

# A Platform ‘Weaponized’:

How YouTube Spreads Harmful Content—And What Can Be Done About It

---

PAUL M. BARRETT AND JUSTIN HENDRIX



# Contents

Executive Summary .....	1
Introduction .....	3
A Platform That Is ‘Almost Inscrutable’ .....	6
The Other 80%: YouTube Outside of the U.S. ....	15
Conclusion and Recommendations .....	19
Endnotes.....	23

## Authors

Paul M. Barrett is deputy director of the NYU Stern Center for Business and Human Rights and an adjunct professor at the NYU School of Law. Lily Warnke, a former fellow at the Center, provided valuable research for this report.

Justin Hendrix is an associate research scientist and adjunct professor at NYU Tandon School of Engineering and the CEO and editor of *Tech Policy Press*, a non-profit media venture concerned with the intersection of technology and democracy.

## Acknowledgments

We are grateful to Craig Newmark Philanthropies and the Open Society Foundations for their continued support of our work on the technology sector.

# Executive Summary

**YouTube has been less scrutinized than other platforms by social scientists, politicians, and civil society groups, including ours. This report seeks to address this knowledge deficit.**

YouTube, the most popular social media platform in the United States, dominates user-generated long-form video-sharing around the world. Like other major platforms, it has a dual nature: It provides two billion users access to news, entertainment, and do-it-yourself videos, but it also serves as a venue for political disinformation, public health myths, and incitement of violence.

YouTube's role in Russia illustrates this duality. Since Russia launched its invasion of Ukraine in February 2022, YouTube has offered ordinary Russians factual information about the war, even as the Kremlin has blocked or restricted other Western-based social media platforms and pressured foreign journalists in the country to silence themselves. But for years before the brutal incursion, YouTube served as a megaphone for Vladimir Putin's disinformation about Ukraine and its relations with the West.

Despite its heft and influence, less is known about YouTube than other major social media sites. A subsidiary of Google, YouTube has been less scrutinized by social scientists, politicians, and civil society groups, including ours. As a result, we know less about YouTube's business practices and the effects they have on the societies in which it operates. This report seeks to address this knowledge deficit.

In Part 2, we examine YouTube's role as the internet's vast video library, one which has contributed to the spread of misinformation and other harmful content. In 2019, for example, YouTube reacted to complaints that its recommendations were pushing impressionable users

toward extremist right-wing views. The company made a series of changes to its algorithms, resulting in a decline in recommendations of conspiratorial and false content.

But recommendations are not the only way that people find videos on YouTube. A troubling amount of extremist content remains available for users who search for it. Moreover, YouTube's extensive program for sharing advertising revenue with popular creators means that purveyors of misinformation can make a living while amplifying the grievances and resentments that foment partisan hatred, particularly on the political right.

In Part 3, we turn our attention to YouTube's role in countries outside of the U.S., where more than 80% of the platform's traffic originates and where a profusion of languages, ethnic tensions, and cultural variations make the company's challenges more complicated than in its home market. Organized misogynists in South Korea, far-right ideologues in Brazil, anti-Muslim Hindu nationalists, and supporters of Myanmar's oppressive military regime have all exploited YouTube's extraordinary reach to spread pernicious messages and rally likeminded users.

The NYU Center for Business and Human Rights began publishing reports on the effects of social media on democracy in the wake of Russia's exploitation of Facebook, Twitter, and YouTube during the 2016 U.S. presidential campaign. We initially advocated for heightened industry self-regulation, in part to forestall government intervention

that could lead to First Amendment complications. As the inadequacy of industry reforms has become clear, we have supplemented our calls for self-regulation with a proposal for [enhancement of the Federal Trade Commission's consumer protection authority](#) to oversee the industry.<sup>1</sup> In Part 4, we offer a concise version of

the FTC proposal, as well as a series of recommendations to YouTube itself.

The report does not address the problem of YouTube hosting [potentially harmful videos](#) aimed at children and teenagers. This [persistent phenomenon](#) deserves continued scrutiny but is beyond the scope of our analysis.<sup>2</sup>

## Recommendations in Brief

### To YouTube:

- 1 Disclose more information about how the platform works:** A place to start is explaining the criteria algorithms use to rank, recommend, and remove content—as well as how the criteria are weighted relative to one another.
- 2 Facilitate greater access to data that researchers need to study YouTube:** The platform should ease its resistance to providing social scientists with information for empirical studies, including random samples of videos.
- 3 Expand and improve human review of potential harmful content:** YouTube's parent company, Google, says that it has more than 20,000 people around the world working on content moderation, but it declines to specify how many do hands-on review of YouTube videos. Whatever that number is, it needs to grow, and outsourced moderators should be brought in-house.
- 4 Invest more in relationships with civil society and news organizations:** In light of their contribution to the collapse of the advertising-based business model of many U.S. news-gathering organizations, the platforms should step up current efforts to ensure the viability of the journalism business, especially at the local level.

### Recommendations to the U.S. government

- 5 Allocate political capital to reduce the malign side effects of social media:** President Biden's off-the-cuff expressions of impatience with the industry aren't sufficient. He ought to make a carefully considered statement and lend his authority to legislative efforts to extend federal oversight authority. Former President Obama's recent [speech at Stanford](#) about disinformation provided a helpful foundation.
- 6 Enhance the FTC's authority to oversee social media:** Some of the issues raised in this report could be addressed by a proposal we made in a February 2022 [white paper](#)—namely, that Congress should authorize the Federal Trade Commission to use its consumer protection authority to require social media companies to disclose more data about their business models and operations, as well as provide procedurally adequate content moderation.

# Introduction

**YouTube reflects the duality characterizing all of the major social media companies: It has served as a vital information pipeline to ordinary Russians during the war in Ukraine but in the past enabled Vladimir Putin's propagandists.**

For at least a decade before Vladimir Putin's invasion of Ukraine in 2022, the Russian ruler relied on one social media platform more than any other to spread disinformation in the West. That platform was YouTube. Over the years, [Kremlin-controlled propaganda outlets](#), spearheaded by RT (formerly Russia Today) and Sputnik News, amassed millions of subscribers and billions of views on YouTube, primarily outside of Russia.<sup>3</sup>

In 2013, Robert Kyncl, then a YouTube vice president and now its Chief Business Officer, [visited an RT studio](#) to celebrate the outlet's accumulation of one billion views. The executive praised RT for offering viewers "authentic" content rather than "agendas or propaganda."<sup>4</sup>

In fact, enabled by YouTube, RT was transmitting a steady stream of Kremlin propaganda. In 2014, RT produced elaborate justifications for Russia's forcible annexation of Ukraine's Crimea region and incitement of insurrection in the Donbas area. After RT participated in a Russian information operation interfering with the 2016 U.S. presidential election, [American intelligence agencies](#) described the Kremlin-affiliated service as Putin's "principal international propaganda outlet."<sup>5</sup> Over the years, YouTube has removed some videos from RT channels, and in 2018 it began labeling RT and Sputnik content as government-funded. But by then, the Russian disinformation purveyors had established their global YouTube following.<sup>6</sup>

Even as it was helping RT expand its Western audience, YouTube emerged

as [the most-visited social media platform](#) in Russia—more popular than VK, the Russian equivalent to Facebook. An estimated 75% of Russians active on the internet now use YouTube. By comparison, 50% use Instagram and just 5%, Twitter.<sup>7</sup> Paradoxically, even as YouTube amplified Kremlin disinformation aimed at the West, it offered ordinary Russians an alternative to state-approved news outlets.

The information environment in Russia shifted drastically when Putin's tanks rolled into Ukraine in February 2022. The Kremlin enacted stiff prison terms for war reporting at odds with Moscow's official line. Foreign journalists in Russia went silent, and many departed. The few independent Russian broadcast and print outlets [shut down](#). Meanwhile, the state media regulator blocked Russian access to Facebook and Instagram and sharply restricted use of Twitter. Amid the methodical censorship, only one major outside source of news remained readily accessible—YouTube.<sup>8</sup>

Russian viewers could go to the video-sharing platform to watch [BBC](#) and

CNN reporting on the civilian devastation in Ukraine and the remarkable resilience of its defenders. On YouTube, Russians could see Ukrainian President Volodymyr Zelensky and exiled allies of imprisoned dissident Alexei Navalny condemn Putin's lawless aggression.<sup>9</sup>

A subsidiary of Google, YouTube is “arguably the largest and most engaging online media-consumption platform in the world.”<sup>10</sup> It helps shape how people understand public events, politics, and culture. Its popularity in Russia explains why Putin has allowed it to remain online there, despite the window it has provided on the carnage in Ukraine. Many of his domestic supporters regularly log onto YouTube and would resent its going dark. “We are not planning to close YouTube,” Maksut Shadaev, Putin’s Minister for digital development told *The Guardian* in May 2022. “Above all, when we restrict something, we should clearly understand that our

users won’t suffer.” Putin continued to tolerate YouTube’s reach into Russia even after the platform [shut down](#) the Western-oriented channels of RT and Sputnik<sup>11</sup> and [removed more than 9,000 pro-Kremlin channels](#) for spreading disinformation about the war.<sup>12</sup>

Other major social media platforms also blocked the Kremlin’s digital megaphones. But as of this writing, YouTube stands out for doing so while continuing to transmit information about Ukraine to Russian citizens. This delivery of uncensored news represents social media companies at their best—akin to how Facebook and Twitter became [organizing tools](#) for protesters during the all-too-brief Arab Spring in the early 2010s. In the years after the Arab democracy movement collapsed, a darker side of Facebook and Twitter became evident as the platforms hosted disinformation undermining elections and racial and religious hatred that

intensified lethal conflict. The same duality characterizes YouTube: It has served as a vital information pipeline to ordinary Russians but in the past enabled Vladimir Putin’s propagandists.<sup>13</sup>

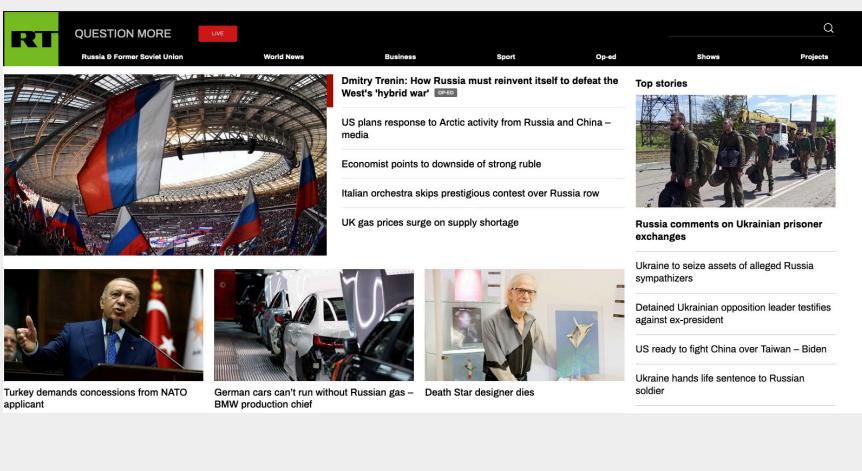
## A platform ‘weaponized’

Just six weeks before Russia’s multi-front invasion of Ukraine, a global coalition of more than 80 fact-checking organizations posted an [open letter](#) to Susan Wojcicki, YouTube’s Chief Executive Officer.<sup>14</sup> “We monitor how lies spread online,” the groups wrote, “and every day, we see that YouTube is one of the major conduits of online disinformation and misinformation worldwide.” The International Fact-Checking Network added that the company “is allowing its platform to be weaponized by unscrupulous actors to manipulate and exploit others, and to organize and fundraise themselves.” The fact-checkers described how YouTube has spread anti-vaccine conspiracies, election disinformation, and hate speech seeking to foment ethnic conflict. “The examples are too many to count,” their letter said. “Many of those videos and channels remain online today, and they all went under the radar of YouTube’s policies, especially in non-English speaking countries and the Global South.” Signing the letter were the most prominent names in independent fact-checking, among them, FactCheck.org and PolitiFact from the U.S., the United Kingdom’s Full Fact, Philippines-based Rappler, India’s BOOM, and Brazil’s Aos Fatos.

