White Paper: March 2022

Rethinking the Freedom & Transparency Act:

Content Moderation, Transparency, and Due Process at YouTube

Intro

How YouTube Works

- Content moderation
- Ranking and recommendations
- Machine learning and automation

Transparency

- Creator notice and appeals
- Creator dashboards, analytics, and controls
- User flagging
- User choice and control over content
- Transparency reports

Intro

At YouTube, our mission is to give everyone a voice and show them the world. We believe that everyone deserves to have a voice, and that the world is a better place when we listen, share, and build community through our stories. In support of this mission, four freedoms comprise the foundation of our philosophy:

- Freedom of Expression: We believe people should be able to speak freely, share opinions, foster open dialogue, and that creative freedom leads to new voices, formats, and possibilities.
- Freedom of Information: We believe everyone should have easy, open access to information, and that video is a powerful force for education, building understanding, and documenting world events, big and small.
- Freedom of Opportunity: We believe everyone should have a chance to be discovered, build a business and succeed on their own terms, and that people—not gatekeepers—decide what's popular.
- Freedom to Belong: We believe everyone should be able to find communities of support, break down barriers, transcend borders, and come together around shared interests and passions.

Over the years, we have seen more and more people come to YouTube to share their experiences and understand their world more deeply. This is especially true when it comes to learning new skills, participating in civic engagement, and developing informed opinions about current events. With so many users around the world looking to YouTube for information, we have a responsibility to provide a quality experience and support an informed citizenry. Over the past several years, responsibility has been a critical area of investment across our company. We have focused extensively on developing policies and building product solutions to live up to this responsibility while preserving the opportunity of an open platform.

Our approach to responsibility consists of four pillars, which we call the 4Rs.

- Remove content that violates our policies as quickly as possible
- Raise up authoritative voices when users seek breaking news and information
- Reduce the spread of content that brushes up against our policy lines, or “borderline content”
- Reward trusted, eligible creators and artists

We share the goals of Brazilian stakeholders to develop a framework that supports freedom of expression and transparency across digital platforms. However, with such a comprehensive undertaking, the slightest detail in text can have significant and potentially unintended consequences for our responsibility efforts, negatively impacting creators and users, and jeopardizing the effectiveness of the legislation. Thus, as we look at the latest version of the Freedom and Transparency Act, we believe it’s critical to help stakeholders better understand how YouTube’s approach to responsibility across several key areas would be negatively impacted by the law. Our hope is that this white paper fosters that understanding such that we can engage in a constructive dialogue to balance policy goals with the realities of operating a safe and open digital platform.

Addressing concerns within the Freedom and Transparency Act

YouTube recognizes the importance of increasing transparency and accountability across digital platforms. We are also taking significant action to address misinformation on our platform and have invested in a comprehensive plan across our 4Rs to support the integrity of the upcoming Presidential election in Brazil. However, we are concerned about the unintended consequences of several provisions, particularly their impact on our ability to use a range of levers and tactics to effectively combat misinformation—one of the original goals of this legislation.

Article 15: Due process for "reduction of reach" and flagging

The law seeks to extend notice and appeal rights to creators when our systems take any action that would result in a "reduction of reach" of that content. This is not clearly defined in the law but can be reasonably interpreted to implicate content demotion. Because our systems are continuously making decisions about what content to show and what content *not* to show to individual users, according to the text as written, we would apparently be required to notify a creator every time we "demote" their content for any reason—including user-related, relevance, quality, and responsibility signals—thus resulting in billions of notifications to creators on a daily basis. To give a sense of volume, we show over 600 million different videos each day in our recommendations. While this may not be the intent of legislators, such a result would be unworkable for creators and make our service impossible to operate. Furthermore, because decisions about what content to show are temporal in nature and change from one user and one moment to another, there's no way to materially reverse a decision to not recommend a piece of content or to rank a piece of content above another, raising a major question about what remedy on appeal would actually be.

