

## MODERATION SYSTEMS & TOOLKIT QUICK LINKS for Facebook Groups

Task	Purpose	How
<b>Post Approval</b>	Approve posts before they go live	Admin Tools → Post Approval → Toggle ON
<b>Keyword Alerts</b>	Flag high-risk terms	Admin Tools → Keyword Alerts → Add Words
<b>Admin Assist Rules</b>	Automate low-risk moderation	Admin Tools → Admin Assist → Add Condition
<b>Group Rules</b>	Create enforceable standards	Admin Tools → Group Rules → Add Rule
<b>Membership Screening Questions</b>	Vet new members before they join	Member Requests → Ask Pending Members Questions
<b>Incident Log</b>	Track harms, repeat users, risky posts	Use platform log or alternative shared file
<b>Moderator Team Roles</b>	Assign clear tasks	Document in internal team file or pinned message

Tool	QUICK LINKS
<b>Moderation Alerts</b>	<a href="https://facebook.com/groups/[GROUPID]/admin_alerts">facebook.com/groups/[GROUPID]/admin_alerts</a>
<b>Keyword Alerts Panel</b>	<a href="https://facebook.com/groups/[GROUPID]/keyword_alerts">facebook.com/groups/[GROUPID]/keyword_alerts</a>
<b>Admin Assist Setup</b>	<a href="https://facebook.com/groups/[GROUPID]/admin_assist">facebook.com/groups/[GROUPID]/admin_assist</a>
<b>Member Requests &amp; Screening</b>	<a href="https://facebook.com/groups/[GROUPID]/member_requests">facebook.com/groups/[GROUPID]/member_requests</a>
<b>Incident Log Template (Google Sheet)</b>	[Insert Link or QR Code]

**ACTION CHECKLIST FOR MODERATORS**

Situation	Action	✓
<b>Harmful Post Detected</b>	Remove immediately → Screenshot → Log it → Optionally DM poster	
<b>Misinformation Detected</b>	Link to facts → Warn user (if repeat) → Reset tone in comments	
<b>Conflict in Comments</b>	Step in with tone comment → Lock thread if needed → Mute aggressors	
<b>Hate Speech or Identity Harm</b>	Remove → escalate if extreme → reference broken rule	
<b>Trauma or Self-Harm Content</b>	Hide temporarily → DM poster with care → Share support links	
<b>Doxxing or Privacy Breach</b>	Remove → Securely log info → Notify member if relevant	
<b>Scam or Fraud</b>	Remove → Ban if repeat → Note scam type (crypto, phishing etc.)	
<b>Gang or Extremist Content</b>	Remove → Screenshot → Escalate to Netsafe or police if threatening	