethical but also goes against the law.

TYPES OF HATE SPEECH



TYPE	DESCRIPTION	EXAMPLE
Disagreement	This involves disagreeing with the ideas or beliefs of a particular group.	Feminism does not exist. All feminists are wrong
Negative Actions	This highlights 'negative' nonviolent actions associated with the group.	Pro-choice activists want to ruin the future of our country by supporting abortions
Negative character	This includes negative characterization or insults towards a particular group.	All homosexuals are pedophiles.
Demonising and dehumanising	This involves belittling groups and equating them to subhuman entities.	Muslim people are pigs or homosexuals are monsters.
Violence	This outrightly calls for violence against the specific group.	Let's kill all Asians. Let's drive them out of our country.



Understanding Non-violent Communication

HOW IS DIGITAL COMMUNICATION DIFFERENT FROM IN-PERSON COMMUNICATION?

1. **Scope and Scale:** One person can send out a hateful or false message that is seen by millions of people all over the world. Hundreds of people can respond via a comment section.
2. **Anonymity:** Digital platforms allow people to communicate more easily with people we have never met and do not know.
3. **Less nonverbal cues:** Nonverbal cues such as eye contact, facial expression, hand gestures, and posture offer more opportunities for humanising and feeling empathy for others. These are missing in digital dialogue.
4. **Less Context:** In-person dialogue often relies on context cues, including ambiance, to help set a positive tone. These are missing in digital dialogue.
5. **Shorter messages:** Social media platforms emphasise short communication. Twitter limits messages to 280 characters. TikTok limits the amount of time to 60 seconds. With less space, people simplify their message to explain what they believe but rarely explain what experiences have led them to those beliefs or any complexity on the issue.
6. **More Emotional:** Emotional content spreads more rapidly. Comments or stories that evoke anger are more likely to receive engagement with “likes” or emoji markers. People may speak in more dramatic terms on social media to make up for a lack of nonverbal cues.
7. **Easier to Leave a Discussion:** Digital dialogue is easier to walk away from when discussions become tense. It can be harder to physically leave an in-person dialogue, so more people may “stay through the hard times.”
8. **More Public Witnesses and Less Privacy:** Digital communication involves silent onlookers, witnesses, or bystanders. A post with communication between 2-10 people who leave comments and respond to others is common on social media. What is distinct is that the post may have hundreds or thousands of silent witnesses who read and observe the interaction.
9. **More Shaming, Humiliation, and Dehumanization and Less Dignity:** People communicate in ways on social media that are rarely seen in physical interactions. It is easier to speak harshly to shame, judge, humiliate, and dehumanise strangers on social media than it is in the physical world. People may be openly attacked and experience humiliation from public shaming on social media.
10. **Bots:** In the physical world, people do not wonder whether they are talking to a robot. On social media, there are thousands of robots (‘bots’) pretending to be people. Social media bots are created for a variety of reasons, both good and bad. When thousands of bots begin sharing a piece of false information, it gives others the false impression that the information is popular.

The above explanation is based on a Toda Institute report on Digital Peacebuilding Communication Skills - Beyond Counter Speech (Schirch 2020).

WHAT ARE COMMON DIGITAL RESPONSES TO PROBLEMATIC SPEECH?

There are at least eight broad patterns of responses in comment sections on social media. These strategies work better or worse depending on different audiences

and contexts. They can be used together or on their own. Each can be done publicly or privately.

SILENT BYSTANDERS AND CONFLICT AVOIDERS

Conflict avoidance or choosing to be a silent bystander are by far the most common approaches people take when they encounter tense, conflicted conversations on religion, politics, health, or other issues. Silent bystanders watch but do not intervene in the digital conflict, hate speech, or disinformation. In the physical world, people may rarely witness abusive behaviour toward others. But online, the scale of conflict, hate, and false information is so great that some may feel overwhelmed by the idea of responding to it. Bystanders may choose silence for fear of making the situation worse or being implicated and pulled into the conflict.

FACT-BASED RESPONSES

Some respond to social media comments spreading false information or conspiracy theories with fact-based arguments. Fact-checking can inadvertently increase the number of people who see false information. But fact checking can work if it creates doubt in some observers so that the sharing of false information declines or is deleted. It seems to work best when done with a group of supportive fact-checkers who reinforce each other.

DISTRACTION, HUMOR, OR POSITIVE RESPONSES

Another type of response to problematic speech is using humour to lighten the mood or even mock a hateful or false comment. Research on counter-speech suggests this may be helpful in some cases. It may also result in more conflict. This sensitive strategy requires care.

SHAMING AND EMOTIONAL RESPONSES

Shaming is a form of “negative counter speech” in which someone observes a comment that they perceive is harmful to others, and they shame the speaker by denouncing the values or harming them. Shaming may mock or ridicule the speaker’s beliefs, demonstrate inconsistencies in a speaker’s thinking, question their goals, or highlight the negative impact of their speech on other people. In their review of organic examples of counter-speech on Twitter, researchers found that rebuking hate speech often led to apologies or deleting the original content.

PRIVATE OR PUBLIC REQUESTS TO REMOVE OR EDIT

Another form of response is to write a public or private message to the person who wrote a problematic comment on a social media platform to make them aware that the comment is viewed as offensive and to explain why it is offensive, and then to ask them to remove or edit their comment.

UPSTANDERS

Upstanding refers to bystanders who are witnessing harassment or hate speech to intervene on behalf of the person being harassed or victimised by hate speech. Upstanders or “cyber-Samaritans” is someone who models upstanding by dissenting to harmful posts by challenging the bully or supporting the victim. When this happens, other people are more likely to join in to support.

HOW TO PRACTISE NON-VIOLENT COMMUNICATION?

If “violent” means acting in ways that result in hurt or harm, then much of how we communicate—judging others, bullying, having racial bias, blaming, finger pointing, discriminating, speaking without listening, criticizing others or ourselves, name-calling, reacting when angry, using political rhetoric, being defensive or judging who’s “good/bad” or what’s “right/wrong” with people—could indeed be called “violent communication.”

Nonviolent Communication (NVC) is sometimes referred to as compassionate communication. Its purpose is to strengthen our ability to inspire compassion and to respond compassionately to others and to ourselves. NVC guides us to reframe how we express ourselves and hear others by focusing our consciousness on what we are observing, feeling, needing, and requesting.

Observations - NVC emphasises observation without judgement. This means presenting the simple facts we have observed. For example, instead of saying, “You have abandoned our group and never post anything anymore,” you can say, “I noticed that you don’t participate in the group as much as you used to.”

Feelings - NVC involves taking responsibility for your feelings. This requires a change in perspective of how others’ words and actions affect our feelings. In NVC, what others say and do is considered the stimulus, but never the cause of feelings. When faced with a negative message from someone else, NVC illuminates four options. To illustrate these options, let’s use the example of criticism, “You’re so selfish”:

- Take it personally: “I really am selfish...”
- Fight back: “I’m not selfish; you’re selfish!”
- Consider your own feelings and needs: Say something like: “When I hear you say that I am selfish, I feel hurt because I need some recognition of the effort I make to consider your preferences.”
- Consider the other person’s feelings and needs: Ask something like: “Are you feeling hurt because you need more consideration for your preferences?”

Needs - Taking the next step, NVC makes the connection between feelings and unmet needs in the individual. These needs are common and fundamental to all human beings. The outer expression of feelings, such as anger and frustration, are seen as indicators of needs, such as love and acceptance, that are unfulfilled. For example, if a moderator of a group is feeling angry at the interactions of some group members, we need to dig deeper and think about what unmet need is causing this feeling. Is the moderator unsupported by the other members? Are they overwhelmed by the responsibilities? Are they not seen and appreciated enough?

Requests - NVC’s final step is to make specific, doable requests for things that enrich the requester’s life. They are made in such a way that it enables the person to respond compassionately to the request. Requests are never demanded. NVC considers demands always to be violent, intimidating, and forceful – the source of many ineffective and unhelpful communication exchanges. Requests in NVC are positive. This means requesting what you want, rather than what you don’t want. An example of this would be saying: “I’d like you to support me in moderating some of the content”, rather than “I don’t want you to ignore me and let me do all the work by myself.”

Managing Conflict in Digital Groups

HOW DO WE MANAGE CONFLICT IN SOCIAL MEDIA GROUPS?

