



6. TUMBLR

- **Release/Launch date:** 2007
- **Number of Active Visitors:** 292 million¹⁸⁷
- **Short Overview of Moderation Process:** Content moderators remove blogs and re-blogs that have been reported by users/visitors or flagged by Artificial Intelligence
- **Signatory to the EU'S Code of Conduct on Illegal Hate Speech Online:** No

¹⁸⁷ "Worldwide visits to Tumblr.com from December 2021 to May 2022," *Statista*, <https://www.statista.com/statistics/261925/unique-visitors-to-tumblrcom/> (Accessed on 21 December 2022).

Key Developments

Terms of Service

The first traceable Tumblr Terms of Service are from March 2007, but they did not include a provision on hate speech until October 2007. At that point, the Terms required subscribers to agree not to post content that is “hateful.”¹⁸⁸ In 2010, Tumblr added an additional note on hate speech: “hate content don’t belong on the web, and certainly don’t belong on Tumblr.”¹⁸⁹ By March 2012, Tumblr had removed both references to hateful content from the Terms and added policies that fall within the scope of this report the company’s new Community Guidelines. In 2022, however, Tumblr added a reference to hate speech back into the Terms of Service, instructing users to “not give tips in exchange for – or to promote or encourage – content that is against our Community Guidelines or that is otherwise illegal, abusive towards others, hateful, or that could result in self-harm.”¹⁹⁰ The hate speech provisions in Tumblr’s Terms of Service have never referenced any specific protected characteristics, however.

Community Guidelines

In March 2012, Tumblr created a Content Policy that prohibited “hate speech and other objectionable content that is unlawful, defamatory, and fraudulent.”¹⁹¹ Within the same month, however, Tumblr launched the Community Guidelines, which began by stating that Tumblr as a “global platform for creativity and self-expression” that was “deeply committed to supporting and protecting freedom of speech.”¹⁹² However, in the same paragraph, the company noted that it drew “lines around a few narrowly-defined but deeply important categories of content and behavior that jeopardize our users, threaten our infrastructure, or damage our community.” One such category of prohibited content was “malicious bigotry,” or actively promoting “violence or extreme hatred against individuals or groups” based on eight different protected characteristics (see Figure 11). This provision implicitly equated hate speech and “malicious bigotry,” defining them as the promotion of violence or hatred against specific identity groups. However, Tumblr did not provide any specifics about what actively promoting hatred would look like.

¹⁸⁸ https://web.archive.org/web/20071229031909/http://www.tumblr.com:80/terms_of_service

¹⁸⁹ https://web.archive.org/web/20100522002538/http://www.tumblr.com/terms_of_service

¹⁹⁰ <https://web.archive.org/web/20220301053331/https://www.tumblr.com/policy/en/terms-of-service>

¹⁹¹ <https://web.archive.org/web/20120319173638/http://www.tumblr.com/policy/en/community>

¹⁹² <https://web.archive.org/web/20120731142112/http://www.tumblr.com/policy/en/community>

*Figure 11*¹⁹³

What Tumblr is not for:

- **Malicious Bigotry.** Don't actively promote violence or extreme hatred against individuals or groups, on the basis of race, ethnic origin, religion, disability, gender, age, veteran status, or sexual orientation. While we firmly believe that the best response to hateful speech is not censorship but more speech, we will take down malicious bigotry, as defined here.

By January 2014, Tumblr had changed the title of this provision from “Malicious Bigotry” to “Malicious Speech,” and lowered the threshold from the “active promot[ion] of violence or extreme hatred” to the “encourage[ment] of violence or hatred.”¹⁹⁴ Tumblr still provided no additional details about what encouraging hatred might look like. Interestingly, Tumblr also removed the reference to the company’s firm belief that the best response to hateful speech is more speech and, instead encouraged users to “dismantle negative speech through argument rather than censorship.” In 2016, Tumblr added more details about what it meant to encourage violence or hatred, telling users not to “make violent threats or statements that incite violence, including threatening or promoting terrorism,” especially if such content threatened people based on nine protected characteristics.¹⁹⁵ Tumblr revised its policy in 2018 (see Figure 12), however, when the company retitled the provision “hate speech” and instructed users not to “encourage violence or hatred,” nor “post content for the purpose of promoting or inciting the hatred of, or dehumanizing, individuals or groups based on race, ethnic or national origin, religion, gender, gender identity, age, veteran status, sexual orientation, disability, or disease.” The 2018 policy also explicitly noted that content “might be offensive without necessarily encouraging violence or hatred.” The policy has not changed since.

