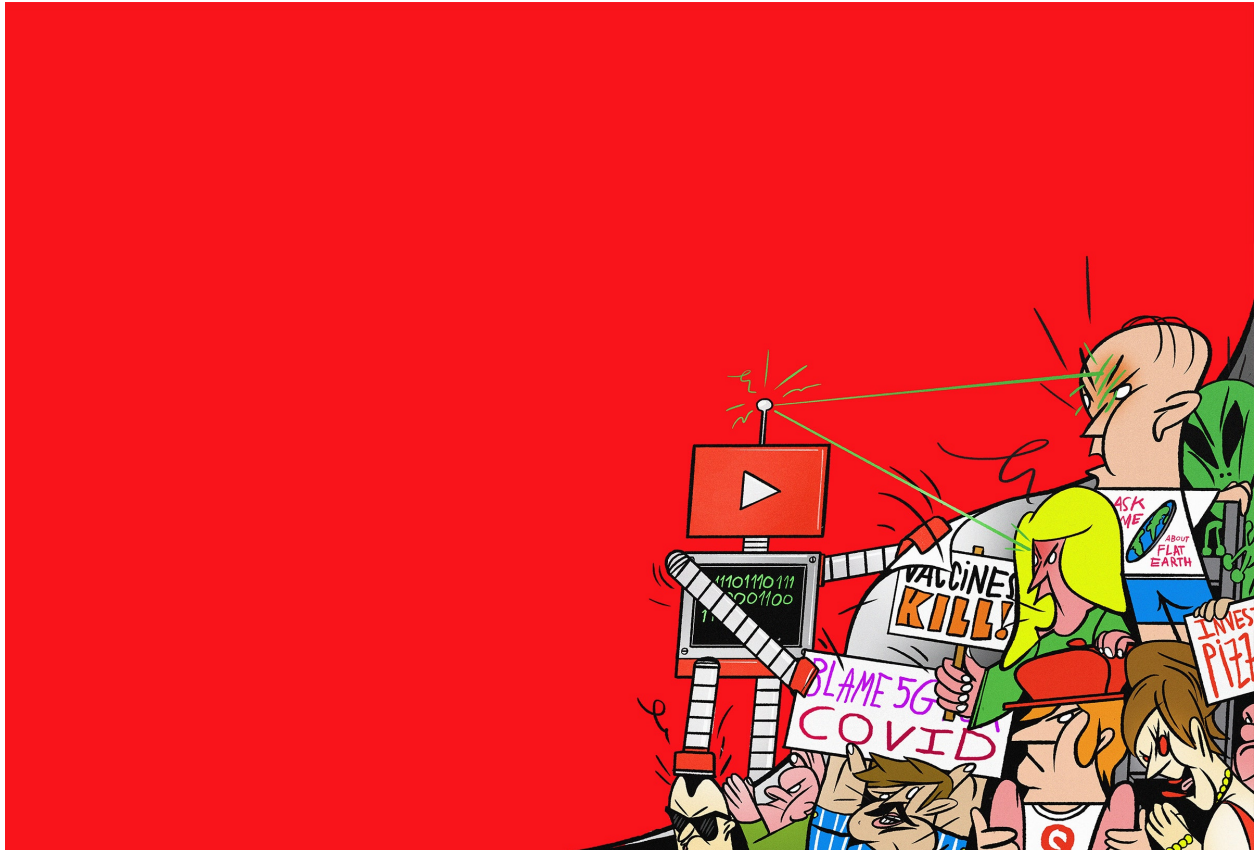# YouTube's Plot to Silence Conspiracy Theories

From flat-earthers to QAnon to Covid quackery, the video giant is awash in misinformation. Can AI keep the lunatic fringe from going viral?

ILLUSTRATIONS: BRAULIO AMADO

MARK SARGENT SAW instantly that his situation had changed for the worse. A voluble, white-haired 52-year-old, Sargent is a flat-earth evangelist who lives on Whidbey Island in Washington state and drives a Chrysler with the vanity plate "ITSFLAT." But he's well known around the globe, at least among those who don't believe they are living on one. That's thanks to YouTube, which was the on-ramp both to his flat-earth ideas and to his subsequent international stardom.