Second, as we've outlined in this paper, YouTube's use of machine classifiers to minimize risk of user exposure to problematic content works in tandem with content removal according to our Community Guidelines. This multi-faceted approach enables us to host more content on the platform but limit proactive recommendations of some content to users who aren't actively seeking it. The content remains visible and available to users across the platform, allowing us to balance freedom of expression with user satisfaction, as well as concerns raised by academics, policymakers, and civil society regarding non-violative content that may be problematic, like misinformation.

As an online platform, we face a perpetual uphill battle against nefarious actors looking to abuse our systems—from individual creators trying to test the boundaries of our hate speech policies to organizations acting at the behest of authoritative regimes to drive large-scale misinformation narratives and shake the foundations of vulnerable democracies. These actors are savvy, and they continuously test and try to game our systems. For example, as explained earlier in this paper, we provide creators with analytics about their videos, including how many views come from recommendations. We have seen bad actors use these analytics to test our systems by uploading multiple copies of the same video with slight tweaks to determine which versions get more views from recommendations. The content remains problematic, it's just not caught by our systems. This, then, results in our need to recalibrate to capture more content.

The due process requirements under Article 15 would effectively provide bad actors with a complete map of what signals we use to lower down bad quality content and they could adjust their content to overcome our systems. This is like giving a burglar a tour of your home, showing them where you hide your spare key, and providing the code to disable your security system. In order to allow our systems to balance free expression with user protection and expectations, we suggest focusing on content moderation actions that would materially restrict the availability of content, such as those we describe in the first section of this paper—e.g., removal and age-gating—and explicitly scope out ranking and recommendations decisions. The latter can and will be addressed through a qualitative transparency obligation as we saw in an earlier draft of the law, which required platforms to report ""information on the employment and operation of automated systems, including the algorithm operation and training basis, as well as the analysis of their impact on the circulation, offer, promotion, reach reduction or removal of contents." This approach increased transparency about how systems work without giving detail such that the systems can be easily abused—including by progenitors of misinformation.

Article 15 also references notice and appeals for flags on content. As discussed earlier in this paper, YouTube allows any user to flag content, and we provide users with a dashboard showing decisions on their flags. However, the volume of user flags and lack of accuracy of those flags creates significant concern about the viability of such obligations. In particular, establishing an out-of-court redress model without sufficient guardrails will create a system overrun with abuse. User flags, while sometimes helpful signals, are extremely unreliable. During the 6 month period ending September 2021, less than 2% of the more than 32M videos flagged globally for review under our Community Guidelines were ultimately removed after human review of that content. Users flag videos for various reasons, including to express displeasure instead of or in addition to clicking the "dislike" button. As an example, the French version of a song from a popular animated children's movie was flagged by French users more than 4,000 times with the top reason for flagging as "Sexual," despite the fact that the video did not involve any sexual content

Article 38: Publisher remuneration

A provision added to the Freedom and Transparency Act in late 2021 would require remuneration to copyright holders for all "journalistic content." YouTube only surfaces content expressly uploaded by users and where users provide us a copyright license. We already implement a revenue sharing arrangement through the YouTube Partner Program (YPP) so copyright holders already have a means of remuneration. Further, "journalistic content" is sufficiently defined. We urge policymakers to consider existing commercial contracts and arrangements like YPP as sufficient "remuneration" under the law, making both that and what constitutes "journalistic content" clear in the bill.

Article 5 and Article 10: Search tool obligations

As a platform with 500 hours of video uploaded every minute, search functionality is a critical way for users to find content they're looking for—whether it's a cooking how-to, a yoga class, a music video from their favorite artist, or information about breaking news. The bill as written creates separate obligations for search tools and for social networks, yet search tools are a fundamental element of social networks. We are concerned that, as drafted, the bill's definition may capture YT's search capability and regulate YouTube as both a "search tool" and "social media" provider.

We encourage legislators to establish definitions of social media provider and search tool that are mutually exclusive so that a platform falls into one category or the other but not both. To do this, we suggest limiting "search tool" to one that indexes online webpages and not "content" or videos, or clearly exempt on-platform search tools that only search and return content hosted by the platform.