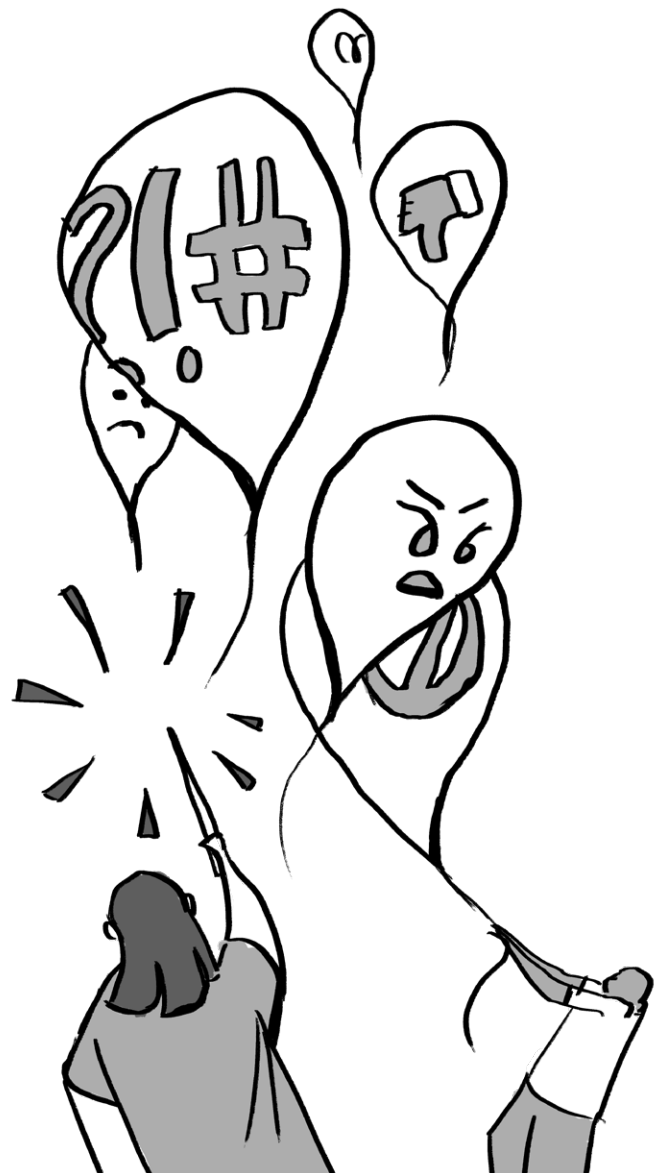
PREVENTING CONFLICT - A QUICK THROWBACK TO GROUP RULES

Many admins say preventing conflict starts with writing great rules, which serve as guidelines for member behaviour. Experienced admins recommend posting these rules long before you think you need them. Clear guidelines are useful for overall group culture when their tone is positive. Describe the behaviours you want to encourage, rather than listing only things members should not do.

REACHING OUT PRIVATELY

Experienced admins recommend privately reaching out to members involved in conflict situations. Taking the time to have a private conversation can make members feel heard and respected. Many admins share that after a private conversation, people have returned to the group as valuable and supportive members. Here is what the digital community stewards can do.

- Get to know the people involved to understand the source of conflict.
- Reach out privately to the people involved in a conflict situation and remind them of the rules.
- Help them understand how they can positively contribute to the group.



MANAGING CONFLICT BETWEEN MEMBERS

From time to time, a social media group may experience conflict between members. Conflict can arise from a difference of opinion, misunderstanding, confusion, or controversy. Respectful disagreement and debate can be part of a healthy community. If a disagreement turns disrespectful or breaks group rules, admin actions can help get the group back on the right track.

When conflict occurs, such as a heated conversation in a thread, experienced admins recommend acting quickly. Get involved before the situation escalates. Speedy action helps reset the tone of the conversation and lets group members know admins care by being present.

- If there is a thread that has gone off track, experienced admins recommend the following:
- Engage with the member in the comments and try to reset the tone of the conversation.
- If needed, restate the rules in the comments of the thread and turn off commenting.
- Leave the thread visible so others can learn what happened.

Some admins temporarily turn on post approvals to calm things down. Once things have cooled down, they'll turn post approvals off again.

If moderation isn't working, consider temporarily muting the member. Allow some time for the discussion (and the member) to cool down. When all else fails, admins should feel empowered to remove members from your group.

Many admins we spoke to told us that they were initially unsure about removing people from the group, but realised it was sometimes necessary. As an admin, you uphold the culture and rules of your community. Members appreciate your moderation, including removing other members who aren't following your community's rules.

In extreme cases, such as something that goes against Community Standards (nudity, hate speech, or threats of violence), you or someone on your team can report the post to the platform.

It is unlikely that admins can watch your group all the time. Experienced admins recommend enlisting members' help by asking them to report heated conversations to an admin so they can take appropriate action.

MANAGING DIFFICULT MEMBERS

A little bit of conflict is inevitable (even healthy) in most groups but managing conflict can be especially challenging when you have a bad actor in your group. Often this can be prevented by establishing great rules and screening new members carefully, but sometimes you'll need to take action against group members who are causing conflict. This could mean working with your team or other group members to resolve the issue, or in some cases, it may be necessary to remove someone from the group.

MANAGING CONFLICT AND NON-VIOLENT COMMUNICATION

A tool that is highly useful when practising NVC - especially with group members who might disagree with admins - is the Change Conversation Pyramid.

- **Comfort** - Make them feel safe enough to talk with you
- **Connection** - Earn their trust so they will take risks
- **Comprehension** - Learn their point of view, so they feel heard
- **Compassion** - Show you care so they will listen to your perspective
- **Cognition** - Gently encourage rethinking so they can update their beliefs

Here are some digital communication tools that will allow you to handle this situation:

1. **Work to prevent conflict before it happens.**

No group is conflict-proof, but if you are proactive, you can work to make sure people know what's allowed and that when conflicts do arise, they're solved in a consistent way. The first step is having clear rules that are very visible in your group. Also, consider keeping a list of group members you're concerned about among your team so you can watch for problem behaviour.

2. **Recognize problems before they get worse.**

Once you have a foundation of rules, the next step is working with your community and your admin team to keep conflict from escalating. Encourage your group members to report bad actors to you or your team. This will allow you to get involved early and contain most problems.

3. **Understand their point of view and diffuse the situation.**

Reach out to members who have broken the rules quickly. One good tactic is to contact these members privately, using chat or even a call, to remind them of the rules. Sometimes a simple misunderstanding can escalate because members feel cornered or ganged up on. Use active listening and try to get to the core feelings behind their statements. Repeat their feelings back to them so you can get to an understanding, i.e., "what I hear you saying...".

4. **Work with your team.**

Your team is there to back you up when things get tricky, so don't forget to use them. When dealing with a bad actor, alert your team as soon as you can. That way, they know that you're on top of it, and they can be there to offer support.

5. **Mute or remove someone and inform them about the rules they have violated.**

Sometimes, if someone is acting out in your group, they just need a little time to calm down. You can use the 'mute' feature to temporarily stop someone from posting or commenting in the group. You'll be able to set the duration of time they're muted, and you can specify to them which group rule they broke. If muting doesn't do it, you can remove them from the group.

6. **Address the situation.**

For tough situations that have escalated in your group, you can make sure everyone understands the actions you've taken by addressing them directly. You can clear the air and control any rumours by going Live, posting, or leaving a comment for the group. Be prepared to answer questions.

Here are some technical tools on Facebook Groups that will allow you to manage bad actors:

1. Post approvals

Post approvals will allow you and your moderation team to screen incoming posts. This is a great way to proactively manage the content that's shared in your group.

2. Maintain an admin activity log

Admin activity log helps you to keep track of admin and moderator activity. Here you can filter and view notes on different actions taken by your team.

3. Set up member questions

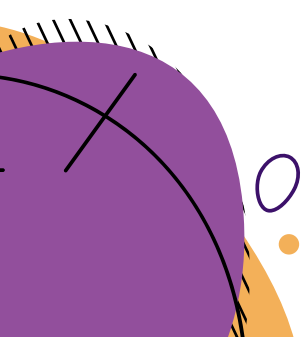
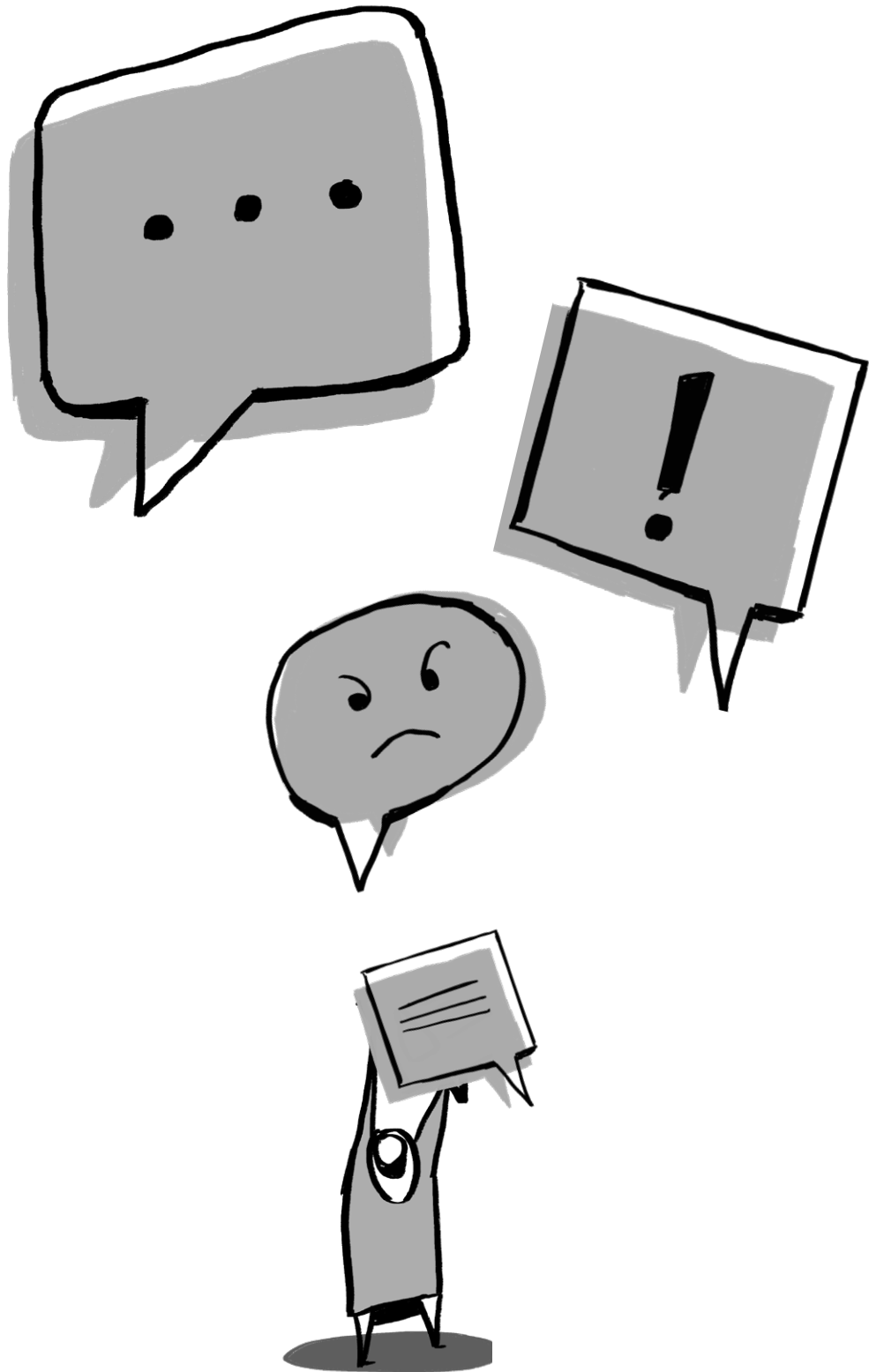
Setting up member questions can help you better screen incoming members to your group.

