Kaltura's AI-powered moderation service

Last Modified on 08/12/2026 6:57 pm IDT

This article is designated for administrators.

About

Kaltura’s AI-powered moderation service helps you review video content more easily by automatically checking it for potentially inappropriate or sensitive material.

Instead of watching every video from start to finish, you get a moderation report that shows what might need attention and where it appears in the video.

The moderation service is available through Kaltura's Agentic Workflows.

How moderation works

The moderation service reviews video content automatically and highlights parts that may need attention. It checks both what appears on screen (visual moderation) and what is being said in the video (verbal moderation).

Visual moderation uses computer vision to analyze images and scenes in the video, while **verbal moderation** evaluates the video transcript using an LLM.

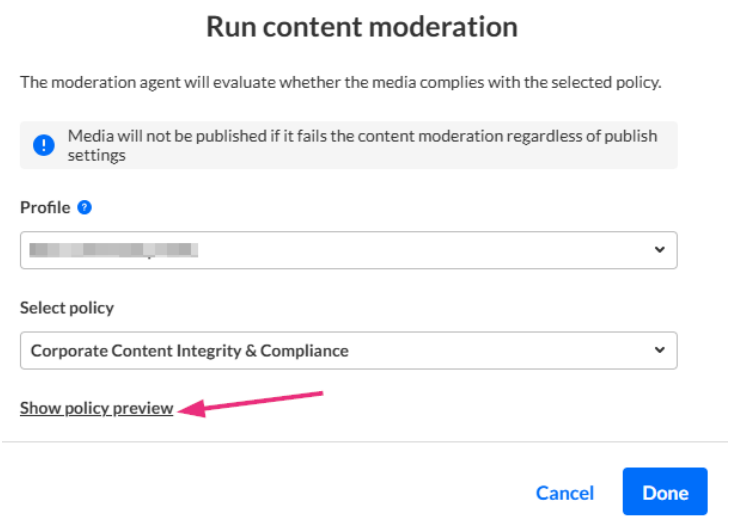
Content is evaluated against moderation policies, either Kaltura’s default policies or your organization’s customized guidelines. Visual and verbal moderation are evaluated separately, and each requires its own policy.

Creating customized policies requires Professional Services hours.

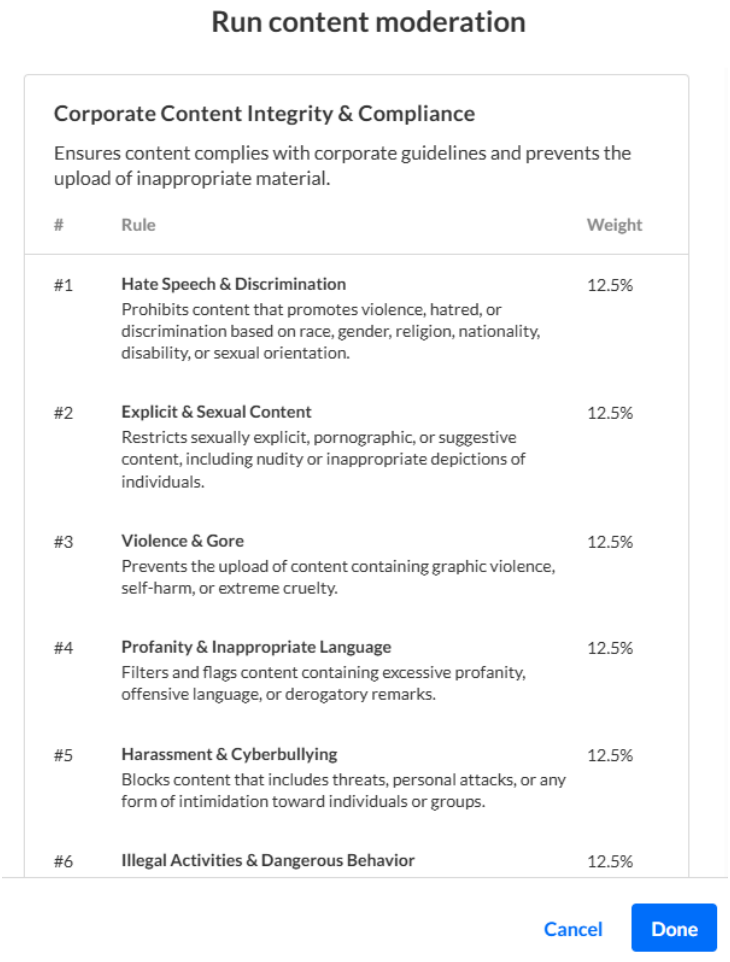
Moderation policy examples

When configuring the **Run content moderation** action, you can click **Show policy preview** to see what the selected policy checks for.

