

Stoatmod and Stoat Trust & Safety

RE: ATTN: Stoat Trust & Safety Review - StoatMod Data Handling

Ticket #7421386 5 Sep to 2 Oct 2026 7 messages

The message text below is unchanged apart from formatting. It's shown oldest first and includes every message in the thread. Screenshots attached to the 14 September email aren't included.

01	Mana to Stoatmod	Saturday 5 September 2026
02	Kurome to Stoat Trust & Safety	Saturday 5 September 2026
03	Mana to Stoatmod	Thursday 10 September 2026
04	Kurome to Stoat Trust & Safety	Thursday 10 September 2026
05	Mana to Stoatmod	Sunday 13 September 2026
06	Kurome to Stoat Trust & Safety	Monday 14 September 2026
07	Kurome to Stoat Trust & Safety	Friday 2 October 2026

Mana, Lead, Trust & Safety | Platform Moderation, Stoat

MESSAGE 01

From: Stoat Trust & Safety
To: StoatMod Team
Date: Saturday 5 September 2026

Hello StoatMod Team,

I am reaching out in regards to StoatMod bot and the current data collection, retention, and processing handling. Following my review, I have concluded that the implementation outlined below, and in StoatMod's Privacy Policy, are not permitted under our Terms of Service and Community Guidelines. Our Community Guidelines explicitly state: "We do not permit the gathering of any personally identifiable information (PII) without end-user consent. This includes, but is not limited to, the automated scraping of messages, member lists, and user profiles without explicit permission from the users whose data is being collected.

Specifically, collecting and retaining user information, including complete message content and identifying information such as user IDs, message IDs, channel information, usernames, and message history, are not permitted under our guidelines for non-consensual collection of user information. This also encompasses storing and retaining previous and edited versions of messages or content and transferring that information to a development server.

Your Privacy Policy for the bot also states that StoaMod may process "message and edited-message text, message and channel IDs, links, domains, mentions, timestamps, image URLs and attachments, moderation categories and confidence scores, and cached before-and-after content used for edit logs." This policy further expands that "Complete content may be retained when a message is deleted, edited, reported, warned or included in a moderation or audit event. Moderation events, including their associated message context and moderation scores, may also be copied to StoaMod's restricted developer server through a private Stoa webhook for service-wide oversight and incident review." This conveys that user content and its associated identifiers are moved outside of the originating server/community.

I have also identified some concern(s) with the collection and retention of IP addresses for extended moderation action(s). In addition, transferring image content and user messages through an external program, such as OpenAI, for classification comprises added processing and handling of information in connection with users, which is not permitted under our Community Guidelines where end-user consent has not been obtained. You can find our Guidelines here: <https://stoa.chat/legal/community-guidelines>

The absence of clear defined periods of retention for said data and the continual preservation for moderation and audit purposes introduce added issues. Information and user data that is not necessary for the bot's functionality must not continue to be kept.

In accordance with our policy, StoaMod must permanently delete the acquired information collected through these methods, and either remove the functionality completely, or alter the implementation to comply with our Community Guidelines.

Please take the necessary steps to address these issues within 30 days and confirm when they have been addressed to ensure StoaMod is compliant. Failure to abide by this notice may result in additional platform-level action being taken, including removal from the platform entirely.

Kind Regards,
Mana - Lead, Trust & Safety | Platform Moderation

--

Stoa Moderation

Email: abuse@stoa.chat - Web: <https://stoa.chat>

Revolt Platforms Ltd. is a company incorporated and registered under the laws of England and Wales.

Registration Number: 16260658

Registered Office:

Suite 5703 Unit 3A, 34-35 Hatton Garden,
Holborn, United Kingdom, EC1N 8DX

From: Kurome (StoaMod)

To: Stoa Trust & Safety

Date: Saturday 5 September 2026

Hello,

Receipt has been confirmed and work will be done before the Monday October 5th 2026 deadline. Please correct me if this date is inaccurate.