4. Turn on keyword alerts

If you know that a certain language is banned or indicative of conflict, you can flag those keywords and get alerted when they come up in your group.

Dealing with bad actors can be a drain on your time and emotional energy. You can often de-escalate a situation by approaching it proactively and with empathy. But at the end of the day, if someone isn't right for your group, you should always feel free to remove them.





MODULE SIX:

Digital Safety

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



Why is Digital Safety Important for Online Communities?

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.

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 criminalized, individuals might feel safer in a digital community than in a physical one. As these communities, and the individuals who belong in these communities, are already vulnerable and at risk, it is crucial that these online communities are safe and well protected.

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 the 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 on 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.

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.

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.

What are the digital safety risks and challenges faced by online ?

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 disincentivizes participation. The following are key risks and challenges you need to know about.

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)."

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.

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).





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.

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 is common among many group leaders. Some common concerns shared by admins of social media groups around the world include "They can just literally wake up and say, 'OK, this isn't going to happen anymore'", "Two years of work and contacts would be lost" and 'Alright, you can't have a group anymore, or you have to start paying'.

PRIVACY

Information collection and sharing may now be so pervasive that privacy concerns are diminishing as a deterrent 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).

Responding to Digital Security Risks and Challenges faced by Online Communities

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.

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.

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 tell us there are different reasons to use post approvals:

- 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's 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.

CREATING 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 tell us 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 about how they feel they might contribute to the 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.

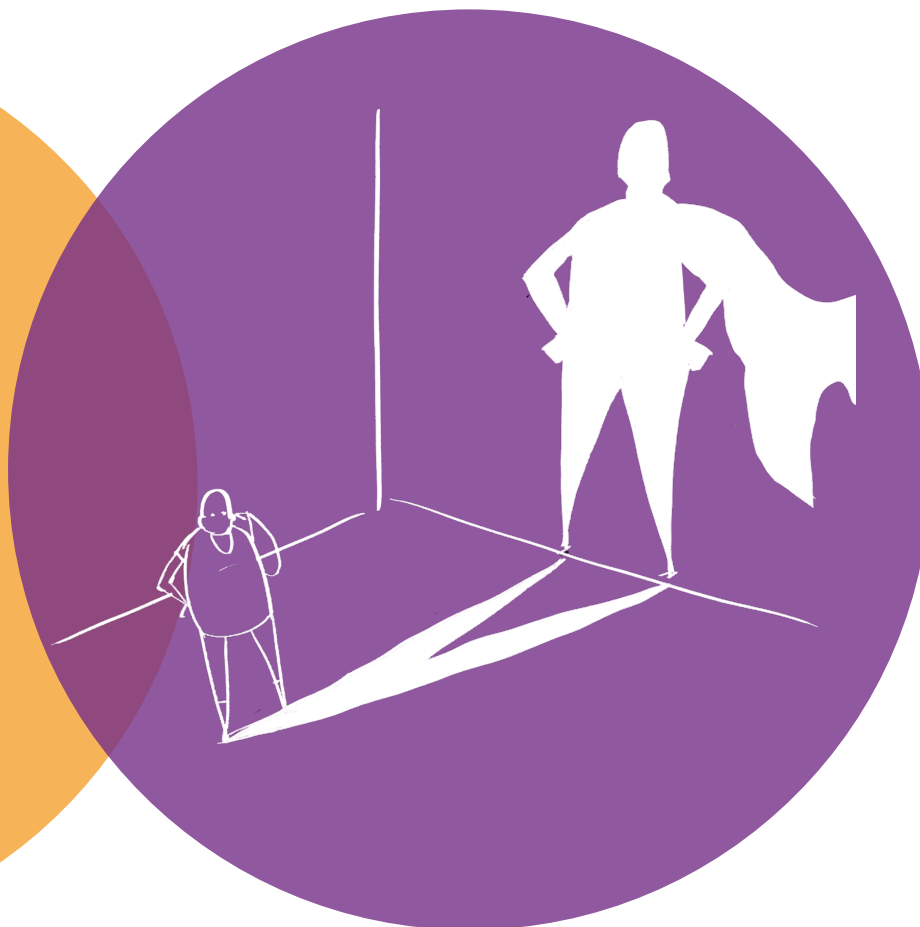
Once your group has grown, and you're receiving numerous membership requests, you may find using membership request filters helpful. With membership request filtering, you can select and sort predefined filters to apply to pending requests. These can be accepted or declined individually or in bulk. The 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.

MODULE SEVEN:

Digital Leadership

Module Objective: Understand how to exercise multi-partiality and practice leadership in digital environments

Module Dilemma: I don't have the leadership skills and confidence to facilitate my group



Who is a Digital Community Steward - An Exploration

HOW DO DIGITAL COMMUNITY STEWARDS EMERGE?

1. Research shows that many group leaders are accidental leaders. They had little idea of how the leadership of their group would unfold when they took on the role. The job is complex and time-consuming and can impose intense pressures for which many leaders are not trained or prepared. For all the focus on leadership in contemporary society, online leadership is an emerging phenomenon that has not yet been well studied or understood.
2. When asked why they lead a group on Facebook, many group leaders said that their main reason for operating on the platform was the extraordinary global reach of Facebook and the fact that the platform and its tools were accessible, easy to use, and free of charge. Many leaders we interviewed also said that they derived significant personal satisfaction and a sense of achievement from their work.



WHAT MAKES A DIGITAL COMMUNITY STEWARD?

In a research study conducted by Governance Lab in 2020, where the responses of over 15,000 social media users were surveyed, the following were identified as the three most important traits in a leader:

- Welcoming differences of opinion among members
- Being visible and communicating well
- Acting ethically at all times

How to be an Efficient and Effective Digital Community Steward - Key Strategies

CHOOSE MULTI-PARTIALITY OVER IMPARTIALITY.

Multi-partiality—developed as a facilitation technique—seeks to level power disparities between individuals by acknowledging the social identities and inequities present within a space. It asks participants to consider how their perspectives are influenced by their social identities, and how these identities influence group dynamics; this can look like facilitators naming who seems to feel most comfortable contributing to the conversation or what perspectives are missing from the

space. It also creates room for facilitators to share their own identities and experiences, something that is often deemed inappropriate in schools.

Overall, multi-partiality falls between impartiality and bias; it does not favour one opinion over another, nor does it completely remove the facilitator from the conversation.

DON'T JUST FACILITATE. REMEMBER TO PARTICIPATE.

Members appreciate it when you are involved as a fellow member of the group, not just as an admin. Being social and interacting with members lets them get to know you and helps build trust. Once you have established this trust, it's easier for members to feel comfortable coming to you with any questions or participating in the group themselves.

As an active admin, you can help encourage a positive group culture by modelling the type of behaviour you want to see in the group. Experienced admins tell us that being active also allows you to notice positive and negative trends in the group. If you notice group members reacting positively to a post, you may want to encourage more of the same.

Admins tell us members like to be recognized, even if it's just with a like or a short comment. Give members positive feedback by actively commenting and reacting to their posts. Acknowledge members and let them know you care.

Posting regularly once or twice a week (or more) lets your members get to know you. Depending on the platform you moderate, consider scheduling posts in advance so you don't need to be online around the clock.



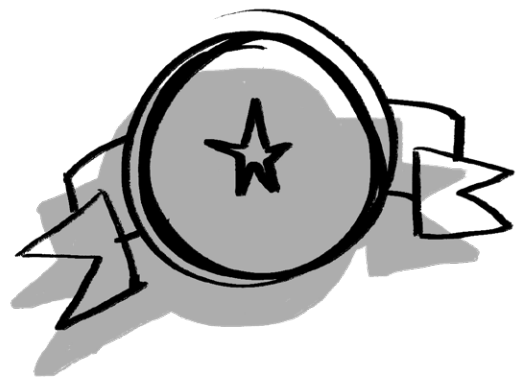
How to be an efficient and effective digital community steward

RELY ON YOUR TEAM FOR HELP

Appoint moderators or admins so that each person can take a day off whenever they're feeling overwhelmed or a little stressed out.

