

The Guardian

'Fiction is outperforming reality': how YouTube's algorithm distorts truth

An ex-YouTube insider reveals how its recommendation algorithm promotes divisive clips and conspiracy videos. Did they harm Hillary Clinton's bid for the presidency?

The methodology behind this story
by Paul Lewis in San Francisco

Main image: Guillaume Chaslot, an ex-Google software engineer. Photograph: Talia Herman for the Guardian

Fri 2 Feb 2018 12.00 GMT

It was one of January's most viral videos. Logan Paul, a YouTube celebrity, stumbles across a dead man hanging from a tree. The 22-year-old, who is in a Japanese forest famous as a suicide spot, is visibly shocked, then amused. "Dude, his hands are purple," he says, before turning to his friends and giggling. "You never stand next to a dead guy?"

Paul, who has 16 million mostly teen subscribers to his YouTube channel, removed the video from YouTube 24 hours later amid a furious backlash. It was still long enough for the footage to receive 6m views and a spot on YouTube's coveted list of trending videos.

The next day, I watched a copy of the video on YouTube. Then I clicked on the "Up next" thumbnails of recommended videos that YouTube showcases on the right-hand side of the video player. This conveyor belt of clips, which auto-play by default, are designed to seduce us to spend more time on Google's video broadcasting platform. I was curious where they might lead.

The answer was a slew of videos of men mocking distraught teenage fans of Logan Paul, followed by CCTV footage of children stealing things and, a few clicks later, a video of children having their teeth pulled out with bizarre, homemade contraptions.

I had cleared my history, deleted my cookies, and opened a private browser to be sure YouTube was not personalising recommendations. This was the algorithm taking me on a journey of its own volition, and it culminated with a video of two boys, aged about five or six, punching and kicking one another.

"I'm going to post it on YouTube," said a teenage girl, who sounded like she might be an older sibling. "Turn around and punch the heck out of that little boy." They scuffled for several minutes until one had knocked the other's tooth out.

There are 1.5 billion YouTube users in the world, which is more than the number of households that own televisions. What they watch is shaped by this algorithm, which skims and ranks billions

of videos to identify 20 “up next” clips that are both relevant to a previous video and most likely, statistically speaking, to keep a person hooked on their screen.

Company insiders tell me the algorithm is the single most important engine of YouTube’s growth. In one of the few public explanations of how the formula works – an academic paper that sketches the algorithm’s deep neural networks, crunching a vast pool of data about videos and the people who watch them – YouTube engineers describe it as one of the “largest scale and most sophisticated industrial recommendation systems in existence”.

Lately, it has also become one of the most controversial. The algorithm has been found to be promoting conspiracy theories about the Las Vegas mass shooting and incentivising, through recommendations, a thriving subculture that targets children with disturbing content such as cartoons in which the British children’s character Peppa Pig eats her father or drinks bleach.

Lewd and violent videos have been algorithmically served up to toddlers watching YouTube Kids, a dedicated app for children. One YouTube creator who was banned from making advertising revenues from his strange videos – which featured his children receiving flu shots, removing earwax, and crying over dead pets – told a reporter he had only been responding to the demands of Google’s algorithm. “That’s what got us out there and popular,” he said. “We learned to fuel it and do whatever it took to please the algorithm.”

Google has responded to these controversies in a process akin to Whac-A-Mole: expanding the army of human moderators, removing offensive YouTube videos identified by journalists and demonetising the channels that create them. But none of those moves has diminished a growing concern that something has gone profoundly awry with the artificial intelligence powering YouTube.

Yet one stone has so far been largely unturned. Much has been written about Facebook and Twitter’s impact on politics, but in recent months academics have speculated that YouTube’s algorithms may have been instrumental in fuelling disinformation during the 2016 presidential election. “YouTube is the most overlooked story of 2016,” Zeynep Tufekci, a widely respected sociologist and technology critic, tweeted back in October. “Its search and recommender algorithms are misinformation engines.”

If YouTube’s recommendation algorithm really has evolved to promote more disturbing content, how did that happen? And what is it doing to our politics?