Article 9: Transparency reporting obligations

We are committed to meaningful transparency. As noted in this paper, our Community Guidelines Enforcement Report has grown significantly since we launched it in 2017, and we continue to use feedback from creators, users, regulators, civil society, and others to evolve our reporting to paint a clear picture of how our actions impact content on the platform. However, the law before us expands transparency obligations in significant and even unclear ways. While many of the requirements are directionally important for users and the public alike, we have concerns with some of the provisions that appear to be reporting for reporting's sake versus providing any

additional meaningful transparency. For example, the current draft would require us to provide data on procedures and decisions regarding "reduction of scope, content flagging and others that restrict freedom of expression, as well as the measures employed to comply with this law." We welcome the opportunity to share our perspective and better understand the goals of stakeholders so that we can reach a compromise that establishes high standards for transparency reporting without compromising platform integrity. To that end, we suggest that reporting obligations regarding procedures and actions should be limited to removals (for both illegal content and breach of terms and conditions), flag volumes, appeals, and reinstatements. We also suggest that legislators revisit earlier text regarding the content moderation workforce data and alter the provision to require "basic information on the team responsible for the enforcement of moderation rules for accounts and contents, as well as information on the circumstances of automated review and review by individuals."

How YouTube Works

This section seeks to provide more detailed information about how we approach three major areas of YouTube's platform that overlap with our responsibility efforts and will be impacted by the Freedom and Transparency Act: content moderation, ranking and recommendations, and machine learning. Throughout this paper we refer to two different groups of people on our platform: creators, who are the people who post content to YouTube, and users, the broader set of people who come to YouTube to find and discover content.

Content moderation

We may remove or limit access to content on YouTube if it's found to violate local laws, copyright law, or YouTube's terms and conditions via our content policies, called Community Guidelines.[1] These Community Guidelines are the foundation of content moderation work on our platform, including the Remove pillar of our 4Rs. Our Community Guidelines provide clear, public-facing guidance regarding content that is not allowed on the platform—e.g., spam, deceptive practices, scams, hate, harassment, and impersonation.

YouTube has always had clear and robust content policies that we make publicly available via our Help Center. For example, we have never allowed pornography, incitement to violence, or content that would harm children. Harmful content on our platform makes YouTube less open, not more, by creating a space where creators may not feel safe to share. That's why our policy development team systematically reviews and updates all of our policies to ensure that they are current, keep our community safe, and preserve openness. We frequently consult outside experts and YouTube creators during the process, and consider regional differences to ensure proposed changes can be applied fairly and consistently around the world.

We remove content that violates our policies as quickly as possible. Violative content represents a fraction of a percent of total views on YouTube, and we track our progress through a metric we made available to the public last year called Violative View Rate (VVR).[2] This metric is an estimate of the proportion of video views that violate our Community Guidelines in a given quarter (excluding spam). Our data science teams have spent more than two years refining this metric, which we consider to be our North Star in measuring the effectiveness of our efforts to fight and

[1] https://www.youtube.com/howyoutubeworks/policies/community-guidelines/

[2] https://blog.youtube/inside-youtube/building-greater-transparency-and-accountability/

reduce abuse on YouTube. As reported in our Community Guidelines Enforcement Report,[3] in Q3 2021, VVR was 0.9-0.11%, meaning that for every 10,000 views, only 9-11 were of violative content. We work continuously to shrink this even further through improved detection and enforcement, relying on a combination of technology and people. We will talk more about automated detection in a later section.

