



## Research Article

Lauren Valentino Bryant\*

# The YouTube Algorithm and the Alt-Right Filter Bubble

<https://doi.org/10.1515/opis-2020-0007>

Received October 31, 2019; accepted April 9, 2020

**Abstract:** The YouTube algorithm is a combination of programmed directives from engineers along with learned behaviors that have evolved through the opaque process of machine learning which makes the algorithm's directives and programming hard to understand. Independent tests to replicate the algorithm have shown that the algorithm has a strong bias towards right-leaning politics videos, including those racist views expressed by the alt-right community. While the algorithm seems to be pushing users towards the alt-right video content merely in an attempt to keep users in a cycle of video watching, the end result makes YouTube a powerful recruiting tool for Neo-Nazis and the alt-right. The filter bubble effect that this creates pushes users into a loop that reinforces radicalism instead of level-headed factual resources.

**Keywords:** YouTube, filter bubble, alt-right, Neo-Nazi, algorithm, racism, search engine

YouTube is a source of not only entertainment, but information and instruction with content that will teach users to do anything from style their hair to replace a car's battery. When asked about her company, YouTube, CEO Susan Wojcicki was eager to say, "we're really more like a library in many ways, because of the sheer amount of video that we have, and the ability for people to learn and to look up any kind of information, learn about it" (Thompson, 2018). While YouTube may be a convenient source of information, YouTube is not like a library in many ways. The taxonomy and organization of YouTube's topics and categorizations are broad and ill defined. Libraries, both physical and digital, are often curated by librarians to include balanced voices that give patrons an accurate view of political, philosophical, literary, and other arguments, while YouTube's content is not curated as much as it is lightly moderated. YouTube does employ content moderators to remove disturbing content such as "terrorism" and "child abuse" from its servers, but this content does so much harm to their employees that "Google is conducting experiments on some moderators to see whether technological interventions, such as allowing workers to watch videos in grayscale, reduce emotional harms" (Newton, 2019). The publishing process that books, journals, magazines, and other publications go through before making their way to a library's shelves or electronic holdings create varying degrees of checks and balances, while YouTube's content is contributed by anyone with an internet connection.

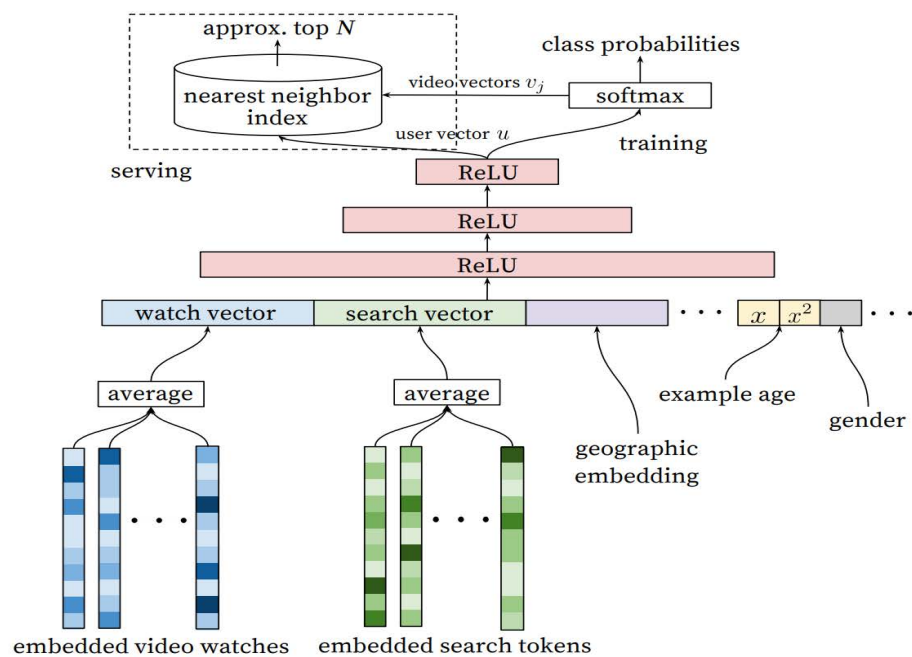
Arguably one of the last impartial spaces, libraries are unique because the information they offer does not come with a hidden consumer or political agenda. YouTube has done little to equip themselves with mission statements, values, and frameworks that would have established best practices for an information commons, a system that does not just offer information in various formats, but builds ways to confirm the validity of that information and preserve it. The business was built upon the goal of making money and not informing or educating. Without regulation, pockets of users with a racist or political agenda found that they could manipulate the algorithm and take control of the way content was presented through both Search Engine Optimization (SEO) and alignment with political structures. Safiya Noble, the author of *Algorithms of Oppression*, reminds her audience that, "Google Search is in fact an advertising platform,