Experienced admins tell us running a successful group takes time and effort. Communities can break apart when an admin becomes drained. When you feel overwhelmed, turn to your team members for help. Set up a schedule, divide responsibilities, and include days off for yourself and your team members. Trust your team to step in when you need to take a break.

Spend time training your team both to help manage the group and uphold the group's culture. Feeling comfortable with your team's decision-making abilities and commitment allows you to take needed breaks from your group.



FIND MODERATORS AND ADMINS

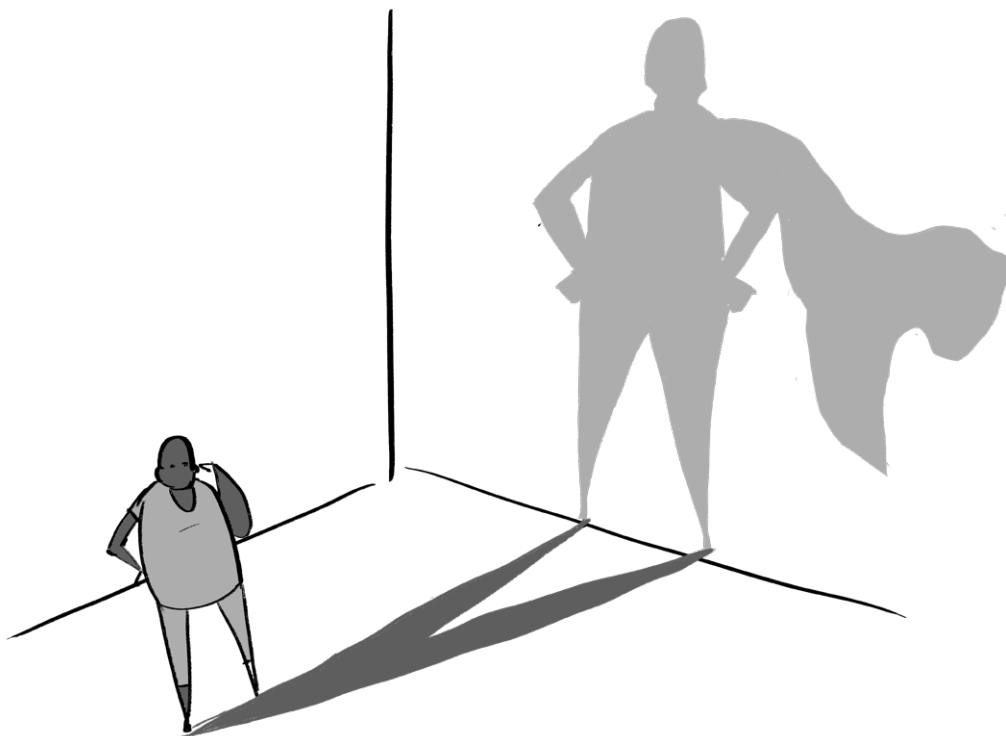
As your community grows, you may find yourself spending more and more time moderating your group. If you feel this way, you're not alone. It may be time to enlist some people to help manage your group.

You may find that certain members already help manage the group. Those members might post great content, comment on other posts or tag you in discussions that need moderation. They understand the culture of the group and feel like a part of the community. Admins we spoke to told us that the best moderators tend to reveal themselves in this manner, so reach out to them and see if they want to take on a more official role.

You can promote any member of your group to one of two roles: admin or moderator. Moderators have fewer permissions than admins, so they can help take some of the work off your plate without giving them more control than you're comfortable with.

Introduce new moderators or admins in a post, so members can get to know them personally and know who to tag in comment threads that need moderation. Also - admin and moderator badges help members quickly and easily identify these leaders. Ideally, your moderators can monitor content and member activity while you're away from your group.

How to be an Efficient and Effective Digital Community Steward - Rely on your Team



UNDERSTAND WHEN YOUR FEEL OVERWHELMED

Do you ever feel like the weight of the world is on your shoulders and that all you can think about is putting out fires in your group? It can actually be exciting to rise to that challenge, but that's also when you can start to burn out. Learn to recognise burnout and when you feel it, know when to take a break.

There are many apps that can help us keep track of our moods and emotions. Mood Meter is one such app you can look into.

WALK AWAY

When you feel anger, or tension or are consistently drawn back into a problem, the first thing you can do is get up, walk away and turn off your notifications. Give yourself some time to disconnect and take a step back. It might be hard to walk away from a problem when it feels like you're the only one who can solve it but looking after yourself is always worth it. The stress can wait.

IDENTIFY AND ADDRESS SENSITIVITIES

Notice the pattern. Do you feel overwhelmed with member requests? Are members messaging you? Are you dealing with conflict in your group?

Learn your triggers, sit with them and process them. You may be able to see a clear source behind your anxious or uneasy feelings. Identify which tasks you might need more help dealing with as well as which issues you may want to avoid entirely.

SET CLEAR EXPECTATIONS ABOUT YOUR AVAILABILITY

Do a Live with your group or have a pinned message to set expectations and limitations. This helps build empathy with your community and models behaviour for your own admin team.

Schedule your posts in advance to make your time off count.

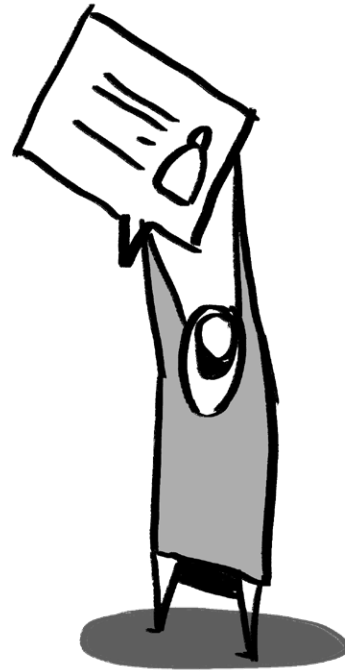
Create office hours that work for you and your team. No one should be expected to be on 24/7.

PRACTICE SELF-CARE

When you feel yourself getting overwhelmed, think about the small things you can do to care for yourself. Maybe it's making yourself a cup of tea, going for a walk, playing with your pet or calling a friend. Sometimes it's the little things that comfort you and that can help you to let go of what's stressing you out. Getting offline is usually a great first step.

BUILD A TRUSTWORTHY TEAM WITH SKILLS AND SENSITIVITIES THAT COMPLEMENT YOURS

Empower your team to make decisions on your behalf. Consider an admin team located in different locales for better coverage day and night. When choosing team members, think about what they can do that you can't. Having a strong network of skills and specialities will help you to distribute responsibilities and save your team from burnout.



MODULE EIGHT:

Growth & Inclusivity

Module Objective: Identify strategies to increase the size of the group while encouraging the group to be more inclusive

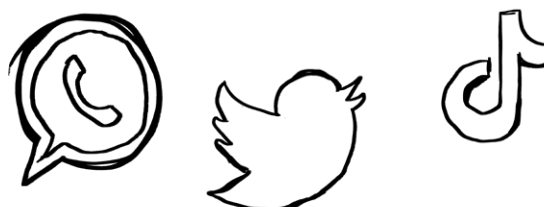
Module Dilemma: I want to expand the numbers and diversity of my group



Expanding your group - Finding new members

CROSS PROMOTING YOUR GROUP

A great way to gain engagement, spread awareness, and attract new members is to cross-promote your group. This could mean building relationships with similar communities on Facebook or expanding your reach to other platforms.



BUILD A NETWORK

Groups can help each other in a big way. Admins can share advice, resources, content, and even audiences – so building a network of like-minded group admins can put you way ahead of the curve.

First, consider where your community fits in the bigger ecosystem of groups. Are there other groups whose members and interests overlap with yours? Look around to see if others run similar or related groups and reach out. It can only help your community (and theirs)!

Connecting with people who are influential in the same space is an easy way to expand your audience. Most admins are eager to make new friends and sharing each other's posts and events is one way to give your members content they'll love while amplifying your own message. You might not have an idea of how to work together right away. Just reach out and introduce yourself!

Here the trainers can ask the participants if any of them collaborate with other groups with similar objectives. Participants can share their experiences and discuss the pros and cons of this strategy.

The trainers can also introduce the participants to opportunities that will help them access other group admins. This includes programs such as: Community Connect, Power Admin Groups, Learning Labs, and Accelerator.

PROMOTE YOUR GROUP ON OTHER PLATFORMS

Many admins use platforms such as TikTok and Instagram to cross promote their group.

Instagram can be a new way to engage potential members with its own hashtags, influencers, and tools. Starting an Instagram for your community is another exciting way to expand your group's reach. The same goes for TikTok! It is important to diversify the platforms we use since it will help us diversify our members too.

See if there's someone in your community who is savvy about using these platforms and ask for their help setting up a profile for your group! If you have photos or other awesome visuals from your group, share them – just make sure you get proper permission before reposting.

Be sure to add a call to action and link to your group from your new profiles – it could be something like “To join us in our group, hit the link in bio!” You can also set up takeovers from community members around the world or “influencers” in your space.

CREATE A FACEBOOK PAGE TO USE AS THE PUBLIC FACE OF YOUR GROUP

Some admins recommend creating a Facebook Page to help promote your group, especially if your group is a private group. These admins use a Page to spread the word about their group and give prospective members a sense of the group's culture. Examples of what you could share on a Page include upcoming public events your group has organised, achievements like successful fundraisers or articles related to your group's subject matter.

