



Google LLC
25 Massachusetts Ave., NW
Ninth Floor
Washington, D.C. 20001

June 12, 2024

United Nations Human Rights
Office of the High Commissioner
ohchr-registry@un.org

Dear Rapporteurs and Independent Experts,

Thank you for your recent correspondence sent pursuant to the communications procedure of the Special Procedures of the United Nations Human Rights Council. We express our sincerest apologies for a clerical error that inadvertently resulted in a lack of formal response to your earlier outreach.

We take our responsibility to respect human rights across all of our services very seriously, and we appreciate the opportunity to discuss Google's approach to removing hate speech and preventing its spread on our platforms. Below you will find information on our cross-company efforts to uphold international human rights standards, as well as details about specific content-related policies and processes.

1. Google's Commitment to Civil and Human Rights

In everything we do, including launching new products and expanding our operations around the globe, we are guided by internationally recognized human rights standards. We are committed to respecting the rights enshrined in the Universal Declaration of Human Rights and its implementing treaties, as well as upholding the standards established in the United Nations Guiding Principles on Business and Human Rights (UNGPs) and in the Global Network Initiative Principles (GNI Principles).

In addition to actively harnessing the power of technology to advance human rights and create new opportunities for people across the globe, we are committed to responsible decision-making around emerging technologies.

Google's work on civil rights and human rights is conducted through our Human Rights Program, a central function responsible for ensuring -- across Google and all its products (such as hardware, Search, Cloud, and YouTube) -- that we are meeting our commitment to the UNGPs, GNI Principles, and other civil and human rights instruments.

The Human Rights Program advances company-wide strategy on civil and human rights, advises product teams on potential civil and human rights impacts, conducts human rights due diligence, and engages external experts and stakeholders.



Our civil and human rights work is integrated into processes and procedures across the company. A network of employees covering product, regional, and functional areas is responsible for the day-to-day work of protecting our users. Employees can also raise concerns about human and civil rights, which are reviewed by the Human Rights Program and senior management, as relevant.

Senior management oversees the implementation of the civil rights and human rights work and provides regular updates to the Audit and Compliance Committee of Alphabet's Board of Directors. In 2020, the Board amended the charter of the Audit and Compliance Committee to explicitly include oversight of civil and human rights issues.

As we grow as a company, we continue to enhance and evolve our implementation of our human rights commitments. For example, we established a Human Rights Executive Council (HREC) to provide oversight and guidance to our Human Rights Program on global human rights and domestic civil rights in the U.S. Council members represent key product areas and functions across Google and regularly hear from the leaders of the Human Rights Program.

a. Engaging with Civil and Human Rights Experts

Engaging with external experts and affected stakeholders is essential to our civil and human rights work. Regular engagement and formal consultation with civil society and other stakeholders informs the development of our products and policies alike.

These engagements help us identify, prioritize, and address existing and potential civil and human rights impacts. They also present important opportunities for Google to get feedback on how and where we should consider improvements to our policies, practices, and services.

b. Transparency

Transparency is core to our commitment to respect human rights. In 2010, we were the first internet company to launch a tool to inform people about government requests for user data or content removal. Today the [Google Transparency Report](#) hub includes transparency reports on requests for user information, government requests to remove content, traffic and disruptions, and many other topics that can potentially impact human rights.

Transparency helps our users, the public, regulators, and policymakers understand our policies and our products. That's why we've created tools like [How Search Works](#), [How Play Works](#), and [How YouTube Works](#) to explain how our products work, and share our approach on specific issues, such as disinformation and information quality.

In 2018, we began publishing a dedicated [YouTube Community Guidelines enforcement report](#). This quarterly report contains data on actions YouTube takes to address content that violates our policies, and includes data specific to its [hate speech](#) policy. Since the report's initial publication, we have continued to expand its scope, incorporating the Violative View Rate, or "VVR," which estimates the percentage of views from violative videos. VVR is discussed in additional detail below.



As a founding member of the Global Network Initiative, Google is subject to a periodic assessment by an independent party to review how the company integrates GNI Principles into our governance, due diligence and risk management, and operational practices. The GNI makes these company assessments [publicly available](#).

2. Google's Policies Prohibiting Hate Speech and Harassment

There is no place for hate speech on any of Google's services, and we have strong policies in place prohibiting such content. Google's [Terms of Service](#) include "basic rules of conduct" for users and apply to each service listed [here](#). Those Terms prohibit users from abusing or harming others, including by bullying or harassment. Many of Google's services also have service-specific policies prohibiting abusive content, and a list of many of those service-specific content policies is available [here](#). These policies are tailored to specific services and may differ to reflect variations across products.

For YouTube, we have clear and public policies prohibiting content that promotes or incites violence or hatred against people based on attributes including race, religion, disability, age, gender, veteran status, caste, ethnicity, immigration status, nationality, sexual orientation, or the victims of violent events or their kin. YouTube's hate speech policies cover a wide range of abuses, including promoting or glorifying violent tragedies, such as school shootings; calling for the segregation or exclusion of people based on protected attributes; or promoting inherently hateful ideology, like Naziism.

