

CHAPTER 7

Moderation tools & systems for Meta Groups

Moderation isn't just about reacting, it's about designing systems that do some of the work for you, so you can protect your energy, be consistent, and keep your community safe and thriving. Different platforms have a range of tools available to help you manage your community smoothly and automate functions to lessen the workload. These are especially important in setting up a 'filter' that stops the most harmful content slipping through, as it's not often possible to monitor a community 24/7.

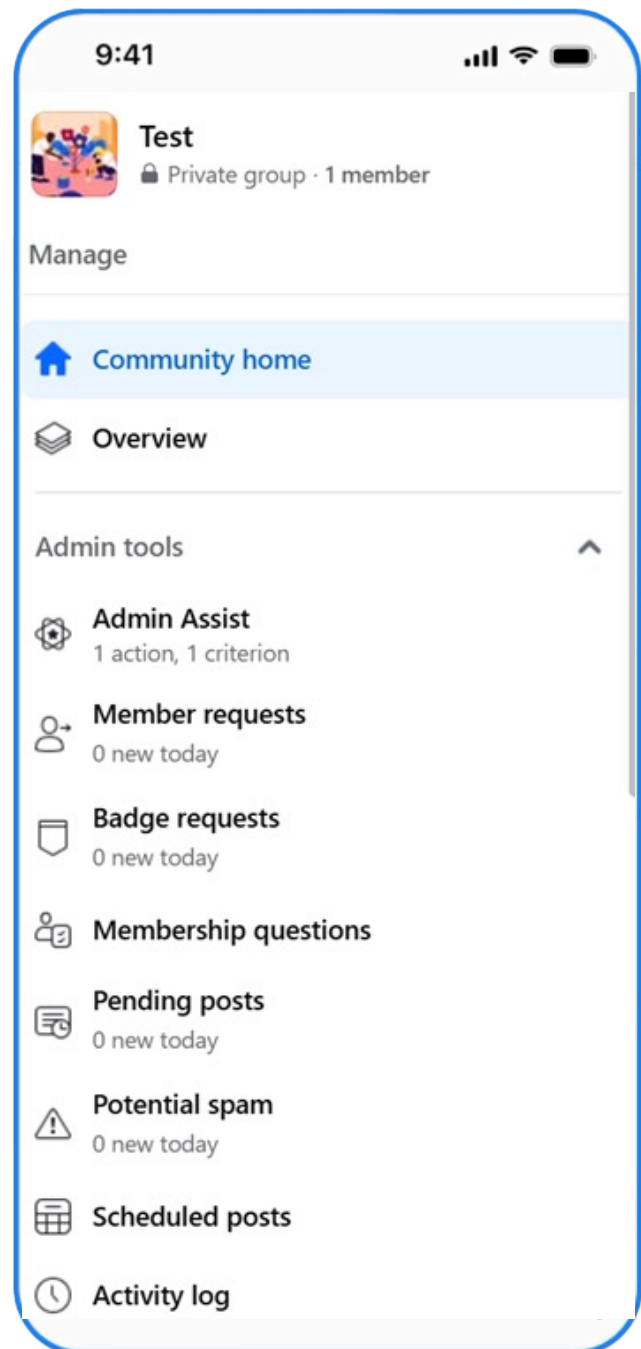
Facebook group tools deep dive for beginners

In New Zealand, Facebook Groups are widely used for everything from local school parents networks, through to shared hobbies or issues based discussions. If you're just starting out as a Facebook Group moderator (mod), you'll need to familiarise yourself with the tools available to help you manage your community.