Formerly a tech-support guy and competitive virtual pinball player, Sargent had long been intrigued by conspiracy theories, ranging from UFOs to Bigfoot to Elvis' immortality. He believed some (Bigfoot) and doubted others ("Is Elvis still alive? Probably not. He died on the toilet with a whole bunch of drugs in his system"). Then, in 2014, he stumbled upon his first flat-earth video on YouTube.

He couldn't stop thinking about it. In February 2015 he began uploading his own musings, in a series called "Flat Earth Clues." As he has reiterated in a sprawling corpus of more than 1,600 videos, our planet is not a ball floating in space; it's a flat, *Truman Show*-like terrarium. Scientists who insist otherwise are wrong, NASA is outright lying, and the government dares not level with you, because then it would have to admit that a higher power (aliens? God? Sargent's not sure about this part) built our terrarium world.

Sargent's videos are intentionally lo-fi affairs. There's often a slide show that might include images of Copernicus (deluded), astronauts in space (faked), or Antarctica (made off-limits by a cabal of governments to hide Earth's edge), which appear onscreen as he speaks in a chill, avuncular voice-over.

Sargent's top YouTube video received nearly 1.2 million views, and he has amassed 89,200 followers—hardly epic by modern influencer standards but solid enough to earn a living from the preroll ads, as well as paid speaking and conference gigs.

Crucial to his success, he says, was YouTube's recommendation system, the feature that promotes videos for you to watch on the homepage or in the "Up Next" column to the right of whatever you're watching. "We were recommended constantly," he tells me. YouTube's algorithms, he says, figured out that "people getting into flat earth apparently go down this rabbit hole, and so we're just gonna keep recommending."

Scholars who study conspiracy theories were realizing the same thing. YouTube was a gateway drug. One academic who interviewed

attendees of a flat-earth convention found that, almost to a person, they'd discovered the subculture via YouTube recommendations. And while one might shrug at this as marginal weirdness—*They think the Earth is flat, who cares? Enjoy the crazy, folks*—the scholarly literature finds that conspiratorial thinking often colonizes the mind. Start with flat earth, and you may soon believe Sandy Hook was a false-flag operation or that vaccines cause autism or that Q's warnings about Democrat pedophiles are a serious matter. Once you convince yourself that well-documented facts about the solar system are a fraud, why believe well-documented facts about anything? Maybe the most trustworthy people are the outsiders, those who dare to challenge the conventions and who—as Sargent understood—would be far less powerful without YouTube's algorithms amplifying them.

For four years, Sargent's flat-earth videos got a steady stream of traffic from YouTube's algorithms. Then, in January 2019, the flow of new viewers suddenly slowed to a trickle. His videos weren't being recommended anywhere near as often. When he spoke to his flat-earth peers online, they all said the same thing. New folks weren't clicking. What's more, Sargent discovered, someone—or something—*was* watching his lectures and making new decisions: The YouTube algorithm that had previously recommended other conspiracies was now more often pushing mainstream videos posted by CBS, ABC, or *Jimmy Kimmel Live*, including ones that debunked or mocked conspiracist ideas. YouTube wasn't deleting Sargent's content, but it was no longer boosting it. And when attention is currency, that's nearly the same thing.

"You will never see flat-earth videos recommended to you, basically ever," he told me in dismay when we first spoke in April 2020. It was as if YouTube had flipped a switch.

In a way, it had. Scores of them, really—a small army of algorithmic tweaks, deployed beginning in 2019. Sargent's was among the first accounts to feel the effects of a grand YouTube project to teach its recommendation AI how to recognize the conspiratorial mindset and demote it. It was a complex feat of engineering, and it worked; the algorithm is less likely now to promote misinformation. But in a country

where conspiracies are recommended everywhere—including by the president himself—<u>even the best AI</u> can't fix what's broken

WHEN GOOGLE BOUGHT YouTube in 2006, it was a woolly startup with a DIY premise: "Broadcast Yourself." YouTube's staff back then wasn't thinking much about conspiracy theories or disinformation. The big concern, as an early employee told me, was what they referred to internally as "boobs and beheadings"—uploads of pornography and gruesome al Qaeda actions.

From the first, though, YouTube executives intuited that recommendations could fuel long binges of video surfing. By 2010, the site was suggesting videos using collaborative filtering: If you watched video A, and lots of people who watched A also watched B, then YouTube would recommend you watch B too. This simple system also up-ranked videos that got lots of views, under the assumption that it was a signal of value. That methodology tended to create winner-take-all dynamics that resulted in <u>"Gangnam Style"-type virality</u>; lesser-known uploads seldom got a chance.