To ensure the revised implementation fully complies with the Stoa Community Guidelines & this notice, would you please confirm a few things:

- What constitutes valid end-user consent? In particular, must every member explicitly opt in, and what notice, recording, withdrawal, and deletion requirements must a consent process meet
- Would anonymous aggregate statistics be allowed? Such as total daily message counts, joins and leaves. Permitted they cannot actually identify individual users and contain any actual message content.
- May a bot send an immediate moderation or audit log to a private channel within the same origin server, or would displaying such information require consent from the message author?
- May StoaMod retain minimum server, channel, message, and user identifiers required to execute a requested moderation action or maintain server configuration? If so, what retention period would be considered acceptable?
- How should user-submitted reports, appeals, supporting evidence, and moderation cases be handled when the submitter knowingly provides the information but the record may also identify other users?
- Is short-lived IP processing permitted for CAPTCHA validation, security rate limiting or abuse prevention? If so, what maximum retention period and safeguards are expected?
- May StoaMod generate and retain keyed HMAC-SHA256 fingerprints of IP addresses without retaining the raw addresses, solely for the purpose for abuse or ban-evasion detection? If permitted, may those fingerprints be linked to Stoa user or server IDs or compared across servers, and what consent, retention, deletion, key-rotation, and access control requirements would apply?
- Is any external content classification permitted with explicit end-user consent, or must message and image transmission to external processors such as OpenAI be removed entirely?
- Are cross-server safety records or alerts ever permitted, and if so, under what consent and retention requirements?
- for existing data, does permanent deletion need to include active databases, application logs, developer-server messages, and all backups? Where selective deletion from backups

is not technically possible, is making the backup inaccessible until it is automatically overwritten acceptable?

- What evidence would you like me to provide to demonstrate completion and compliance with the notice. For example: A deletion statement, revised data map, retention schedule, processor list, or technical summary?

On a separate point without suggestion that another bot's practice may affect or excuse my/our obligations: I would like to raise a concern about the consistent application of these requirements across the platform. Another bot (01J4Q9DS6CV4DNCW6AZRMV349A) "log_it" has a published privacy policy which states that it collects and stores server IDs, selected Channel IDs, and event activity. It's transparency page states that this activity can include deleted and edited message content that the information is backed up into a locally stored log.txt file. It's deletion policy also seems to exclude information remaining in that file

- <https://mu-mega-bots.github.io/log-it/information/privacy.html>
- <https://mu-mega-bots.github.io/log-it/information/transparency.html>

Could you confirm whether the same end-user consent, retention and deletion requirements apply to "log_it" and other logging bots on the platform that process or retain deleted and edited message content? I am not requesting any information or confidential enforcement action. I am seeking a clarification on the Community Guidelines. If this concern should be submitted as a separate report please let me know and I shall do so.

These questions above are not intended to dispute or delay compliance with the notice. It is intended to ensure that a future revised implementation follows Stoa's requirements accurately.

Thanks,
Kurome#0606
StoaMod

From: Stoa Trust & Safety
To: Kurome (StoaMod)
Date: Thursday 10 September 2026

Hi Kurome,

Thank you for confirming the receipt of my notice and advising the revisions/rework would be completed by the October 5th, 2026 deadline. The date you have is correct.

Let me provide some clarification to the question(s) you raised:

- When it comes to StoaMod's reliance on consent to collect or process Personal Identifiable Information (PII), that consent must come from the affected end-user as per our Community Guidelines. A server owner's decision to implement the bot into their server should not, by itself, be treated as a blanket consent for collecting user information, such as message content, IPs, user profiles, and other PII.

StoaMod implementation should provide clear notice to users of the exact information being collected, the reason it is being collected, and where the information is processed/stored. Users should also have a clearly defined avenue for withdrawing their consent and the ability to have that information deleted should they choose to exercise these rights.

- Aggregate statistic collection would need to be genuinely anonymous and cannot reasonably be used/collected to identify individual users. The statistics also cannot contain message content or user-level records.
- Moderation and audit logs within a Stoa server that are necessary for the operation of the implementation and functionality can continue as long as it is access-controlled, limited to information only necessary for that purpose. However, permission and functionality for audit/moderation logs do not constitute consent to collect and retain unrelated message content or historical message content and store that information on another server outside of the originating Stoa community/server.
- StoaMod may be allowed to retain minimum identifiers should they be considered reasonably necessary to ongoing and active moderation action(s) or to help maintain server configuration. These identifiers should never be retained indefinitely for the reason(s) that they may potentially be useful in the future. Retention of this information should therefore be limited to what is necessary, with all unneeded historical data removed and/or deleted, and a clearly defined time-frame for retention outlined in your Privacy Policy.
- Information that is willingly and knowingly provided by a user as a part of an appeal or report may be processed for the purposes of investigating and resolving said appeal or report. When evidence submitted contains information in relation to outside users, StoaMod should limit the retention and access to information necessary for the

investigation and should avoid holding on to or storing third-party information unnecessarily.