Admin tool panel

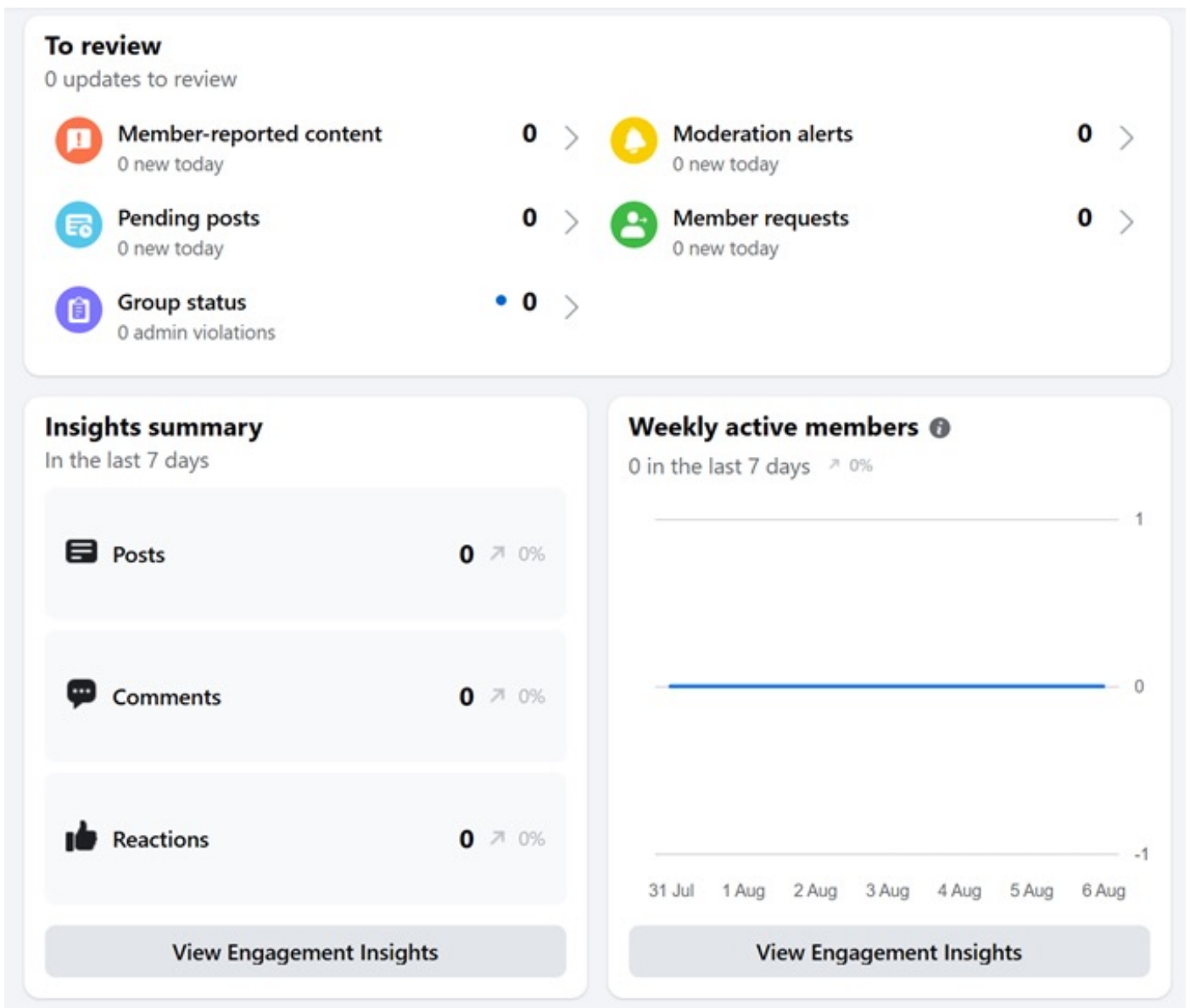
As a mod, you will be able to see an Admin Tool panel on the left side of your screen when logged into your group. This is where you'll find all the controls and insights required to manage and respond to activity.

IT LOOKS LIKE THIS:



Overview dashboard

If you're after a quick insight into what requires your attention at a glance, the overview dashboard is a useful function and shows pending posts, membership requests, moderation alerts, conflict warnings, etc, all in one tidy place.