In 2011, Google tapped Cristos Goodrow, who was then director of engineering, to oversee YouTube's search engine and recommendation system. Goodrow noticed another problem caused by YouTube's focus on views, which was that it encouraged creators to use misleading tactics—like racy thumbnails—to dupe people into clicking. Even if a viewer immediately bailed, the click would goose the view count higher, boosting the video's recommendations.

Goodrow and his team decided to stop ranking videos based on clicks. Instead, they focused on "watch time," or how long viewers stayed with a video; it seemed to them a far better metric of genuine interest. By 2015, they would also introduce neural-net models to craft recommendations. The model would take your actions (whether you'd finished a video, say, or hit Like) and blend that with other information it had gleaned (your search history, geographic region, gender, and age, for example; a user's "watch history" became increasingly significant too). Then the model would predict which videos you'd be most likely to

actually watch, and presto: recommendations, more personalized than ever.

The recommendation system became increasingly crucial to YouTube's frenetic push for growth. In 2012, YouTube's vice president of product, Shishir Mehrotra, declared that by the end of 2016 the site would hit a billion hours of watch time per day. It was an audacious goal; at the time, people were watching YouTube for only 100 million hours a day, compared to more than 160 million on Facebook and 5 billion on TV. So Goodrow and the engineers began thirstily hunting for any tiny tweak that would bump watch time upward. By 2014, when Susan Wojcicki took over as CEO, the billion-hour goal "was a religion at YouTube, to the exclusion of nearly all else," as she later told the venture capitalist John Doerr. She kept the goal in place.

The algorithmic tweaks worked. People spent more and more time on the site, and the new code meant small creators and niche content were finding their audience. It was during this period that Sargent saw his first flat-earth video. And it wasn't just flat-earthers. All kinds of misinformation, some of it dangerous, rose to the top of watchers' feeds. Teenage boys followed recommendations to far-right white supremacists and Gamergate conspiracies; the elderly got stuck in loops about government mind control; anti-vaccine falsehoods found adherents. In Brazil, a marginal lawmaker named Jair Bolsonaro rose from obscurity to prominence in part by posting YouTube videos that falsely claimed left-wing scholars were using "gay kits" to convert kids to homosexuality.

In the hothouse of the 2016 US election season, observers argued that YouTube's recommendations were funneling voters into ever-more-extreme content. Conspiracy thinkers and right-wing agitators uploaded false rumors about Hillary Clinton's imminent mental collapse and involvement in a nonexistent pizzeria pedophile ring, then watched, delightedly, as their videos lifted off in YouTube's Up Next column. A former Google engineer named Guillaume Chaslot coded a web-scraper program to see, among other things, whether YouTube's algorithm had a political tilt. He found that recommendations heavily favored Trump as

well as anti-Clinton material. The watch time system, in his view, was optimizing for whomever was most willing to tell fantastic lies.

As 2016 wore on and the billion-hour deadline loomed, the engineers went into overdrive. Recommendations had become the thrumming engine of YouTube, responsible for an astonishing 70 percent of all its watch time. In turn, YouTube became a key source of revenue in the Alphabet empire.

Goodrow hit the target: On October 22, 2016, a few weeks before the presidential election, users watched 1 billion hours of videos on YouTube.

AFTER THE 2016 election, the tech industry came in for a reckoning. Critics laced into Facebook's algorithm for boosting conspiratorial rants and hammered Twitter for letting in phalanxes of Russian bots. Scrutiny of YouTube emerged a bit later. In 2018 a UC Berkeley computer scientist named Hany Farid teamed up with Guillaume Chaslot to run his scraper again. This time, they ran the program daily for 15 months, looking specifically for how often YouTube recommended conspiracy videos. They found the frequency rose throughout the year; at the peak, nearly one in 10 videos recommended were conspiracist fare.

"It turns out that human nature is awful," Farid tells me, "and the algorithms have figured this out, and that's what drives engagement." As Micah Schaffer, who worked at YouTube from 2006 to 2009, told me, "It really is they are addicted to that traffic."