---

\*Corresponding author, Lauren Valentino Bryant, Ray W. Howard Library, Shoreline Community College, Shoreline 98133 WA USA, E-mail: lbryant@shoreline.edu; Laurenvbryant@gmail.com

not intended to solely serve as a public information resource in the way that, say, a library might” (Noble 2019). Noble and her ground-breaking research has paved the way for a larger conversation from many angles which have brought inequities to light in burgeoning technologies. As a for-profit company, YouTube is not in a position to be an impartial third party, and indeed are acting in exactly the opposite of that role, pushing and prodding its viewers indiscriminately towards its advertisements. This manipulative system is not the root of the problem, but rather the directive that the algorithm had been told to aim for, resulting in an unintended consequence on its own. Youtube’s video recommendation system may be promoting racist viewpoints which is distorting the overall perception of content on YouTube as a whole, a dangerous misunderstanding since the platform has taken on the responsibility of providing not only amusement and entertainment for the masses, but informing and educating them as well.

The measure of success for the YouTube algorithm is convincing the user to watch an additional video after the end of the first video has finished. The default behavior of the YouTube player is to immediately play the suggested video, an issue in itself with consent. The algorithm improves through machine learning which means every time it has a successful interaction, and a user allows one of the suggested videos to be played, the algorithm learns that there is a relationship between the video watched and the video suggested. Exactly how the algorithm works is a bit of a black box, some of its internal logic is opaque even to its engineers. The algorithm’s learned behavior, a process that takes place without human intervention, is internal. Google did publish a white paper in 2016 that reveals the engineering intent behind the design. The formula incorporates every moment that a video is watched as a positive number while videos that were not clicked end up as negative numbers, failures adding to the negative watch time (Covington, Adams, & Sargin, 2016). A user’s demographics which we see in Figure 1 as “example age” and “gender” is mentioned as a factor along with “items in a user’s history” which is supposed to accurately determine what kind of videos the person might want to watch next (Covington, Adams, & Sargin, 2016). The other factor mentioned is “search query tokens” which would imply that the keywords a user types into a search box would follow them around and inspire future recommendations (Covington, Adams, & Sargin, 2016).



**Figure 1.** The YouTube algorithm is designed to use “example age” and other demographic information to anticipate what the user may want to watch next. Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations. RecSys ‘16 Proceedings of the 10th ACM Conference on Recommender Systems, pp. 191-198. <https://doi.org/10.1145/2959100.2959190>

The intent of the engineers was to maximize exposure of advertisements on YouTube which is problematic, but not as malicious as the actual outcome we see the algorithm performing. The Google engineers measure their success by “watch-time” or “click-through rate” which tells us that their goal is to make sure the user watches more videos which creates more opportunities for users to click on videos or advertisements (Covington, Adams, & Sargin, 2016). Since YouTube’s main source of revenue is through their advertisers, the most obvious goal is to encourage users to click on the advertisements.