Given YouTube’s global impact—for good and ill—there’s something anomalous about the platform: Much less is known about it than other major social media sites, especially Twitter and Facebook, which along with Instagram is owned by Meta Platforms. YouTube has been less scrutinized by social scientists, policy makers, and politicians. NGOs and academic-based centers, including ours, have likewise had more to say about Facebook and Twitter than

### RT: Amplifying the Putin Line

In response to Russia’s invasion of Ukraine, YouTube and Facebook blocked the Western-targeted accounts of RT, a news outlet affiliated with the Putin regime and formerly known as Russia Today. (This image is from RT’s website.) But for years before the shut-down, YouTube facilitated the rapid growth of the Russian disinformation channel, which helped Putin spread falsehoods about Ukraine and seek to drive a wedge between the U.S. and its European allies. At times, YouTube even [celebrated](#) RT’s success.



YouTube. As a result, we understand less about how its business model and related practices affect the more than 100 countries where it operates localized versions of its platform. This report aims to redress the lack of public understanding of YouTube.

As in Russia, YouTube is the most popular social media site in the U.S. Four out of five American adults [say they use YouTube](#); two-thirds say the same thing about Facebook. And YouTube's share is increasing, while Facebook's is decreasing.<sup>15</sup> Globally, YouTube says that it has more than two billion users who watch more than a billion hours of video every day.<sup>16</sup> But its reach goes far beyond its regular visitors. Links to YouTube videos are routinely [among the most shared](#) on other major platforms, including Facebook, Twitter, and Reddit. YouTube operates a search engine second in volume only to Google's, and it enjoys a [near monopoly](#) of the online user-generated long-form video market.<sup>17</sup> Its advertising business continues to grow, delivering revenue in 2021 of [nearly \\$29 billion](#), or about 11% of what Google's holding company, Alphabet, reported.<sup>18</sup>

YouTube from time to time has acknowledged its amplification of harmful content and taken remedial steps. Since 2018, its parent, Google, has doubled the number of people worldwide who review and remove content to 20,000, although the company won't say how many focus primarily on YouTube or are direct employees, as opposed to outsourced workers.<sup>19</sup> Over the past five years, YouTube has made considerable progress addressing an earlier profusion of [videos promoting Islamist terrorism](#).<sup>20</sup> Working with Google, it also has identified and taken action against government-backed "[coordinated influence operations](#)," and it labels government-funded news outlets.<sup>21</sup> In 2019, YouTube adjusted its automated recommendation system to deemphasize conspiratorial and extremist falsehoods.

But people who already hold extreme views don't necessarily need automated recommendations to find hateful or misleading content. They can locate it by searching or following links to videos they find on Facebook or Twitter or right-wing platforms like [Gab](#) or [Gettr](#). Based on his [research](#), Dartmouth political scientist Brendan Nyhan writes that YouTube continues to facilitate "alarming levels of [user] exposure to potentially harmful content."

## Too little, too late

In a [February 2022 blog post](#), YouTube's Chief Product Officer, Neal Mohan, wrote expansively about more changes the company is exploring. These include blocking users on other platforms from embedding links to YouTube videos that contain false, conspiratorial material but aren't quite ominous enough to cause YouTube to remove them.<sup>22</sup> But to date, the reforms that YouTube has actually adopted have not been adequate.

The platform says it has introduced contextual "information panels" in "dozens of countries and languages" to accompany videos on climate change, flat-earth conspiracies, and other topics prone to misinformation. But localized versions of YouTube are available in more than 100 countries and 80 languages. (In certain other countries, users have access to a generic global version of YouTube.) "We continue our work to expand availability [of information panels] across global markets and topics," the company says. It has also experimented with "fact-checking information panels" that assess videos on more particularized subjects. So far, these are available in only six countries.

Insufficient attention to misuse of the platform is more pronounced outside of the U.S., where language and cultural barriers make it more difficult for a Silicon Valley-based company to filter content. A bizarre [international network of conspiracy-minded physicians](#) that

started in Germany, spread to Spain, and then surfaced in Latin America has used YouTube to undermine Covid-19 policies, dismissing government actions as a "plan of global domination."<sup>23</sup> [Other videos in Greek and Arabic](#) have encouraged millions of YouTube viewers to boycott vaccinations or use phony cures.<sup>24</sup> Some Brazilians have used the platform to [intensify distrust of elections](#) and promote disinformation favoring the country's right-wing president, Jair Bolsonaro.<sup>25</sup> In Ethiopia, YouTube videos have [amplified incitement](#) to ethnic strife, while in the Philippines, backers of Ferdinand "BongBong" Marcos, Jr., who won a presidential election in May 2022, used YouTube and other platforms to [obscure the government corruption](#) and violence that characterized the regime of his autocratic father.<sup>26</sup>

Social media companies are not the sole cause of democratic backsliding, eroding trust in commonly held facts, and surging nativist resentment. Other forces—including economic globalization, heightened inequality, racism, hyper-partisan cable television, and a decades-long trend toward extreme political polarization—are also roiling the U.S. and many other developed societies. What social media companies have done is act as a powerful accelerant, pouring fuel on the flames and amplifying users' worst instincts. In this fraught environment, YouTube, like other influential platforms, has an obligation to do more to counter exploitation of its prodigious capacity to spread harmful content.<sup>27</sup>

# A Platform That is ‘Almost Inscrutable’

**YouTube serves as the internet’s vast video library. Many users check out recipes, musical performances, or do-it-yourself tutorials. A small percentage—but one that may constitute a meaningful absolute number—looks for more alarming content.**

There are several reasons that less is known about YouTube than Facebook or Twitter. First, it is more difficult and expensive to analyze a large volume of videos than it is to search for words or phrases in a text data set of Facebook or Twitter posts. Budget-conscious academic researchers weigh the feasibility of competing projects; dissecting video costs a lot more in human hours and computer time.

A second reason experts have less insight into YouTube is that the company makes itself, in the words of Kate Starbird, a researcher at the University of Washington, “[almost inscrutable](#).” Compared to its counterparts—especially Twitter—YouTube provides fewer application programming interfaces, or APIs, which outsiders can use to obtain large amounts of data, and it limits the volume of information available via its APIs. “We can see pieces here and there,” says Starbird, “but we can’t systematically collect large-scale data from YouTube to use in our research.”<sup>28</sup> Brendan Nyhan, a political scientist at Dartmouth College, told us that YouTube has managed to keep enforcement of its content rules “totally opaque.”

Legal scholar Evelyn Douek identifies “a singular moment that defines YouTube’s intentional opacity and the lack of accountability this facilitates.”<sup>29</sup> The moment came [in 2018](#), when Google (and therefore YouTube) provided the most limited data set of the three companies to the independent researchers tasked by the Senate Select Committee on Intelligence

with preparing reports analyzing the nature and extent of Russian interference in the 2016 U.S. election.<sup>30</sup> An affiliate of research centers at Columbia and Harvard Universities, Douek notes that journalists, civil society analysts, and academics who “live on Twitter,” have shown a shortsightedness about YouTube’s importance.

Members of Congress share this myopia. Beginning in 2018, congressional panels repeatedly grilled Mark Zuckerberg, CEO of Meta Platforms (formerly Facebook), and his then-counterpart at Twitter, Jack Dorsey. But YouTube CEO [Susan Wojcicki never appeared before lawmakers](#).<sup>31</sup> Sundar Pichai, Chief Executive of YouTube’s parent, Google, has testified several times, but lawmakers [mostly asked](#) questions about the parent and overlooked its subsidiary.<sup>32</sup> (Wojcicki “regularly meets with lawmakers around the world—most recently with the U.S. Congressional Hispanic Caucus to discuss addressing harmful Spanish-language misinformation,” the company told us.)

YouTube often has remained below the radar simply by keeping its [corporate](#)

mouth shut. It followed both Facebook and Twitter in announcing special policies during the chaotic post-election period in 2020 and in suspending then-President Donald Trump after the January 6, 2021, U.S. Capitol insurrection.<sup>33</sup> (YouTube contests this description, saying: “We use a variety of ways to convey our efforts to the public, from top executives having [tough conversations](#) with the press to [blog posts](#) providing overviews on how YouTube works.”)