Our harassment policy on YouTube prohibits content that targets an individual with prolonged or malicious insults based on intrinsic attributes, including their protected group status or physical traits. This policy also includes harmful behavior such as deliberately insulting or shaming minors, threats, bullying, doxxing, or encouraging abusive fan behavior. We also prohibit conspiracy theory content that is used to justify violent acts.

Our advertising guidelines also prohibit monetization of content that incites hatred against, promotes discrimination, disparages, or humiliates an individual or group of people, or that is gratuitously incendiary, inflammatory, or demeaning.

a. Developing and Updating Our Policies

We have always been proud that we are a company that celebrates a broad and diverse set of voices, and we have taken many steps over the years to help protect diverse communities from hate and harassment across our services. Because the world moves quickly and online hate is always evolving, we regularly review our policies to make sure that they reflect the changes occurring both on and off our platforms.

For example, in 2019, we updated our hate speech policy on YouTube to specifically ban videos alleging that a group is superior based on qualities like race, gender, religion, or sexual orientation in order to justify discrimination, segregation or exclusion. In 2020, YouTube updated its hate and



harassment policies to prohibit content targeting an individual or group with conspiracy theories that have been used to justify real-world violence.

Our enforcement evolves as well, and we routinely make updates to track changes in things like new hateful slang or terminology. We consult with NGOs, academics, and other external experts in areas such as hate speech when developing new content policies or considering certain product features. Experts help us forecast how global events could cause harmful content to spread across our platform, including by uncovering gaps in our systems that might be exploited by bad actors, or by providing recommendations for new updates. As we work to ensure that our policies evolve with the current landscape, our guiding focus is around one major goal: preventing egregious real-world harm.

b. Protecting Openness and Accessibility

We aim to provide access to an open and diverse information ecosystem, but that does not mean that anything goes on our services. Removal of content is an important lever we use to address hate speech and harassment. It is not the only lever at our disposal, however, and we constantly balance our responsibility to protect our users with our commitment to openness.

We are especially attentive to educational, scientific, artistic, or documentary contexts, including journalistic content, where the content might otherwise violate our policies. Such content would include, for example, a documentary on YouTube about a hate group. While the documentary may contain hate speech that would otherwise violate our policies, we may allow it to remain on the platform if it condemns or decries the group and their hateful ideas. This context must appear in the images or audio of the video itself. However, this exception is not a free pass. For example, a documentary promoting violence or hatred wouldn't be allowed. Any user can [flag](#) content for review if they believe the content violates YouTube's hate speech policies.

3. Enforcing Our Hate Speech and Harassment Policies

Hate speech is a complex policy area to enforce at scale, as decisions require nuanced understanding of local languages and contexts. To help us consistently enforce our policies, we work with subject matter experts across multiple languages. We also deploy machine learning to better detect potentially hateful content to send for human review, applying lessons from our enforcement against other types of content, like violent extremism.

To complement those enforcement efforts, YouTube also has a [Priority Flagger program](#) in certain jurisdictions, which provides robust content reporting processes for government agencies and non-governmental organizations (NGOs) that can notify YouTube of content that they believe violates its Community Guidelines. Participants in the Priority Flagger program receive training in enforcing YouTube's Community Guidelines, and their flags are prioritized by YouTube for review. In addition, participants of the Priority Flagger program have a direct line of communication with YouTube's Trust & Safety teams to facilitate more prompt issue resolutions.



If users see content on any of Google's products that they believe violates the law or their rights, we provide a mechanism for them to report it. We will review the reported material and consider blocking, limiting, or removing access to it.

Taken together, our enforcement tools are designed to remove violative content as quickly as possible, and importantly before it has gained a large number of viewers.

a. Measuring the Violative View Rate

As noted above, in 2020, YouTube added a new data point to the enforcement report called Violative View Rate (VVR), which is an estimate of the proportion of video views that violate the Community Guidelines in a given quarter. VVR is YouTube's North Star for measuring its progress in combating harmful content on the platform, and YouTube believes sharing it with the public is an important way to create transparency and accountability. In order to calculate VVR, YouTube takes a sample of the views on YouTube and sends the sampled videos for review. Once YouTube receives the decisions from reviewers about which videos in the sample are violative, YouTube aggregates these decisions in order to arrive at an estimate. In Q1 2024, the VVR was 0.12-0.13%, meaning that out of every 10,000 views on YouTube, only 12-13 come from violative content.

4. Promoting Responsible AI

Google is committed to the responsible development and use of AI, including protecting against bias. In 2018, we announced [Google's AI Principles](#), which govern our research and product development and impact our business decisions in this area. Specifically, Google believes AI should: be socially beneficial; avoid creating or reinforcing unfair bias; be built and tested for safety; be accountable to people; incorporate privacy design principles; uphold high standards of scientific excellence; and be made available for uses that accord with these principles.