‘Like reality, but distorted’

Those are not easy questions to answer. Like all big tech companies, YouTube does not allow us to see the algorithms that shape our lives. They are secret formulas, proprietary software, and only select engineers are entrusted to work on the algorithm. Guillaume Chaslot, a 36-year-old French computer programmer with a PhD in artificial intelligence, was one of those engineers.

During the three years he worked at Google, he was placed for several months with a team of YouTube engineers working on the recommendation system. The experience led him to conclude that the priorities YouTube gives its algorithms are dangerously skewed.

“YouTube is something that looks like reality, but it is distorted to make you spend more time online,” he tells me when we meet in Berkeley, California. “The recommendation algorithm is not optimising for what is truthful, or balanced, or healthy for democracy.”

Chaslot explains that the algorithm never stays the same. It is constantly changing the weight it gives to different signals: the viewing patterns of a user, for example, or the length of time a video is watched before someone clicks away.

The engineers he worked with were responsible for continuously experimenting with new formulas that would increase advertising revenues by extending the amount of time people watched videos. "Watch time was the priority," he recalls. "Everything else was considered a distraction."

Chaslot was fired by Google in 2013, ostensibly over performance issues. He insists he was let go after agitating for change within the company, using his personal time to team up with like-minded engineers to propose changes that could diversify the content people see.

He was especially worried about the distortions that might result from a simplistic focus on showing people videos they found irresistible, creating filter bubbles, for example, that only show people content that reinforces their existing view of the world. Chaslot said none of his proposed fixes were taken up by his managers. "There are many ways YouTube can change its algorithms to suppress fake news and improve the quality and diversity of videos people see," he says. "I tried to change YouTube from the inside but it didn't work."

YouTube told me that its recommendation system had evolved since Chaslot worked at the company and now "goes beyond optimising for watchtime". The company said that in 2016 it started taking into account user "satisfaction", by using surveys, for example, or looking at how many "likes" a video received, to "ensure people were satisfied with what they were viewing". YouTube added that additional changes had been implemented in 2017 to improve the news content surfaced in searches and recommendations and discourage the promotion of videos containing "inflammatory religious or supremacist" content.

It did not say why Google, which acquired YouTube in 2006, waited over a decade to make those changes. Chaslot believes such changes are mostly cosmetic, and have failed to fundamentally alter some disturbing biases that have evolved in the algorithm. In the summer of 2016, he built a computer program to investigate.

The software Chaslot wrote was designed to provide the world's first window into YouTube's opaque recommendation engine. The program simulates the behaviour of a user who starts on one video and then follows the chain of recommended videos - much as I did after watching the Logan Paul video - tracking data along the way.

It finds videos through a word search, selecting a "seed" video to begin with, and recording several layers of videos that YouTube recommends in the "up next" column. It does so with no viewing history, ensuring the videos being detected are YouTube's generic recommendations, rather than videos personalised to a user. And it repeats the process thousands of times, accumulating layers of data about YouTube recommendations to build up a picture of the algorithm's preferences.

Over the last 18 months, Chaslot has used the program to explore bias in YouTube content promoted during the French, British and German elections, global warming and mass shootings, and published his findings on his website, Algotransparency.org. Each study finds something different, but the research suggests YouTube systematically amplifies videos that are divisive, sensational and conspiratorial.

When his program found a seed video by searching the query “who is Michelle Obama?” and then followed the chain of “up next” suggestions, for example, most of the recommended videos said she “is a man”. More than 80% of the YouTube-recommended videos about the pope detected by his program described the Catholic leader as “evil”, “satanic”, or “the anti-Christ”. There were literally millions of videos uploaded to YouTube to satiate the algorithm’s appetite for content claiming the earth is flat. “On YouTube, fiction is outperforming reality,” Chaslot says.