There are two sides to the YouTube algorithm: one facet represents the design and intent of the algorithm by its engineers. The other facet is represented as an unknown factor because the algorithm is a learning neural network and has created some connections on its own through machine learning. Machine learning is distinct from a programmed response in that it is a behavior that the computer has improved upon on its own, often using pattern recognition to determine a better or faster way to attain the original directive which has been set by a human engineer. There are educated guesses supported by data that conclude that this second facet of the YouTube algorithm may be operating in some ways that were unintended by its creators. An independent test was done, “each query in a fresh search session with no personal account or watch history data informing the algorithm, except for geographical location and time of day, effectively demonstrating how YouTube’s recommendation operates in the absence of personalization” (O’Donovan, Warzel, McDonald, Clifton, & Woolf, 2019). After more than 140 such tests, the observers decided, “YouTube’s recommendation engine algorithm isn’t a partisan monster — it’s an engagement monster. [...] Its only governing ideology is to wrangle content — no matter how tenuously linked to your watch history — that it thinks might keep you glued to your screen for another few seconds” (O’Donovan et. al., 2019). Indeed, this is a popular theory that YouTube’s algorithm is not trying to teach or convince the user of a certain truth, but simply wants to convince them to continue to watch the videos on the site.

Zynep Tufekci, a Turkish techno-sociologist, has studied YouTube ever since she had an unexpected experience with it as she was researching the 2016 presidential election campaign. While doing research for an article, she watched several “Donald Trump rallies on YouTube” which led the recommendation engine to autoplay “white supremacist rants, Holocaust denials and other disturbing content” (Tufekci, 2018). Even though racist content was the trigger that caused Tufekci to dig deeper, she concluded that the algorithm’s overall intent was not racist in nature. “For all its lofty rhetoric, Google is an advertising broker, selling our attention to companies that will pay for it. The longer people stay on YouTube, the more money Google makes” (Tufekci, 2018). YouTube is a for-profit company driven by online advertisements and we know that the goal of the algorithm is to drive users to use the service as much as possible, optimizing the chance that the user will click on an ad, therefore generating revenue for the website.

YouTube’s recommendation system makes complex, goal-based decisions, using a set of independently operating computer programs that mimic the human brain, often called a neural network. YouTube uses a neural network learning algorithm that perpetuates content to users. This algorithm may have found an unexpected relationship between racism and the right amount of curiosity that prompts a person to continue to watch YouTube videos. Two academic researchers made a visualized data map of 13,529 YouTube channels, starting with the top most popular channels from opposite political perspectives, recreating the YouTube algorithm in an attempt to figure out what was happening when it was recommending increasingly extreme political content (Kaiser & Rauchfleisch, 2018). They found tightly bound relationships between the right-leaning content on YouTube, specifically “Fox News,” “Alex Jones,” and “white nationalists,” along with some conspiracy theories, anti-feminist, anti-political correctness channels, and a channel called the Manosphere (Kaiser & Rauchfleisch, 2018). The connection between these channels was more tightly knit and closer together on the right-leaning side than it was on the left-leaning one. The authors found this, “highly problematic” because a user who pursues even mildly conservative content is “only one or two clicks away from extreme far-right channels, conspiracy theories, and radicalizing content” (Kaiser & Rauchfleisch, 2018). There was a multitude of other data included in the visualization data, including non-political channels such as video games, guns, music, and tech, which are individually popular on their own but not interconnected in the way the right-wing communities within YouTube are connected (Kaiser & Rauchfleisch, 2018). Since the algorithm has made this unlikely connection, it has a bias of recommending racist or white supremacist videos more often to users. The surprising outcome of this machine learning is

that the algorithm is showing an unexplained bias to suggest alt-right content to users, even promoting it to an inflated presence on the website without prior prompting of this preference.

