

California Terms of Service Report

January 1, 2024 (reporting period July 1 – September 30, 2023)

At Microsoft, we believe digital safety is a shared responsibility requiring a whole-of-society approach. This means that the private sector, academic researchers, civil society, and governmental and intergovernmental actors all work together to address challenges that are too complex – and too important – for any one group to tackle alone.

For our part, we prohibit certain content and conduct on our services, and we enforce rules that we’ve set to help keep our customers safe. We use a combination of automated detection and human content moderation to remove violating content and suspend accounts. Additional information is available on [Microsoft’s Digital Safety site](#).

Microsoft provides the information below for GroupMe and the consumer version¹ of Microsoft Teams (“Teams”) in response to California Business & Professions Code Section 22677(a).

1. The current version of the terms of service.

The current versions of Microsoft’s terms of service, as defined under Section 22676(b), are available at the following links:

- Microsoft Service Agreement (“MSA”): <https://www.microsoft.com/en-us/servicesagreement>
- Microsoft Code of Conduct (“Code of Conduct”): https://www.microsoft.com/en-us/servicesagreement#3_codeOfConduct
- Microsoft Content and Conduct Policies (“Policies”): <https://www.microsoft.com/en-us/DigitalSafety/policies>
- How we enforce our Code of Conduct: <https://www.microsoft.com/en-us/DigitalSafety/moderation-and-enforcement>
- How to report inappropriate content, messages, or safety concerns: <https://www.microsoft.com/en-us/DigitalSafety/report-a-concern>

The Code of Conduct, which is part of the MSA, explains what content and conduct are prohibited when using GroupMe, Teams and other in-scope Microsoft services. Policies provides users additional detail on the content prohibited by the Code of Conduct.

GroupMe and Teams have additional terms, guidelines and policies governing user content and conduct. GroupMe’s and Teams’ additional terms (collectively, “Community Guidelines”) are available at the following links:

¹ Note: [placeholder for clarification note on how exactly the business describes Teams Consumer]

- **GroupMe Community Guidelines:** <https://support.microsoft.com/en-us/office/learn-more-about-groupme-community-guidelines-26e268c8-e941-4b5a-9bd7-711559a70be8?culture=en-us&country=us>
 - **Teams Community Guidelines:** <https://support.microsoft.com/en-us/office/learn-more-about-microsoft-teams-free-community-standards-dfcf8b69-6566-4182-a9aa-89f523714f84>
2. **If the company has filed its first report, a complete and detailed description of any changes to the terms of service since the previous report.**

N/A

3. **A statement of whether the current version of the terms of service define each of the following categories of content, and, if so, the definition of those categories: (A) Hate speech or racism; (B) Extremism or radicalism; (C) Disinformation or misinformation; (D) Harassment; (E) Foreign political interference.**

A. *Hate speech or racism*

The category of “Hate speech and racism” is prohibited in the Code of Conduct, which reads:

Don’t engage in activity that is harmful to you, the Services, or others (e.g., transmitting viruses, stalking, trying to generate or sharing content that harasses, bullies or threatens others, posting terrorist or violent extremist content, communicating hate speech, or advocating violence against others).

Our Policies provide additional details and examples:

Hate speech

Microsoft wants to create online spaces where everyone can participate and feel welcome.

We do not allow hateful content that attacks, insults, or degrades someone because of a protected trait, such as their race, ethnicity, gender, gender identity, sexual orientation, religion, national origin, age, disability status, or caste.

Hate speech includes:

- Promoting harmful stereotypes about people because of a protected trait.
- Dehumanizing statements, such as comparing someone to an animal or other non-human, because of a protected trait.
- Encouraging or supporting violence against someone because of a protected trait.
- Calling for segregation, exclusion, or intimidation of people because of their protected trait.
- Symbols, logos, or other images that are recognized as communicating hatred or racial superiority.

GroupMe and Teams also prohibit “Hate speech and racism” in their Community Guidelines, building on the information provided in the Code of Conduct and Policies:

- **Our Shared Values:** “Hate has no place here.”

- **Do your part to keep everyone safe:** “For example, don’t: ... [c]reate or operate a “hate group” aimed at a particular individual or group.”

Support a welcoming and inclusive community

Harassment and hate take many forms, but none have a home here. To make GroupMe a place for everyone, and to prevent people from feeling uncomfortable or unwelcome, we all need to be stewards. This means more than just not harassing other users—it means embracing them. It means taking particular care to keep in mind how other users might interpret your content.

For example, don’t:

- Use derogatory language to refer to others.
- Make fun of other people’s identities or personal traits.
- Use language that operates to exclude or demean others.
- Turn what should be a comment about a user’s content into a personal attack.
- Create a group dedicated to tearing down another individual or group.