A voter in Ohio. Trump won the election by just 80,000 votes spread across three swing states. Photograph: Ty Wright/Getty Images

He believes one of the most shocking examples was detected by his program in the run-up to the 2016 presidential election. As he observed in a short, largely unnoticed blogpost published after Donald Trump was elected, the impact of YouTube’s recommendation algorithm was not neutral during the presidential race: it was pushing videos that were, in the main, helpful to Trump and damaging to Hillary Clinton. “It was strange,” he explains to me. “Wherever you started, whether it was from a Trump search or a Clinton search, the recommendation algorithm was much more likely to push you in a pro-Trump direction.”

Trump won the electoral college as a result of 80,000 votes spread across three swing states. There were more than 150 million YouTube users in the US. The videos contained in Chaslot’s database of YouTube-recommended election videos were watched, in total, more than 3bn times before the vote in November 2016.

Even a small bias in the videos would have been significant. “Algorithms that shape the content we see can have a lot of impact, particularly on people who have not made up their mind,” says Luciano Floridi, a professor at the University of Oxford’s Digital Ethics Lab, who studies the ethics of artificial intelligence. “Gentle, implicit, quiet nudging can over time edge us toward choices we might not have otherwise made.”

Promoting conspiracy theories

Chaslot sent me a database of more YouTube-recommended videos his program identified in the three months leading up to the presidential election. It contained more than 8,000 videos – all of them detected by his program appearing “up next” on 12 dates between August and November 2016, after equal numbers of searches for “Trump” and “Clinton”.

It was not a comprehensive set of videos and it may not have been a perfectly representative sample. But it was, Chaslot said, a previously unseen dataset of what YouTube was

recommending to people interested in content about the candidates - one snapshot, in other words, of the algorithm's preferences.

Jonathan Albright, research director at the Tow Center for Digital Journalism, who reviewed the code used by Chaslot, says it is a relatively straightforward piece of software and a reputable methodology. "This research captured the apparent direction of YouTube's political ecosystem," he says. "That has not been done before."

I spent weeks watching, sorting and categorising the trove of videos with Erin McCormick, an investigative reporter and expert in database analysis. From the start, we were stunned by how many extreme and conspiratorial videos had been recommended, and the fact that almost all of them appeared to be directed against Clinton.

Some of the videos YouTube was recommending were the sort we had expected to see: broadcasts of presidential debates, TV news clips, Saturday Night Live sketches. There were also videos of speeches by the two candidates - although, we found, the database contained far more YouTube-recommended speeches by Trump than Clinton.

But what was most compelling was how often Chaslot's software detected anti-Clinton conspiracy videos appearing "up next" beside other videos.

There were dozens of clips stating Clinton had had a mental breakdown, reporting she had syphilis or Parkinson's disease, accusing her of having secret sexual relationships, including with Yoko Ono. Many were even darker, fabricating the contents of WikiLeaks disclosures to make unfounded claims, accusing Clinton of involvement in murders or connecting her to satanic and paedophilic cults.

One video that Chaslot's data indicated was pushed particularly hard by YouTube's algorithm was a bizarre, one-hour film claiming Trump's rise was predicted in Isaiah 45. Another was entitled: "BREAKING: VIDEO SHOWING BILL CLINTON RAPING 13 YR-OLD WILL PLUNGE RACE INTO CHAOS ANONYMOUS CLAIMS". The recommendation engine appeared to have been particularly helpful to the Alex Jones Channel, which broadcasts far-right conspiracy theories under the Infowars brand.



The conspiracy theorist and talkshow host Alex Jones. Photograph:
Brooks Kraft/Getty Images

There were too many videos in the database for us to watch them all, so we focused on 1,000 of the top-recommended videos. We sifted through them one by one to determine whether the content was likely to have benefited Trump or Clinton. Just over a third of the videos were either unrelated to the election or contained content that was broadly neutral or even-handed. Of the remaining 643 videos, 551 were videos favouring Trump, while only 92 favoured the Clinton campaign.

The sample we had looked at suggested Chaslot's conclusion was correct: YouTube was six times more likely to recommend videos that aided Trump than his adversary. YouTube presumably never programmed its algorithm to benefit one candidate over another. But based on this evidence, at least, that is exactly what happened.