Some employees within YouTube's company "raised concerns about the mass of false, incendiary and toxic content that the world's largest video site surfaced and spread" (Bergen, 2019). A user-run video database without moderation is dangerous on its own, but that is not exactly what was happening here; the algorithm was interfering with peoples' preferences and seemed to be pushing racist and alt-right propaganda to the surface. An employee at YouTube proposed a new "vertical" in 2018, "a category that the company uses to group its mountain of video footage," suggesting that this section of videos should be dedicated to the alt-right (Bergen, 2019). "Based on engagement, the hypothetical alt-right category sat with music, sports and gaming as the most popular channels at YouTube, an attempt to show how critical these videos were to YouTube's business" (Bergen, 2019). While this one employees' efforts may not reflect the values of YouTube's company as a whole, they were not the first one to notice a trend in the massive amount of white supremacist videos on the video sharing site. The Southern Poverty Law Center had articles as early as 2007 that mentioned how compared to dropping pamphlets on lawns, "posting video footage [on video sharing sites] is vastly less difficult, expensive, risky and time-consuming—and it can be done anonymously with virtually no effort" (Mock, 2007). YouTube has always had hate speech policies, but recently updated these policies of June 2019 to specifically target white nationalists by condemning behavior that might, "[c] all for the subjugation or domination over individuals or groups" or "[d]eny that a well-documented, violent event took place" such as the Holocaust which is a conspiracy theory that is popular with many members of the alt-right ("Hate speech policy", 2019). There was evidence that YouTube knew about the problematic content in 2017 because Susan Wojcicki mentioned in a 2018 Wired interview that "we started working on making sure we were removing violent extremist content. And what we started doing for the first time last year was using machines to find this content. So we built a classifier to identify the content and lo and behold, the machines found a lot more content than we had found" (Thompson, 2018). That Wojcicki both admits that the extremist content is problematic and claims that there was a lot more of it than the human searchers were able to find means that YouTube does not seem to have some secret alt-right agenda, but is just preoccupied and does not consider it a top concern. During that same Wired interview, Wojcicki was asked about morality and responsibility, but she said the company is not sure where they stand with moral concerns, adding an answer in the form of an analogy: "we don't want it to be necessarily us saying, we don't think people should be eating donuts, right. It's not our place to be doing that" (Thompson, 2018).

Online hate groups have unified on the internet and even though they span across many spaces such as Reddit, 8chan, 4chan, Twitter, Discord channels, a large majority of them claim that the content on YouTube with its constant stream of recommendations contributed to their recruitment. An internet investigative journalist on Bellingcat, an organization of independent online investigators, tracked down and interviewed 75 white supremacist fascists to find out about each person's "red-pilling," a term to explain "converting someone to fascist, racist and anti-Semitic beliefs" (Evans, 2018). When spaces are created on the internet for hate groups, the concern has been raised that the groups will echo hateful messages back to each other, preventing the influence from the outside world to create a more realistic perspective for these individuals. The term "filter bubble" was coined by the internet activist Eli Pariser to describe a phenomenon in which search algorithms can contribute to surrounding a user with their own viewpoints, sending their own search terms back at them in the form of results, ensuring that they rarely come into contact with an opposing source. In Pariser's book, *The Filter Bubble: How the New Personalized Web is Changing What We Read and How We Think*, he points out one of the main problems with the filter bubble: "But the filter bubble isn't tuned for a diversity of ideas or of people. It's not designed to introduce us to new cultures. As a result, living inside it, we may miss some of the mental flexibility and openness that contact with difference creates" (Pariser, 2011). Unfortunately one of the problems with internet subcultures is that they create an artificial space that guarantees the absence of diversity and conflict, even if the belief system is illogical or harmful. A dramatic contrast to the filter bubble may be the ideal library space, where each individual can encounter intelligent, well-worded perspectives that are opposed to one's own.

The Bellingcat journalist published findings that are similar to Pariser's "filter bubble" theory, and mentioned that "Infowars reached the height of its influence as a result of sites like Facebook and YouTube.

By the time they were kicked off YouTube, Infowars had more than 2.4 million followers and 1.6 billion page views across 36,000 videos” (Evans, 2018). Like YouTube, Facebook uses a unique algorithm that serves up content to the user in their news feed that is predictably similar to their own searches and browsing clicks. The two online environments provide spaces where they could continue to explore and browse for hours and not encounter an opposing perspective. Similar to the findings of the researchers who visualized the data points of YouTube’s political channels, the Bellingcat investigator gives an example of one person who identified as “‘moderate republication’ before ‘Steven Crowder, Paul Joseph Watson, Milo Yiannopolos, Black Pidgeon Speaks,’ and other far-right YouTubers slowly red-pilled him. Over time he ‘moved further and further right until he could no longer stand them. That’s why he likes those groups even still, because if we just had the Fascists, we’d never convert anyone’” (Evans, 2018). American politics plays into this situation a great deal as we have seen with both the Harvard researcher’s data and Bellingcat. The racist viewpoints are connected inextricably to the topics of video games, feminism, LGBTQ, and a multitude of other political talking points.