## Addressing the ‘rabbit hole’

An exception to YouTube’s avoidance of outside scrutiny has been the “rabbit hole” controversy. Beginning in 2018, [journalists](#) and [academics](#) reported episodes when YouTube fed users a series of recommendations of increasingly extreme political content. Describing her own YouTube surfing, [Zeynep Tufekci](#), a sociologist who now heads the [Craig Newmark Center for Ethics and Security in Journalism at Columbia University](#), noted that when she clicked on videos of Trump rallies, “YouTube started to recommend and ‘autoplay’ videos for me that featured white supremacist rants, Holocaust denials, and other disturbing content.” After creating a fresh account and seeking out videos of Hillary Clinton and Senator Bernie Sanders, Tufekci received recommendations involving “secret government agencies and allegations that the United States government was behind the attacks of September 11.” YouTube, she concluded, “leads viewers down a rabbit hole of extremism.”<sup>34</sup>

YouTube’s recommendations are responsible for [70% of all time spent](#) on the platform, and the rabbit-hole phenomenon could have serious consequences.<sup>35</sup> The government of New Zealand identified YouTube as a key [contributor to the radicalization](#) of a lone terrorist who massacred 51 worshippers in two Christchurch mosques in March 2019.<sup>36</sup>

YouTube terminated the hate-spewing channels frequented by the New Zealand shooter. And even before the slaughter in Christchurch, the company had begun responding to the recommendations predicament. In December 2019, it announced that more than 30 adjustments to its recommendation system had reduced the prevalence of “borderline content,” meaning misinformation, conspiracy theories, and other material that brushes up against, but doesn’t quite violate, the platform’s rules. Over the course of 2019, [YouTube said](#), there was “a 70% average drop in watch time of this content coming from non-subscribed recommendations in the U.S.”<sup>37</sup>

YouTube hasn’t been forthcoming about specific examples of borderline content that it is now suppressing or about whether the same degree of

improvement has occurred beyond the U.S., which accounts for only 20% of its user base.<sup>38</sup> It also hasn’t clarified whether the 70% average drop has proven sturdy. As of May 2020, conspiratorial recommendations had rebounded, diminishing the drop to only 40%, according to [researchers at the University of California, Berkeley](#), who nevertheless called the reduction “an encouraging trend.”<sup>39</sup> Without offering any updated statistics, YouTube told us that while “you might see consumption of borderline content go up and down,” the company continues to strive to reduce viewing of such material in the U.S. and globally.

Another study, published in 2021 by [researchers based at the University of Pennsylvania](#), found “little evidence that the YouTube recommendation algorithm is driving attention” to far-

## Subway Shooter: Missed Warning Signs

Frank James, the man accused of shooting and wounding 10 people in a New York subway car in April 2022, had posted hundreds of videos on YouTube and Facebook, many of which depicted him making violent threats and other hateful statements. Despite his talk about wanting “to kill people” and “watch people die right in front of my f—king face,” the platforms apparently did not bring James’ videos to the attention of authorities before the mass shooting. His accounts were removed after the attack.



right or “anti-woke” content.<sup>40</sup> Noting that YouTube is “under-studied relative to other social media platforms,” the Penn team scrutinized the browsing behavior, on and off YouTube, of more than 300,000 Americans from January 2016 through December 2019. Rather than recommendations heightening user engagement with far-right content, the Penn team concluded that “consumption of political content on YouTube appears to reflect individual preferences that extend across the web as a whole.”<sup>41</sup>

## Different realities'

Whatever its precise measure, the reduction in conspiratorial recommendations “does not make the problem of radicalization on YouTube obsolete nor fictional,” Hany Farid, a Berkeley

computer scientist, and his fellow researchers argued. “Aggregate data hide very different realities for individuals.”<sup>42</sup> Radicalization may be relevant for only a small fraction of visitors to the site, but given the gargantuan size of YouTube’s user base, that fraction represents a sizable absolute number. What’s more, those individuals are the ones society should be most worried about.

Researchers at NYU’s [Center for Social Media and Politics](#) scrutinized recommendations received by 760 Americans in the fall of 2020. After classifying the recommendations by ideology, they found no evidence that users on average were guided toward extreme content, defined as the right- or left-most 5% of videos. But in a manuscript pending publication,

they offered several important qualifications: YouTube recommendations *did* nudge both Republicans and Democrats toward more conservative and more homogenous content, although this push was “very mild on average.” And the platform’s “conservative bias” increased as recommendations accumulated.

“Our aggregate findings mask important and meaningful exceptions” to the tendency of YouTube recommendations to exert only a mild push to the ideological right, the NYU researchers said. “In a small fraction of cases,” they added, “users did ‘fall down a rabbit hole’ in which they were only shown relatively extreme recommendations for the duration of their time on the platform.”

## To Maximize Engagement, You Maximize Outrage’

YouTube doesn’t intentionally spread racist hatred or anti-vaccine misinformation. To the contrary, it has rules banning such harmful material, and it uses automated and human content moderation to try to enforce its rules. But YouTube’s content moderation function resembles an ambivalent arsonist frantically trying to extinguish conflagrations that he himself helped start. The same is true of Facebook, Instagram, Twitter, and TikTok.

Our Center has described this paradoxical situation in [past reports](#), but it’s worth reiterating here. The platforms collect the lion’s share of their considerable revenue from advertisers seeking the attention of consumers. To demonstrate that users remain glued to their screens, platforms prioritize “engagement,” a metric reflecting user watch time and the volume of “likes,” shares, retweets, and comments.

But here’s the problem: Content that tends to [heighten engagement](#) often does so because it provokes emotions like anger, fear, and resentment.<sup>1</sup> So, when software engineers design algorithms to rank and recommend content, the automated systems [favor posts that stir these negative emotions](#).<sup>2</sup>

Eric Schmidt, the former chief executive officer of Google, YouTube’s parent, and executive chairman of Alphabet, the holding company for both Google and YouTube, candidly explained the phenomenon during an [interview on the Armchair Expert podcast](#) in January 2022. “Corporations, at least in social media land, are optimizing, maximizing revenue,” he said. “You maximize revenue, you maximize engagement. To maximize engagement, you maximize outrage.”<sup>3</sup>

<sup>1</sup> <https://www.pnas.org/doi/10.1073/pnas.2024292118>

<sup>2</sup> <https://science.sciencemag.org/content/370/6516/533>

<sup>3</sup> <https://armchairexpertpod.com/pods/eric-schmidt>

## The internet's video library

Brendan Nyhan, the Dartmouth political scientist, worries less about unwitting users tumbling down rabbit holes than about small numbers of extremists who know what they're looking for and find it by means of YouTube's search function or by subscribing to channels dedicated to conspiratorial or hateful content. "I'm much more concerned about people who already have highly resentful views becoming more resentful or being mobilized to violent action or some kind of supporting [of] violent action," Nyhan told us.

In this sense, YouTube serves as the internet's vast video library. Many users check out recipes, musical performances, or do-it-yourself tutorials. A small percentage—but one that may constitute a meaningful absolute number—looks for more alarming content.

With his Dartmouth colleague Annie Y. Chen, Nyhan oversaw [a study published in 2021](#) by the Anti-Defamation League that investigated racist or otherwise hateful channels on YouTube. They found little systemic evidence that the platform guided unsuspecting individuals to harmful content. Still, their data did "indicate that exposure to videos from extremist or white supremacist channels on YouTube remains disturbingly common."<sup>43</sup>

One such channel is that of conspiracy theorist [Mike Cernovich](#), a regular guest host of the far-right [Alex Jones Show](#) on the InfoWars website. Although YouTube had removed some high-profile channels before the study period in 2020—including that of former Ku Klux Klan leader David Duke—about one in ten subjects viewed at least one video from an extremist channel, the study found. More than two in ten viewed at least one video from a channel classified as "alternative," meaning that it served as a potential gateway to more extreme far-right fare. Alternative YouTube channel hosts include Candace Owens, who has [echoed Russian propaganda](#) justifying the invasion of Ukraine.<sup>44</sup>

In a separate report published in 2021, researchers with the [Tech Transparency Project](#) reinforced concerns about YouTube users who readily find the disturbing material they seek. When the Tech Transparency team used searches and other online behavior to simulate interest in militant themes, YouTube recommended videos such as, "[5 Steps to Organizing a Successful Militia](#)" and "[So You Want to Start a Militia.](#)"<sup>45</sup>

The danger lies not in the average user experience but the ability of people inclined toward extremism to easily find what they're looking for or project violent intentions. Payton Gendron, the white 18-year-old accused of killing 10 African-American shoppers in a Buffalo, N.Y., grocery store in May, [went to YouTube](#) to watch videos about mass shootings, police gunfights, and tips

on firearm use. It wasn't until after the Buffalo massacre that YouTube removed three of the gun-related videos that Gendron mentioned in a diary.<sup>46</sup> In another case, Frank James, who allegedly shot and wounded 10 people in a New York subway car in April 2022, had posted [hundreds of videos of himself on YouTube and Facebook](#), many of them racist, threatening, and conspiratorial. In one YouTube video, James, an African American in his early 60s, said, "And so the message to me is: I should have gotten a gun, and just started shooting motherf—ers."<sup>47</sup>

James' videos didn't gain a large audience on YouTube. But Kate Starbird, the University of Washington researcher, points out that many YouTube users routinely copy the links of potentially harmful videos and plant them on other

### 'Young Turks': Commenters Advocate to '#ArmTheLeft'

While right-leaning views dominate extreme content on YouTube, some creators on the left also promote polarization. The Young Turks (TYT) channel has nearly 5.2 million subscribers, making it "one of the most popular political coverage channels" on the platform, according to the London-based [Institute for Strategic Dialogue \(ISD\)](#). The mob attack on the U.S. Capitol in January 2021 elicited Super Chat comments on Young Turks that promoted "the use of firearms to achieve political aims," according to the ISD. One comment, which a Super Chat participant paid \$20 to boost, included the hashtag "#ArmTheLeft." Another called for the "arming and training" of "leftist militias."



platforms. Referring to her research on election disinformation, the Syrian civil war, and conspiratorial explanations of mass shootings, [Starbird told NPR](#) in 2021: “Over and over again, YouTube is the dominant domain in those conversations [on social media]. It’s not Facebook. They’re all pulling in content from YouTube.”<sup>48</sup>

## Monetizing content

Whether one seeks to hawk workout tips or an ideological agenda, certain YouTube features, or affordances, facilitate finding a loyal audience—and potentially profiting in the process. YouTube “creators” can use an ordinary smart phone to record themselves and then upload the video to a self-created YouTube channel. Other YouTube users may subscribe to the channel

free of charge. After accumulating 1,000 subscribers and 4,000 annual “watch hours,” a creator may apply to [YouTube’s Partner Program](#) and, if approved, begin receiving a portion of revenue generated by advertising. Typically, the creator receives up to 55% of the revenue, with the company retaining about 45%. The ads are sold and distributed by Google’s AdSense service and placed immediately before or after the creator’s videos.<sup>49</sup> The extent to which creators can “monetize” video content sets YouTube apart from other platforms, which provide money-making opportunities but on a smaller scale. YouTube creators who have tens of thousands of subscribers can clear five figures a year; those with hundreds of thousands or millions can pocket six figures.