A Page can also help you vet potential members. People who have seen your Page before joining your group might have a better sense of whether the group is a good fit for them. By reading people's comments on your Page, you may be able to tell which folks are a good fit for your group, too. You can link the Page to your group so people can easily request to join, or you can proactively invite new members you've seen commenting on the Page.

If you have multiple groups, you can link them all and use your page as a hub. Moreover, using a page also means that you can boost your posts or create ads on both Facebook and Instagram - which is an effective way to get your community seen across the platforms.

BRANDING YOUR COMMUNITY

BRAND YOUR GROUP VISUALLY

A great way you can reinforce your brand and create a familiar space for members is with visuals. Use a consistent colour palette, choose cover photos that reflect your purpose, and think about your logo. Logos can represent your members, the community's purpose, and how you want to express your group mission visually to the world.

Some groups change their cover photo each month based on community suggestions, polls, contests, or submissions. When in doubt, ask your members! How do they view the community? Giving options instead of making them open-ended will help to provide structure.

BE UNIQUE AND CREATIVE

Everybody wants to be a part of something special. Cultivating inside jokes, catchphrases, and nicknames can reinforce a sense of community and culture.

Give your community members a nickname that they can connect with and reinforce it in your posts. Make them feel like they're part of a community, but don't force it – keep it authentic to your brand and involve the community as much as possible.

Example from an admin: "In my community, we refer to each other as GGIs (Girls Gone International) - this brings us together around a shared identity."

Having fun with symbols, emojis and catchphrases can also bring members together and help them participate. Start by thinking of an image or symbol that represents your community and encourage members to use it by using it in your posts.

Example: "In Girl Gone International, our symbol is a woman with a suitcase. We use that imagery in our logo and also think of emojis that might capture our community's brand (we often use a world emoji or female dancer to show our personality as a brand)."



DEFINE YOUR MISSION

Every group has a purpose – a reason for existing. By writing a formal mission statement, you decide your purpose and what you and your members stand for. You can post your mission publicly or use it within your team to guide whether a decision is right for the group. Even if you don't write it down at all, it's useful to think about! To create a mission statement, take a few minutes and write down what your group is all about. Why did you create it? What do you hope people will use it for? What do you hope it will give people? Now write that down.

One way to create your mission statement is to use the following template:

"[YOUR COMMUNITY'S NAME] is a group for [WHO ARE YOUR MEMBERS] that [WHAT YOU DO] and want to [YOUR VISION]."

An example of a mission from a group admin:

"Girl Gone International is a group of women, living, loving, and traveling abroad that want to connect with globally minded international women around the world online and face-to-face."

MAKE YOUR GROUP EASY FOR PEOPLE TO FIND WITH KEYWORDS AND TAGS

Admins recommend using popular words related to your group's subject matter in your group's name since potential group members might search for those common terms. For example, a parenting group might want to include “parents,” “moms” or “dads” in their group name. This will help interested people find your group when searching for those words on Facebook.

Here the trainers can ask the participants if any of them have changed the group name since they established the group. If yes, the participants can share their experience regarding this decision: What was the old name? What's the new name? Did this change make the group more visible and accessible?



Note: Consider using group tags to make your group easier to find. Location tags help people search for and discover groups that are relevant to their area. Trainers can also do a quick demonstration to show how to activate group and location tags on Facebook.

USE MARKETING AND ADVERTISING TO GROUP YOUR GROUP

Advertise your group offline with unique pieces of merchandise. Create t-shirts, cups, stickers, or other items and include your group's name and information about the group. Some admins reward top contributors by offering a t-shirt or sticker as a surprise gift.

Some admins leave business cards, flyers, or other items with interested people at local events, trade shows, or conventions. Wear a shirt with your group's name, logo, and other information when attending events as a conversation starter. You can encourage people to join your group right away or have a sign-up sheet and contact interested people later.

Reach out to businesses related to your group's theme to see if they'd be interested in partnering with your group.

A QUICK NOTE ON ORGANIC GROWTH



It may seem challenging to find new members for your group at first. Many successful groups started out small and gradually increased their membership.

If you are starting a group about a topic you are passionate about, you might already know some people who share that passion. Get them involved early and encourage them to actively participate and invite others to the group.

Many experienced admins told us that members who really like your group, often invite other interested people to join. If those new members also invite people, your group can grow pretty fast. All the admins we have spoken to told us that word-of-mouth is a great way to grow your group.

To enable word-of-mouth growth, you'll need to encourage the creation of great posts and foster a sense of community. The type of content created, and the value people receive from the group will vary depending on the type of group you run.

Public, content-driven groups can grow by encouraging members to post great content in the group. Members will share popular group posts to News Feed to share with their friends. Those friends can request to join the group to see more content.

Smaller, closed groups may not have posts shared with others outside the group. The value of this kind of group is in its supportive community.

Creating a welcoming, positive community empowers members to want to get the word out about their experience, whether it be to let people know about the group's great content, its supportive community, or both.

Expanding your group - diversity and inclusion

DESIGN JUSTICE

Design justice rethinks design processes, centres people who are normally marginalised by design, and uses collaborative, creative practices to address the deepest challenges our communities face. It is an approach to design that is led by marginalised communities, and that aims explicitly to challenge, rather than reproduce structural inequalities. It has emerged from a growing

community of designers in various fields who work closely with social movements and community-based organisations around the world. Design justice takes the concepts of diversity, equity, and inclusion a step further when designing tech solutions, physical products but also social movements, and community development endeavours.

The following are some key principles design justice focuses on when developing and designing solutions for communities. It is highly important for digital community stewards to be aware of and practise the following principles when designing solutions and introducing strategies to their online communities:

- We use design to **sustain, heal, and empower** our communities, as well as to seek liberation from exploitative and oppressive systems.
- We **center the voices of those who are directly** impacted by the outcomes of the design process.
- We **prioritize design's impact** on the community over the intentions of the designer.
- We view **change as emergent from an accountable, accessible, and collaborative process**, rather than as a point at the end of a process.
- We see the role of the **designer as a facilitator rather than an expert**.
- We believe that **everyone is an expert based on their own lived experience** and that we all have unique and brilliant contributions to bring to a design process.
- We share **design knowledge and tools** with our communities.
- We work towards **sustainable, community-led, and -controlled** outcomes.
- We work towards **non-exploitative solutions** that reconnect us to the earth and to each other.
- Before seeking new design solutions, **we look for what is already working** at the community level. We honor and uplift traditional, indigenous, and local knowledge and practices.

CREATING AN INCLUSIVE COMMUNITY

An inclusive community is one that champions members from all backgrounds, provides a safe space for them to share their experiences and encourages them to learn from one another.

Remember, diversity can come through no matter what your group is about - even if it is a female-only community, hyper-local, or already a community where diverse identities come together, there are many ways that your members differ. Think about who your members are and the intersectionality they represent. Some examples are gender identity, sexual orientation, ethnicity, or socioeconomic status.

Here are the main ways through which we can create an inclusive community.

MAKE INCLUSION AND DIVERSITY A PART OF YOUR VALUES AND STATEMENTS

Your group values are the foundation of your community. This is where you express your group's purpose, what you value and what you and your members stand for. Having strong values can help guide your members and shape the culture of your community.

When considering how your community supports diversity and inclusion, you can start by looking at how you express this in your group statements. Think about the different places where you share your values and purpose - this could be your group description, group rules, in boarding materials and anywhere you speak to members.

When describing who your members are or what your group does, you can add in relevant ways that you support diverse voices. The trainers can ask the participants to revisit their mission statements and group rules to see if these reflect the values of diversity and inclusion.

MAKING YOUR BRANDING AND IMAGERY INCLUSIVE

Visually representing your community is important to communicate your mission and who your members are. Showcasing the diversity of members in your community's cover photo, logo, and across all of your group's imagery can help communicate who your community is.

Some ways you can do this are through:

- **Illustrations** - Illustrations are often the easiest way to represent your members and community's diversity. Look inside your community to see if there are artists or if they have referrals or recommendations. You can also check out Canva or Adobe Stock for illustrations you can use in your designs.
- **Photography** - Photos can help to show real members of your community. You can ask members to submit photos of themselves and other members or use images from events. Make sure to always get permission from everyone in the photo before you share it!
- **Text and other graphics** - There are many more ways to represent your community's members visually. You can create maps to show where members are from, showcase imagery that is representative of their backgrounds or the languages they speak, and more.

You can even rotate your community's imagery every month and ask members to share important dates to celebrate, like Pride Month, Black History Month, International Women's Day, Indigenous Peoples Day - the list goes on! You and your community can use imagery to celebrate your members' diversity and learn how they want to be represented.

Here are some tools you can use to implement the above strategies:

- **Cover photo** - Your cover photo is a great place to highlight your group's brand. Choose an image or logo that represents your community.
- **Messenger** - Messenger can help to keep your team connected. You can use chat to escalate issues, share best practices, and stories from the community, and work together to build your community.
- **Messenger Rooms** - Messenger Rooms is a great way to get your team together to talk through important topics to your community.

PROMOTING INCLUSION IN YOUR COMMUNITY

You have a thriving community filled with diverse voices, perspectives, experiences, stories, learnings, and so much more. Sharing the voices of your community and creating space for members to feel seen and be referred to in the proper way can provide insight to many of your members and strengthen their bonds.

Here are the main ways through which you can promote inclusivity within your community.