'Leading people down hateful rabbit holes'

"We have a great deal of respect for the Guardian as a news outlet and institution," a YouTube spokesperson emailed me after I forwarded them our findings. "We strongly disagree, however, with the methodology, data and, most importantly, the conclusions made in their research."

The spokesperson added: "Our search and recommendation systems reflect what people search for, the number of videos available, and the videos people choose to watch on YouTube. That's not a bias towards any particular candidate; that is a reflection of viewer interest."

It was a curious response. YouTube seemed to be saying that its algorithm was a neutral mirror of the desires of the people who use it - if we don't like what it does, we have ourselves to blame. How does YouTube interpret "viewer interest" - and aren't "the videos people choose to watch" influenced by what the company shows them?

Offered the choice, we may instinctively click on a video of a dead man in a Japanese forest, or a fake news clip claiming Bill Clinton raped a 13-year-old. But are those in-the-moment impulses really a reflect of the content we want to be fed?

Tufekci, the sociologist who several months ago warned about the impact YouTube may have had on the election, tells me YouTube's recommendation system has probably figured out that edgy and hateful content is engaging. "This is a bit like an autopilot cafeteria in a school that has figured out children have sweet teeth, and also like fatty and salty foods," she says. "So you make a line offering such food, automatically loading the next plate as soon as the bag of chips or candy in front of the young person has been consumed."

Once that gets normalised, however, what is fractionally more edgy or bizarre becomes, Tufekci says, novel and interesting. "So the food gets higher and higher in sugar, fat and salt - natural human cravings - while the videos recommended and auto-played by YouTube get more and more bizarre or hateful."

But why would a bias toward ever more weird or divisive videos benefit one candidate over another? That depends on the candidates. Trump's campaign was nothing if not weird and divisive. Tufekci points to studies showing that "field of misinformation" largely tilted anti-Clinton before the election. "Fake news providers," she says, "found that fake anti-Clinton material played much better with the pro-Trump base than did fake anti-Trump material with the pro-Clinton base."

She adds: “The question before us is the ethics of leading people down hateful rabbit holes full of misinformation and lies at scale just because it works to increase the time people spend on the site - and it does work.”

Tufekci was one of several academics I shared our research with. Philip Howard, a professor at the Oxford Internet Institute, who has studied how disinformation spread during the election, was another. He questions whether a further factor might have been at play. “This is important research because it seems to be the first systematic look into how YouTube may have been manipulated,” he says, raising the possibility that the algorithm was gamed as part of the same propaganda campaigns that flourished on Twitter and Facebook.

In testimony to the House intelligence committee, investigating Russian interference in the election, Google’s general counsel, Kent Walker, played down the degree to which Moscow’s propaganda efforts infiltrated YouTube. The company’s internal investigation had only identified 18 YouTube channels and 1,100 videos suspected of being linked to Russia’s disinformation campaign, he told the committee in December - and generally the videos had relatively small numbers of views. He added: “We believe that the activity we found was limited because of various safeguards that we had in place in advance of the 2016 election, and the fact that Google’s products didn’t lend themselves to the kind of micro-targeting or viral dissemination that these actors seemed to prefer.”



General counsels for Twitter, Facebook and Google prepare to testify before the House intelligence committee hearing on Russia’s use of social media to influence the election. Photograph: Shawn Thew/EPA

Walker made no mention of YouTube recommendations. Correspondence made public just last week, however, reveals that Senator Mark Warner, the ranking Democrat on the intelligence committee, later wrote to the company about the algorithm, which he said seemed “particularly susceptible to foreign influence”. The senator demanded to know what the company was specifically doing to prevent a “malign incursion” of YouTube’s recommendation system. Walker, in his written reply, offered few specifics, but said YouTube had “a sophisticated spam and security breach detection system to identify anomalous behavior and malignant incursions”.