Creators have money-making opportunities other than splitting ad revenue. In the Super Chat program, for example, they can offer viewers of live-stream videos the chance to pay modest amounts to have their written comments pinned to the top of the live-chat feed for a longer period of time. Creators typically keep 70% of this cash stream; YouTube takes 30%. “More creators than ever are earning money from our non-ads products like Super Chat,” Pichai, Google’s CEO, said during a [conference call with industry analysts](#) in February 2022.<sup>50</sup>

The typical YouTube aesthetic is do-it-yourself TV: a performer who speaks directly to the audience with memes and video clips interspersed. Incurring negligible production costs, creators can foster an air of intimacy and credibility that is largely absent from platforms like Facebook and Twitter. Rebecca Lewis, a Ph.D candidate at Stanford in communications theory, has [identified](#) a populous network of conservative YouTube creators, ranging from mainstream Republicans to overt white supremacists, whose audiences overlap and share agendas such as attacking “woke” liberals and revealing imaginary “deep state” conspiracies. They visit like-minded channels and [amplify each other’s content](#), Lewis observes.<sup>51</sup>

YouTuber Tim Pool exemplifies the “micro-celebrity” methods Lewis describes. A creator in his mid-thirties who initially gained attention by live-streaming the 2011 left-wing Occupy Wall Street protests, Pool in recent years has often used his [YouTube appearances and interviews](#) to amplify divisive right-wing conspiracy theories.<sup>52</sup> Appearing in a trademark black knit hat, Pool operates a network of at least five YouTube channels, which in total boast millions of subscribers. Recent [video titles](#) include: “Leaked Emails Implicate Biden Family in UKR [Ukraine] Biolab Scandal, Media Caught Lying” and “Vaccine Passport Chip Implants Are Real, Alex Jones Right AGAIN As More Cities Declare Vax Mandates.”<sup>53</sup> He has hosted [Jones](#) and [Enrique Tarrio](#), the former chairman of the Proud Boys, a violent hard-right group. Tarrio and

### Tim Pool: Hosting Far-Right Headliners

YouTube creator Tim Pool provides a platform to far-right figures such as Alex Jones of Infowars and Enrique Tarrio of the Proud Boys. YouTube shares ad revenue with Pool, whose channels collectively have millions of subscribers. He also generates income by selling merchandise and holding Super Chats, which allow viewers of live-stream videos to pay to have their comments posted prominently. The day before the January 6 assault on the Capitol, fans of Timcast IRL [paid modest amounts](#) to elevate comments such as, “Prepare yourselves - the wicked and immoral will be annihilated — this is the only way” (\$6.66) and “The time of war is here. Only a fool still has faith in the courts” (\$4.99). In an email, Pool told us: “We don’t control what others say, and we can’t do anything about people’s free speech.” He added that his team has “worked with YouTube to deter calls for violence and extremism.”



four other Proud Boys members have been indicted on charges of seditious conspiracy for allegedly planning the January 6 attack on the U.S. Capitol. Lawyers for the defendants have said there is no evidence that their clients engaged in such a plot.<sup>54</sup>

## Carrot and stick

The data-analysis site [Social Blade estimates](#) that just one of Pool's channels, [Timcast IRL](#), brings in thousands of dollars a month in ad revenue shared by YouTube.<sup>55</sup> In a [Twitter exchange](#) in 2021, Pool indicated his familiarity with the nuances of YouTube's rules and his belief that he avoids violations that could lead to the company cracking down on his channels.<sup>56</sup> Supplementing his advertising revenue, Pool steers loyalists to [timcast.com](#), an off-YouTube site where he sells t-shirts, mugs, posters, and satiric "Proof of Gun" stickers, which resemble pandemic vaccination cards.<sup>57</sup> Memberships costing \$10 to \$1,000 a month entitle Pool enthusiasts to a range of exclusive events such as a recent podcast hosted by conservative journalist [Amber Athey](#) entitled, "Leftist Says She Is Purposefully Grooming Her Kids to Be Gay."<sup>58</sup>

In an email, Pool told us that in preparing his commentaries he relies exclusively on sources approved by [NewsGuard](#), a service that rates the reliability of news and information sites. All of his claims "are backed by data and certification," he said, and he posts his source material openly. For interview subjects, Pool said, "We seek out relevant people in news and culture to discuss issues, just like CNN or other mainstream outlets." He noted that he has hosted non-conservative guests such as technology writer Matt Binder and essayist Stephen Marche. Pool added that he publicly maintains "the position that political violence doesn't work." As for his ad revenue, Pool said "Social Blade's information is incorrect and unreliable."

Troubling content that YouTube does not remove under its Community Guidelines may nevertheless violate

its more stringent [Advertiser-Friendly Content Guidelines](#) and, as a result, be disqualified from carrying ads. Creators who repeatedly violate either set of policies may be excluded altogether from the YouTube Partner Program. The threat of "demonetization" constitutes another means, in addition to content down-ranking and removal, of protecting users from harmful material and assuring advertisers that their brands won't be sullied by association with such material. At the same time, however, demonetization of an individual video or an entire channel typically does not stop a creator from continuing to post content on YouTube. This carrot-and-stick approach reflects the company's determination to enforce its rules while trying not to alienate popular creators.

Increasingly, however, creators are circumventing the prospect of demonetization by using their YouTube channels to generate alternative cash flows on *other* platforms. Employing what's known as "alternative monetization," creators direct their YouTube loyalists to separate sites where the creators sell merchandise, collect donations, and charge for access to "exclusive" live-streamed events.

Especially popular among extreme right-leaning creators, this strategy undercuts, at least to some extent, YouTube's notion that demonetization constitutes a potent deterrent to online misbehavior, according to a [recent study](#) by researchers at Cornell Tech. YouTube's "moderation through demonetization is not likely to be an effective tool" for discouraging production of

## 'The Replacement': A Deadly Myth

YouTube hosts an array of heavily trafficked channels that aim to whip up opposition to immigration, according to Define American, a pro-immigrant group. Some of these channels amplify "replacement theory," a pernicious myth that Democrats aim to shift political power from white Americans to immigrants and other people of color who reliably support liberal elites. This [PragerU](#) video, entitled "Illegal Immigration: It's About Power," is narrated by Tucker Carlson, who has popularized the replacement concept on his Fox News show. The video has tallied 5.2 million views since 2018. The 18-year-old white shooter accused of targeting African Americans during a massacre in May 2022 in Buffalo, N.Y., cited the replacement idea in an online manifesto.



“problematic content,” the Cornell Tech team concluded.<sup>59</sup> Based on analysis of 136,000 channels and 71 million videos, the researchers found that 61% of “alt-right” and misogynist creators are pointing their YouTube followers elsewhere—for example, to buy politically themed products and make direct contributions via fundraising sites such as [patreon.com](https://www.patreon.com).<sup>60</sup>

## Election falsehoods

The extraordinarily bitter 2020 presidential election and its aftermath, combined with disruptions related to Covid-19, have provided opportunities to assess YouTube’s efforts to reduce various forms of harmful misinformation. Like its social media peers, YouTube has stepped up content moderation activities on both fronts—but not enough to prevent a continuing flow of false and conspiratorial content.

The Election Integrity Partnership, which brought together academic, civil society, and private sector researchers, published a [revealing assessment](#) of how misinformation spread from a variety of sources and was then further amplified on Twitter. YouTube was the third most-linked domain in the partnership’s sample of consequential falsehoods shared on Twitter, behind only far-right sites [The Gateway Pundit](#) and [Breitbart](#). Starbird, the University of Washington researcher and one of the partnership’s contributors, [described](#) YouTube’s role as “a place for misinformation to hide and be remobilized later.”<sup>61</sup>

One example mentioned in the partnership’s final report in 2021 was a YouTube livestream by right-wing creator Steven Crowder titled, “[Democrats Try to Steal Election!?](#)” The video, which was posted on November 4, 2020, and has tallied more than five million views, remains available. YouTube appended a text panel—one which appeared with all 2020 election-related content—saying, “Robust safeguards help ensure the integrity of election results.” The company also provided a link to the

## YouTube’s Data Gaps

Social media companies collect users’ personal data so that marketers can target specific types of consumers. The platform companies also analyze prodigious amounts of content in an effort to police their platforms for material that violates their rules. But for all of this data analysis, Google, YouTube, and other platforms withhold key information that the public and policy makers need to form sound opinions about the effectiveness of content moderation.

Google [discloses](#) statistics about the number of YouTube channels and individual videos that are removed for repeated violation of its guidelines on violence, hate speech, harassment, and the like. But it’s difficult to know what to make of these numbers because they represent the numerators of fractions lacking denominators.

Consider channel removals. In the final three months of 2021, YouTube took down 3.9 million channels, 90% of which were removed because they trafficked in spam, misleading content, or scams, according to Google’s latest [Transparency Report](#). The next-largest category, with about 4%, were channels featuring sexual content or nudity.<sup>1</sup> But how impressive is the removal of 3.9 million channels? It’s hard to say, since Google doesn’t estimate how many rule-breaking channels there were in all. That larger universe would provide the denominator for the fraction, constituting the platform’s success rate.

The company reports that of those videos taken down, about one third are removed by its automated screening system before anyone sees them, roughly another third are caught after they’ve been viewed no more than 10 times. The final third come down after having been viewed 11 or more times. So, what can we tell from that? Again, not a lot.

Artificial intelligence has improved enough to nab millions of rule-breaking videos before they do any harm. But most of those automated removals probably concern garden variety spam—relatively easy to detect, potentially annoying, but not exactly corrosive of democracy. Of the one-third slice that do get seen, we don’t know whether they are viewed 11 times or 11 million times. Google doesn’t say. Another set of missing data are statistics on false positives: how often YouTube’s moderation systems mistakenly remove videos, such as those meant to document human rights abuses or war crimes.

In April 2021, Google began releasing a new metric: YouTube’s “[violative view rate](#).” VVR estimates the proportion of video views of content that violates platform rules (excluding spam). In other words, VVR measures how often users encounter harmful content. Google says that VVR is moving in the right direction, which is to say downward, by 81% since the fourth quarter of 2017.

<sup>1</sup><https://transparencyreport.google.com/youtube-policy/removals?hl=en>

<sup>2</sup><https://transparencyreport.google.com/youtube-policy/views?hl=en>

To calculate VVR, human reviewers manually analyze a sample of actual YouTube views to determine how often violative videos slipped past automated and human content moderation filters. In late 2021, the estimated rate was 0.12% to 0.14%.<sup>2</sup>

But once more, it's not clear what those percentages mean. If there were 10,000 total views of content on YouTube in a given quarter, 0.14% would suggest a modest-sounding 14 instances of violative views. But users watch a billion hours of video a day. At that scale, the number of daily views of harmful content would be measured in the millions.

## Self-Policing on YouTube

### Removals (October 1 - December 31, 2021)

The platform uses automated and human review to filter out harmful content but doesn't disclose the percentage of content that it takes down.

 **Channels**  
**3,850,275**

 **Videos**  
**3,754,215\***

\*YouTube removed an additional 77,327,364 videos associated with channels that it took down.  
Source: [Google](#)

### Declining 'Violative View Rate'

Google estimates the percentage of YouTube views of rule-breaking content (fourth quarter of each year).