YouTube executives deny that the billion-hour push led to a banquet of conspiracies. "We don't see evidence that extreme content or misinformation is on average more engaging, or generates more viewership, than anything else," Goodrow said. (YouTube also challenged Farid and Chaslot's research, saying it "does not accurately reflect how YouTube's recommendations work or how people watch and interact with YouTube.") But, within YouTube, the principle of "Broadcast Yourself," without restriction, was colliding with concerns about safety and misinformation.

On October 1, 2017, when a man used an arsenal of weapons to fire into a crowd of people at a concert in Las Vegas, YouTube users immediately began uploading false-flag videos claiming the shooting was orchestrated to foment opposition to the Second Amendment.

Just 12 hours after the shooting, Geoff Samek arrived for his first day as a product manager at YouTube. For several days he and his team were run ragged trying to identify fabulist videos and delete them. He was, he told me, "surprised" by how little was in place to manage a crisis like this. (When I asked him what the experience felt like, he sent me a clip of Tim Robbins being screamed at as a new mailroom hire in *The Hudsucker Proxy*.) The recommendation system was apparently making things worse; as BuzzFeed reporters found, even three days after the shooting the system was still promoting videos like "PROOF: MEDIA & LAW ENFORCEMENT ARE LYING."

"I can say it was a challenging first day," Samek told me dryly. "Frankly, I don't think our site was performing super well for misinformation ... I think that kicked off a lot of things for us, and it was a turning point."

YouTube already had policies forbidding certain types of content, like pornography or speech encouraging violence. To hunt down and delete these videos, the company used AI "classifiers"—code that automatically detects potentially policy-violating videos by analyzing, among other signals, the headlines or the words spoken in a video (which YouTube generates using its automatic speech-to-text software). They also had human moderators who reviewed videos the AI flagged for deletion.

After the Las Vegas shooting, executives began focusing more on the challenge. Google's content moderators grew to 10,000, and YouTube created an "intelligence desk" of people who hunt for new trends in disinformation and other "inappropriate content." YouTube's definition of hate speech was expanded to include Alex Jones' claim that the murders at Sandy Hook Elementary School never occurred. The site had already created a "breaking-news shelf" that would run on the

homepage and showcase links to content from news sources that Google News had previously vetted. The goal, as Neal Mohan, YouTube's chief product officer, noted, was not just to delete the obviously bad stuff but to boost reliable, mainstream sources. Internally, they began to refer to this strategy as a set of R's: "remove" violating material and "raise up" quality stuff.

But what about content that wasn't *quite* bad enough to be deleted? Like alleged conspiracies or dubious information that doesn't advocate violence or promote "dangerous remedies or cures" or otherwise explicitly violate policies? Those videos wouldn't be removed by moderators or the content-blocking AI. And yet, some executives wondered if they were complicit by promoting them at all. "We noticed that some people were watching things that we weren't happy with them watching," says Johanna Wright, one of YouTube's vice presidents of product management, "like flat-earth videos." This was what executives began calling "borderline" content. "It's near the policy but not against our policies," as Wright said.

By early 2018, YouTube executives decided they wanted to tackle the borderline material too. It would require adding a third R to their strategy—"reduce." They'd need to engineer a new AI system that would recognize conspiracy content and misinformation and down-rank it.

IN FEBRUARY, I visited YouTube's headquarters in San Bruno, California. Goodrow had promised to show me the secret of that new AI.

It was the day after the Iowa caucuses, where a vote-counting app had failed miserably. The news cycle was spinning crazily, but inside YouTube the mood seemed calm. We filed into a conference room, and Goodrow plunked into a chair and opened his laptop. He has close-cropped hair and sported a normcore middle-aged-dad style, wearing a zip-up black sweater over beige khakis. A mathematician by training, Goodrow can be intense; he was a dogged advocate of the billion-hour project and neurotically checked view stats every single day. Last winter he mounted a brief and failed run in the Democratic primary for his San

Mateo County congressional district. Goodrow and I were joined by Andre Rohe, a dry-witted German who came to YouTube in 2015 to be head of Discovery engineering after three years heading Google News.

Rohe beckoned me to his screen. He and Goodrow seemed slightly nervous. The inner workings of any system at Google are closely guarded secrets. Engineers worry that if they reveal too much about how any algorithm works—particularly one designed to down-rank content—outsiders could learn to outwit it. For the first time, Rohe and Goodrow were preparing to reveal some details of the recommendation revamp to a reporter.