In addition to removal, we may take other actions to restrict the availability of content to some or all users. For example, we use interstitials to warn viewers that certain videos may not be appropriate for certain audiences or may contain objectionable content that some may find offensive, such as graphic footage of a war zone. This allows us to restrict visibility without removing content from the platform. In addition, we provide user controls that filter certain types of content based on appropriateness for certain audiences. Users may choose to browse YouTube in Restricted Mode[4], which filters out videos containing adult content like drugs, alcohol, sexual activity, violence, profanity. Restricted Mode is regularly used in education settings, like classrooms or libraries. We also require creators to indicate whether their content is specifically "made for kids," and this designation along with rigorous reviews determines whether content can be shown in our separate YouTube Kids service designed for children under 13.

We review each complaint regarding allegedly illegal content carefully. If the content is in violation of local law, we will locally block the content that we identify as illegal. In cases where that same content violates our Community Guidelines, we remove it globally according to our terms. For example, we have seen significant overlap between content reported as illegal under Germany's NetzDG and content prohibited under our Community Guidelines. In the first half of 2021, we globally removed more than 16,000 pieces of content reported under NetzDG[5] as hate speech or political extremism for violating our Community Guidelines, and only locally blocked 134. Thus, our Community Guidelines provide significant coverage for content considered illegal according to local laws.

There are limited cases where we automatically remove illegal content, such as child sexual abuse material (CSAM), where context is not relevant to the decision whether to remove. We use hash-matching software called CSAI Match[6] to remove content previously identified as CSAM and machine learning technology to identify CSAM that hasn't previously been identified. In both

[3] https://transparencyreport.google.com/youtube-policy/views

[4] https://support.google.com/youtube/answer/174084

[5] https://transparencyreport.google.com/netzdg/youtube

[6] https://protectingchildren.google/

cases, we report CSAM to the National Center for Missing and Exploited Children (NCMEC), which coordinates with relevant local authorities.

Ranking and recommendations

With 500 hours of video uploaded to YouTube every minute, the number of videos in our corpus is massive. But like any platform, we operate within the confines of pixels available on a user's laptop, tablet, or mobile device. There is no way for us to show every video hosted on YouTube to every user—just as a bookstore couldn't keep every book available in print on its shelves. So in order to make videos in that corpus discoverable to users, we have to make choices about how to organize and display information.

This is where ranking and recommendations come into play. Our ranking and recommendations systems use a combination of signals relating to user satisfaction, user engagement, relevance, quality, and platform responsibility to determine what videos appear in results and recommendations, and how they are ordered. But our ultimate goal is to help people find the videos they want to watch and that will give them value. We provide search functionality for people who seek content related to a particular topic, as well as recommendations on a user's homepage and on the video watch page, called "Watch next." A user's homepage is what they see when they first open YouTube—it displays a mixture of personalized recommendations, subscriptions, and the latest news and information. The Up Next panel appears when a user watches a video and suggests additional content based on that video, alongside other videos that the user may be interested in.

Recommendations are a tool both for organization and discovery, surfacing content from a massive corpus of content based on the value it is likely to provide to consumers. Our system sorts through billions of videos to recommend content tailored to a user's specific interests. Unlike other platforms, we don't connect users to content through their social network. Instead, the success of YouTube's recommendations depends on accurately predicting the videos that a viewer will enjoy watching.

To provide such a custom experience, our recommendation system doesn't operate off of a "recipe book" of what to do. Our systems are constantly evolving, learning every day from over 80 billion pieces of information we call signals. A number of signals build on each other to help inform our system about what a user finds satisfying: clicks, watchtime, survey responses, sharing, likes, and dislikes. Like a user's recommendations, the importance of each signal depends on that user, the context the user is in at the time (for example, watching from a mobile device in the morning versus on a television in the evening), and the system adapts dynamically as a user's viewing habits change. For example, if a user tends to share any video that they

watch—including the ones rated one or two stars—our system will learn not to heavily factor in shares when recommending content to that user.

In addition to signals driven by user activity, our systems also take into account signals relating to our responsibility efforts. In recent years, and in response to the growing crisis of misinformation online, we have begun to use classifiers to identify whether a video is "authoritative" or "borderline." These classifications rely on human evaluators who assess the quality of information in a channel or video. These evaluators hail from around the world and are trained through a set of detailed, publicly available rating guidelines.[7] We also rely on certified experts, such as medical doctors, when content involves health information.