<b>2017</b>	<b>0.63% - 0.72%</b>	
<b>2018</b>	<b>0.23% - 0.27%</b>	
<b>2019</b>	<b>0.16% - 0.18%</b>	
<b>2020</b>	<b>0.16% - 0.18%</b>	
<b>2021</b>	<b>0.12% - 0.14%</b>	

Source: [Google](#)

website of the federal Cybersecurity and Infrastructure Security Agency, which debunked the myth that the election had been "stolen."

As former President Trump and his supporters promoted that myth in late 2020, YouTube [moved more tentatively](#) than Facebook or Twitter to counter disinformation about rogue ballots and rigged voting machines. On December 9, the deadline for states to certify their election results, [YouTube announced](#) that it would start removing misleading allegations that "widespread fraud or errors changed the outcome." But the new policy applied only to videos uploaded after December 9. An untold number of videos such as Crowder's remained on the site, contributing to the erosion of trust in elections and democracy itself.<sup>62</sup>

During the January 6 melee at the U.S. Capitol, Trump [used his YouTube account](#) to reassure the mob. "Go home, we love you, you're very special ....I know how you feel," he said in a video from the White House. Six days later, after [both Twitter and Facebook](#) had barred Trump, [YouTube blocked him](#) as well, citing an "ongoing potential for violence." Since then, YouTube hasn't specified when Trump's account might be reactivated, saying only that [he would be allowed to return](#) "when the risk of violence has decreased." Meanwhile, on January 13, 2022, the Select House Committee investigating the Capitol insurrection [alleged in a subpoena](#) to Google that YouTube "was a platform for significant communications by its users that were relevant to the planning and execution of the January 6th attack on the United States Capitol."

Defending its actions related to the election fallout, YouTube told us in May 2022 that "over the last year, we've removed tens of thousands of videos for violating our U.S. elections-related policies, the majority before hitting 100 views. In addition, our systems actively point to high-

authority channels and limit the spread of harmful misinformation for election-related topics. We remain vigilant ahead of the 2022 elections.”

Improvements on YouTube appear to have a ripple effect elsewhere. Researchers with NYU’s Center for Social Media and Politics [found](#) that the platform’s December 2020 policy change to prohibit stolen-election claims was followed by a reduction in the prevalence of videos containing such false claims on Facebook and Twitter. The finding, described in a forthcoming paper, indicates once again the enormous influence that YouTube has across the rest of the internet.

## Video and the virus

Beginning early in the pandemic, [YouTube](#), along with Facebook and Twitter, banned coronavirus misinformation pushing phony cures, minimizing the seriousness of the disease, and undercutting public health directives on masking and distancing. These steps represented a reversal of the platforms’ previous insistence that they were not “[arbiters of the truth](#)” and therefore should not police misinformation. On at least some truths, YouTube and the others now conceded that they had to take a stand. They would favor fact-based content and point users to authoritative medical information.

But even as YouTube took down hundreds of thousands of specious coronavirus videos—a total of more than 1.5 million by December 31, 2021—more kept surfacing. Playing whack-a-mole with dangerously false content often didn’t work, at least in part because YouTube allowed some of the most prolific promoters of such misinformation to keep their accounts. When YouTube has disciplined channels for disseminating coronavirus misinformation, it often has imposed only short suspensions, such as the [week-long banishment](#) of the pro-Trump outlet One America News Network.<sup>63</sup>

Sixteen months into the pandemic, the YouTube accounts of half of the 12 anti-vaccine activists named publicly by the [Center for Countering Digital Hate](#) as being responsible for creating more than 50% of all of the anti-vaccine content shared on social media remained easily searchable—and were still posting videos.<sup>64</sup> It wasn’t until September 2021, when the company broadened its Covid-19 misinformation policy to [ban all vaccination misinformation](#), that YouTube finally removed several channels associated with two of the leading names on the top-dozen list: Joseph Mercola and Robert F. Kennedy Jr., both of whom said they were unfairly censored.<sup>65</sup> “YouTube is the vector for a lot of this misinformation,” [Lisa Fazio](#), an associate professor of psychology at Vanderbilt, told *The Washington Post* in September 2021. “If you see misinformation on Facebook or other places, a lot of the time it’s YouTube videos. Our conversation often doesn’t include YouTube when it should.”<sup>66</sup>

“

**Sixteen months into the pandemic, the YouTube accounts of half of the 12 anti-vaccine activists named publicly by the Center for Countering Digital Hate as being responsible for creating more than 50% of all of the anti-vaccine content shared on social media remained easily searchable—and were still posting videos.**

”

# The Other 80%: YouTube Outside of the U.S.

Even as YouTube serves as a major vector in the U.S. for mis- and disinformation about Covid-19 and election fraud, activists who track social media around the world say that the video platform plays an even more problematic role in other countries.

Vladimir Putin's use of online disinformation over many years to try to justify Russia's aggression against Ukraine provides a grim reminder of how violent actors can exploit social media platforms. In the most egregious example of this phenomenon, Facebook employees failed beginning as early as 2014 to act on waves of [online misinformation and hateful attacks](#) in Myanmar that contributed to the ethnic cleansing of more than 700,000 Muslim Rohingya by that country's military and allied Buddhist groups.

In response to a global outcry, Meta Platforms (then Facebook) belatedly acknowledged that it did not do enough to prevent its platform from being used to pursue what the [U.S. has recognized as genocide](#). The company has substantially expanded its content moderation efforts in Myanmar and several other politically volatile countries.<sup>67</sup>

In an [open letter](#) published in April 2022, a group of more than 30 civil society groups argued that in the years leading up to the Russian invasion of Ukraine, the major social media platforms repeated some of the mistakes Meta made in Myanmar. The platforms tolerated the spread of Russian disinformation about Ukraine and allowed themselves to be manipulated by coordinated Kremlin campaigns to mass-report content critical of Russia or Putin as violating platform rules.<sup>68</sup>

While all of the platforms need to be more vigilant, a distinction has emerged, Dia Kayyali, one of the organizers of the letter, told [Tech Policy Press](#). An associate director for advocacy at Mnemonic, a group that archives and investigates

digital information about human rights violations, Kayyali said that recently, "Meta and Twitter have made significant efforts in these areas." Conversely, the activist added, YouTube's parent, Google, "has deeply underinvested in civil society engagement and due diligence, especially considering what a huge company it is."<sup>69</sup> Beyond its failure to sufficiently police harmful content, Kayyali emphasizes, YouTube frequently removes violent videos intended to preserve evidence of war crimes.

Even as YouTube serves as a major vector in the U.S. for mis- and disinformation about Covid-19 and election fraud, Kayyali and others who track social media around the world say that the video platform plays an even more problematic role in other countries. More than 80% of YouTube's traffic [comes from outside the U.S.](#), and even less is known about its business practices and impact on the information environment beyond its home market.<sup>70</sup>

YouTube disputes these criticisms. Its human content moderators "support numerous languages, including Rus-

sian and Ukrainian,” while its automated review system “is what allows us to enforce our policies at scale,” the company told us. “We engage regularly with civil society, human rights, and advocacy organizations around the world,” and more than 300 NGOs and government agencies have joined its [Trusted Flagger program](#), which allows participants to efficiently report they believe YouTube should remove.

Considerable anecdotal evidence, however, suggests that problems persist and are widespread. In South Korea, an “anti-feminist” group that [once used](#) the motto “Till the day all feminists are exterminated” runs a YouTube [channel](#) with 475,000 subscribers.<sup>71</sup> Research by Jonas Kaiser of Suffolk University and Adrian Rauchfleisch of National Taiwan University shows that in Germany, far-right groups such as [Pegida](#) and the [Identitarian movement](#), as well as the

right-wing political party [Alternative for Germany \(AfD\)](#), have built an increasingly centralized network of conspiracy and alternative news channels on YouTube, often exploiting resentment targeting refugees.<sup>72</sup>

The following look at India, Brazil, and Myanmar illustrates the challenges that arise from YouTube’s lack of transparency and its apparent dearth of regional and local expertise in identifying misinformation and hate speech.

## India

India’s more than 450 million YouTube users make it the platform’s [biggest market](#)—nearly double the size of its American user base. The most troubling abuse of YouTube in India involves the targeting of Muslims by backers of the ruling Bharatiya Janata Party and other right-leaning Hindu nationalist groups.

One [pernicious conspiracy theory](#) recently amplified via video asserts that Muslims purposefully spread coronavirus in India as a form of jihad. Mundane rivalries between groups of street vendors have turned violent after YouTube-fueled campaigns [singled out Muslim merchants](#) for attack.<sup>73</sup> In a 2021 report called “How Anti-Muslim Content Spreads on YouTube in India,” First Draft News, a nonprofit that counters hate speech and misinformation, identified a number of Islamophobic channels, each with more than a million subscribers, that demean Muslims and incite violence against them. Religious intolerance long predated the arrival of YouTube in India, but widespread social media use has intensified the hostility. First Draft questioned whether the platform’s content moderation efforts in the country—which YouTube has refused to describe with specificity—have any hope of slowing the proliferation of bigoted content.<sup>74</sup>

Menacing online attacks on women often blend with anti-Muslim themes in India. A [spate of misogynistic rants](#) by nationalistic Indian YouTube influencers have made such invective popular on the platform. The diatribes, many of which include physical threats, are often delivered as selfie videos. In 2019, YouTube reportedly bestowed upon one of these video makers—an individual who at the time had 800,000 subscribers—a “Silver Creator” award. When YouTube deleted some of the misogynistic accounts, creators simply started new ones.<sup>75</sup> Another genre of hate videos features photos of Muslim women gleaned from public sources and facetiously puts the subjects “[up for sale](#),” sometimes leading to abusive comments and talk of rape.<sup>76</sup>

Karen Rebelo, an investigative reporter with the Indian fact-checking organization BOOM, told us that the video platform has brought benefits as well as problems. “Like other social media, YouTube has democratized the news publishing landscape, giving space to independent journalists and small and local organizations with limited resources,” she said. “But it has also diluted the quality of journalism and allowed

### Hindu Hostility: Fueling Anti-Muslim Hatred in India

News Nation, a national Hindi news outlet with more than three million subscribers on YouTube, has spread anti-Muslim animus, such as this video in which Raj Thackeray, a prominent politician, suggests that certain members of Tablighi Jamaat, an orthodox Islamic missionary group, should be “shot dead.” The non-profit First Draft News, which researches online disinformation, [pointed out](#) in a 2021 report that other popular Indian outlets, including ABP News and Zee News, have used their YouTube channels to amplify “misleading and inflammatory claims” blaming Muslims for the spread of Covid-19, among other societal ills.



the rise of propaganda to masquerade as news.” One now-common tactic, she said, occurs “during major news events, natural calamities, terrorist attacks, and military conflicts, where old and unrelated videos are shared out-of-context on YouTube.”<sup>77</sup>

The platform faces a difficult task in monitoring and taking action against such activity, according to Prateek Waghre, a researcher with the Technology and Policy Program at the Bangalore-based Takshashila Institution. “You will rarely find a YouTube content creator who sticks to just one language, or especially just English. They are most likely switching between at least English, Hindi, probably a bunch of other languages as well,” Waghre told us. The linguistic jumps occur not just within videos, but sometimes within a single sentence. Complicating these challenges, the Indian government enacted a law in 2021 giving itself the authority to demand removal or reinstatement of content on social media platforms.<sup>78</sup>

## Brazil

With more than 100 million YouTube users in a nation of 215 million, Brazil is the platform’s fourth-largest market, behind only India, Indonesia, and the U.S. By many accounts, the video-sharing service has helped enable the rise of the hard-right political movement led by Jair Bolsonaro. A former military officer and obscure legislator who leapt to prominence and was elected president in 2018, Bolsonaro and his supporters have harnessed the extraordinary reach of YouTube and other social media platforms. Bolsonaro, whose harsh tone and savvy communication tactics are often compared to those of former U.S. President Donald Trump, was an early adopter of YouTube, using the platform to spread videos that, according to *The New York Times*, “falsely accused left-wing academics of plotting to force schools to distribute ‘gay kits’ to convert children to homosexuality.”