We identify four areas we will not pursue—technologies: (1) that are likely to cause overall harm; (2) whose principal purpose is to cause or directly facilitate injury to people; (3) that gather or use information for surveillance violating internationally accepted norms; and (4) whose purpose contravenes widely-accepted principles of international law and human rights. We continue to be one of the leading companies investing heavily in responsible AI research and implementation to ensure we can live up to our AI Principles. This includes hundreds of people working in culture, education, and participation; technical progress; internal review processes; and community outreach and exchange.

5. Providing Researchers Access to YouTube Data

At Google, we believe it is important to provide meaningful transparency and to work with researchers studying YouTube. In July of 2021, we launched a new initiative for researchers affiliated with accredited, higher-education institutions that supports research-specific access to YouTube's global Data API. We offer eligible academic researchers:

- Scaled access to YouTube's public data corpus with as much quota as required for their research;
- The opportunity to derive insights from global YouTube data; and
- Support and technical guidance from YouTube.



YouTube's Researcher Program has no requirements or stipulations about which topics researchers are allowed to study. We encourage academics from diverse backgrounds, studying a range of topics, to apply.

In accordance with Article 40 of the EU Digital Services Act, Google also recently introduced a metadata scraping solution for YouTube as part of the [Google Researcher Program](#), which is available to researchers based in the EU who meet certain statutory requirements. Researchers may access the program through an [application process](#) that confirms their eligibility under Article 40. Qualifying researchers from diverse disciplines can obtain access to data from Google Maps, Google Play, Google Search, Google Shopping or YouTube to study those platforms from a wide range of perspectives and research frameworks, including scraping publicly accessible metadata on YouTube's user interface.

6. Advertising

With respect to your specific question regarding ads, we have strict ads [policies](#) and [advertiser friendly guidelines](#) that govern the types of ads and advertisers we allow on our platforms and the YouTube content that we allow to monetize. For ads, we do not allow ads on our platform that incite hatred against, promote discrimination of, or disparage an individual or group on the basis of their race or ethnic origin, religion, or nationality ([Dangerous and Derogatory Content Policy](#)). We also prohibit ads on our platform that contain violent language, gruesome or disgusting imagery, or graphic images or accounts of physical trauma ([Shocking Content Policy](#)).

For YouTube, we have [Advertiser-Friendly Guidelines](#) that prohibit videos that contain violence or shocking content from monetizing. YouTube also has [Community Guidelines](#) that govern all videos posted on the platform, regardless of monetization. Across Google Ads and YouTube, we have implemented a [sensitive event](#) to enhance our efforts to prevent inappropriate ads or video monetization. We often institute "sensitive events" following conflicts or other tragic occurrences. We do not allow ads that exploit or capitalize on designated "sensitive events," nor do we allow YouTube videos to monetize on them. We may allow monetization if there's crucial context, such as authoritative news reporting.

We also have [publisher policies](#) that prevent publishers from monetizing on content that incites hatred against, promotes discrimination of, or disparages an individual or group on the basis of their race or ethnic origin, religion or nationality. We provide transparency about who our advertisers are, where they are located and which creatives they have served through Google. Our [Ads Transparency Center](#) is a searchable hub of verified advertisers where users can view basic information about the advertiser and see the ads they are running on our platforms. We invest heavily in the enforcement of our policies. We have a team of thousands working around the clock to create and enforce our policies at scale. To provide a sense of the scale of our [enforcement efforts in 2022](#), we removed over 5.2 billion ads, restricted over 4.3 billion ads and suspended over 6.7 million advertiser accounts.



7. Google's Response to the Israel-Hamas Conflict

We are committed to ensuring that YouTube remains a safe space for our users, creators, and partners following the attack in Israel on October 7th and during the ensuing conflict in Gaza. Content moderation on our platforms is a critical focus of our efforts toward these ends. As noted above, hate speech such as content targeting any religious or ethnic communities is not allowed on our products that host user-generated content. As of April 30, 2024, YouTube had removed more than 110,000 videos, terminated more than 5,000 channels, and removed more than 160 million comments specific to these events. In addition, we demonetized over 300,000 videos for violating our Advertiser Friendly Guidelines on violence, terrorism, and sensitive events, and age-gated thousands of videos not appropriate for viewers under 18.

8. Conclusion

Google is proud of the work we have done to protect our users and maintain safe and trusted services, but we recognize that our work is never done. We will continue to learn and to improve our services while being guided by our belief in technology's power and potential to have a profoundly positive impact across the world.

Thank you for your correspondence.

Sincerely,

A handwritten signature in black ink that reads "A. Walden". The signature is written in a cursive, flowing style.

Alexandria Walden
Global Head of Human Rights