To determine authoritativeness, evaluators answer a few key questions. Does the content deliver on its promise or achieve its goal? What kind of expertise is needed to achieve the video goal? What's the reputation of the speaker in the video and the channel it's on? What's the main topic of the video (e.g., News, Sports, History, Science, etc)? Is the content primarily meant to be satire? These answers and more determine how authoritative a video is. The higher the score, the more the video is promoted when it comes to news and information content.

To determine borderline content, evaluators assess factors that include, but aren't limited to, whether the content is: inaccurate, misleading or deceptive; insensitive or intolerant; and harmful or with the potential to cause harm. The results are combined to give a score for likelihood that the video contains harmful misinformation or is borderline. We then limit recommendations of that content to users who do not subscribe to the channel on which that content appears.

Our search and recommendations systems are also designed to raise up authoritative voices in response to user queries, especially those that are "news-y" or related to topics prone to misinformation. In 2017, we started to prioritize authoritative voices such as local and national news sources for information queries in search results and "watch next" recommendation panels. Millions of search queries are getting this treatment today and we continue to expand to more topics and countries. In addition, the rise of misinformation in recent years led us to further expand the ways we use our recommendation system to address problematic misinformation and borderline content—that is content that comes close to, but doesn't quite violate our Community Guidelines. This includes conspiracy theory videos ("the moon landing was faked") or other content that spreads misinformation ("orange juice can cure cancer"). We discuss how this works in the following section.

[7] https://static.googleusercontent.com/media/guidelines.raterhub.com/en//searchqualityevaluatorguidelines.pdf

Because of the dynamic nature of our systems and the range of signals taken into account, providing more transparency isn't as simple as listing a formula for recommendations. This also makes it impossible for a platform like YouTube to provide notice to a content creator any time we take action that could be interpreted as a "reduction in reach." Every time a user visits their YouTube home page or watches a video, YouTube's recommendation systems weigh user-based signals, responsibility signals, and other signals—e.g., relevance, quality, language, and time of day in order—to recommend content in the space provided in the user interface. Because of the volume of content on YouTube and the limited amount of space available to show recommendations, any decision not to show a video in the "top" spot, for any reason, could be technically considered to reduce its reach.

Given that decisions are decentralized and unique to each individual YouTube user in a given moment, the number of users and the amount of content on our platform would trigger billions of notices to creators on a daily basis. The more content a creator has on the platform, the more notifications they'd receive, making them so noisy as to be meaningless. Thus, instead of notifications about what's *not* recommended, we provide creators with analytics about traffic generated by recommendations so they can understand how this aspect of discovery on our platform contributes to their overall traffic.

Machine learning and automation

YouTube has over 2 billion monthly logged-in users and every day people watch over a billion hours of video. Because of the scale at which YouTube operates, we would not be able to sustain any of our responsibility efforts to protect users without the support of machine learning and automation. For example, machine learning plays a critical role in content moderation on YouTube, and we deploy it in two key ways: proactively identify and flag harmful content to our human reviewers, and automatically remove certain types of content very similar to what has been previously removed, such as spam or sexually graphic material. In both cases, data inputs are used to train the systems to identify patterns in content—both the rich media content in videos, as well as textual content like metadata and comments—and our systems then use those patterns to make predictions about new examples to match.

Machine learning is well-suited to detect patterns, which also helps us to find content similar to other content we have already removed, even before it is ever viewed. We use hashes (or "digital fingerprints") to catch copies of known violative content before they are even made available to view. The systems then automatically remove content only where there is high confidence of a policy violation—e.g., spam—and flag the rest for human review. In the third quarter of 2021, of the 6.2 million videos removed for violating our Community Guidelines, 94% of those videos were first flagged by our automated systems.