Bolsonaro’s supporters rely heavily on messaging apps WhatsApp and Telegram to attack his opponents

and amplify his factually challenged positions on topics such as Covid-19 vaccinations, which the Brazilian head of state has suggested are dangerous. These efforts often involve the circulation of videos initially posted on YouTube. Ricardo Gandour, the São Paulo-based former editor-in-chief of the newspaper *O Estado de S.Paulo*, told us that Bolsonaro’s ascendancy underscores the power of social media. And he adds: “YouTube is no doubt the most dangerous platform for disinformation in Brazil.”

Tai Nalon, chief executive of the Brazilian fact-checking organization Aos Fatos, concurs that YouTube plays a central role as the video depository for users of other platforms and that the “Bolsonarista disinformation agenda occupies a disproportionate space on social media.” He describes the president’s “network of influence” as grounded in “the exaltation of masculinity, fascist nostalgia, and conspiracy paranoia.” Aos Fatos has documented that, as of March 21, 2022, Bolsonaro has made at least 5,084 false or misleading statements since taking office. The president, for his part, has lashed out at fact-checkers and their “absurd arbitrariness.”<sup>80</sup>

Brazilian misinformation on YouTube goes beyond the political realm. A study published in 2021 in the academic journal *Media, Culture & Society* concluded that 80% of a sample of nearly 600 YouTube videos about a 2019 oil spill along Brazil’s northeast coast consisted of misinformation. Among the consequences of this kind of pollution of the information environment, the study’s authors wrote, are “algorithmic resistance” to fact-based journalism and erosion of “public debate about environmental catastrophes in favor of private interests (fake politics).”<sup>81</sup>

YouTube is buffeted by conflicting political and legal forces in Brazil. It confronted Bolsonaro in 2021 by removing videos posted on his channel that made false claims about Covid-19 cures and a supposed link between coronavirus vaccines and AIDS.

In the summer of 2021, the country’s federal electoral court ordered the platform to suspend shared-revenue payments to more than a dozen channels accused of spreading election disinformation. Then, in the fall, Bolsonaro pushed in the other direction, issuing a temporary ban on social media companies removing certain claims, including his statements that if he fails to win reelection this year, voting fraud will be the reason.<sup>82</sup> Nevertheless, the platform has promised that it would apply to Brazil the same policies it began enforcing in the U.S. in 2020, removing baseless videos seeking to undermine trust in elections and democracy.<sup>83</sup>

Google has funded a variety of initiatives in Brazil and elsewhere in Latin America to promote digital media literacy and the dissemination of accurate information about politics and public events. But some who are familiar with the practical effects of such efforts remain skeptical. For such a wealthy company, it is a “cup of coins—you know, pocket money,” Gandour, the São Paulo journalist, told us.

## Myanmar

YouTube is a distant second to Facebook in terms of popularity among social media users in Myanmar. But the video-sharing site’s expanding role as a conduit for political disinformation in the troubled Southeast Asian country was becoming clear to fact-checking organizations and other civil society groups even before Myanmar’s military retook power in February 2021.

The army coup was cheered by dozens of pro-military YouTube channels, some with millions of views, that spread falsehoods undermining the election in 2020 of a fragile civilian government, which was then swept aside by the military just a few months later. Some of the channels posed as news outlets, and collectively the bogus YouTube videos often had greater reach than actual journalism organizations.<sup>84</sup> (The company hasn’t launched a local iteration of the platform in Myanmar; users there see a generic global version.)

Victoire Rio, an adviser at the Myanmar Tech Accountability Network, told us that well before the election in 2020, her group was raising red flags to YouTube about problematic activity on the platform, including the proliferation of channels amplifying pro-military disinformation. In one instance, she says, videos hosted on YouTube were linked to an email campaign impersonating, or “spoofing,” the accounts of election commission officials to add credibility to phony voter-fraud claims. Though YouTube had adopted a policy against [false claims of widespread electoral fraud](#) in the U.S., which the company has since expanded to Germany and Brazil, it has not applied the policy in Myanmar.

Rio also expressed concern about the role of financially motivated actors who exploit YouTube videos about Myanmar and as a result spread falsehoods across the broader information ecosystem. Her research on a network of Cambodian and Vietnamese clickbait purveyors informed an extensive [report by MIT Technology Review](#) in 2021, which described how Google and Facebook monetization programs had the effect of amplifying disinformation: Opportunistic Cambodians and Vietnamese repost viral YouTube videos from Myanmar on blogs and Facebook, not for political purposes, but to monetize views. In some cases, footage that originally had been uploaded to YouTube as evidence of human rights violations was repurposed in this manner and, in the process, misled viewers into thinking that the abuses were happening again. This kind of profit-motivated manipulation of video content has inadvertently intensified the confusion and fear felt by Myanmar citizens who are bombarded with propaganda from their military rulers and other murky sources.<sup>85</sup>

YouTube says it is responding to problems in Myanmar when it discovers channels or content that violate its terms of service. It terminated a number of channels that *MIT Technology Review* brought to its attention in 2021. In the wake of the February 2021 coup and subsequent army crackdowns, YouTube told us that it has “removed

thousands of videos and hundreds of channels” for promoting violence or ethnic hatred. For example, in August 2021, [it terminated 137 channels](#) that were part of a coordinated Burmese-language campaign that supported the military coup.<sup>86</sup>

But Rio says that in Myanmar, the military and its allies often simply start new channels, or continue to exploit others that the company has not yet removed. “Instead of rooting them out, they are just addressing them one channel at a time or one video at a time,” says Rio. Like Mnemonic’s Kayyali, Rio believes that YouTube has invested inadequate resources to address these complex challenges.

Dangerous exploitation of YouTube crops up in unexpected ways—in one instance, whipping up hostility to migrant workers from Myanmar who seek employment in neighboring Thailand. During the pandemic, a Thai civil society group called Social Media Monitoring for Peace pointed out a [wave of hateful attacks](#) by Thai people on YouTube and other platforms. One Thai YouTube commenter said: “Wherever you see Myanmar people, shoot them down.”<sup>87</sup>

“

**The army coup was cheered by dozens of pro-military YouTube channels, some with millions of views, that spread falsehoods undermining the election in 2020 of a fragile civilian government, which was then swept aside by the military just a few months later.**

”

# Conclusion and Recommendations

**Enactment of limited government regulation would not relieve Silicon Valley of the need to take action of its own. Because of the First Amendment, the U.S. government cannot seek to cleanse online platforms of the many forms of expression that, while not illegal, are still harmful.**

In 2017, when the NYU Stern Center began publishing reports about social media's effects on democracy, we stressed a [preference for industry self-regulation](#) as a way to avoid the First Amendment complications that would accompany government intervention. Those free speech complications remain: Government must not set content policy or participate in substantive content decisions. But some time ago, we realized that the continuing insufficiency of self-regulation meant that government would have to step in, cautiously.

In the recommendations that follow, we discuss as an initial matter the responsibilities of YouTube and Google—and by extension, of all of the major platforms. These obligations include far more transparency about how their technology works and how it interacts with human decisions that determine what content users see, which items get amplified, which get down-ranked or removed, and why. We also reiterate a proposal we published in early 2022 for legislation that would enhance the Federal Trade Commission's consumer protection authority to oversee social media companies. We believe lawmakers and regulators can put this idea into practice without violating the First Amendment's wise constraint on government interference with speech.

Enactment of limited government regulation would not relieve Silicon Valley of the need to take action of its own. Because of the First Amendment, the U.S. government cannot seek to cleanse online platforms of the many forms of expression that, while not illegal, are still harmful. Hateful speech that doesn't directly incite imminent

violence comes to mind, as does most mis- and disinformation, even on vital topics such as public health requirements in the midst of a lethal pandemic. Government can mandate disclosure of data relevant to such harms, and it can require that companies allocate resources necessary to fulfill promises they make about mitigating them.

But as a practical matter, the platforms themselves inevitably will bear the bulk of the responsibility for policing expression they host. The scale and speed of their operations, their immediate access to content on their sites, and the complexity of their technology dictate that the platforms cannot be overseen the way a government inspector scrutinizes a meatpacking or pharmaceutical plant. Regulation of social media, by necessity, will be a hybrid affair.

# Recommendations

## To YouTube

### 1 Disclose more information about how the platform works

For all of the controversy in recent years about the detrimental side effects of widespread use of social media, the public, researchers, and policymakers still know very little about the inner workings of YouTube and other major platforms. Without endangering user privacy or proprietary trade secrets, YouTube could win greater public trust and better inform political debates if, for the first time, it explained the specific criteria its algorithms use to rank, recommend, and remove content—as well as how often and why those criteria change and how they are weighted relative to one another.

To what degree, for example, have product designers prioritized user engagement, a factor that can lead to the amplification of content that stokes anger and fear? What attributes are likely to cause a video to “go viral”? And beyond the 2019 recalibration of the platform’s recommendation system, what mechanisms have designers incorporated to slow virality if there are indications that content could stir ethnic animosity or incite violence?

Other topics that deserve more sunlight include the size and nature of YouTube’s content moderation corps. The company has resisted revealing basic information such as the proportion of reviewers who are outsourced, meaning they are not directly employed by YouTube but instead by third-party contractors. As we discuss in more detail in recommendation 3, outsourced workers are less likely to be effective in this critical function because they do not receive direct supervision from senior YouTube employees and executives.