Admin Assist (Facebook group automation tool)

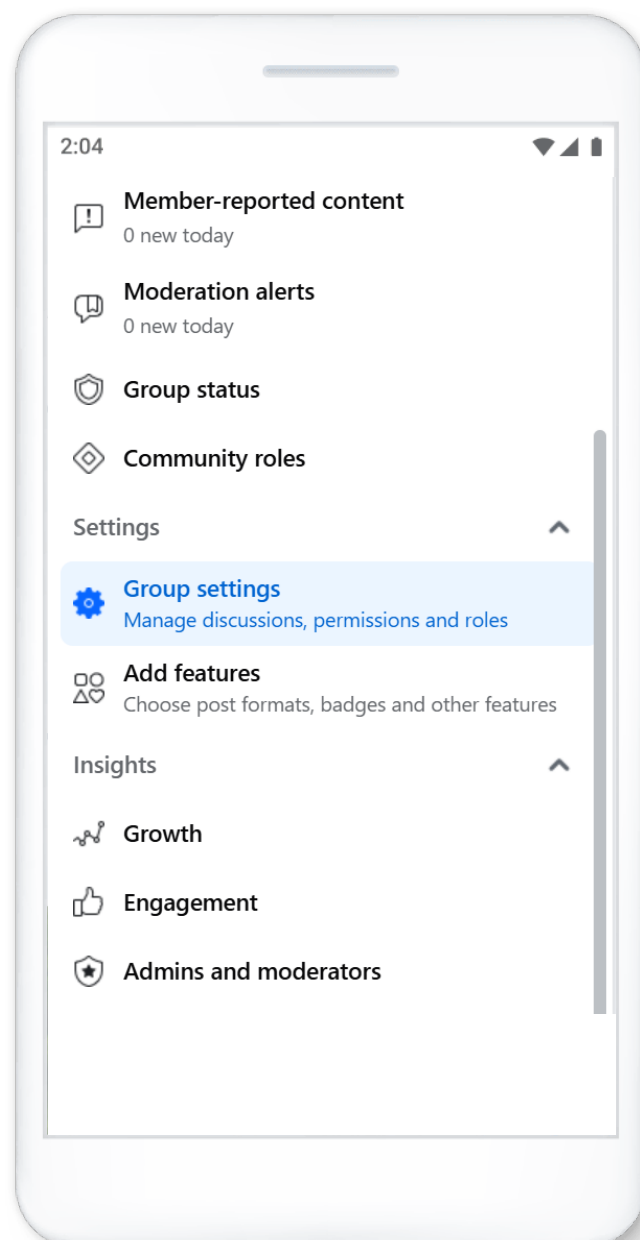
This is a really helpful function with the ability to set multiple rules for automatic member, post or comment approval, decline, or review based on rules you set up. This is often the first defence against spam or fake accounts, harmful content or repeat offenders.

There are plenty of options based on your group requirements, but as a minimum consider setting up automation to:

- Screening new members
- Auto-removing posts with banned domains or phrases
- Routing posts that contain specific tags (e.g., “CW: trauma”) for admin approval

TIP:

Combine Admin Assist + keyword alerts for a low-effort, high-control moderation system.