Once our machine learning systems flag potentially violative content, human reviewers then remove content to determine if the content violates our policies, taking into account context—e.g., possible educational, documentary, scientific, or artist value (EDSA)[8]—and nuance that machines cannot evaluate. If human review finds the content to violate our policy without EDSA, we remove the content globally. If the content does not violate our policies or is deemed to meet our EDSA bar, the content remains live. These decisions are in turn used as inputs to improve the accuracy of our technology so that we are constantly updating and improving the system's ability to identify potentially violative content. In addition, when we introduce a new policy or alter an existing one, it takes our systems a bit of time to catch up and begin to detect relevant content so our enforcement of new policies improves continuously.

Our responsibility efforts relating to ranking and recommendations would not be possible without the support of automation. YouTube has limited the dissemination of low-quality content via recommendations since 2011, when we built classifiers to identify videos that were racy or violent and prevented them from being recommended. In 2015, we noticed that sensationalistic tabloid content was appearing on homepages and took steps to limit viewers' exposure. A year later, we started to predict the likelihood of a video to include minors in risky situations and removed those from recommendations. And in 2017, to ensure that our recommendation systems were not biased against content belonging to individuals or groups based on attributes like gender or sexual orientation, we began evaluating the machine learning that powers our systems to ensure that unintended algorithmic bias such as gender bias isn't present.

We use machine learning to train our system to model the evaluators' decisions regarding authoritative and borderline content (explained in the previous section) so that we can scale their assessments to all videos on YouTube. The more authoritative a video, the more it's promoted in recommendations on sensitive topics. The more "borderline" a video, the less frequently it is recommended. Connecting viewers to high-quality information and minimizing the chances they'll see problematic content is not just important from a platform safety perspective, it is also paramount to our goal of recommending content that delivers value. These efforts complement the work done by our robust Community Guidelines, by allowing content that some may find objectionable to remain visible and accessible to users on the platform who wish to find and view it, and are critical to our responsibility efforts.

---

8 https://blog.youtube/inside-youtube/look-how-we-treat-educational-documentary-scientific-and-artistic-content-youtube/

Transparency

YouTube is committed to providing meaningful transparency and accountability on the impact of platforms' decisions for users, citizens, and other businesses. We believe that providing users with information regarding how we treat their information and actions we take with content, can empower them to better understand their experience on the platform. This section provides an overview of the main ways we do this today.

Creator notice and appeals

We notify creators when they violate our terms and conditions by posting content prohibited under our Community Guidelines or when their content is deemed to violate local law. For violation of our Community Guidelines, we rely on a 3-strike system.[9] Each creator receives one warning in the lifetime of an account for violating our policies where we remove content but do not issue a strike. Subsequent violations result in strikes. A first strike in a 90-day period results in suspension of upload privileges for 7 days. A second strike in that same period would result in a 14-day suspension. A third strike in a 90-day period results in termination. Additionally, severe misconduct, such as child sexual abuse material, has a zero tolerance policy and results in immediate termination. When a creator's channel is terminated, we consider any attempt by that creator to create or use another channel to upload content circumvention and will terminate those channels as well. Note that we have a different strike system[10] in place for copyright violations.

As with any system, we sometimes make mistakes, which is why we allow appeals.[11] Creators are notified when their video is removed, and we provide a link with simple steps to appeal the decision. In this process, creators are able to provide additional context about the content, such as explaining EDSA purposes. If a creator chooses to submit an appeal, it goes to human review, and the decision is either upheld or reversed. We also allow creators to appeal actions on illegal content, actions that materially restrict the availability of content (e.g., age-gating), decisions to demonetize content, and termination or suspension of a creator's YouTube channel.

[9] https://support.google.com/youtube/answer/2802032

[10] https://support.google.com/youtube/answer/2814000

[11] https://support.google.com/youtube/topic/9153827

Creator dashboards, analytics, and controls

All YouTube creators have access to a dashboard and a robust set of controls over their content. The dashboard provides analytics[12] about their content, including how much traffic comes from recommendations. With regard to specific videos, creators may choose different visibility settings for their own content, including unlisted (so it won't appear in search results) and private (where only the creator and specific people they choose can view the content). We also offer creators tools and settings[13] to moderate comments on their videos. For example, they can toggle comments on or off across their channel or on individual videos; create "blocklists" of words or phrases and enqueue comments containing these words and phrases for their review prior to being public; or manually moderate individual comments posted on their videos.