Copyright © 2025 Kaltura Inc. All Rights Reserved. Designated trademarks and brands are the property of their respective owners. Use of this document constitutes acceptance of the Kaltura Terms of Use and Privacy Policy.

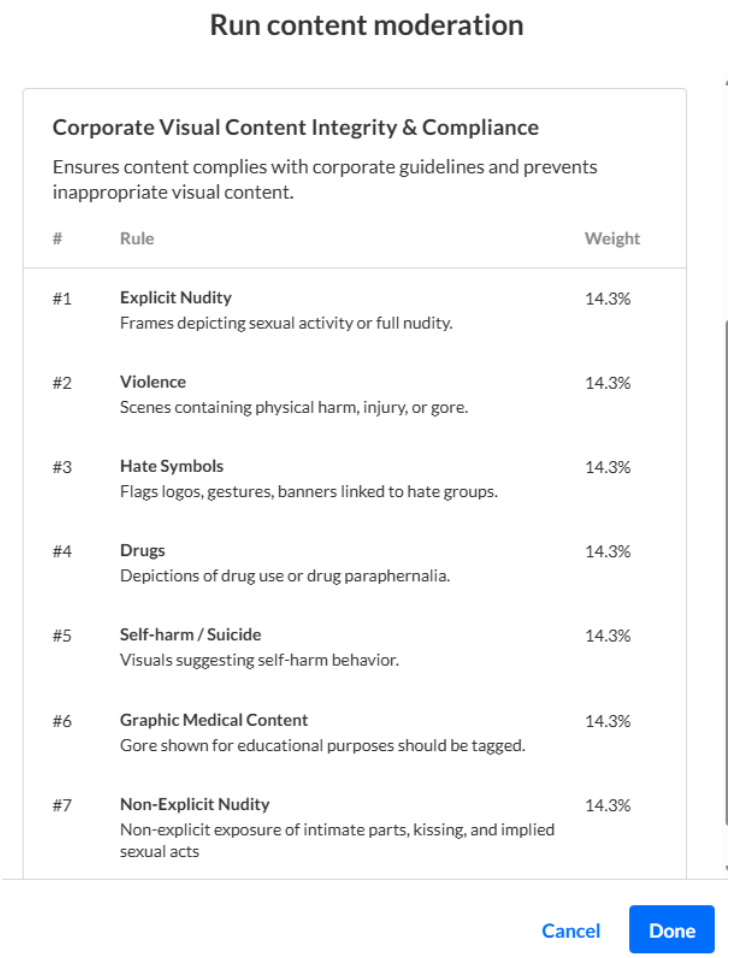


Below is a preview of a default **Corporate Content Integrity & Compliance** policy.



Below is a preview of a default **Corporate Visual Content Integrity & Compliance** policy.

Copyright © 2025 Kaltura Inc. All Rights Reserved. Designated trademarks and brands are the property of their respective owners. Use of this document constitutes acceptance of the Kaltura Terms of Use and Privacy Policy.



Reviewing results

After a video is reviewed, the service generates a moderation report.

The report includes:

- Up to 10 findings per policy rule
- Timestamps showing where each issue starts in the video
- Results for visual moderation, verbal moderation, or both

This helps reviewers jump directly to the relevant parts of the video.

Based on the moderation report, you can decide whether to approve or block the content manually or apply an automated workflow if one is configured.

- Moderation results are generated automatically by AI and may contain errors. The service highlights parts of the video that may need review, but a human reviewer always makes the final decision.
- Responsibility for approving or blocking content rests with the licensee.

Copyright © 2025 Kaltura Inc. All Rights Reserved. Designated trademarks and brands are the property of their respective owners. Use of this document constitutes acceptance of the Kaltura Terms of Use and Privacy Policy.

How to use the moderation service

The moderation service is run through Kaltura's Agentic Workflows.

You can:

- add 'Run content moderation' as a workflow action
- choose the relevant moderation policy
- review results after the workflow runs

For step-by-step instructions, see Create an agentic workflow.

Copyright © 2025 Kaltura Inc. All Rights Reserved. Designated trademarks and brands are the property of their respective owners. Use of this document constitutes acceptance of the Kaltura Terms of Use and Privacy Policy.