Post approval

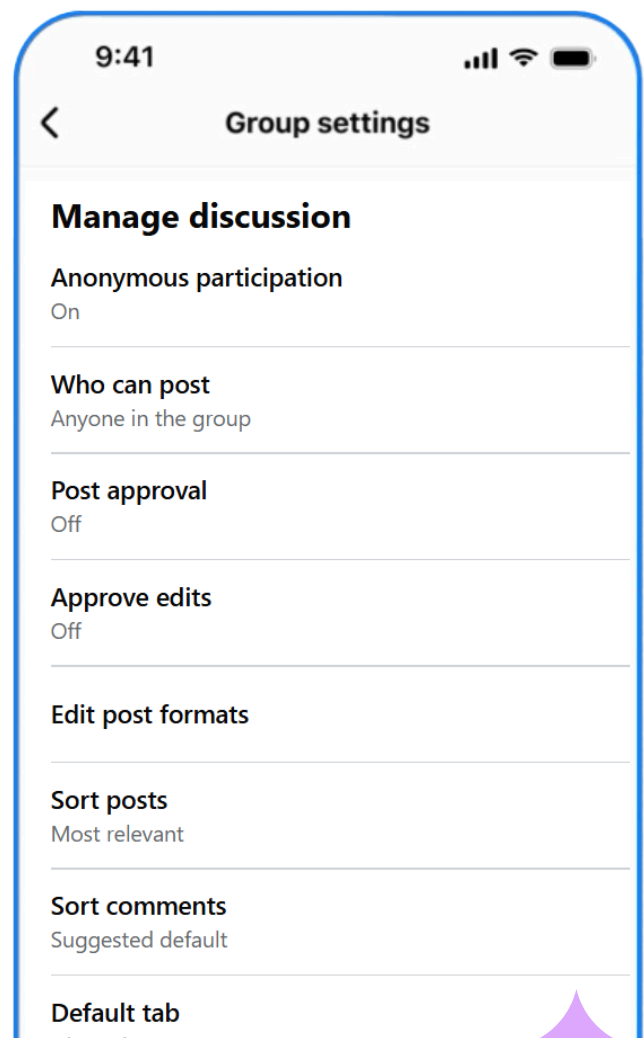
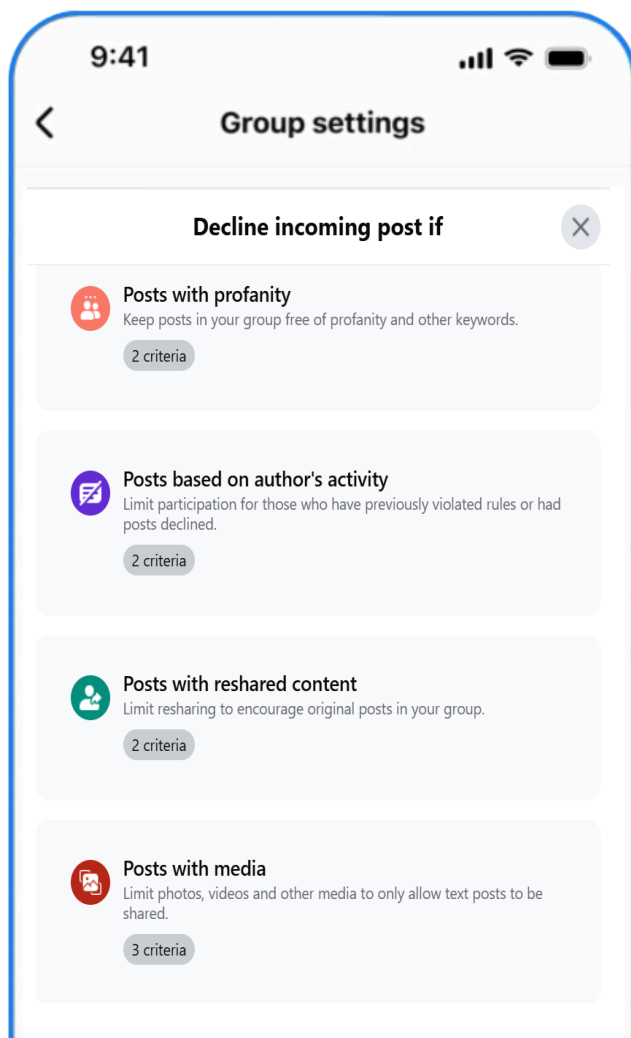
There are times when you might need to turn on post approval for your group, to manage tricky situations and protect members. You can do this in your Group Setting function.

Consider using for:

- New members
- High-conflict threads (e.g. During protests, elections or national events)
- Sensitive group topics (health, trauma, activism)

TIP:

Approve posts in batches daily to reduce burnout. Let members know in advance: “This group now uses post approval during high-volume weeks to keep things safe and on-topic.”



Keyword alerts

Found under the Moderation Alerts heading, Keyword Alerts will flag content being posted that might need closer attention. These can be a combination of ‘always on’ watch words, or trending controversial topics that require monitoring.

