



#### 4. SNAPCHAT

- **Release/Launch Date:** July 8, 2011
- **Number of Active Users:** 557 million <sup>157</sup>
- **Short Overview of Content Moderation Process:** Snapchat's content disappears within a 24-hour period. No information on whether this platform outsources content moderation is available. Content moderators review posts that have been flagged by AI, reported by users or non-users. Non-users can file a report available on Snapchat's Help Centre.
- **Signatory to the EU'S Code of Conduct on Illegal Hate Speech Online:** Yes

<sup>157</sup> "Most popular social networks worldwide as of January 2023, ranked by number of monthly active users," *Statista*, <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/> (accessed on April 30, 2023).

## Key Developments

### *Terms of Use*

Snapchat's first Terms of Use date to 2011, and they included a provision that addressed racist content. Under the heading "what you cannot do with the application," Snapchat prohibited "transmitting any material that could be considered racist, threatening or unlawful in any way."<sup>158</sup> According to the coding rules for this report, this provision addresses hate speech, because it prohibits material that discriminates on the basis of race – i.e. "racist" material. This provision was relatively narrow, since race was the only identity protected from discriminatory content. In 2013, Snapchat added another relevant provision to the Terms of Use, which directed users not to "send snaps that (their friends) don't want to receive (threats, harassment, racism etc.)."<sup>159</sup>

In 2014, Snapchat eliminated these provisions from its Terms of Use. The company removed the suggestion to avoid sending friends "threats, harassment, and racism" and deleted the prohibition on transmitting any material that could be considered racist. Instead, the company added a section titled "Prohibited Activities," which prohibited harassment and intimidation but made no reference to harassment or intimidation of individuals based on identity characteristics. The next year, however, Snapchat added a provision to the Terms of Use that explicitly prohibited content containing "hate speech," though it offered no definition of the concept.<sup>160</sup> In September 2021, Snapchat removed the hate speech provision from the Terms. Since then, Snapchat's Terms of Use have not addressed hate speech.

### *Community Guidelines*

Snapchat also has Community Guidelines, which date back to 2014. They provide a general overview of prohibited content on the platform. The initial versions of these Guidelines prohibited harassment, bullying, and threats, but there was no mention of prohibiting such content if it targeted individuals based on their identity. Thus, these provisions did not address hate speech specifically. By 2017, however, the company had an explicit prohibition on hate speech in the Guidelines, which read: "Hate Speech: Don't post content that demeans, defames, or promotes discrimination on the basis of race, ethnicity, national origin, religion, sexual orientation, gender, disability, or veteran status."<sup>161</sup>

Snapchat has expanded the scope of its hate speech provisions in recent years, however. In 2021, Snapchat began addressing hate speech under a Guideline titled "Terrorism, Hate Groups, and Hate Speech." The rule stated: "Hate speech or content that demeans, defames, or promotes discrimination or violence on the basis of race, color, caste, ethnicity, national origin, religion,

<sup>158</sup> <https://web.archive.org/web/20120711233922/http://www.snapchat.com/terms>

<sup>159</sup> <https://web.archive.org/web/20130417050031/http://www.snapchat.com/terms>

<sup>160</sup> <https://web.archive.org/web/20151122121643/https://www.snapchat.com/terms>

<sup>161</sup> <https://web.archive.org/web/20170127184718/https://support.snapchat.com/en-US/a/guidelines>

sexual orientation, gender identity, disability, or veteran status, immigration status, socio-economic status, age, weight or pregnancy status is prohibited.” While this revision did not update the 2017 definition of hate speech, it did cover seven additional protected characteristics. The rule also prohibited “hate groups,” though the company did not provide a definition of such organizations.

In early 2023, Snapchat published an in-depth explanation of its hate speech, terrorism, and violent extremism policy, as part of a series of Community Guidelines explainers.<sup>162</sup> In this brief, the company expanded upon its definition of hate speech. Snapchat explained that, in addition to content that demeans or promotes discrimination against individuals on the basis of protected characteristics, “hate speech also extends to the valorization of perpetrators—or the denigration of victims—of human atrocities (such as genocide, apartheid, or slavery)” and “the use of hate symbols, which means any imagery that is intended to represent hatred or discrimination toward others (including those featured in the [hate symbols database](#) maintained by the Anti-Defamation League).”

### **Analysis of Policy Scope**

Table 7 reveals that the scope of Snapchat’s hate speech policy has increased substantially over the past ten years. Until 2023, Snapchat appeared to define hate speech as content that discriminates, defames, or demeans based on protected characteristics, but the recent explainer suggests the company also considers content that praises the perpetrators of genocide, apartheid, or slavery to be hate speech, which is broader than the previous conceptualization. This definition is much broader than the definition implied by Article 20.