### 2 Facilitate greater access to data that researchers need to study YouTube

The video-sharing platform has not provided researchers with the sort of data access that, for example, Twitter has. To do more thorough empirical studies, researchers have told us that they need more information via YouTube’s API, or application programming interface. Currently, YouTube provides access only to content and metadata that are available at the time researchers connect to the API. To understand how a given video gained attention over time, they need access to historic data as well.

Social scientists tell us that another potentially fruitful feature would be the ability for researchers to retrieve random samples of YouTube content, a need that may arise when a given search query otherwise returns an unmanageable number of results. Random sampling would allow researchers to make inferences from a subset of data and would better support the tracking of important trends. Some disclosures go beyond data access via the API. At present, YouTube doesn’t systematically disclose when it makes policy changes that may bear on the availability of problematic content.

We understand that YouTube currently is in touch with certain academic researchers about these issues, and we hope this contact leads to progress. Researchers and the public at large would benefit from a greater understanding of how the platform functions.

### 3 Expand and improve human review of potentially harmful content

The greater transparency discussed in the first two recommendations would help clarify how YouTube can expand and improve its content moderation system, especially in the design and operation of the automated filtering responsible for the vast majority of video removals. But it’s a safe bet that the company also can do a better job deploying the human reviewers who are essential to the policing of the platform. Their role is critical because machines typically lack the ability to distinguish between, say, the glorification of Nazism, which should be removed, and legitimate historical or satirical material about Nazism—a distinction that [YouTube’s algorithm has at times struggled with](#).<sup>88</sup>

This is a massive task that will require Google to invest significantly greater resources, especially in countries in the Global South. Given that YouTube users watch a billion hours of video a day, the company needs more than

# Recommendations

the 20,000 people it says are currently involved in content moderation. That figure includes engineers, product specialists, policy experts, and others whose jobs relate to moderation but do not themselves review videos. YouTube declined to provide us with the number of hands-on reviewers.

It's difficult to say how much larger the reviewer workforce ought to be because we know so little about how these people interact with the platform and the automated portion of its content moderation effort. For example, YouTube will not say what percentage of its reviewers are outsourced employees of third-party vendors, as opposed to direct employees of the social media platform.

As we discussed in a [report published in 2020](#), the industry-wide practice of outsourcing a majority of moderation duties tends to have two unfortunate effects: First, it results in reviewers receiving inadequate on-site psychological counseling and other health benefits needed in light of the traumatizing videos to which many are repeatedly exposed. Second, reviewers tend not to get the kind of intensive, skilled supervision that they more likely would receive if they were full-time platform employees. Better internal oversight would lead to more effective performance and a reduction in levels of harmful content.<sup>89</sup>

## 4

### Invest more in relationships with civil society and news organizations

Responding to criticism that YouTube has not done enough to counter the spread of misinformation, the company told us that it had contributed \$1 million to the International Fact-Checking Network. We commend Google for supporting this worthy cause. But corporate generosity requires context. Alphabet, the holding company for Google and YouTube, reported [2021 revenue of \\$258 billion](#), a 47% jump from 2020. With resources of that magnitude, the combined company can and should do much more.

Since 2018, Google has [committed more than \\$300 million](#) to provide training and grants to news organizations struggling to adjust to the digital business environment. The central economic characteristic of this environment is that social media platforms like YouTube and search engines like Google have siphoned off much of the advertising revenue that once supported traditional news-gathering. In light of these circumstances, YouTube and its corporate affiliates need to step up their efforts to support competent journalists, editors, and organizations capable of reporting and analyzing local, national, and international events.

Deeper, more substantive engagement between the industry and civil society groups about both the diminishment of harmful content levels and preservation of evidence of human rights abuses in conflict zones would generate greater public trust and loyalty.

## To the U.S. government

## 5

### Allocate political capital to reduce the detrimental side effects of social media

President Biden has periodically indicated irritation about social media platforms, but usually in [cryptic, non-specific](#) comments.<sup>90</sup> Now is the time for him to make a substantial statement on the topic in a sit-down interview or a speech. YouTube's dual role in Russia—providing news to ordinary Russians but also serving as a conduit to the West for Putin's propaganda—provides Biden with a timely illustration of the good and the bad that social media can deliver. Former President Obama's recent [speech at Stanford](#) about disinformation offered a useful model for the incumbent.

In terms of a substantive agenda, Biden should provide leadership in two ways: using his influence to exhort the social media industry to self-regulate more vigorously and, at the same time, helping shape Congressional debate to improve the chances that lawmakers eventually adopt legislation providing for effective federal oversight of the industry. Both approaches are necessary. The government must act because to date the industry has not sufficiently policed itself. But because of the First Amendment, government has only limited room to maneuver. Government must refrain from interfering with either users' right to express themselves or platforms' right to decide what content they host. Below, we outline some of the options available to the government.

# Recommendations

## 6 Enhance federal consumer protection authority to oversee social media

In a February 2022 [white paper](#), our Center made the case for legislation authorizing the Federal Trade Commission to use its consumer protection authority to provide, for the first time, systematic oversight of social media companies. The watchwords of consumer protection enforcement are deterring “unfair or deceptive” practices. For generations, the FTC has stopped businesses in a wide array of industries from using false promises about products, intimidation, and phony offers of “free” service. The agency, which in recent decades has been starved for funds and staff, needs a Congressional directive empowering it to impose carefully circumscribed regulation of social media. The FTC would have to do so without dictating substantive content policies or decisions, which would violate the First Amendment’s ban on government interference with free speech.

The two central elements of this proposal, which draws on ideas in several pending bills, are transparency and procedurally adequate content moderation.<sup>91</sup> Congress should authorize the FTC to oversee mandatory disclosure of currently secret information about how the companies’ automated systems rank, recommend, and remove content: what criteria are programmed into the algorithms that on a daily basis decide whether billions of written communications and images “go viral” or are relegated to the bottom of user feeds? How do these artificial intelligence-driven systems interact with human content moderators? If companies do not voluntarily expand the range of data available to outside social scientists (see recommendations 1 and 2), the legislation would define this recalcitrance as an “unfair or deceptive” practice, forcing disclosure.

The requirement of procedurally adequate moderation would oblige platforms to fulfill the promises they make to users in terms of service and community standards. The FTC would ensure that the standards are clear and internally consistent, that enforcement decisions are explained in a way that affected users can understand, and that users have ready access to an appeals process. The agency also would have authority to assess whether content moderation resources—budgets, personnel, and management attention—are commensurate with the daunting task.

One question would be whether the common industry practice of outsourcing the vast majority of human review has undercut effectiveness and should be ended, with all moderators brought in-house. Additional questions include whether platforms employ sufficient people with the language skills and cultural awareness to allow for meaningful internal oversight of whether sites are being exploited by malign actors in the more than 150 countries where social media services are available. By limiting its attention to procedural issues—such as whether platforms follow through on promises they make about protecting users from harmful content—the FTC can avoid inhibiting the First Amendment rights enjoyed by both users and the platforms themselves.

Far ahead of the U.S. on the regulatory front, the European Union was finalizing its [Digital Services Act](#) in the spring of 2022. Although fidelity to free speech principles is strong in most of the 27-nation union, the E.U. has more room to legislate in the absence of a stringent provision comparable to the First Amendment. Despite this difference, American lawmakers and regulators ought to look to the DSA for painstakingly crafted requirements on platform transparency, internal risk assessment, and independent auditing of these assessments.<sup>92</sup>