With President Donald Trump supporting groups like the alt-right, many hate groups have grown emboldened and more active during Trump’s presidency. Groups online that may have been niche, hidden, and remote have found footholds in lax hate speech policies such as Twitter’s and until recently, YouTube’s. These groups have created echo chambers where it is difficult to hear anything outside their own voices, espousing their hate speech in the anonymous, risk-free, public forum that the internet has provided. In a grouping of pure data taken from YouTube, a pair of researchers claimed, “we were able to identify a YouTube-created right-wing filter bubble. [...]YouTube’s algorithm connects them visibly via recommendations. It is, in this sense, an algorithmic version of the Thomas theorem, which famously suggested that, ‘If men define situations as real, they are real in their consequences’” (Kaiser & Rauchfleisch, 2018). If a community such as the alt-right does have influence and control over a large chunk of YouTube’s content, does this reflect the views and beliefs of the people in the United States? After all, YouTube is as ubiquitous as the internet itself, the YouTube app coming standard on nearly every cell phone, tablet, and web enabled device, the videos available in the results screen of any Google search.

The Bellingcat investigator shares some pop culture references that are used by the alt-right groups online and then reminds readers that, “it is important to remember that these groups have a body count and represent a real threat. Their absurdity does not negate their danger” (Evans, 2018). Many of the mass shooting attacks worldwide have been traced back to a small, thriving online community, the 8chan website, where users reinforce each others’ racist beliefs and claim violence as the only remedy. “The El Paso shooting follows a pattern carried out in Christchurch, New Zealand in March and in Poway, California in April. In both attacks, the suspects published manifestos to 8chan. Both manifestos were saturated with white nationalist talking points, portraying whites as the victims of a plan for elimination” (Hayden, 2019). 8chan, the anonymous message board site that is proud to be moderation free and allow their users completely free speech, even if those things are illegal. Because of this lenient policy, 8chan has attracted violent extremists from many groups that use the platform to organize and encourage each other’s malicious crimes. After the El-Paso shooting of a Walmart where 26 were injured and 22 people died, authorities reported that they were, “working to confirm the authenticity of, and any links between, the suspect and a manifesto published to the fringe internet platform 8chan in advance of the attack. The apparent manifesto refers to the ‘Hispanic invasion of Texas’” (Hayden, 2019). When the news of the El Paso shooting was released, one 8chan user wrote simply, “ACCELERATE, ACCELERATE,” while another joked, “Clean up in aisle 4!” (Hayden, 2019). The comment about cleaning up, can only be interpreted as the user’s attempt to dehumanize the victim and compare them to a broken or spilled item. The coaxing comments seen on 8chan could be exactly what a hesitant individual needs to enact real-life violence.

In 2017 an independent survey found that only 9% of Americans found it acceptable to hold alt-right or Neo-Nazi views, while 50% of Americans found these views unacceptable (Langer, 2017). The surprising truth is that there appears to be small, vocal right-leaning groups online that produce a great deal of online content and much of that content is for YouTube. However, the majority of Americans are not supportive of alt-right, racist ideologies. There is danger in allowing our media to define a political spectrum that is not representative of the actual country. It allows the views of a vocal minority to be represented as a majority

which is a danger when so many people find the appearance of consensus so persuasive. political situations for the rest of the country that are not supported by the majority of the citizens because perceptions, even illusions, can be deceiving. If our main information source is viewed on a skewed angle from an alt-right perspective, it can lend malicious influence to the general population. If the alt-right continues to be allowed to dominate our major media platform that 9% can only grow and the country may come to resemble the content on Youtube if Youtube isn't changed to better reflect its viewers.