Set these up to flag:

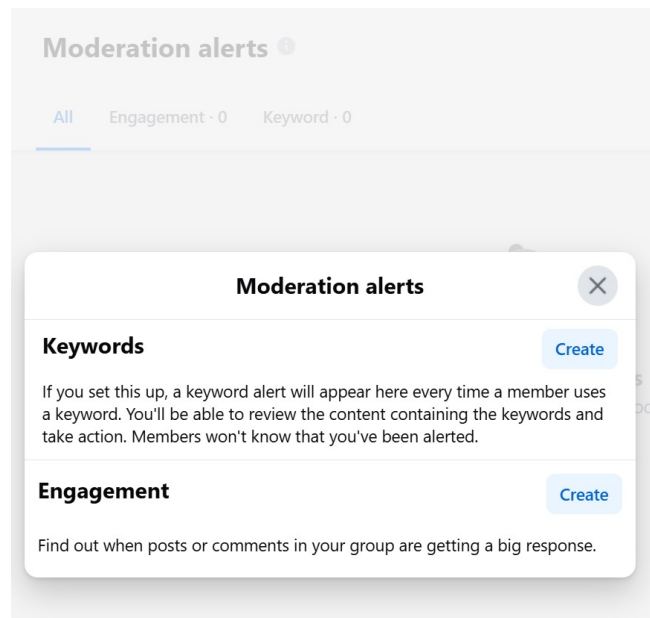
- Hate crimes slurs
- Crisis terms (e.g., “kill myself,” “want to die”)
- Coded language (e.g.”nazi terms” ”)
- Trigger words (e.g., “rape,” “sexual abuse”)
- Other controversial hot button issues

You can review the list of posts/ comments with keywords flagged here too.

- Click each one to review in context.
- Choose to:
 - Leave as is
 - Remove
 - Message member

TIP:

Review alerts weekly. Don't rely solely on automation, they're your early warning system, not your gatekeeper.



How to pause posting or comments

Useful during heated debates, trauma incidents, or spam floods.

To Pause Posting:

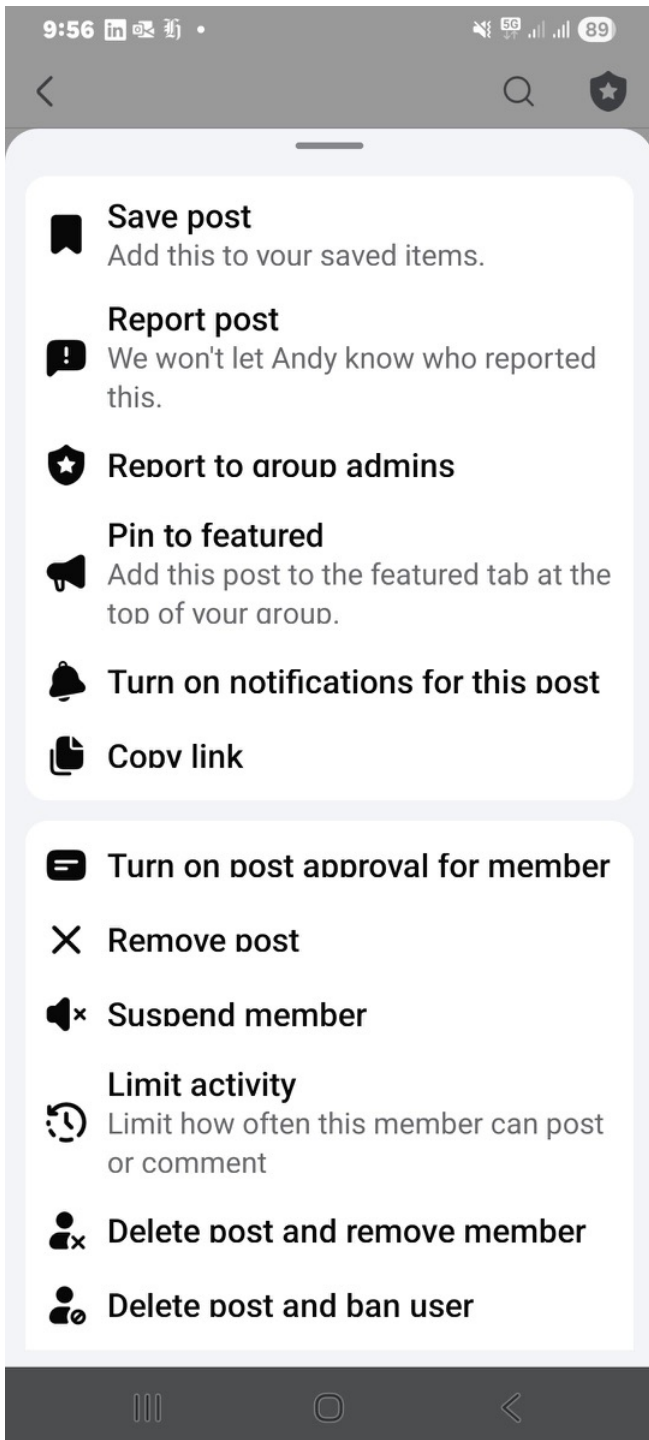
- Go to Admin Tools.
- Scroll down to Group Settings.
- Use ‘Post Approval’ or temporarily set the group to ‘Announcements Only.’