To create an AI classifier that can recognize borderline video content, you need to train the AI with many thousands of examples. To get those training videos, YouTube would have to ask hundreds of ordinary humans to decide what looks dodgy and then feed their evaluations and those videos to the AI, so it could learn to recognize what dodgy looks like. That raised a fundamental question: What is "borderline" content? It's one thing to ask random people to identify an image of a cat or a crosswalk—something a Trump supporter, a Black Lives Matter activist, and even a QAnon adherent could all agree on. But if they wanted their human evaluators to recognize something subtler—like whether a video on Freemasons is a study of the group's history or a fantasy about how they secretly run government today—they would need to provide guidance.

YouTube assembled a team to figure this out. Many of its members came from the policy department, which creates and continually updates the rules about the content YouTube bans outright. They developed a set of about three dozen questions designed to help a human decide whether content moved significantly in the direction of those banned areas, but didn't quite get there.

These questions were, in essence, the wireframe of the human judgment that would become the AI's smarts. These hidden inner workings were listed on Rohe's screen. They allowed me to take notes but wouldn't give me a copy to take away.

One question asks whether a video appears to "encourage harmful or risky behavior to others" or to viewers themselves. To help narrow down what type of content constitutes "harmful or risky behavior," there is a set of check boxes pointing out various well-known self-harms YouTube has grappled with—like "pro ana" videos that encourage anorexic behaviors, or graphic images of self-harm.

"If you start by just asking, 'Is this harmful misinformation?' then everybody has a different definition of what's harmful," Goodrow said. "But then you say, 'OK, let's try to move it more into the concrete, specific realm by saying, is it about self-harm? What kinds of harm is it?' Then you tend to get higher agreement and better results." There's also an open-ended box that an evaluator can write in to explain their thinking.

Another question asks the evaluators to determine whether a video is "intolerant of a group" based on race, religion, sexual orientation, gender, national origin, or veteran status. But there's a supplementary question: "Is the video satire?" YouTube's policies prohibit hate speech and spreading lies about ethnic groups, for example, but they can permit content that mocks that behavior by mimicking it.

Rohe pointed to another category, one that asks whether a video is "inaccurate, misleading, or deceptive." It then goes on to ask the evaluator to check all the possible categories of factual nonsense that might apply, like "unsubstantiated conspiracy theories," "demonstratively inaccurate information," "deceptive content," "urban legend," "fictional story or myth," or "contradicts well-established expert consensus." The evaluators each spend about 5 minutes assessing each video, on top of the time it takes to watch it, and are encouraged to do research to help understand its context.

Rohe and Goodrow said they had tried to reduce potential bias among the human evaluators by choosing people who were diverse in terms of age, geography, gender, and race. They also made sure each video was rated by up to nine separate evaluators so that the results were

subject to the "wisdom of a group," as Goodrow put it. Any videos with medical subjects were rated by a team of doctors, not laypeople.

This diversity among the evaluators' views can pose problems for training the AI, though. If evaluators are too divided over whether a video is deceptive or factually misleading, then their responses won't provide a clear signal. As Woojin Kim, a vice president of product management, pointed out, "If we're talking about a contentious political topic, where you do have multiple perspectives ... those would oftentimes end up being marked not as borderline content." When the AI classifier was trained on those examples, it absorbed the same divided mentality. If it encountered a new video with the same characteristics, it would, metaphorically, shrug and not classify it as borderline either.

The evaluators processed tens of thousands of videos, enough for YouTube engineers to begin training the system. The AI would take data from the human evaluations—that a video called "Moon Landing Hoax— Wires Footage" is an "unsubstantiated conspiracy theory," for example—and learn to associate it with features of that video: the text under the title that the creator uses to describe the video ("We can see the wires, people!"); the comments ("It's 2017 and people still believe in moon landings ... help ... help"); the transcript ("the astronaut is getting up with the wire taking the weight"); and, especially, the title. The visual content of the video itself, interestingly, often wasn't a very useful signal. As with videos about virtually any topic, misinformation is often conveyed by someone simply speaking to the camera or (as with Sargent's flat-earth material) over a procession of static images.

Another useful training feature for the AI was "co-watches," or the fare users typically watch before or after the video in question. In a sense, it was a measure of the company a video keeps. If National Geographic posts a video titled "Round Earth vs. Flat Earth," an AI might recognize it as having words very similar to a flat-earth video. But the co-watches would likely be an interview with the astrophysicist Neil deGrasse Tyson or a scientist's TED talk, while a flat-earth conspiracy video might pair with a rant on the CIA's UFO cover-up.