Creators who monetize their content via the YouTube Partner Program[14] have visibility into the monetization status of their videos. We have a separate set of policies that specify what content is not appropriate for monetization, called our Advertiser-Friendly Content Guidelines.[15] We inform creators in writing by email or in product when we enforce those policies, and let them know what options are available. In addition, when a creator turns on ads, they see a monetization icon[16] next to each video on their channel indicating that video's eligibility for advertising. For example, we show a user that monetization is "limited" if that video doesn't meet all of our monetization policies and let them know if they can appeal for human review, if human review is underway, or if human review has already been conducted.

User flagging

In order to help identify violative content that our systems may not catch, we enable users to flag content that may potentially violate our Community Guidelines via an in-product reporting tool[17] and provide users with a dashboard where they can track whether we take action in response to

[12] https://support.google.com/youtube/answer/9002587

[13] https://support.google.com/youtube/answer/9483359

[14] https://support.google.com/youtube/answer/72851

[15] https://support.google.com/youtube/answer/6162278

[16] https://support.google.com/youtube/answer/9208564

[17] https://support.google.com/youtube/answer/2802027

their flags. When flagging a video, human flaggers can select a reason they are reporting the video and leave comments or video timestamps for YouTube's reviewers. Reviewers evaluate flagged videos against all of our Community Guidelines and policies, regardless of why they were originally flagged. Flagging a video does not necessarily result in it being removed, and while we regularly receive a very high volume of user flags—17,920,560 in Q3 2021 alone—user flags are unfortunately not very accurate, as users often use the reporting tool simply because they don't like a video rather than because it violates any policy. We discuss this in detail in the following section.

User choice and control over content

YouTube offers users a range of tools which enable them to both fine tune their experience on the platform and decide what personal information is used to influence recommendations. For example, users can subscribe to individual channels to see more content on those channels in recommendations and, if the user chooses, receive notifications when a creator posts new content to the channel.

YouTube offers a range of privacy choices for users, including browsing YouTube while signed into a Google account, while signed out, or via Incognito Mode.[18] Signed-in users can view, pause, edit, or clear their watch history and clear their search history, remove individual search entries from search suggestions, or pause search history at any time through the YouTube history settings. Users may also delete YouTube search and watch history automatically after a certain amount of time through the Google Account settings.[19] In-product controls enable users to remove recommended content—including videos and channels—from their Home pages and Watch Next. We also ask users directly about their experience with individual videos and our recommendation systems using random surveys that appear on their home page and elsewhere throughout the app. We use this direct feedback to fine-tune and improve these systems for all users.

Transparency reports

As a part of Google, we have led the way for the industry in terms of reporting on content removal at the request of governments and according to our own Community Guidelines, as well as information about government requests for information about users. We continue to expand our

[18] https://support.google.com/youtube/answer/9040743

[19] https://www.google.com/account/about/

initiatives and the information we share and in 2021, released our first ever copyright transparency report.[20] Last year, we added the Violative View Rate (VVR) metric[21] described earlier in this paper, to our quarterly Community Guidelines Enforcement Report. We also worked with MIT Professor of Statistics Arnold Barnett last year to review and validate our methodology[22]. Sharing this metric and the other data we have made available through the report helps us not only to hold ourselves accountable for our progress but also to help the public better understand how we think about what it means to maintain a safe platform.

[20] https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-22_2021-1-1_2021-6-30_en_v1.pdf

[21] https://transparencyreport.google.com/youtube-policy/views

[22] https://www.analysisgroup.com/globalassets/insights/publishing/2021-youtube-vvr-methodology-a-statistical-assessment-arnold-barnett.pdf