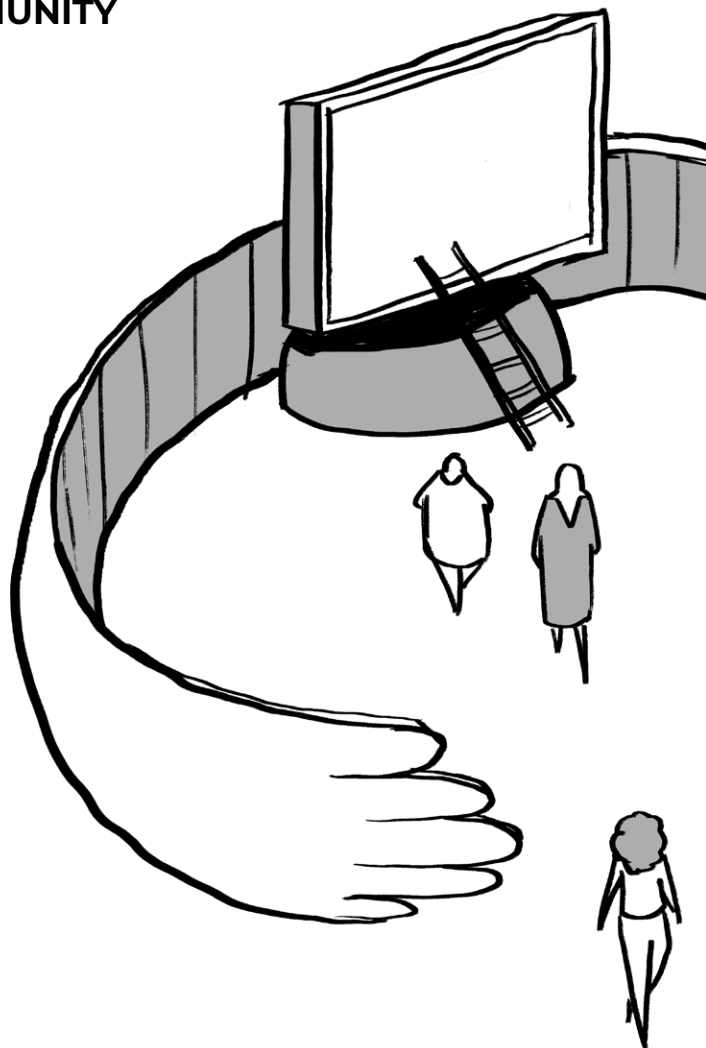
ASK YOUR MEMBERS TO SELF-IDENTITY BY SHARING PRONOUNS

As an optional membership question, you can ask members to self-identify by sharing their pronouns (ex. she/her, they/them). This can help set the tone in your community for everyone to have themselves be correctly addressed. You can include this in your Welcome Post with new members and set the tone by sharing more about yourself, this could be where you are from, your pronouns, and other information you are comfortable sharing.

Again, this is an optional opportunity for your members to share more about themselves if they chose, so representing that in your question is important (ex. “Optional: Would you like to let us know your pronouns?”). Keep in mind that this is very personal to people so making it optional with the expectation that people may or may not want to share is important. Ensuring that your community members are in a safe space and will be protected is also key, again you want to make sure you are setting a tone for inclusion and not exposing your members to harassment.

HIGHLIGHT DIVERSE STORIES FROM WITHIN YOUR COMMUNITY

Learning more about your members and sharing it with the rest of the community is a great way to create bonds and help members find others that they identify with. Highlighting diverse stories from within your community i.e. “Spotlight/story/member of the week” posts, can bring your members closer together and get them excited to learn more about each other.



Think about members who are of different races and ethnicities, sexual orientations, genders, ages, socioeconomic backgrounds, or people with disabilities. Ask members to nominate each other or check your “Top Contributors” in your “Insights” tool to get started! At the end of your spotlight, you can ask, “Who would you like to nominate for the next spotlight?” This can also include your community’s allies.

Here are some tools you can use to implement the above strategies:

- **Member questions** - Setting up member questions can give your members the chance to share their pronouns and helps set the tone of your community.
- **Welcome post** - Welcome posts can help you set the tone for your members, sharing their pronouns and creating a space where members can connect.

PROVIDING SAFE SPACES FOR DIVERSE MEMBERS OF YOUR COMMUNITY

Your community is a place where people are coming together to share, build and have conversations. As with every community, it is important to have all voices represented and heard. As a community leader, you can help create a place where your members feel safe to share their experiences openly. Here are the main ways through which we can provide safe spaces for the members of your community.

CREATING STRONG RULES WITH STRICT MODERATION

As a community leader, you know the importance of rules and guidelines for your community. Having strong rules against hate speech, misinformation and discrimination are clear but sometimes they take form in other ways.

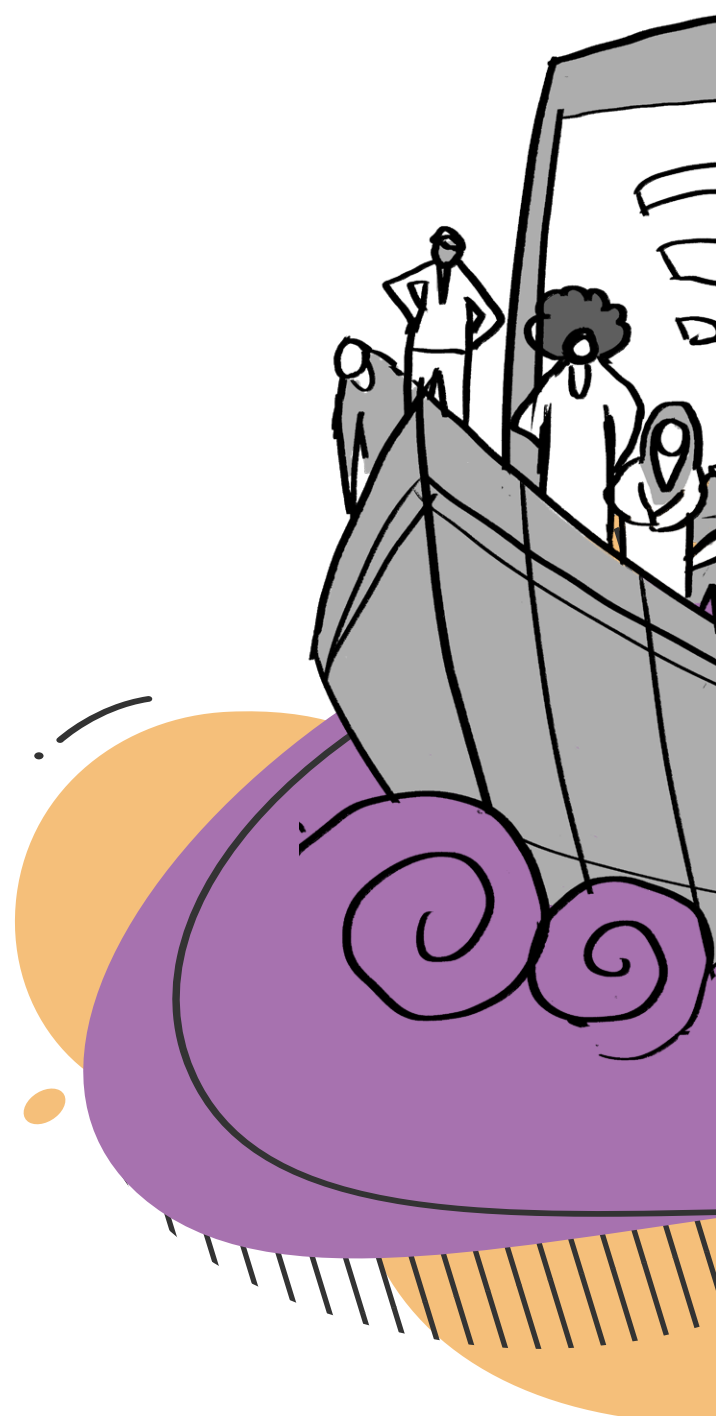
For example, “All Lives Matter.” This seems like an innocuous way of expressing that everyone should be considered when talking about rights. But this is disrespectful to members of the Black community. Proclaiming that “All Lives Matter” negates the “Black Lives Matter” movement and experience of Black people. There are many ways that this can be explained to community members, and resources can be shared with infographics, analogies, etc. Negating the experience of others is not providing a safe space for those members of your community.