- Briefly processing network data for the purpose such as CAPTCHA/Turnstile verification, rate limiting, and prevention of abuse is a separate use case from collecting and retaining full IP addresses on a continual basis. Any network data/information that is processed for verification reasons should be limited to what is necessary for that verification.

As it is currently written in the Privacy Policy, permanently storing complete IP addresses and generating fingerprints to identify potential alternate accounts cannot continue without appropriate safeguards and compliance with end-user consent where it is applicable.

- Removing the raw IP address data from the equation does not in itself resolve the core concern if a fingerprint continues to function as a persistent identifier.

Specifically, generated fingerprints should not be used to create a persistent database of cross-server user profiles or to correlate those users between separate communities on the platform just because the source is based off of cryptography methods. Any security methods retained or captured for prevention of abuse should be access controlled, be subject to limits on the data/information retention, and complete removal of said data when it is no longer necessary.

- External classification usage, such as OpenAI or other third-party services, may continue only where end-user consent is explicitly obtained from each user before their content is transmitted to an external service. Disclosure of these practices through a Privacy Policy does not constitute consent by itself.
- Having cross-server processing cannot be used to create a general profile for users across unrelated community servers. Safety-related information shared across servers should be strictly limited to legitimate safety reason(s) with access controls, and only kept/retained for as long as necessary to fulfil the functionality. The current implementation should be revised with this in mind.
- To fulfil this request, the removal of information should be completed from active databases and any accessible application(s) or development infrastructure. Backups containing this data and information must also be destroyed in their entirety.

Please ensure that the revised implementation(s) are in place by the October 5th, 2026 deadline. If you should need further assistance, please don't hesitate to reach out.

Kind Regards,

Mana - Lead, Trust & Safety | Platform Moderation

--

Stoat Moderation

Email: abuse@stoat.chat - Web: <https://stoat.chat>

Revolt Platforms Ltd. is a company incorporated and registered under the laws of England and Wales.

Registration Number: 16260658

Registered Office:

Suite 5703 Unit 3A, 34-35 Hatton Garden,
Holborn, United Kingdom, EC1N 8DX

From: Kurome (StoaMod)
To: Stoa Trust & Safety
Date: Thursday 10 September 2026

Hi Mana,

Thank you for your response and for confirming the October 5th deadline.

However I need clearer answers about whether StoaMod's core functionality can continue and how these requirements apply consistently across the platform:

1. I understand your current position that sending content to OpenAI or any other external service requires specific consent from the end-user. Is that an absolute requirement for external AI moderation, or would stoat consider a specifically authorised agreement with defined purposes, restricted processing, enforceable retention limits and appropriate safeguards? If so, what conditions would it need to meet?
2. On another point: Would an AI model running entirely on StoaMod's own infrastructure be permitted to assess incoming messages without individual opt-in, solely for the purpose to enforce rules inside of it's origin community, provided the message content is discarded after assesement, no routine message archive or cross-server profile is created, and no content is transmitted to an external classifier
3. If individual consent is required, then how can moderation operate effectively?: If every member must opt-in before any AI assesement, users intending to abuse a community could decline. What model does Stoa permit for providing automatic protection across a community?
4. Would a clear, separate, affirmative consent step during server onboarding be considered acceptable?
5. Please also clarify whose consent is actually required when a message contains a quotation, screenshots or information about others. The sender's consent may not cover everyone represented.
6. Please address my last concern: My pervious email raised log_it, bot ID "01J4Q9DS6CV4DNCW6AZRMV349A", and linked their privacy policy & transparency page. I identified statements concerning edited/deleted message content and locally stored "log.txt" role. Your response did not address this example.

Automod, bot ID "01FHGJ3NPP7XANQQH8C2BE44ZY" does not have a public privacy policy and also publicly documents automatic message filtering using user message content without end-user consent and logging of message edits and deletions

<https://automod.vale.rocks/docs/automod/commands/configuration/filter>

<https://automod.vale.rocks/docs/automod/commands/configuration/logs>

I understand that overlapping features do not establish identical storage practices or authorisation. Nevertheless being instructed to remove features that are comparable to

bots that are publicly advertising without an explanation creates concerns about uneven treatment.