B. Extremism or radicalism

The category of “Extremism or radicalism” is prohibited in the Code of Conduct, which reads:

Don’t engage in activity that is harmful to you, the Services, or others (e.g., transmitting viruses, stalking, trying to generate or sharing content that harasses, bullies or threatens others, **posting terrorist or violent extremist content**, communicating hate speech, or advocating violence against others).

Our Policies provide additional details and examples:

Terrorism and Violent Extremism

At Microsoft, we recognize that we have an important role to play in preventing terrorists and violent extremists from abusing online platforms. We do not allow content that praises or supports terrorists or violent extremists, helps them to recruit, or encourages or enables their activities. We look to the United Nations Security Council’s [Consolidated List](#) to identify terrorists or terrorist groups. Violent extremists include people who embrace an ideology of violence or violent hatred towards another group.

In addressing terrorist and violent extremist content, we also work to ensure that people can use our services to talk about terrorism or violent extremism, share news or research about it, or express opposition to it.

GroupMe and Teams also prohibit this content in their Community Guidelines, building on the information provided in the Code of Conduct and Policies:

- **Keep it legal:** “For example, don’t: ... [p]romote or praise terrorists. Organize violent gatherings.”
- **Harmful behavior has no place on GroupMe:** “For example, don’t: ... [e]ncourage terrorism (such as by posting instructions for making bombs or other weapons.”

C. *Disinformation or misinformation*

The Code of Conduct does not explicitly define disinformation or misinformation, but prohibits the following:

Don't engage in activity that is fraudulent, false or misleading (e.g., asking for money under false pretenses, impersonating someone else, creating fake accounts, automating inauthentic activity, generating or sharing content that is intentionally deceptive, manipulating the Services to increase play count, or affect rankings, ratings, or comments).

The Community Standards discourage users from sharing disinformation or misinformation while using the services:

- “[m]isinformation has no place” on the services
- “[i]f you’re looking for a place on the internet ... to share misinformation, GroupMe [and Teams] isn’t the place for you.”

D. *Harassment*

The Code of Conduct prohibits harassment:

Don't engage in activity that is harmful to you, the Services, or others (e.g., transmitting viruses, stalking, **trying to generate or sharing content that harasses, bullies or threatens others**, posting terrorist or violent extremist content, communicating hate speech, or advocating violence against others).

Our Policies provide additional details and examples:

Bullying and Harassment

Microsoft seeks to create a safe and inclusive environment where you can engage with others and express yourself free from abuse. We do not allow content or conduct that targets a person or group with abusive behavior. This includes any action that:

- Harasses, intimidates, or threatens others.
- Hurts people by insulting or belittling them.
- Continues contact or interaction that is unwelcome, especially where contact causes others to fear injury.

GroupMe and Teams also address “Harassment” in their Community Guidelines, building on the information provided in the Code of Conduct and Policies:

- **Conduct:** “Each of you deserves a place to be yourself with confidence, free from bullying, hatred, and harassment.”

Support a welcoming and inclusive community

Harassment and hate take many forms, but none have a home here. To make GroupMe a place for everyone, and to prevent people from feeling uncomfortable or unwelcome, we all need to be stewards. This means more than just not harassing other users—it means embracing them. It means taking particular care to keep in mind how other users might interpret your content.

For example, don't:

- Use derogatory language to refer to others.
- Make fun of other people's identities or personal traits.
- Use language that operates to exclude or demean others.
- Turn what should be a comment about a user's content into a personal attack.
- Create a group dedicated to tearing down another individual or group.

(Substantially similar provisions are found in Teams' Community Standards)

Be respectful with your words

It's okay to disagree with fellow members of your group but keep your tone respectful and your language clean. Hate has no place here, and it's not ok for your sharing of opinions to turn into harassment. Harassment includes any negative behavior that's personalized, disruptive, or likely to make someone feel unwelcome or unsafe. To qualify as harassment, the behavior doesn't have to be drawn-out or persistent. Even a single abusive message could harm someone's experience. Know when to draw the line, and when to back off. Know and respect your group members.

For example, acceptable sharing of a difference in opinion might include:

- Not sure how you came to this conclusion, bro. Try again.
- Thanks for sharing your opinion, but I don't agree.
- This does not seem accurate. Could you please share your sources? Misinformation has no place here.
- This group is not for me; I am going to remove myself.