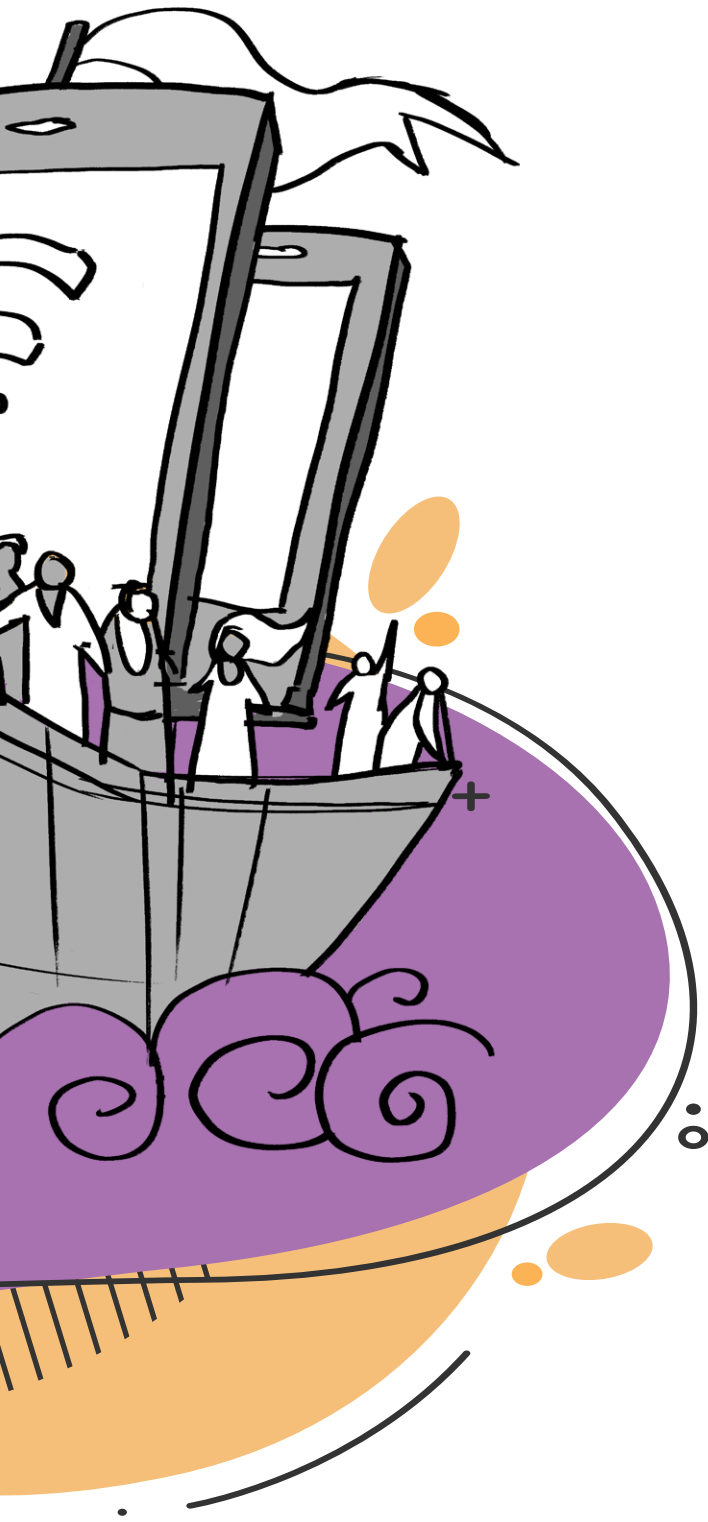
Strong rule enforcement for inflammatory language is important when creating a safe space for important conversations. As we approach things with empathy, it can be challenging to strike a balance between what is an “education” moment and what is a “delete and ban this person” moment. Admins often employ a strike system for members, automatic deletion, muting of members, and/or posting themselves to address certain violations.

LISTEN TO YOUR MEMBERS THROUGH CONTINUOUS FEEDBACK

Listen to your members! How can you improve? How can you create a better community for everyone to connect? How can you make your community the best possible community for your members?

Send out surveys! There are many, simple, anonymous survey platforms (such as Survey Monkey and Google Forms) you can find to ask your members about themselves, what you can improve, what issues can be addressed, what rules can be better enforced, and what changes to your mission can be made. Make it a point to do these often - some admins do them every three-six months (ex. mid-year, end of year). You can even do these around major cultural moments (ex. BLM movement, COVID-19, elections) to understand how you can address the needs of your members through these times.





When discussing topics such as inclusivity and safety, admins need to be self-aware of the power they hold and exercise over these online communities. This is why listening and reflecting are essential to make members feel valued and heard. As a community leader and member, you are there to listen as much as you are there to share. Being an active member of a community is often about learning, listening, and implementing, and sharing what you've learned. Remember that member feedback and active listening are crucial in improving trust and overall group dynamics!

EDUCATE YOURSELF AND PROVIDE MENTORING AND SUPPORT WHEN POSSIBLE

Helping to educate yourself and others with reputable resources on topics like anti-racism, discrimination against minorities, the history of oppressed groups, immigration policies, etc., is the first step toward creating a healthy dialogue. Use your voice to share information, resources, organisations, and places to donate, and educate others on how they can do the same. Providing a safe space for fellow community members to interact and share their diverse experiences is an important part of leading and interacting in communities. Lifting the voices of your fellow community members and calling out discriminatory and disrespectful behaviour is our obligation as community members and allies.

Many communities are dedicated to a certain interest, identity or stage of life. Working with other communities (whether within the same ecosystem or on a different topic) can help to bring a richer experience to you and your members. Reach out to other community leaders or influential people in your space and ask them to join your community for a panel, interview or more! Remember, this is a collaboration, so be open and frame your ask in a way that is reciprocal. Creating bonds with other community leaders can help both of your communities.

Here are some tools you can use to implement the above strategies:

- **Rules enforcement** - Rules enforcement lets you inform your members of the rule they violated. This can be done through Admin Tools when you are muting a member or approving posts.
- **Turn off commenting** - Sometimes, a post's comments can become unproductive. You can turn off the comments to prevent additional comments to the thread. It also helps to let your community know why you are turning off the comments.
- **Keyword alerts** - When you want to keep an eye out for when certain words or phrases are used in your group, or quickly receive notifications about it, you can set up Keyword Alerts using these phrases, admin team names, or hashtags that you use for members to self-moderate.
- **Polls** - Polling your members is a great way to learn what matters most to your community and get suggestions. For example, you can poll your members on how you can help support diverse businesses or ask for feedback or ideas on how to continue to support diverse voices.

MODULE NINE:

Digital Rights

Module Objective: Understand and reinforce digital rights to respect freedom of expression and discourage hate and harassment

Module Dilemma: I want to respect freedom of expression, but the type of hate I'm seeing in my group seems to surpass the limit of acceptability



Community Standards - What do we need to know?

WHAT ARE COMMUNITY STANDARDS?

THE KEY VALUES

The 22 community standards of Facebook, which the trainers will discuss shortly, are based on feedback from people and the advice of experts in fields such as technology, public safety and human rights. Furthermore, to ensure that everyone's voice is valued, these standards include different views and beliefs, especially from people and communities that might otherwise be overlooked or marginalised.

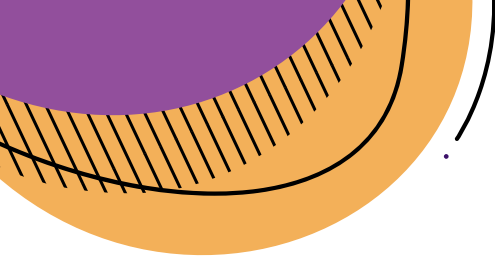
The goal of these community standards is to create a place for expression and give people a voice. Meta wants people to be able to talk openly about the issues that matter to them, even if some may disagree or find them objectionable. In some cases, they allow content – which would otherwise go against these standards – if it's newsworthy and in the public interest. But they only do this only after weighing the public interest value against the risk of harm and looking to international human rights standards to make these judgments.

However, it is important to recognise that the internet creates new and increased opportunities for abuse. For these reasons, when these platforms limit expression, they do it in service of four values. The trainers will introduce the following four values, which frame the community standards.

- **Authenticity** - We want to make sure that the content people see on Facebook is authentic. We believe that authenticity creates a better environment for sharing, and that's why we don't want people using Facebook to misrepresent who they are or what they're doing.
- **Safety** - We're committed to making Facebook a safe place. We remove content that could contribute to a risk of harm to the physical security of persons. Content that threatens people has the potential to intimidate, exclude or silence others and isn't allowed on Facebook.



- **Privacy** - We're committed to protecting personal privacy and information. Privacy gives people the freedom to be themselves, choose how and when to share on Facebook, and connect more easily.
- **Dignity** - We believe that all people are equal in dignity and rights. We expect that people will respect the dignity of others and not harass or degrade others.



WHAT ARE THE COMMUNITY STANDARDS?

For successful moderation of content, it is crucial for digital community stewards to be aware of the community standards which guide moderators to understand what content is acceptable and what is not.



The following are the 22 community standards of Facebook.

Violence and Criminal Behavior

1. Violence and incitement
2. Dangerous individuals and organisations
3. Coordinating harm and promoting crime
4. Restricted goods and services
5. Fraud and deception

Safety

6. Suicide and self-injury
7. Child sexual exploitation, abuse, and nudity
8. Adult sexual exploitation
9. Bullying and harassment
10. Human exploitation
11. Privacy violations

Objectionable Content

12. Hate speech
13. Violent and graphic content
14. Adult nudity and sexual activity
15. Sexual solicitation

Integrity and Authenticity

16. Account integrity and authentic identity
17. Spam
18. Cybersecurity
19. Inauthentic behaviour
20. Misinformation
21. Memorialization

Respecting Intellectual Property

22. Intellectual property

HOW DO WE ENFORCE COMMUNITY STANDARDS IN GROUPS?



HOW DO WE ENFORCE COMMUNITY STANDARDS IN GROUPS?

Why are community standards important for my group?

Being committed to maintaining groups as safe place where meaningful connections happen is paramount. It is important for groups to remain a safe space for people to connect. Groups are proactively monitored for hate speech and calls for violence using a combination of the latest technology and human review.

Posts that violate community standards on things like hate speech are removed, and Facebook removes any groups that consistently violate those standards. This enforcement policy ensures that when reviewing a group to decide whether or not to take it down, we now look at admin and moderator content for violations. This includes posts by members that they have approved.