7. What exactly should the deletion statement and completion statements cover? I understand that all effected backups must be destroyed in full. Please confirm whether the deletion scrope also includes previously posted origin-server moderation logs and existing user-submitted reports or appeals

Given the deadline: I'd appreciate answers to each answers to each question. If a question requires a wider policy decision please refer it to the person responsible and identify which points remain unresolved.

Thanks,
Kurome#0606
StoatMod

From: abuse@revolt.chat

To: Kurome (StoaMod) · admin@modstoa.site

Date: Sunday 13 September 2026, 20:02

Hi Kurome,

I have provided additional clarification to answer your questions below:

1. Yes. Explicit end-user consent is an absolute requirement where message content and other information is transmitted through OpenAI or other third-party services for AI moderation purposes.

Stoa will not be considering any specialized agreement between our service and third-party bot developers/operators at this time.

2. Yes, this may be allowed. Operating an AI model hosted entirely on StoaMod infrastructure would not, by itself, be considered a transmission to a third-party service. However, the model should remain in scope with our Community Guidelines/Terms of Service, and the model is solely used to assess incoming messages for moderation/verification within the originating community server on Stoa. The content outlined may be processed without individual user opt-in as long as the content is destroyed after the initial assessment has finished, and it cannot be retained to create an archive.

This AI system also cannot, under any circumstances, be used to train any sort of AI model/LLM.

3. All bot operations on the platform must comply with our Terms of Service and Community Guidelines; these policies do not prevent the use of automatic moderation tools. The requirements must be followed regardless of whether individual users would otherwise be able to opt out of the process. As previously clarified, individual user opt-in is not solely required because an AI system/tool is being used for necessary internal moderation. A system that is hosted internally may assess messages for the purpose of enforcing rules within the originating Stoa server, as long as they follow the requirements provided.
4. Yes. A clearly outlined and separate consent step may be acceptable where end-user consent is required. The process for obtaining consent must be clearly defined and explain the data/information practices, which should include what information is processed or collected, the reason for which it is used, and for how long the information is retained. Information cannot be stored indefinitely.
5. Consent must be obtained from the user(s) who submitted the message(s) for consent-based processing. This is the only consent required at this time. Where consent is required, it applies to content they have submitted to/on our service. Processing should continue to remain limited to what is necessary and should not be used to create records or profiles about individual users.

6. I will not disclose or discuss any details of enforcement actions or compliance issues involving outside bot operators with an outside party. Each bot service on the platform must comply with our policies independently, and any other potential infringement by another operator does not change the requirements enforced for StoaMod.
7. The specific details and provided proof of compliance are left to your discretion, provided that you have demonstrated completion of our outlined requirements in this notice. Existing moderation logs retained within Stoa services itself are allowed to remain. Appeals and reports that have been submitted to StoaMod directly are considered intentional interactions with the bot and are allowed to remain as long as they are operationally necessary.

Kind Regards,

Mana - Lead, Trust & Safety | Platform Moderation

--

Stoa Moderation

Email: abuse@stoa.chat - Web: <https://stoa.chat>

Revolt Platforms Ltd. is a company incorporated and registered under the laws of England and Wales.

Registration Number: 16260658

Registered Office:

Suite 5703 Unit 3A, 34-35 Hatton Garden,

Holborn, United Kingdom, EC1N 8DX

From: admin@modstoaMod.site
To: Stoa Trust & Safety · abuse@revolt.chat
Date: Monday 14 September 2026, 08:28

Hi Mana,

Thanks to your answers to my questions work is underway and a planned phased plan has been announced to the support server of a timeline of events and actions which will be implemented before a planned date and a response will be submitted before the final date of October 5th.

Just two more questions:

1. Attached is an image of a "members tab" which shows server moderators an overview of members who have joined their server. Is this currently acceptable within Stoa's guidelines? I assume not since it is displaying user IDs and information of a user and taking it off the platform
2. Attached is an image of a case overview. We're planning on moving over to a locally hosted AI model on our own infrastructure. However you mentioned that the message content must be destroyed after an initial assessment. Would a user need to consent to keep the message content for server moderators to see why the message was flagged? How will moderators know why a message was flagged/what the user actually sent? Would it be acceptable to send an embed of the message content in a designated channel of the origin server (which does not store on our server) and not include it on the web dashboard? Or can it be displayed on the web dashboard as long as it is only shown with the relevant parties?

Thanks,

Kurome#0606

This email had two screenshots attached (the Members tab and a case overview). They aren't included here.