Harassment looks like:

- This is the [profanity] idea I've ever heard. Get a brain, [racial slur].
- How about we meet in person to settle this one? I know where you live.
- I cannot believe someone as [slur against those with disabilities] as you are allowed to be in this group. KYS.
- This [profanity] university is a joke. You will all be sorry soon.
- These [homophobic slurs] are ruining this country. Anyone who agrees, come join my group; your "offensive" jokes will be safe there.

(Substantially similar provisions are found in Teams' Community Guidelines.)

E. Foreign political interference

There is no policy expressly defining foreign political interference.

4. Description of the content moderation practices used.

Microsoft uses a combination of automated technology, user reporting and trained human reviewers to find or respond to reports of, and take action against, content or conduct in GroupMe and Teams that violates the Code of Conduct, Policies or Community Guidelines.

Microsoft has systems and processes in place to make sure we are fair and consistent when enforcing our collective terms. This helps protect our users while also respecting their rights. For more information on how Microsoft detects, responds to reports, reviews and actions potentially prohibited content for its services like GroupMe and Teams, please see "How Microsoft enforces its policies."

A. Any existing policies intended to address the categories of content described in Cal. Business & Professions Code Section 22677(a)(3).

Please see the response in Section 3 above.

B. How automated moderation systems enforce terms of service of the social media platform and when these systems involve human review.

GroupMe

GroupMe uses a hash-matching technology known as PhotoDNA to proactively identify known instances of terrorist and violent extremist content ("TVEC"), which is prohibited by the Code of Conduct, Policies and GroupMe's Community Guidelines. If PhotoDNA detects potential TVEC, it is subsequently reviewed by a human reviewer to determine the appropriate response action. For more information on PhotoDNA, please see: <https://www.microsoft.com/en-us/photodna>.

Teams

Teams use PhotoDNA to proactively identify known instances of TVEC, which are prohibited by the Code of Conduct, Policies and Teams' Community Guidelines. If PhotoDNA detects potential TVEC, it is subsequently reviewed by a human reviewer to determine the appropriate response action.

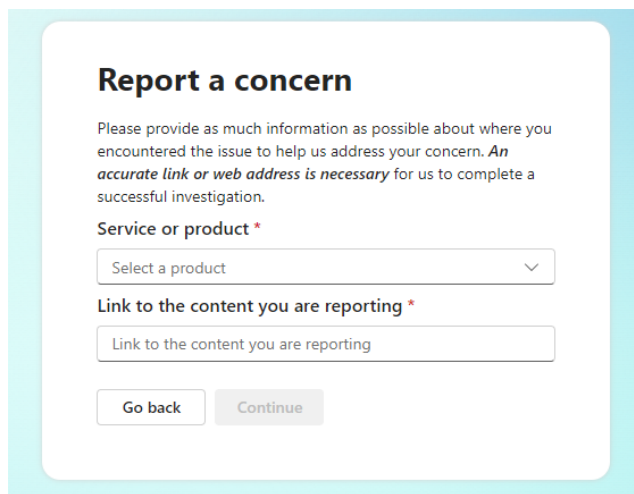
Teams uses additional automated content detection systems to detect other potential violations of the Code of Conduct and Teams' Community Guidelines (e.g., adult content and gore) that may be found in a Community's name or description, or user generated content posted within a Community.

C. How the company responds to user reports of violations of the terms of service.

GroupMe: users can report potentially prohibited content/conduct in the GroupMe service. For more information on how users can report in GroupMe, please see: <https://support.microsoft.com/en-us/office/how-do-i-report-abuse-in-groupme-0b825cde-a73a-4576-830d-4597289d78ca>.

Teams: users can report potentially prohibited content/conduct in the Teams service. For more information on how users can report in Teams, please see: <https://support.microsoft.com/en-us/office/report-a-concern-in-microsoft-teams-free-2e2ea20c-2866-4b65-a979-8132c02dc231>.

Microsoft Report a Concern: If a user visits [Microsoft's centralized Report a Concern portal](#) to report potentially prohibited content or conduct, they are redirected to the reporting portals for GroupMe and Teams referenced above.



The screenshot shows a web form titled "Report a concern". Below the title is a paragraph of instructions: "Please provide as much information as possible about where you encountered the issue to help us address your concern. An accurate link or web address is necessary for us to complete a successful investigation." There are two required input fields: "Service or product *" which is a dropdown menu with "Select a product" and a downward arrow, and "Link to the content you are reporting *" which is a text input field with the placeholder text "Link to the content you are reporting". At the bottom of the form are two buttons: "Go back" and "Continue".

When users report content or conduct, those reports are sent for human review to determine whether it is violative. If a violation is found, Microsoft actions the content consistent with its policies.

