AI FOR COMMUNITY MODERATORS

For time-poor moderators and those filling the role as a volunteer, AI tools can help lighten the workload and respond faster. Built-in AI enabled moderation tools are built into most platforms now and help mods set up a first line of defence against online harm. Other Generative AI tools can also be a useful helper for community moderators, but AI doesn't understand your group's kaupapa or tone unless you tell it, and it can make mistakes.

What you can use it for:

- Set up keyword filters, member screening and other back-end systems
- Drafting messages & holding statements – calm, neutral, supportive
- Summarising long threads – get the gist quickly without wading through
- Translating tone (for example, from too harsh to gentle, or too formal to friendly)
- Brainstorming tone-setting responses – when you're stuck for words
- Training support – practice scenarios, inclusive phrasing ideas

AI should help you hold the tone and save time, not replace your judgement.

Getting better results

AI can only work with what you give it. Vague prompts give vague results. The more context and instruction you include, the better the output will match your group's tone, kaupapa, and needs.

FOR EXAMPLE:

Poor prompt:
"Write a group post about respect."

Better prompt:
"Write a short, friendly post reminding members in a private Aotearoa parenting group to keep kōrero respectful during heated debates. Include a gentle reminder about our group rules and use plain New Zealand English."

"I'm a moderator of a private mental health support group in Aotearoa. Write a short message asking members to stop arguing and refocus on respectful kōrero. Keep it kind, calm, and mana-enhancing."

TIPS:

- Include context (who you are, what kind of group it is)
- Specify tone (calm, inclusive, mana-enhancing)
- Tell it how long or what format you need (two sentences, list, announcement)
- Mention what not to do (for example, "avoid emojis," "keep it neutral")

Iterating for better results

AI improves when you give feedback. Try adding:
“Make it sound more natural.”
“Use lighter language.”
“Shorten to two sentences.”
“Rewrite with a te reo Māori greeting and New Zealand spelling.”
“Keep it professional but friendly.”

Cultural and contextual awareness

AI tools are mostly trained on global data, not Aotearoa communities, so guide them to localised language and context.

Safe AI use

- Don’t share names, screenshots, or private details in your prompts Instead: paraphrase (“a member made a racist comment” rather than pasting their post).
- Don’t copy-paste confidential moderation discussions.
- Don’t rely on AI to decide what is harmful or illegal, escalate to Netsafe or Police if you’re unsure.
- Use trusted, reputable AI platforms.
- Avoid random plug-ins or bots that may store data unsafely.

Final check before posting

Before you post or send an AI-drafted message, check:

- Does it reflect your group’s kaupapa?
- Does it match your group’s tone?
- Is it accurate and neutral?
- Have you checked any links or sources of information for accuracy?
- Would you stand by it if someone asked how it was written?

AI can be a helpful co-pilot for moderators, not a replacement for empathy, cultural care, or accountability. Used wisely, it can help you focus on what matters most: guiding kōrero that keeps your community safe, balanced, and respectful. Always check, edit, and own what you post. You’re the human voice of your community.

For more in-depth information on safe and responsible AI use, see **netsafe.org.nz/artificial-intelligence**