Table 8 illustrates that the scope of characteristics protected under Snapchat’s hate speech provisions has also expanded. In 2011, the platform simply banned racist content, but by 2021, the platform banned content that discriminated against or demeaned someone based on any one of fifteen different characteristics. Compared to the ICCPR, Snapchat’s hate speech prohibition covers a broad range of characteristics. Specifically, Snapchat’s protected characteristics include color, immigration status, gender identity, sexual orientation, age, disability, veteran status, socio-economic status, weight, and pregnancy, which are not covered in Article 20 (2). Arguably characteristics such as “socio-economic status”, “veteran status”, “pregnancy”, and “weight” would not satisfy the necessity test required to pass muster under Article 19 (3), just as “demeaning” speech and the “denying” of historical crimes and atrocities constitute protected speech under ICCPR.

<sup>162</sup> “Hateful Content, Terrorism, and Violent Extremism: Community Guidelines Explainer Series,” *Snap Privacy and Safety Hub, Transparency*, January 2023, <https://values.snap.com/privacy/transparency/community-guidelines/hateful-content> .

Table 7

<b>Content Explicitly Covered by Snapchat's Hate Speech Policies</b>		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	
Promotion of Hatred															
Support for Organized Hate (Including Symbols)														X	
<i>On the basis of protected characteristics</i>	Incitement to or Threats of Violence								X	X	X	X	X	X	
	Attacks														
	Statements of inferiority or content that demeans							X	X	X	X	X	X	X	
	Dehumanization														
	Expressions of contempt or disgust														
	Calls for exclusion or segregation														
	Discrimination	X	X	X	X	X		X	X	X	X	X	X	X	X
	Denying or mocking historical atrocities, or valorizing the perpetrators														X
	Slurs														
	Harmful Stereotypes														
	Conspiracy Theories														
	Cursing														

Table 8

Characteristics Protected in Snapchat's Hate Speech Policies													
	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
<b>Total</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>8</b>	<b>9</b>	<b>8</b>	<b>8</b>	<b>15</b>	<b>15</b>	<b>15</b>
Race	X	X	X	X			X	X	X	X	X	X	X
Ethnicity							X	X	X	X	X	X	X
National Origin							X	X	X	X	X	X	X
Religion							X	X	X	X	X	X	X
Gender							X	X					
Color											X	X	X
Immigration Status											X	X	X
Sex													
Gender Identity								X	X	X	X	X	X
Sexual Orientation							X	X	X	X	X	X	X
Age											X	X	X
Disability							X	X	X	X	X	X	X
Disease/ Medical Condition													
Veteran Status							X	X	X	X	X	X	X
Occupation													
Weight											X	X	X
Pregnancy											X	X	X
Caste											X	X	X
Victims of a Major Event													
Socio-Economic Status											X	X	X
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

In Q2 2019, Snapchat began reporting information on the volume of content actioned for violations of the Community Guidelines. While the scope of the content covered by Snapchat's policy did not change dramatically after this point, the number of protected characteristics rose from 8 to 15 in January 2021. According to Snapchat's Transparency Report, the company took enforcement action on 77,587 Snaps, or 1.4% of total content actioned, between July and December 2020.<sup>163</sup> Between January and July 2021, Snapchat took enforcement action on 121,639 Snaps, or 1.9% of total content actioned during the period.<sup>164</sup> It is possible that the expansion in the scope of Snapchat's policies is related to the increase in content actioned in Q1 2021. However, the content actioned number fell to 93,341, or 1.5% of all content actioned, in Q2 2021,<sup>165</sup> suggesting the bump in Q1 did not persist. This observation suggests changes in the policy scope alone cannot account for the increase in enforcement action between Q2 2020 and Q1 2021. As with the data from the other platforms, this data underscores the need to give external researchers access to platform data, so they can rigorously analyze the impact of changes in policy scope on content removals and other enforcement actions.

<sup>163</sup> "Transparency Report: July 1, 2020 - December 31, 2020," *Snap Inc.*, July 1, 2021, <https://www.snap.com/en-US/privacy/transparency/2020-12-31>.

<sup>164</sup> "Transparency Report: January 1, 2021 - June 30, 2021," *Snap Inc.*, November 22, 2021, <https://www.snap.com/en-US/privacy/transparency/2021-06-30>.

<sup>165</sup> "Transparency Report: July 1, 2021 - December 31, 2021," *Snap Inc.*, April 1, 2022, <https://www.snap.com/en-US/privacy/transparency/2021-12-31>.