Tristan Harris, a former Google insider turned tech whistleblower, likes to describe Facebook as a “living, breathing crime scene for what happened in the 2016 election” that federal investigators have no access to. The same might be said of YouTube. About half the videos Chaslot’s program detected being recommended during the election have now vanished from YouTube - many of them taken down by their creators. Chaslot has always thought this suspicious. These were

videos with titles such as “Must Watch!! Hillary Clinton tried to ban this video”, watched millions of times before they disappeared. “Why would someone take down a video that has been viewed millions of times?” he asks.

I located a copy of “This Video Will Get Donald Trump Elected”, a viral sensation that was watched more than 10m times before it vanished from YouTube. It was a benign-seeming montage of historical footage of Trump, accompanied by soft piano music. But when I played the video in slow motion, I saw that it contained weird flashes of Miley Cyrus licking a mirror. It seemed an amateurish and bizarre attempt at inserting subliminal, sexualised imagery. But it underscored how little oversight we have over anyone who might want to use YouTube to influence public opinion on a vast scale.

I shared the entire database of 8,000 YouTube-recommended videos with John Kelly, the chief executive of the commercial analytics firm Graphika, which has been tracking political disinformation campaigns. He ran the list against his own database of Twitter accounts active during the election, and concluded many of the videos appeared to have been pushed by networks of Twitter sock puppets and bots controlled by pro-Trump digital consultants with “a presumably unsolicited assist” from Russia.

“I don’t have smoking-gun proof of who logged in to control those accounts,” he says. “But judging from the history of what we’ve seen those accounts doing before, and the characteristics of how they tweet and interconnect, they are assembled and controlled by someone - someone whose job was to elect Trump.”

Chaslot and some of the academics I spoke to felt this social media activity was significant. YouTube’s algorithm may have developed its biases organically, but could it also have been nudged into spreading those videos even further? “If a video starts skyrocketing, there’s no question YouTube’s algorithm is going to start pushing it,” Albright says.

YouTube did not deny that social media propaganda might have influenced its recommendations, but played down the likelihood, stressing its system “does not optimise” for traffic from Twitter or Facebook. “It appears as if the Guardian is attempting to shoehorn research, data and their conclusions into a common narrative about the role of technology in last year’s election,” the spokesperson added. “The reality of how our systems work, however, simply don’t support this premise.”

After the Senate’s correspondence with Google over possible Russian interference with YouTube’s recommendation algorithm was made public last week, YouTube sent me a new statement. It emphasised changes it made in 2017 to discourage the recommendation system from promoting some types of problematic content. “We appreciate the Guardian’s work to shine a spotlight on this challenging issue,” it added. “We know there is more to do here and we’re looking forward to making more announcements in the months ahead.”

Content creators



A video by the Next News Network fed on debunked allegations against Bill Clinton. Photograph: YouTube

With its flashy graphics and slick-haired anchor, the Next News Network has the appearances of a credible news channel. But behind the facade is a dubious operation that recycles stories harvested from far-right publications, fake news sites and Russian media outlets.

The channel is run by anchor Gary Franchi, once a leading proponent of a conspiracy that claimed the US government was creating concentration camps for its citizens. It was the Next News Network that broadcast the fabricated claims about Bill Clinton raping a teenager, although Franchi insists he is not a fake news producer. (He tells me he prefers to see his channel as “commentating on conservative news and opinion”.)

In the months leading up to the election, the Next News Network turned into a factory of anti-Clinton news and opinion, producing dozens of videos a day and reaching an audience comparable to that of MSNBC’s YouTube channel.

Chaslot’s research indicated Franchi’s success could largely be credited to YouTube’s algorithms, which consistently amplified his videos to be played “up next”. YouTube had sharply dismissed Chaslot’s research.

I contacted Franchi to see who was right. He sent me screen grabs of the private data given to people who upload YouTube videos, including a breakdown of how their audiences found their clips. The largest source of traffic to the Bill Clinton rape video, which was viewed 2.4m times in the month leading up to the election, was YouTube recommendations.