# Endnotes

- 1 <https://bhr.stern.nyu.edu/ftc-whitepaper>
- 2 <https://www.techtransparencyproject.org/articles/guns-drugs-and-skin-bleaching-youtube-kids-still-poses-risks-children>; <https://www.theguardian.com/technology/2022/may/05/youtube-kids-shows-videos-promoting-drug-culture-firearms-toddlers>
- 3 <https://www.washingtonpost.com/technology/2022/03/16/facebook-youtube-russian-bans/>
- 4 <https://www.nytimes.com/2017/10/23/technology/youtube-russia-rt.html>
- 5 [https://www.dni.gov/files/documents/ICA\\_2017\\_01.pdf](https://www.dni.gov/files/documents/ICA_2017_01.pdf)
- 6 <https://www.politico.com/news/magazine/2022/03/29/law-counter-putins-propaganda-unused-00020805>
- 7 <https://time.com/6156927/youtube-russia-ukraine-disinformation/>
- 8 In April, Roskomnadzor accused YouTube of spreading “fake” videos to discredit the Russian military. But rather than shut down the platform, the regulator imposed only mild discipline, such as fining Google and arranging for Russian search engines to attach a warning to YouTube videos saying that the company was violating Russian law. Separately, some digitally sophisticated Russians were able to use virtual private networks (VPNs) or the anonymizing website Tor to access versions of Facebook and Twitter.
- 9 Zelensky: <https://www.youtube.com/watch?v=VLJGVJrprs0>; Navalny's allies: <https://www.youtube.com/watch?v=p2gNZrfmcVU>. In an April 2022 interview with CNN, a top Navalny lieutenant called YouTube Russia's main source of “the actual real news.” As of mid-May, YouTube said it had removed more than 8,000 channels and more than 60,000 videos related to the war for violating its content policies on misinformation, hate speech, and graphic violence, among others.
- 10 <https://www.pnas.org/content/118/32/e2101967118>
- 11 <https://www.theguardian.com/technology/2022/may/22/youtube-ukraine-invasion-russia-video-removals>
- 12 <https://www.washingtonpost.com/technology/2022/03/16/facebook-youtube-russian-bans/>
- 13 YouTube has stumbled during the Russian invasion. Early in the war, it excluded independent broadcaster Echo of Moscow before criticism prompted it to restore the station.
- 14 <https://www.poynter.org/fact-checking/2022/an-open-letter-to-youtubes-ceo-from-the-worlds-fact-checkers/?s=03>
- 15 <https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/>
- 16 <https://blog.youtube/press/>
- 17 YouTube's near-monopolistic position in the long-form video market: <https://www.semanticscholar.org/paper/A-Longitudinal-Analysis-of-YouTube's-Promotion-of-Faddoul-Chaslot-e9b8266ec5d94660dde5f89b4e01586dcddf2515>
- 18 [https://abc.xyz/investor/static/pdf/20220202\\_alphabet\\_10K.pdf?cache=fc81690](https://abc.xyz/investor/static/pdf/20220202_alphabet_10K.pdf?cache=fc81690). TikTok has carved out a significant portion of the short-video market, especially among younger users: <https://www.emarketer.com/content/tiktok-youtube-marketers>.
- 19 In December 2017, YouTube said it would expand its content moderator workforce to 10,000: <https://blog.youtube/news-and-events/expanding-our-work-against-abuse-of-our/>. Today, the company told us, that figure stands at 20,000.
- 20 <https://www.theguardian.com/technology/2017/mar/25/google-youtube-advertising-extremist-content-att-verizon>
- 21 <https://blog.google/threat-analysis-group/tag-bulletin-q1-2022/>
- 22 <https://blog.youtube/inside-youtube/inside-responsibility-whats-next-on-our-misinfo-efforts/>
- 23 <https://chequeado.com/investigaciones/the-international-scheme-of-doctors-for-the-truth-a-denialist-trademark-registered-by-natalia-prego/>
- 24 <https://www.ellinikahoaxes.gr/2020/10/18/fuellmich-video-with-false-coronavirus-info/>
- 25 <https://www.nytimes.com/2019/08/11/world/americas/youtube-brazil.html>
- 26 Ethiopia: <https://www.accord.org.za/conflict-trends/the-interaction-of-mass-media-and-social-media-in-fuelling-ethnic-violence-in-ethiopia/>; Phillipines: <https://www.wired.com/story/youtube-philippines-election/>.
- 27 <https://bhr.stern.nyu.edu/polarization-report-page>
- 28 <https://www.npr.org/2021/04/13/986678544/exploring-youtube-and-the-spread-of-disinformation>
- 29 <https://www.wired.com/story/why-isnt-susan-wojcicki-getting-grilled-by-congress/>
- 30 <https://www.lawfareblog.com/documents-senate-intelligence-committee-publishes-two-reports-internet-research-agency>
- 31 Wojcicki does occasionally surface publicly, as in a 2019 interview with “60 Minutes” and, in 2021, a Business Insider profile and at the Atlantic Council, a think tank in Washington.
- 32 <https://tcrn.ch/3EWYVlz>
- 33 <https://www.washingtonpost.com/technology/2021/08/25/youtube-content-moderation-strategy/>
- 34 <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>
- 35 <https://www.tubefilter.com/2018/01/11/youtube-most-watch-time-driven-by-recommendations/>
- 36 <https://christchurchattack.royalcommission.nz/the-report/>
- 37 <https://blog.youtube/inside-youtube/the-four-rs-of-responsibility-raise-and-reduce/>
- 38 In a blog post, YouTube explained that it uses algorithms called “classifiers” to identify “borderline content.” The classifiers rely on human evaluators who use publicly available guidelines to rate channels and videos based on factors such as accuracy, deceptiveness, intolerance, and “potential to cause harm.”
- 39 <https://arxiv.org/abs/2003.03318>
- 40 <https://www.pnas.org/content/118/32/e2101967118>
- 41 In 2021, researchers at NYU's Center for Social Media and Politics published a study finding that YouTube's diminished recommendations of conspiratorial and “politically extreme” videos had a salutary ripple effect on other platforms: Sharing of such YouTube videos decreased on both Twitter and Reddit. But the researchers also determined that, in the wake of YouTube's intervention, sharing of videos from conspiracy-oriented YouTube channels increased significantly on Reddit, with no meaningful change on Twitter. Other academics are skeptical that YouTube's algorithms play a role in radicalization, emphasizing that the supply of extremist content merely corresponds to user demand.
- 42 “There has been, and there continues to be, user radicalization on YouTube,” according to a research team based at the Swiss Federal Institute of Technology in Lausanne. After studying user behavior as seen in more than 72 million comments on two million YouTube videos in 2020, ranging from mildly conservative to the far-right fringe, the Swiss-based group found that “users consistently migrate from milder to more extreme content.” But the researchers couldn't determine the mechanism causing this migration.
- 43 <https://www.adl.org/resources/reports/exposure-to-alternative-extremist-content-on-youtube>

- 44** The Nyhan-Chen team refined its analysis without changing its central findings in a [forthcoming paper](#).
- 45** <https://www.techtransparencyproject.org/articles/youtubes-filter-bubble-problem-worse-fox-news-viewers>
- 46** <https://www.nytimes.com/2022/06/01/technology/fringe-mainstream-social-media.html>
- 47** <https://www.justice.gov/usao-edny/pr/frank-james-charged-brooklyn-mass-shooting-new-york-subway>
- 48** <https://www.npr.org/2021/04/13/986678544/exploring-youtube-and-the-spread-of-disinformation>
- 49** <https://support.google.com/youtube/answer/72857?hl=en>
- 50** [https://abc.xyz/investor/static/pdf/2021\\_Q4\\_Earnings\\_Transcript.pdf?cache=0118641](https://abc.xyz/investor/static/pdf/2021_Q4_Earnings_Transcript.pdf?cache=0118641)
- 51** <https://journals.sagepub.com.proxy.library.nyu.edu/doi/pdf/10.1177/1527476419879919>
- 52** <https://www.youtube.com/channel/UCG749Dj4V2fKa143f8sE60Q>
- 53** Leaked emails: <https://www.youtube.com/watch?v=5IM4rnNcvPl>; vaccine passport chip: <https://www.youtube.com/watch?v=YRjKdziEewc>
- 54** <https://www.nytimes.com/2022/06/06/us/politics/proud-boys-charged-sedition-capitol-attack.html>. YouTube banned Jones in 2018 for violating its guidelines against promoting hatred and violence. Jones did not respond to emails seeking his comment.
- 55** <https://socialblade.com/youtube/user/timcasts>
- 56** <https://archive.ph/HhBrz>.
- 57** <https://timcast.com/members-area/section/timcast-irl/page/3/>
- 58** <https://spectatorworld.com/author/amber-athey/>
- 59** <https://arxiv.org/abs/2203.10143>
- 60** Disputing a central premise of the Cornell Tech study, YouTube maintains that demonetization is merely the withdrawal of a privilege, not a form of content moderation.
- 61** <https://www.platformer.news/p/how-youtube FAILED THE 2020 ELECTION>
- 62** Crowder did not respond to emails seeking his comment.
- 63** [https://www.hollywoodreporter.com/business/digital/youtube-suspends-oann-channel-over-covid-19-misinformation-4097052/](https://www.hollywoodreporter.com/business/digital/youtube-suspends-oann-channel-over-covid-19-misinformation-4097052). YouTube uses a three-strike system of escalating penalties for violation of its Community Guidelines.
- 64** [https://252f2edd-1c8b-49f5-9bb2-cb57bb47e4ba.filesusr.com/ugd/f4d9b9\\_b7cedc0553604720b7137f8663366ee5.pdf](https://252f2edd-1c8b-49f5-9bb2-cb57bb47e4ba.filesusr.com/ugd/f4d9b9_b7cedc0553604720b7137f8663366ee5.pdf)
- 65** <https://blog.youtube/news-and-events/managing-harmful-vaccine-content-youtube/>
- 66** <https://www.washingtonpost.com/technology/2021/09/29/youtube-ban-joseph-mercola/>
- 67** <https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html>
- 68** <https://www.eff.org/files/2022/04/13/openletterukrainesocialmediaplatforms16.pdf>
- 69** <https://techpolicy.press/human-rights-groups-call-on-tech-platforms-to-reform-practices-in-crisis-zones/>
- 70** <https://arxiv.org/abs/2203.10143>
- 71** <https://www.nytimes.com/2022/01/01/world/asia/south-korea-men-anti-feminists.html>
- 72** <https://www.tandfonline.com/doi/full/10.1080/08838151.2020.1799690>
- 73** <https://www.theguardian.com/world/2020/apr/13/coronavirus-conspiracy-theories-targeting-muslims-spread-in-india>; <https://thewire.in/communalism/hindutva-ecosystem-muslim-fruit-sellers-threat-india>
- 74** <https://firstdraftnews.org/articles/talibhi-jamaat-blamed-for-covid-19-spread-on-india-youtube/>
- 75** <https://foreignpolicy.com/2020/08/25/india-youtube-stars-hatemongers-nationalism/>
- 76** <https://www.bbc.com/news/world-asia-india-57764271>
- 77** <https://www.boomlive.in/fact-check/video-from-syria-peddled-as-fatal-crash-that-killed-cds-bipin-rawat-15932>; <https://www.boomlive.in/fact-check/indian-media-taliban-afghanistan-united-states-old-photos-videos-14578>
- 78** <https://www.cnbc.com/2021/04/20/indias-social-media-law-puts-big-techs-power-into-states-hands-critics-say.html>
- 79** <https://www.nytimes.com/2019/08/11/world/americas/youtube-brazil.html>
- 80** <https://www.poynter.org/ifcn/2022/aos-fatos-brazil-bolsonaro-disinfo-machine/>
- 81** <https://journals.sagepub.com/doi/full/10.1177/0163443720977301>
- 82** <https://www.nytimes.com/2021/09/09/world/americas/bolsonaro-social-networks.html>
- 83** <https://blog.youtube/intl/pt-br/news-and-events/youtube-brasil-e-international-center-for-journalists-icfj-anunciam-programa-de-apoio-ao-jornalismo-no-combate-a-desinformacao/>
- 84** <https://www.reuters.com/article/us-youtube-myanmar-misinformation-youtube-faces-complaints-of-lax-approach-on-overseas-election-misinformation-idUSKBN28S0QE>
- 85** <https://www.technologyreview.com/2021/11/20/1039076/facebook-google-disinformation-clickbait/>
- 86** <https://blog.google/threat-analysis-group/tag-bulletin-q3-2021/>
- 87** <https://www.reuters.com/business/media-telecom/anti-myanmar-hate-speech-flares-thailand-over-virus-2020-12-24/>
- 88** <https://www.techdirt.com/2021/05/12/content-moderation-case-study-youtubes-new-policy-nazi-content-results-removal-historical-education-videos-2019/>
- 89** <https://bhr.stern.nyu.edu/tech-content-moderation-june-2020>
- 90** <https://www.bbc.com/news/technology-55762104>; <https://www.nytimes.com/2021/07/16/us/politics/biden-facebook-social-media-covid.html>
- 91** These bills include, in the House, the Democratic-sponsored [Digital Services Oversight and Safety Act](#) and [Online Consumer Protection Act](#) and, in the Senate, the bipartisan [Platform Accountability and Transparency Act](#) and [Platform Accountability and Consumer Transparency Act](#).
- 92** [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_22\\_2545](https://ec.europa.eu/commission/presscorner/detail/en/ip_22_2545)

NYU Stern Center for Business and Human Rights  
Leonard N. Stern School of Business  
44 West 4th Street, Suite 800  
New York, NY 10012  
+1 212-998-0261  
[bhr@stern.nyu.edu](mailto:bhr@stern.nyu.edu)  
[bhr.stern.nyu.edu](http://bhr.stern.nyu.edu)

© 2022 NYU Stern Center for Business and Human Rights  
All rights reserved. This work is licensed under the  
Creative Commons Attribution-NonCommercial 4.0  
International License. To view a copy of the license,  
visit <http://creativecommons.org/licenses/by-nc/4.0/>.



**NYU | STERN**

Center for Business  
and Human Rights