The AI classifier does not produce a binary answer; it doesn't say whether a video is or isn't "borderline." Instead, it generates a score, a mathematical weight that represents how likely the video is to approach the borderline. That weight is incorporated into the overall recommendation AI and becomes one of the many signals used when recommending the video to a particular user.

IN JANUARY 2019, YouTube began rolling out the system. That's when Mark Sargent noticed his flat-earth views take a nose dive. Other types of content were getting down-ranked, too, like moon-landing conspiracies or videos perseverating on chemtrails. Over the next few months, Goodrow and Rohe pushed out more than 30 refinements to the system that they say increased its accuracy. By the summer, YouTube was publicly declaring success: It had reduced by 50 percent the watch time of borderline content that came from recommendations. By December it reported a reduction of 70 percent.

The company won't release its internal data, so it's impossible to confirm the accuracy of its claims. But there are several outside indications that the system has had an effect. One is that consumers and creators of borderline stuff complain that their favorite material is rarely boosted any more. "Wow has anybody else noticed how hard it is to find 'Conspiracy Theory' stuff on YouTube lately? And that you easily find videos 'debunking' those instead?" one comment noted in February of this year. "Oh yes, youtubes algorithm is smashing it for them," another replied.

Then there's the academic research. Berkeley professor Hany Farid and his team found that the frequency with which YouTube recommended conspiracy videos began to fall significantly in early 2019, precisely when YouTube was beginning its updates. By early 2020, his analysis found, those recommendations had gone down from a 2018 peak by 40 percent. Farid noticed that some channels weren't merely reduced; they all but vanished from recommendations. Indeed, before YouTube made its switch, he'd found that 10 channels—including that of David Icke, the British writer who argues that reptilians walk among us—comprised 20 percent of all conspiracy recommendations (as Farid defines them);

afterward, he found that recommendations for those sites "basically went to zero."

Another study that somewhat backs up YouTube's claims was conducted by the computer scientist Mark Ledwich and Anna Zaitsev, a postdoctoral scholar and lecturer at Berkeley. They analyzed YouTube recommendations, looking specifically at 816 political channels and categorizing them into different ideological groups such as "Partisan Left," "Libertarian," and "White Identitarian." They found that YouTube recommendations mostly now guide viewers of political content to the mainstream. The channels they grouped under "Social Justice," on the far left, lost a third of their traffic to mainstream sources like CNN; conspiracy channels and most on the reactionary right—like "White Identitarian" and "Religious Conservative"—saw the majority of their traffic slough off to commercial right-wing channels, with Fox News being the hugest beneficiary.

If Zaitsev and Ledwich's analysis of YouTube "mainstreaming" traffic holds up—and it's certainly a direction that YouTube itself endorses—it would fit into a historic pattern. As law professor Tim Wu noted in his book _The Master Switch_, new media tend to start out in a Wild West, then clean up, put on a suit, and consolidate in a cautious center. Radio, for example, began as a chaos of small operators proud to say anything, then gradually coagulated into a small number of mammoth networks aimed mostly at pleasing the mainstream.

For critics like Farid, though, YouTube has not gone far enough, quickly enough. "Shame on YouTube," he told me. "It was only after how many years of this nonsense did they finally respond? After public pressure just got to be so much they couldn't deal with it."

Even the executives who set up the new "reduce" system told me it wasn't perfect. Which makes some critics wonder: Why not just shut down the recommendation system entirely? Micah Schaffer, the former YouTube employee, says, "At some point, if you can't do this responsibly, you need to not do it." As another former YouTube

employee noted, determined creators are adept at gaming any system YouTube puts up, like "the velociraptor and the fence."

Still, the system appeared to be working, mostly. It was a real, if modest, improvement. But then the floodgates opened again. As the winter of 2020 turned into a spring of pandemic, a summer of activism, and another norm-shattering election season, it looked as if the recommendation engine might be the least of YouTube's problems.