The same was true of all but one of the videos Franchi sent me data for. A typical example was a Next News Network video entitled “WHOA! HILLARY THINKS CAMERA’S OFF... SENDS SHOCK MESSAGE TO TRUMP” in which Franchi, pointing to a tiny movement of Clinton’s lips during a TV debate, claims she says “fuck you” to her presidential rival. The data Franchi shared revealed in the month leading up to the election, 73% of the traffic to the video - amounting to 1.2m of its views - was due to YouTube recommendations. External traffic accounted for only 3% of the views.

Franchi is a professional who makes a living from his channel, but many of the other creators of anti-Clinton videos I spoke to were amateur sleuths or part-time conspiracy theorists. Typically, they might receive a few hundred views on their videos, so they were shocked when their anti-Clinton videos started to receive millions of views, as if they were being pushed by an invisible force.

In every case, the largest source of traffic – the invisible force – came from the clips appearing in the “up next” column. William Ramsey, an occult investigator from southern California who made “Irrefutable Proof: Hillary Clinton Has a Seizure Disorder!”, shared screen grabs that showed the recommendation algorithm pushed his video even after YouTube had emailed him to say it violated its guidelines. Ramsey’s data showed the video was watched 2.4m times by US-based users before election day. “For a nobody like me, that’s a lot,” he says. “Enough to sway the election, right?”

Daniel Alexander Cannon, a conspiracy theorist from South Carolina, tells me: “Every video I put out about the Clintons, YouTube would push it through the roof.” His best-performing clip was a video titled “Hillary and Bill Clinton ‘The 10 Photos You Must See’”, essentially a slideshow of appalling (and seemingly doctored) images of the Clintons with voiceover in which Cannon speculates on their health. It has been seen 3.7m times on YouTube, and 2.9m of those views, Cannon said, came from “up next” recommendations.

Chaslot has put a spotlight on a trove of anti-Clinton conspiracy videos that had been hidden in the shadows – unless, that is, you were one of the millions YouTube served them to. But his research also does something more important: revealing how thoroughly our lives are now mediated by artificial intelligence.

Less than a generation ago, the way voters viewed their politicians was largely shaped by tens of thousands of newspaper editors, journalists and TV executives. Today, the invisible codes behind the big technology platforms have become the new kingmakers.

They pluck from obscurity people like Dave Todeschini, a retired IBM engineer who, “let off steam” during the election by recording himself opining on Clinton’s supposed involvement in paedophilia, child sacrifice and cannibalism. “It was crazy, it was nuts,” he said of the avalanche of traffic to his YouTube channel, which by election day had more than 2m views.

“Breaking news,” he announced in one of his last dispatches before the vote: the FBI, he said, had just found graphic images of Clinton and her aide in “sexually compromising positions” with a teenager. “It seems to me, with Bill Clinton’s trips to paedophile island a number of times, that what we have here is nothing short of the Clinton paedophile ring,” he declared.

Todeschini sits in his darkened living room in New Jersey, staring into his smartphone. “I’ll tell you what: the rabbit hole just got a couple of yards deeper.”

Contact the author: paul.lewis@theguardian.com.

A full description of the methodology Chaslot used to detect YouTube recommendations (and an explanation of how the Guardian analysed them) is available here.

Since you’re here...

... we have a small favour to ask. More people are reading and supporting our independent, investigative reporting than ever before. And unlike many news organisations, we have chosen an approach that allows us to keep our journalism accessible to all, regardless of where they live or what they can afford.

The Guardian is editorially independent, meaning we set our own agenda. Our journalism is free from commercial bias and not influenced by billionaire owners, politicians or shareholders. No one edits our editor. No one steers our opinion. This is important as it enables us to give a voice to those less heard, challenge the powerful and hold them to account. It's what makes us different to so many others in the media, at a time when factual, honest reporting is critical.

Every contribution we receive from readers like you, big or small, goes directly into funding our journalism. This support enables us to keep working as we do - but we must maintain and build on it for every year to come. **Support The Guardian from as little as \$1 - and it only takes a minute. Thank you.**

Support The Guardian



Topics

- Silicon Valley
- US politics
- US elections 2016
- Donald Trump
- Hillary Clinton
- features