To turn off comments:

1. Click the three dots (...) on a specific post.
2. Select ‘Turn Off Commenting.’

TIP:

Add a “Tone Reset” comment before pausing or locking.



Reported by members

Communities that have a strong, engaged membership and a clear set of rules will often have members that will back up the admin by helping to enforce rules and etiquette. This is either through comments that they leave in response to emerging issues on the page, or by reporting content that could be harmful or against the rules. Anything that is reported, will show up on your dashboard as 'Reported by Members' for your consideration and action.

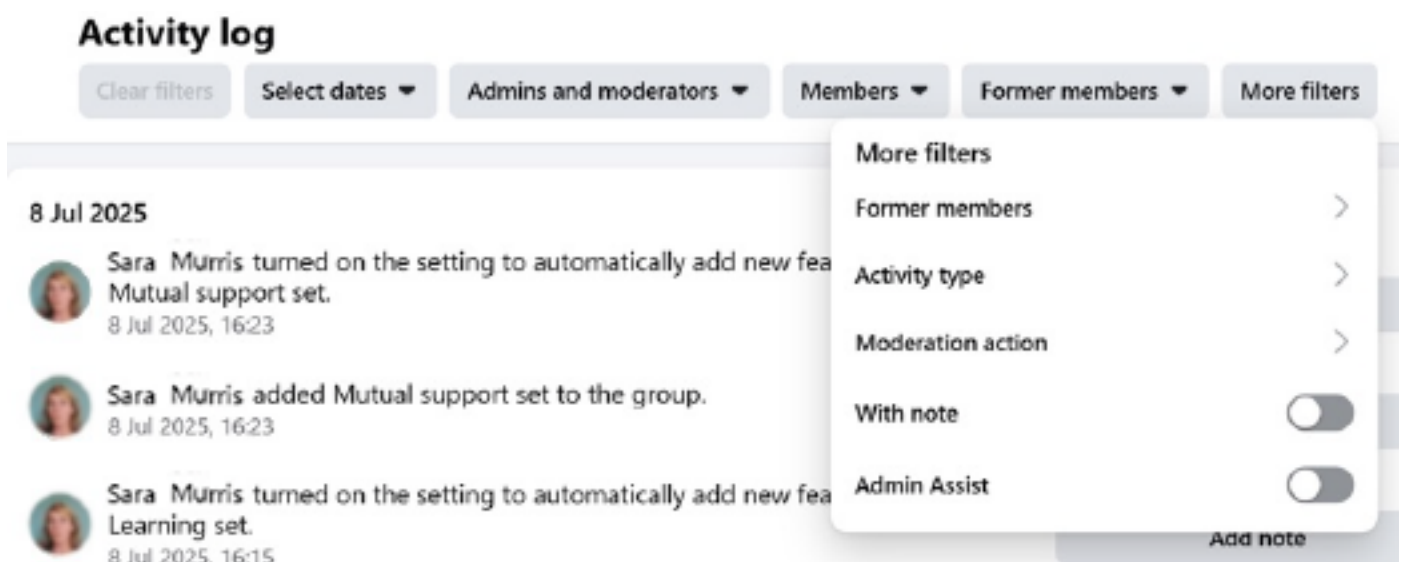


Activity log

A key tool for any moderation team is the Activity Log, as this is your go-to audit trail for everything admins and mods have done, including deleting posts, approving members, removing comments, issuing suspensions, etc.