How can I use it in my group?

These community standards, as demonstrated in the previous activity, need to be taken into account when moderating the group content. If the content posted by group members violates these community standards, the admin should take the necessary steps to remove it. Community standards can also be used in your online

community to help build your rules. The participants should go back to the rules they developed in the beginning of the training and see if the community standards of the platform are reflected in the rules they have created for their group. This will help you identify the connection between the standards for the platform and the standards for your group.

What is group quality?

Group Quality is an overview of content removed in your group for violating certain community standards, including a section for false news found in your group. If a group is in violation of community standards, Facebook may do any of the following:

- Disable the group if there are sufficient admin and moderator violations, which include:
 1. If an admin or moderator creates content (for example: posts, comments, rooms) that violates our Community Standards.
 2. If an admin or moderator approves violating content from a group member.
- Take down the violating content.
- Turn on temporary post approval for members who repeatedly post violating content in the group.
- Show a group's content lower in the News Feed if a group repeatedly shares false news. Facebook may also stop suggesting that people join the group.

With the Group Quality tool, you can see exactly when Facebook removed a piece of content from your group along with which community standard it was violating. And, for most violation types, what the actual content was. This is to provide more context on the decision that was made.

Admins can find this tool in the Admin Tools section of your group, under “Group Quality”. Note that Group Quality is only available to admins, not moderators.

Digital Rights - What do we need to know?



LET'S UNDERSTAND UNIVERSAL DIGITAL RIGHTS

The term digital rights describe the human rights that allow individuals to access, use, create, and publish digital media or to access and use computers, other electronic devices, or communications networks. The term is particularly related to the protection and realisation of existing rights, such as the right to privacy or freedom of expression, in the context of new digital technologies, especially the Internet.

- UDHR - Universal Declaration of Human Rights
- ICCPR - International Covenant on Civil & Political Rights
- ICESCR - International Covenant on Economic, Social & Cultural Rights
- CRC - Convention on the Rights of the Child
- CRPD - Convention on the Rights of Persons with Disabilities

There are multiple international laws, treaties, covenants, and principles that focus on promoting and protecting human rights – which includes digital rights. This includes:

Based on the above and more, let's look at important digital rights and principles that are essential for ensuring the safety and participation of all digital citizens.

The following was obtained from the charter of human rights and principles for the internet developed by the Internet Rights and Principles Coalition.

Set A:

UNIVERSALITY: All humans are born free and equal in dignity and rights, which must be respected, protected, and fulfilled in the online environment.

ACCESSIBILITY: Everyone has an equal right to access and use a secure and open Internet.

NEUTRALITY: Everyone must have uniform access to the Internet's content, free from prioritisation, discrimination, censorship, filtering, or traffic control.

RIGHTS: The Internet is a space for the promotion, protection, and fulfilment of human rights. Everyone has the duty to respect the rights of all others in the online environment.

EXPRESSION: Everyone has the right to hold and express opinions and to seek, receive, and impart information on the Internet without arbitrary interference or surveillance. Everyone has the right to communicate anonymously online.

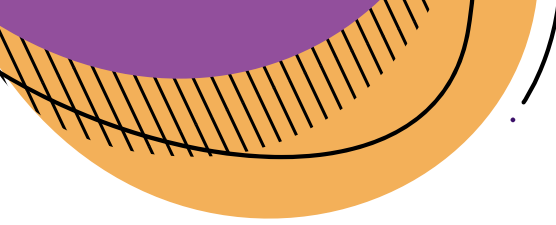
LIFE, LIBERTY, AND SECURITY: The rights to life, liberty, and security must be respected, protected, and fulfilled online. These rights must not be infringed upon or used to infringe other rights in the online environment.

PRIVACY: Everyone has the right to privacy online, free from surveillance, including the right to control how their personal data is collected, used, disclosed, retained, and disposed of.

DIVERSITY: Cultural and linguistic diversity on the Internet must be promoted, and technical and policy innovation should be encouraged to facilitate a diversity of expression.

STANDARDS AND REGULATION: The Internet's architecture shall be based on open standards that facilitate interoperability and inclusion of all for all.

GOVERNANCE: Rights must form the legal and normative foundations upon which the Internet operates and is governed. This shall happen in a transparent and multilateral manner, based on principles of openness, inclusive participation, and accountability as prescribed by law.



Set B: Another way to think about digital rights can be clarified through the descriptions below:

Access and nondiscrimination - Access to the Internet enables you to exercise your human rights. As a general principle, you should not be disconnected from the Internet against your will, except when it is decided by a court. Moreover, Internet access should be affordable and non-discriminatory.

Freedom of expression and information - You are free to express yourself online and to access information and opinions, including those that may offend, shock or disturb, whilst respecting others' reputations and privacy.

Freedom of assembly, association and participation - You have the freedom to use any website, application, or another service to associate with your peers. You also enjoy the right to protest peacefully online. However, you should be aware that you may face legal consequences if online protest leads to blockages, disruption of services or damage to the property of others.

Privacy and data protection - Your personal data should only be processed with your consent or if it is laid down by law. You should be informed if your personal data is processed or transferred to other parties and when by whom, and for what purpose.

Education and literacy - You should have online access to education and knowledge in order to exercise your rights and freedoms on the internet.

Decolonising Digital Rights

A QUICK NOTE ON DECOLONISING DIGITAL RIGHTS

The objective of this is to help digital community stewards understand that digital technologies have the potential not only to reproduce but also to amplify existing forms of oppression, such as racism, sexism, ableism, homophobia, and transphobia.

The growing use and deployment of digital technologies have the potential to affect almost every aspect of our lives, as they become involved in everything from hiring processes to the operation of the “welfare state” and the criminal justice system.

The digital rights field exists to promote and protect rights and freedoms in the digital sphere. In order to do

so, it is crucial that the field reflects the society it works to safeguard. Here, the field must do better and ensure there are no blind spots in our work so that the digital rights of marginalised groups are upheld.

By a decolonising process, we mean a process that acknowledges that these forms of oppression have their roots in a history of domination and colonisation and are maintained by structural forces. Our goal is to initiate a process that challenges the structural causes of oppression in order to work towards a digital rights field in which all groups in society have their voices heard and which works to protect the digital rights of all.

You can also watch [this video](#) to the participants to provide further context to the work that is being done globally to mainstream the decolonisation of digital rights.

SOURCES & RESOURCES FOR PARTICIPANTS

Note: The following sources were used in developing the content for this training. While this serves as a list of sources, we also encourage participants to look further into these resources for further learning.

RESOURCE NAME AND LINK	RESOURCE TYPE AND SOURCE	TOPIC
The power of virtual communities	Report, Governance Lab	General
Establishing membership and rules	Assorted Articles, Facebook Community	Membership and Engagement
New Facebook Group Insights:	Video Tutorial, YouTube	Membership and Engagement
How to Improve Your Facebook Group Engagement	Video Tutorial, YouTube	Membership and Engagement
Understanding Information Disorder	Guide, First Draft News	Information Disorder
An Online Course from First Draft	Online Course, First Draft News	Information Disorder
Online Course on Misinformation and Disinformation	Online Course, BBC Media Action	Information Disorder
Navigating Disinformation	Online Course, UN Women	Information Disorder
Online Course on Journalism: 'Fake News' & Misinformation	Online Course, UNESCO	Information Disorder
Remote Learning Course on Dealing with Disinformation Amidst the Infodemic	Online Course, Out of The Box Media Literacy Initiative	Information Disorder
A Citizen's Guide to Information Disorder	Guide, CSU Political Science Capstone	Information Disorder

Module on Hate Speech	Training Module, Media Defence	Hate Speech
Media Against Hate Speech Training Module	Training Module, European Federation of Journalists	Hate Speech
Guide on Positive and Nonviolent Communication	Guide, Positive Psychology	Non-Violent Communication
Managing Conflict	Assorted Articles, Facebook Community	Digital Peacebuilding Communication Skills
Digital Peacebuilding Communication Skills - Beyond Counter Speech	Report, Lisa Schirch Toda Peace Institute	Digital Peacebuilding Communication Skills
Introduction to Smart Politics	Powerpoint Presentation, Smart Politics	Digital Peacebuilding Communication Skills
Understanding Privacy Settings	Article, Facebook Community	Digital Safety
Building a Team and Managing Self-Care	Article, Facebook Community	Digital Leadership
Growing Your Group	Assorted Articles, Facebook Community	Digital Community Growth, Diversity and Inclusion
Diversity & Inclusion	Assorted Articles, Facebook Community	Digital Community Growth, Diversity and Inclusion
Design Justice	Organisational Website and Principles, Design Justice Network	Digital Community Growth, Diversity and Inclusion
The Charter of Human Rights and Principles for the Internet	Report, Internet Rights and Principles Coalition	Digital Rights

Introduction to Digital Rights	Report, Share Foundation	Digital Rights
Decolonizing Digital Rights	Project/Initiative, Freedom Fund	Digital
Training Manual on Digital Rights and Freedom of Expression Online	Training Manual, Media Legal Defense Initiative	Digital Rights
Module 04 – My Rights Online	Training Module, WebWise	Digital Rights
Facebook Community Standards	Facebook Policy, Transparency Center, Facebook	Digital Safety and Rights



Digital Community Stewards:

BUILDING SOCIAL COHESION IN ONLINE
COMMUNITIES



<https://www.sfcg.org/>



PARTICIPANT'S GUIDE