D. How the company would remove individual pieces of content, users, or groups that violate the terms of service, or take broader action against individual users or against groups of users that violate the terms of service.

Violating Microsoft's Code of Conduct or Policies, or service-specific Community Guidelines may result in a range of actions against specific content and/or a responsible Microsoft account holder. For example, we may remove user content or place restrictions on a Microsoft account holder's access to service features. For egregious or repeat violations, a Microsoft account may be suspended temporarily or permanently. For certain egregious violations of our policies (e.g., CSEAI), we may permanently suspend a Microsoft account after a single violation.

In making enforcement decisions, we strive to apply our policies objectively and proportionately, which means we work to make the corrective action fit the violation. For example, some violent or graphic images may be newsworthy or important for educational or research purposes, and we take into account these factors when reviewing content.

More information on how Microsoft enforces its terms and policies is included in the Community Guidelines, and also published at <https://www.microsoft.com/en-us/DigitalSafety/moderation-and-enforcement>

- **The languages in which the platform does not make terms of service available, but does offer product features, including but not limited to, menus and prompts.**

Teams: The Code of Conduct and Policies are offered in all major world languages, including English and Spanish. Community Standards identified in response to Question 1 are available in English. Teams is offered in multiple languages, globally.

GroupMe: The Code of Conduct and Policies are offered in all major world languages, including English and Spanish. Community Standards identified in response to Question 1 are available in 36 languages, including English and Spanish. GroupMe is offered in multiple languages, globally.

5. Content moderation activity information.

Below is the content moderation activity data Microsoft possesses that is responsive to Section 22677(a)(5). The data addresses tracked content moderation activity for GroupMe and Teams² worldwide,³ during the reporting period (July 1 – September 20, 2023).

A. Number of flagged items of content.

The table below reports the number of flagged items of content during the reporting period, disaggregated by the manner in which the content was flagged (i.e., by user reporting or by automated content detection), applicable policy, and media type.

Table 1

Product	Method of Flagging	Policy	Number of flagged items of content					Total
			Image	Text	Video	URLs	Other*	
TEAMS	User reported	Hate speech	88	597	8	-	12	705
		Terrorism and violent extremism	49	173	3	-	3	228
		Bullying and harassment	148	488	11	-	6	653
	Automated Detection Tool	Hate speech	-	8,158	-	-	-	8,158
		Terrorism and violent extremism	-	58	-	-	-	58
		Bullying and harassment	-	3,898	-	-	-	3,898
GroupMe	User reported	Hate speech	-	-	-	-	-	-
		Terrorism and violent extremism	16,098	16,533	165	12	-	32,808
		Bullying and harassment	48,167	47,424	434	18	-	96,043

*Other refers to instances of user reported flagged content where we were unable to determine the content type, or it was a non-supported type like PDF, Word, and Excel.

B. Number of actioned items of content.

The table below reports the number of items of content GroupMe and Teams actioned under the applicable policies during the reporting period. For Teams, there were no actioned items as a result of content reported by users. The actioned items for Teams were both flagged and actioned by automated content detection tools. The actioned items for GroupMe were originally flagged by user reporting and actioned by human reviewers.

Table 2

² Note: for Teams, the data below constitutes content moderation activity from Teams Community.

³ GroupMe and Teams presently are unable to break down content moderation activity by geography.

Product	Policy	Number of Actioned items of content					Total
		Image	Text	Video	URLs	Other	
TEAMS	Hate speech	2	8,037	-	-	-	8,039
	Terrorism and violent extremism	-	28	-	-	-	28
	Bullying and harassment	-	17	-	-	-	17
GroupMe	Hate speech	-	-	-	-	-	-
	Terrorism and violent extremism	58	491	10	-	-	559
	Bullying and harassment	120	1,782	21	-	-	1,923

C. Number of actioned items of content that resulted in action taken against the user or group of users responsible for the content.

Teams and GroupMe is not aware of any action taken against the user or group of users responsible for the content included in Table 2 above. For the policies referenced in Table 1 and Table 2, if content is determined to be violative, it can be removed in accordance with our policies.

D. Number of actions items of content that were removed, demonetized, or deprioritized.

All content items referenced in Table 2 above were removed. GroupMe and Teams does not demonetize or deprioritize content items. Both services are provided to consumers free of charge. Neither service contains advertising, nor do they recommend or suggest user content to users.

E. Number of views, viewers, and reshares of the actioned items of content.

As legacy communications systems that were not traditionally social media, neither GroupMe nor Teams has tracked views, viewers or reshares.

F. Number appeals and reversals of actioned items of content.

Teams and GroupMe are not aware of any user appeals as a result of actioned items of content included in Table 2 above.