From: admin@modstoaMod.site
To: StoaMod Trust & Safety · abuse@revolt.chat
Date: Friday 2 October 2026, 03:02 BST (02:02 UTC)

Hi Mana,

I am writing to confirm the changes made in response to your notice.

The developer audit channel, ID 01KZ86J5D0GAMB7AKPP9A1HDP8, and its associated webhook, ID 01KZSADVTD8KFYYR2MVJC04ZHV, have been permanently deleted. This system previously received moderation logging information relating to message edits and deletions, joins, and members leaving. All queued webhook payloads and copies under StoaMod's control have also been deleted.

All data falling within the scope of the notice has been removed from StoaMod's active databases, application files, logs, developer-server copies, and backups under my control. This includes historical message content, member information collected by the retired systems, and hashed IP identifiers.

OpenAI and other third-party AI processing have been removed. Moderation now uses a model hosted entirely on StoaMod's own infrastructure. No user content has been used to train or fine-tune this model. I am also planning to publish the model and relevant technical information openly for further transparency.

Join CAPTCHA is disabled by default. If a server enables it, the user must accept a separate privacy notice by ticking a required checkbox before submitting the form. The CAPTCHA no longer processes IP addresses or performs alt-account, VPN, proxy, Tor, data-centre, or network-abuse detection.

The CAPTCHA records only:

- StoaMod user ID
- Originating server ID
- Privacy-policy version and acceptance time
- A hashed, one-time verification token
- Creation and expiry times
- Failed-attempt count

These records are deleted within 30 days.

Cross-server join-history sharing, moderation-history sharing, and safety alerts have been removed. The associated sharing records have been deleted. Global bans and warnings remain available for direct service-wide enforcement, but they are no longer presented to server moderators as cross-server member history or alerts.

Individual user statistics, popular-word records, and user activity leaderboards have also been removed and their historical data deleted. StoaMod now retains only anonymous aggregate statistics, such as total daily messages, joins, and leaves. These statistics do not contain message content or identify individual users.

The application system now requires applicants to accept a privacy notice before answering any questions. Application information is available only to authorised moderators of the originating server and is retained for no more than 14 days after submission. An undecided application is automatically denied at that deadline. All application content is deleted at the deadline whether the application was accepted, denied, or automatically denied.

The appeal system also requires a privacy-policy agreement before an appeal can be submitted. The separate option allowing an appellant to contribute their own moderated message to improving the local model is optional and unchecked by default. It cannot be used for another person's message.

No human user data has been used for model training. If a user selects this option in the future, only their own moderated message may be retained, for no more than 90 days after the appeal decision, after which it is deleted.

Content from an automatically moderated message is currently retained for up to 30 days so authorised moderators can see what was acted upon and why, and so the affected user can appeal. If an appeal is submitted before the 30-day deadline, retention extends until the appeal is decided. An appeal that remains undecided is automatically rejected after 10 days, so an appeal submitted on day 29 could extend the maximum period to 39 days. Once a decision is made, or the 10-day appeal period expires, the content is deleted unless the user gave the separate AI-improvement permission described above.

Automated moderation decisions and new appeals are recorded in the originating server's configured moderation-log channel where applicable. StoaMod does not include the original moderated message in moderator notification DMs.

The Members tab remains restricted to authorised moderators of the originating server. It displays account and server-membership information already available to those moderators through Stoa. It is not public and is not shared with moderators of other servers. I have retained this feature because I have not yet received a response on the member-profile question raised in my previous correspondence.

The revised privacy policy has been published at <https://modstoa.site/privacy>. It now explains the information used by CAPTCHA, moderation, appeals, applications, statistics, and other operational features, along with their applicable retention periods.

When a server configures StoaMod, the dashboard explains the processing performed by relevant optional features. Where individual consent is required, StoaMod obtains it within the applicable end-user flow and does not treat server-owner configuration as a substitute for that consent.

The two points on which I would still appreciate written confirmation are:

1. Whether the moderated-content schedule described above—30 days ordinarily and up to 39 days when an appeal is submitted on day 29 is acceptable for moderator review and appeals.
2. Whether the restricted, origin-server Members tab may remain in its current form.

If either implementation is not acceptable, please state the required change or maximum retention period and I will address it promptly.

Please confirm whether this email provides sufficient evidence of completion.

Kind regards,

Kurome#0606

StoatMod