In Facebook Groups you can filter by date, by specific moderator, by type of action or by member and former members.

Automated actions set up with Admin Assist will also get logged here, so you can check in and reverse actions if need be.



Mod resources you can create that might be helpful

- “Mod Incident Log” (online spreadsheet or shared form)
- Shared Folder (screenshots, past statements)
- “New Mod Welcome Pack” (with tone guide, rules, scripts)
- “Debrief Chat” for after big moderation calls or trauma content

BONUS: save time with quick links in your mod chat

Send these direct links to fellow admins:

- **Moderation Alerts:** facebook.com/groups/YOURGROUPID/admin_alerts
- **Keyword Alerts:** facebook.com/groups/YOURGROUPID/keyword_alerts
- **Admin Assist Rules:** facebook.com/groups/YOURGROUPID/admin_assist
- **Member Requests:** facebook.com/groups/YOURGROUPID/member_requests

AI-enabled features

Many of the automated moderation tools use AI in the background to power results, including the Admin Assist suite and Moderation Assist for Pages. There are other ways moderators can use Meta's AI tools to improve efficiency and lessen manual tasks.

Meta AI (Beta) can help you rewrite moderator announcements in different tones (friendly, professional, concise) or suggest new post ideas to keep communities engaged. You can also use this to quickly get a summary of post comments, to check how your community is responding to certain content and take a 'pulse check'

There are plenty of AI tools available from a range of providers to either help with moderation tasks or generate content quickly. They can be a great time-saver, but it's important to understand how to use these safely and responsibly, particularly in terms of accuracy and privacy. See page 57 in the Appendix for more AI guidance.

Using guides

Guides are one of the most powerful but under-used moderation tools. They let you turn scattered posts into a structured, always-available handbook inside your group. Think of them as a bookshelf for the important stuff, rules, safety info, FAQs, and cultural values, all neatly organised so members can find them anytime.

You can:

- Move any post into a Guide (before or after publishing).
- Organise Guides into themes (Welcome, Safety, Culture, Topical Posts, FAQs, Emergencies).
- Best for permanent reference material members will need again and again.
- Helps new members get up to speed quickly, reduces repeat questions on the same topic, and keeps your group's culture consistent.

Featured posts

- Featured posts are the "noticeboard" of your group. They sit right at the top of the feed where everyone sees them first. Unlike Guides, they're not permanent, they shine a spotlight on what matters most right now.

Details:

- Pin important posts so they appear at the top of the group feed.
- Use for time-sensitive or urgent updates (events, announcements, campaigns).
- Only a few can be featured at once — rotate them often.
- Keeps members focused on what's most relevant today.