The relationship of racist content equating to increased ad clicks to the algorithm is one that may do harm, not only to those users that are exposed to YouTube, but to future applications of the algorithm. The YouTube algorithm itself was not programmed with the intent to cause racial bias in the video it recommends. It is likely that the algorithm's design will be used to produce future technologies, which is why I hope the company of YouTube, Google, and the parent company Alphabet consider making the current algorithm more transparent to those that might study it, so that it can be improved upon. YouTube is beginning to consider aspects of itself that libraries have long figured out: collection development policies, standards of ethics, freedom of information and its boundaries, along with cataloging and categorization. Instead of making generalizations and vague comparisons of YouTube to a library, the leaders at YouTube need to take responsibility for the fact that their media has become an influential factor not just on the way the world entertains itself, but the way we educate and inform ourselves as well.

## References

- Bergen, M. (2019, April 2). YouTube executives ignored warnings, letting toxic videos run rampant. *Bloomberg*. Retrieved from <https://www.bloomberg.com/news/features/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant>
- Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations. *RecSys '16 Proceedings of the 10th ACM Conference on Recommender Systems*, pp. 191-198. <https://doi.org/10.1145/2959100.2959190>
- Evans, R. (2018, October 11). From memes to Infowars: How 75 fascist activists were "red-pilled" [Blog post]. Retrieved from Bellingcat website: <https://www.bellingcat.com/news/americas/2018/10/11/memes-infowars-75-fascist-activists-red-pilled/>
- Hate speech policy. (2019, June 5). Retrieved October 28, 2019, from YouTube Help website: <https://support.google.com/youtube/answer/2801939?hl=en>
- Hayden, M. E. (2019, August 04). White nationalists praise El Paso attack and mock the dead. *Southern Poverty Law Center*. Retrieved from: <https://www.splcenter.org/hatewatch/2019/08/04/white-nationalists-praise-el-paso-attack-and-mock-dead>
- Kaiser, J., & Rauchfleisch, A. (2018, April 11). Unite the right? How YouTube's recommendation algorithm connects the U.S. far-right. Retrieved October 28, 2019, from Medium website: <https://medium.com/@MediaManipulation/unite-the-right-how-youtubes-recommendation-algorithm-connects-the-u-s-far-right-9f1387ccfabd>
- Langer, G. (2017, August 21). Trump approval is low but steady; On Charlottesville lower still. *ABC News/Washington Post*. Retrieved from <https://www.langerresearch.com/wp-content/uploads/1190a1TrumpandCharlottesville.pdf>
- Mock, B. (2007, April 20). Neo-Nazi groups share hate via YouTube. *Southern Poverty Law Center*. Retrieved from <https://www.splcenter.org/fighting-hate/intelligence-report/2007/neo-nazi-groups-share-hate-youtube>
- Newton, Casey. (2019, December 16). The terror queue: These moderators help keep Google and YouTube free of violent extremism—and now some of them have PTSD. *The Verge*. Retrieved from <https://www.theverge.com/2019/12/16/21021005/google-youtube-moderators-ptsd-accenture-violent-disturbing-content-interviews-video>
- O'Donovan, C., Warzel, C., McDonald, L., Clifton, B., & Woolf, M. (2019, January 24). We followed YouTube's recommendation algorithm down the rabbit hole. Retrieved October 28, 2019, from BuzzFeed News website: <https://www.buzzfeednews.com/article/carolineodonovan/down-youtubes-recommendation-rabbit-hole>
- Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Retrieved from <https://books.google.com/books?id=wcalrOI1YbQC&pg=PP1&pg=PP1#v=onepage&q&f=false>
- Thompson, N. (2018, March 15). Susan Wojcicki on YouTube's fight against misinformation. *Wired*. Retrieved from <https://www.wired.com/story/susan-wojcicki-on-youtubes-fight-against-misinformation/>
- Tufekci, Z. (2018, March 10). YouTube, the great radicalizer. *The New York Times*, Opinion. Retrieved from <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>