

Microsoft EU Code of Conduct on Countering Illegal Hate Speech Online +

Introduction

Microsoft Ireland Operations Limited (“Microsoft”) – the provider of Microsoft services in the European Union (EU) – welcomes the opportunity to submit its first report under the EU Code of conduct on countering illegal hate speech online + (the “Code”). This report is aligned with Regulation (EU) 2022/2065 (Digital Services Act) and describes Microsoft’s policies, processes, and measures for addressing hateful conduct (which Microsoft defines as content or conduct that attacks, denigrates, intimidates, or degrades individuals or groups because of one or more of their protected traits) on the following Microsoft Consumer Hosted Services : Xbox, GroupMe, and Microsoft Teams Consumer (“Teams”). This report covered the period from 1 July 2025 to 31 December 2025.

1. Terms and conditions for addressing hateful conduct

Microsoft’s approach to addressing hateful conduct is grounded in its publicly available [Code of Conduct](#), as set out in the Microsoft Services Agreement (“MSA”), which applies to all users and sets out prohibited and unlawful content and conduct, and expectations for user behavior when accessing Microsoft services, like Xbox, GroupMe, and Teams. More information about Microsoft’s policies may be found [here](#). “Hate speech” is expressly prohibited by the Code of Conduct. “Hate speech” is also expressly prohibited by the clear and publicly available service-specific terms and Community Standards for Xbox, GroupMe, and Teams. No significant changes were made to the relevant terms and conditions during the reporting period.

- **Xbox:** [Xbox Community Standards | Xbox](#)
- **GroupMe:** [GroupMe Community Guidelines](#)
- **Teams:** [Microsoft Teams Community Standards](#)

2. Review and possible removal of or disabling access to hateful conduct

Identification of Hateful Conduct

Microsoft has a robust reporting mechanism permitting any person to quickly and easily report user content or conduct for a variety of potential harms, including hateful conduct. Microsoft's [centralized Report A Concern function](#) allows users to report content that violates Microsoft's Code of Conduct, as well as any service-specific policies and/or community standards. Users can report hateful conduct directly within in-service experiences across supported devices and access points for Xbox, GroupMe, and Teams without contacting the other party.

Xbox: Players can report hateful conduct using built-in reporting tools available on console, PC, and mobile, directly within the chat experience, for individual messages and other in-scope content. Users can select "hate speech" as the "report category," then complete the remainder of the Report A Concern form to report the potential violative content or conduct. More information is available [here](#). Xbox also uses automated detection to identify, block, and address potentially violative content. Automated tools classify violating text and shared media and may prevent it from appearing on the service to reduce user exposure.

GroupMe and Teams: Users can report hateful conduct via Report a Concern dialogs on mobile applications (iOS and Android) and browser, directly from chats, for individual messages, profiles, or other in-scope content. Users right-click the message to be reported, select "Report a Concern," select "hateful conduct" as the type of concern, then click "send." More information on GroupMe is available [here](#); more information on Teams is available [here](#). Additionally, Teams is piloting an automated text-detection capability, beginning in Communities, to help surface potentially violative content for further review.

Moderation and Enforcement

Hateful conduct identified by user reporting or proactive, automated-content detection is reviewed by human moderators as part of Microsoft's centralized moderation operations. Human reviewers consider context, intent, and applicable legal thresholds. Moderation operations are centralized, with support for reports submitted in languages commonly used across EU Member States, and escalation paths are available for complex or sensitive cases. Where content is found to violate the MSA, service-level policies and Community Standards, or applicable local law, proportionate enforcement action are taken, which may include content removal, account level enforcement, or other appropriate actions. More information about Microsoft's moderation and enforcement policies can be found [here](#).

Xbox: All user reports go to an Xbox moderation team, where they are reviewed against Xbox's Community Standards and the MSA, including hateful conduct. Violating content is

removed and appropriate actions are taken against the player. Players are informed about enforcement and removal, with instructions for appeal for eligible decisions through a dedicated channel. Content and conduct flagged by automated content detection is prioritized through structured workflows and undergoes a similar review process. Automation supports triage, while trained trust and safety reviewers make context-dependent decisions and assess content against Xbox's Community Standards and the MSA, as well as relevant EU law, where applicable. Review is supported across multiple EU languages and by tools that improve reviewer efficiency and reduce exposure to harmful content, while preserving due process and human judgment.

GroupMe and Teams: All user reports enter a centralized moderation pipeline. All reports are assessed against the applicable services' Community Standards and Guidelines and the MSA. Moderators assess whether the content constitutes prohibited content/conduct, including hateful conduct and whether enforcement is required under the relevant policy. Complainants are notified of enforcement decisions (irrespective of whether it is found to be hateful conduct or not). The reported user is notified of enforcement decisions if an action is taken. Appeal mechanisms are available in accordance with each service's moderation processes. Content surfaced through Teams' automation content detection is reviewed by trained trust and safety reviewers against applicable policies and standards.

Detection and Review Metrics (EU)

During the period covered by this report (1 July 2025 - 31 December 2025)

	# of Content automatically detected and prevented	# of content reported by players and removed	# of Content appeals submitted for content action	# of appealed content that were restored	Average review completion time from notice received
Xbox	1,817,013	12,958	2,586	520	5.25 hours

	# of content detected via automated means	# of content removed via automated means or following a user report	# of Content appeals submitted for content action	# of appealed content that were restored	Average review completion time from notice received
GroupMe	0	2	0	0	8.89 hours
Teams	2	243	5	0	15.96 hours

3. Transparency, accountability, and monitoring

Microsoft did not receive any reports from Monitoring Reporters under the Code during the monitoring period.

4. Intra-industry and multi-stakeholder cooperation to address and prevent risks of the spread of hateful conduct

Microsoft has long worked with civil society organisations across Europe to better understand and address online harms, including hateful conduct. One such partner is the **International Network Against Cyber Hate (INACH)**, a European NGO that brings together organisations from across the region to monitor, analyse, and counter illegal and harmful hateful conduct online, while promoting human rights and freedom of expression. We have consistently supported INACH's work by sponsoring the annual INACH summits, engaging in dialogue, contributing to multi-stakeholder discussions, and supporting spaces where civil society, policymakers, and platforms can exchange evidence and best practices.

No discussions relating to a 'crisis' as defined under Article 36 of the DSA took place for the in-scope services during the reporting period.

5. Awareness raising, online civility and counter narrative initiatives

Microsoft promotes safe use of its services, including Xbox, GroupMe, and Teams, through clear and publicly available Community Standards or Guidelines grounded in the MSA, which prohibit hateful conduct and set expectations for respectful behavior. These standards are referenced within each service's reporting experience to help users understand prohibited content and the reporting options available to them. Microsoft also provides in-service reporting tools across these services that enable users to report harmful content when encountered, supporting community-led safety and enforcement practices. Currently, Xbox, GroupMe, and Teams do not operate standalone awareness-raising campaigns or counter-narrative initiatives specific to hateful conduct.

Microsoft also supports a range of digital literacy and citizenship initiatives, including Minecraft Education's [Cyber Safe](#), which aims to cultivate online safety and citizenship skills for young people through gaming. For example, CyberSafe: Good Game is focused on amplifying positive connections online. We also provide a version of our policies aimed at [young people](#).