A MONTH AFTER I visited YouTube, the new coronavirus pandemic was in full swing. It had itself become a fertile field for new conspiracy theories. Videos claimed that 5G towers caused Covid-19; Mark Sargent had interrupted his flat-earth musings to upload a few videos in which he said the pandemic lockdown was an ominous preparation for social control. He told me the government would use a vaccine to inject everyone with an invisible mark, and "then it goes to the whole Christian mark of the beast," the prophesy from the Book of Revelations.

On March 30, I talked to Mohan again, but this time on Google Hangouts. He was ensconced in a wood-paneled room at his home, clad in a blue polo shirt, while the faint sounds of his children echoed from elsewhere in the house.

YouTube, he told me, had been moving aggressively to clamp down on disinformation about the pandemic and to counteract it. The platform created an "info panel" to run under any video mentioning Covid-19, linking to the Centers for Disease Control and other global and local health officials. By late August, these panels had received more than 300 billion impressions. YouTube had been removing videos with dangerous "medical" information every day, including those promoting "harmful cures," as Mohan says, and videos telling people to flout stay-at-home rules. To raise up useful information, the company arranged for several popular YouTubers to interview Anthony Fauci, the director of the National Institute of Allergy and Infectious Diseases who had become a regular presence on TV and a voice of scientific reason.

Mohan had also been meeting with YouTube's "intel desk," whose researchers had been trying to root out the latest Covid conspiracies. Goodrow and Rohe would use those videos to help update their AI classifier at least once a week, so it could help down-rank new strains of borderline Covid content.

But even as we spoke, YouTube videos with wild-eyed claims were being uploaded and amassing views. An American chiropractor named John Bergman got more than a million views for videos suggesting that hand sanitizer didn't work and urging people to use essential oils and vitamin C to treat the contagion. On April 16, a conspiracy channel named the Next News Network uploaded a video claiming that Fauci was a "criminal," that coronavirus was a false-flag operation to impose "mandatory vaccines," and that if anyone refused to be vaccinated, they'd be "shot in the head." It racked up nearly 7 million views in two weeks, before YouTube finally took it down. Then came ever more unhinged uploads, including the infamous "Plandemic" video—alleging a conspiracy to push a vaccine—or the so-called "white coat summit" of July 27, in which a group of doctors assembled in front of the Supreme Court to falsely claim that hydroxychloroquine could cure Covid and that masks were unnecessary.

YouTube was playing a by-now familiar game of social media whack-a-mole. A video that violated YouTube's rules would emerge and rapidly gain views, then YouTube would take it down. But it wasn't clear that recommendations were key to these sudden viral spikes. On August 14, a 90-minute video by Millie Weaver, a contributor to the far-right conspiracist site Infowars, went online, filled with claims of a deep state arrayed against President Trump. It was linked and shared in a number of right-wing circles. Dozens of Reddit threads passed it on ("Watch it before it's gone," one redditor wrote), and it was shared more than 53,000 times on Facebook, as well as on scores of right-wing YouTube channels, including by many followers of QAnon, one of the fastest-growing—and most dangerous—conspiracy theories in the nation. YouTube took it down a day later, saying it violated its hate-speech rules. But within that 24 hours, it amassed over a million views.

This old-fashioned spread—a mix of organic link-sharing and astroturfed, bot-propelled promotion—is powerful and, say observers, may sideline any changes to YouTube's recommendation system. It also suggests that users are adapting and that the recommendation system may be less important, for good and ill, to the spread of misinformation today. In a study for the think tank Data & Society, the researcher Becca Lewis mapped out the galaxy of right-wing commentators on YouTube who routinely spread borderline material. Many of those creators, she says, have built their often massive audiences not only through YouTube recommendations but also via networking. In their videos they'll give shout-outs to one another and hype each other's work, much as YouTubers all enthusiastically promoted Millie Weaver's fabricated musings.

"If YouTube completely took away the recommendations algorithm tomorrow, I don't think the extremist problem would be solved. Because they're just entrenched," Lewis tells me. "These people have these intense fandoms at this point. I don't know what the answer is."

One of the former Google engineers I spoke to agreed: "Now that society is so polarized, I'm not sure YouTube alone can do much," as the engineer noted. "People who have been radicalized over the past few years aren't getting unradicalized. The time to do this was years ago."

Let us know what you think about this article. Submit a letter to the editor at mail@wired.com.

Clive Thompson ([@pomeranian99](https://twitter.com/pomeranian99)) is a WIRED contributing editor and author of *Coders: The Making of a New Tribe and the Remaking of the World*.