Moderation Assist for Facebook pages

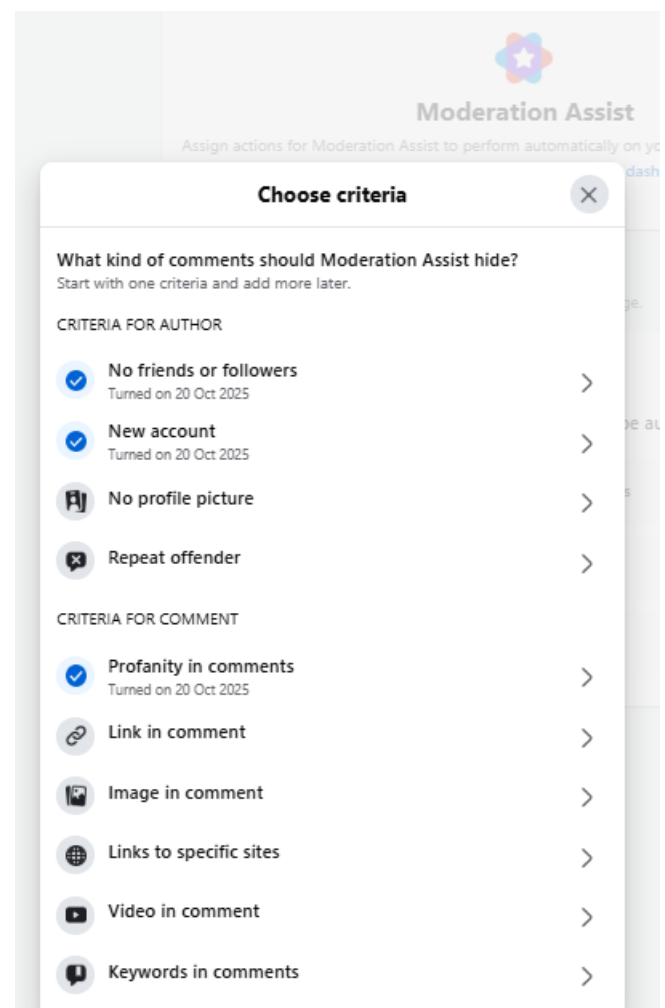
Most public-facing organisations in Aotearoa (schools, councils, community services, NGOs) now rely on Pages to communicate updates and receive community feedback, sometimes hundreds of comments a day. Moderating these spaces well is essential for protecting staff, audiences, and reputations. As Pages are public, comments appear instantly and can spread fast. Meta’s Moderation Assist tools (found under the Professional Dashboard) help automatically hide or review harmful or spam content before it’s seen.

Access:

Go to your Page > Professional Dashboard > Moderation Assist

There are an array of options within the dashboard to hide or flag comments based on certain criteria, including:

- **Keyword filtering:** Add custom words or phrases (e.g. “DM me,” “crypto,” “giveaway,” “scam”) to automatically hide comments or send them for review
- **Profanity in comments:** Turn on the “Hide Comments Containing Commonly Reported Words”
- **URL & spam filters:** Automatically hide comments with external links, repeated emojis, or bot-style behaviour
- **Auto restrictions:** Limit or ban users who repeatedly have comments hidden
- **Inbox alerts:** Flag messages containing keywords like “abuse,” “help,” or “urgent”



Best practice

Publish a clear commenting policy on your Page -“We remove abusive, defamatory, or spam comments to protect our community.”

Review hidden comments regularly, automation isn't perfect.

Meta Support Pros for Pages

Any Meta user managing a government, politics, nonprofit page or Business Account has access to support via Meta Support Pros. This could include providing guidance on day-to-day best practices to answering technical questions you may encounter on Meta platforms. See more on how to access this service.

Meta guidelines and actions

Even if you stay within NZ law, Meta can pause, restrict, or remove your Facebook group/page if its Community Standards are repeatedly broken, and it doesn't matter if the breach comes from a member, not you.

Meta community standards

Meta may remove or restricts content, accounts, pages and groups for:

- Illegal activity -sale of drugs, weapons, counterfeit goods, hacking services
- Violence & incitement - threats, coordination of harm
- Dangerous organisations and

individuals - bans content that praises, supports or represents entities tied to violence or extremism

- Bullying & harassment - targeted abuse, doxxing, shaming
- Child sexual exploitation, abuse, and nudity - all content that sexualises, exploits, or endangers children
- Hate crimes - attacks based on race, ethnicity, religion, gender, sexual orientation, disability
- Misinformation in specific high-harm categories
- Adult sexual content -nudity, pornography, sexual solicitation and exploitation

[Find out more about Meta Community Standards and view the full list here.](#)

Meta is ending the third-party fact-checking program and moving to a Community Notes model, starting in the United States. Community Notes lets people add more context to Facebook, Instagram and Threads posts that are potentially misleading or confusing. At the time of writing, third party fact checking on Meta platforms is still active in Australasia.