¹⁹³ <https://web.archive.org/web/20120731142112/http://www.tumblr.com/policy/en/community>

¹⁹⁴ <https://web.archive.org/web/20140702023313/http://www.tumblr.com/policy/en/community>

¹⁹⁵ <https://web.archive.org/web/20160702110539/https://www.tumblr.com/policy/en/community>

Figure 12¹⁹⁶

- **Hate Speech.** Don't encourage violence or hatred. Don't post content for the purpose of promoting or inciting the hatred of, or dehumanizing, individuals or groups based on race, ethnic or national origin, religion, gender, gender identity, age, veteran status, sexual orientation, disability or disease. If you encounter content that violates our hate speech policies, please report it.

[Report hate speech](#)

Keep in mind that a post might be mean, tasteless, or offensive without necessarily encouraging violence or hatred. In cases like that, you can always block the person who made the post—or, if you're up for it, you can express your concerns to them directly, or use Tumblr to speak up, challenge ideas, raise awareness or generate discussion and debate.

Analysis of Policy Scope

Over time, Tumblr has added more details to its hate speech policies, but they have also increased in scope. The relevant provisions in the Terms of Service prohibited hateful content but never defined it, suggesting the policy could cover a wide range of content deemed to involve hatred. This lack of clarity is problematic in light of the legality requirement in Article 19 of the ICCPR. While the first version of the Community Guidelines prohibited the active promotion of hatred, by 2014, Tumblr had shifted to prohibiting the encouragement of hate, possibly representing a broader range of covered content. In 2018, Tumblr added a prohibition on content that dehumanizes individuals on the basis of protected characteristics, representing an additional expansion in the covered content. This prohibition is much broader than the mandatory prohibition set out Article 20 (2), as well as the permitted restrictions under Article 19, under the three-part test.

Additions to the list of protected characteristics have also represented expansions to the scope of Tumblr's hate speech policy. The hate speech provisions in the Terms of Service did not mention any protected characteristics, but the initial relevant policy in the Community Guidelines listed eight. Tumblr added "gender identity" in 2016 and "national origin" and "disease" in 2018. Tumblr's list is broader than the characteristics covered by Article 20(2), due to the inclusion of gender, gender identity, age, veteran status, sexual orientation, disability, and disease. Arguably specific restrictions of speech based on categories such as veteran status and age are not in conformity with Article 19.

¹⁹⁶ <https://web.archive.org/web/20181001144639/https://www.tumblr.com/policy/en/community>

Table 11

<i>Content Explicitly Covered by Tumblr 's Hate Speech Policies</i>		2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
	Hate(ful) speech/content	X	X	X	X	X	X	X					X	X	X	X	X	X
	Promotion of Hatred						X	X	X	X	X	X	X	X	X	X	X	X
	Support for Organized Hate (Including Symbols)																	
<i>On the basis of protected characteristics</i>	Incitement to or Threats of Violence						X	X	X	X	X	X	X					
	Attacks																	
	Statements of inferiority or content that demeans																	
	Dehumanization												X	X	X	X	X	X
	Expressions of contempt or disgust																	
	Calls for exclusion or segregation																	
	Discrimination																	
	Denying or mocking historical atrocities, or valorizing the perpetrators																	
	Slurs																	
	Harmful Stereotypes																	
	Conspiracy Theories																	
	Cursing																	

Table 12

Characteristics Protected in Tumblr's Hate Speech Policies																	
	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Total	0	0	0	0	0	8	8	8	8	9	9	11	11	11	11	11	11
Race						X	X	X	X	X	X	X	X	X	X	X	X
Ethnicity						X	X	X	X	X	X	X	X	X	X	X	X
National Origin												X	X	X	X	X	X
Religion						X	X	X	X	X	X	X	X	X	X	X	X
Gender						X	X	X	X	X	X	X	X	X	X	X	X
Color																	
Immigration Status																	
Sex																	
Gender Identity										X	X	X	X	X	X	X	X
Sexual Orientation						X	X	X	X	X	X	X	X	X	X	X	X
Age						X	X	X	X	X	X	X	X	X	X	X	X
Disability						X	X	X	X	X	X	X	X	X	X	X	X
Disease/ Medical Condition												X	X	X	X	X	X
Veteran Status						X	X	X	X	X	X	X	X	X	X	X	X
Occupation																	
Weight																	
Pregnancy																	
Caste																	
Victims of a Major Event																	
Socio-economic Status																	
Culture																	
Tribe																	

Notes: An X indicates the company's hate speech policies covered that protected characteristic for at least one month during the given year.

Changes in Enforcement Volume

Tumblr does not provide public information about content removals due to its Terms of Service or Community Guidelines.