



FEBRUARY 27, 2019

## Facebook Insider Leaks Docs; Explains “Deboosting,” “Troll Report,” & Political Targeting in Video Interview

(San Francisco) Project Veritas has obtained and published documents and presentation materials from a former Facebook insider.



Insider, Formerly Responsible for Content Review in Facebook’s Intellectual Property Dept Speaks Out, Loses Job

Facebook Engineers Plan to “demote bad content”

Conservative Facebook Page Livestreams Secretly “deboosted,” No Notice to Page Owners

Facebook Can Classify Users as Trolls Based on Their Vocabulary, Then Punish By Limiting Bandwidth, Blocking Comments...

Facebook Engineer: “hateful’ content is coming from right-leaning sites.”

“Special features” Triggered “leading up to important elections”

[View the documents here.](#)

(Update: In the video, we said Steven Crowder's “livestream was dethrottled,” when in fact his entire page was “dethrottled.”)

(UPDATE 2: [FACEBOOK RESPONDS](#) “We fired this person a year ago for breaking multiple employment policies and using her contractor role at Facebook to perform a stunt for Project Veritas,” a spokesperson told The Verge. “Unsurprisingly, the claims she is making validate her agenda and ignore the processes we have in place to ensure Facebook remains a platform to give people a voice, regardless of their political ideology.”)

(Update 3: Project Veritas has included captions on original images in this article to avoid reader confusion.)

**(San Francisco)** Project Veritas has [obtained and published documents and presentation materials](#) from a former Facebook insider. This information describes how Facebook engineers plan and go about policing political speech. Screenshots from a Facebook workstation show the specific technical actions taken against political figures, as well as “[e]xisting strategies” taken to combat political speech.

[\(Other brave individuals who feel compelled to expose wrongdoing they witness can contact Project Veritas here.\)](#)

Project Veritas founder James O’Keefe said that to expose dishonesty and censorship in big tech companies, he will be relying upon more insiders, informants and leakers in the future:

“Our future depends on those who are willing to give up everything for what they believe.”

---

# Strategies we use today

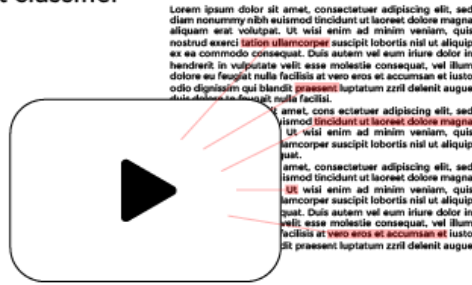
- Existing strategies
  - suspend or delete bad accounts
  - delete bad content
  - demote bad content
  
- 3. Trigger the “special features” when
  - Raids are detected, e.g. when known troll associates are content
  - User uploads a comment or post with a match on our ba Appendix)
  - **Leading up to important elections (pre-cautionary)**

To gain a better understanding of the documents, Project Veritas spoke with the Facebook insider in an interview. The insider separated from Facebook in 2018 and was later hired by Project Veritas.

“I saw things that were going on that I personally found to be troubling.”

## How does “deboosting” work?

Sensitive keywords in videos are identified by a “text classifier”



A program called “Sigma” will limit the visibility of the video



According to the insider, the documents revealed a routine suppression of the distribution of conservative Facebook pages. The technical action she repeatedly saw, and for which Project Veritas was provided documentation, was labeled *ActionDeboostLiveDistribution*. Said the insider, “I would see [this term] appear on several different conservative pages. I first noticed it with an account that I can’t remember, but I remember once I started looking at it, I also saw it on Mike Cernovich’s page, saw it on Steven Crowder’s page, as well as the Daily Caller’s page.”

Conservative commentator Steven Crowder’s page [had been suppressed](#) before in April 2016, and Crowder told Project Veritas they settled a dispute related to the issue with Facebook out of court. Asked for comment on this story, Steven Crowder’s attorney Bill Richmond said:

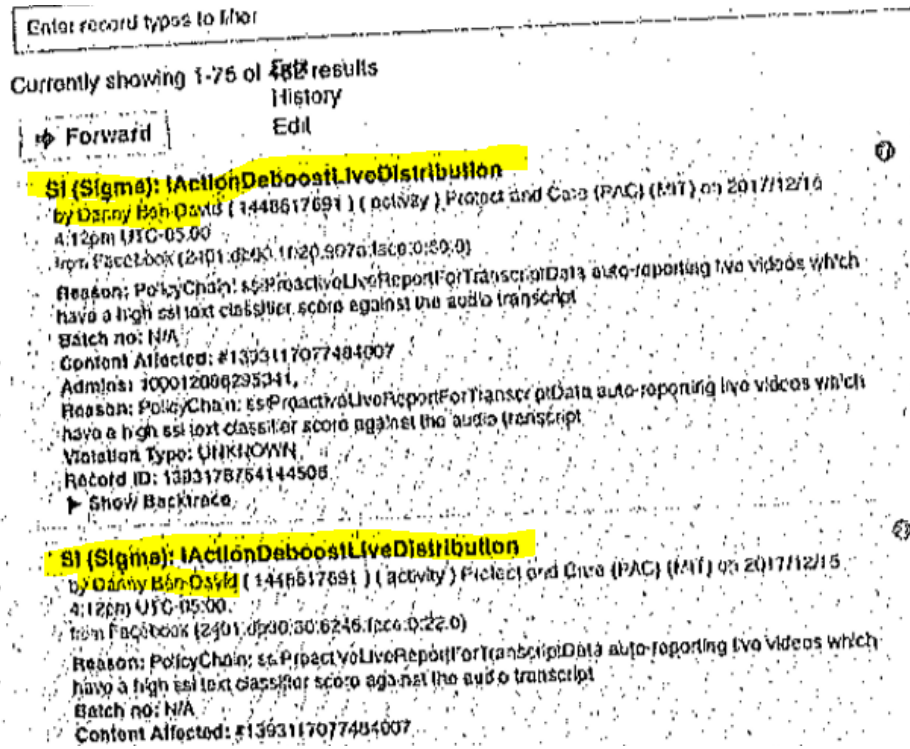
*“Louder With Crowder is investigating the allegations of concealed stream throttling by Facebook. The accusations are deeply troubling given the previously settled dispute with Facebook uncovered by Gizmodo.com, which found the show was targeted by Facebook workers with secret audience restrictions on political grounds alongside other prominent conservative voices.”*

A screenshot of an action log on Mike Cernovich’s Facebook page provided by the insider, shows the tag. The insider believes that the “deboost” code suppresses the distribution of livestream videos on Facebook. Project Veritas spoke to a current Facebook employee off the record who said that the code could limit a video’s visibility in news feeds, remove sharing features, and disable interactive notifications.

When approached for comment, author and filmmaker Mike Cernovich said the troubling issue is that Facebook could just “make stuff up” about people through these systems. “Facebook, or an individual at Facebook, has the unilateral power to create false allegations against someone he or she doesn’t

like. The person accused not only can't do anything about the allegation, they don't even have an idea the allegation was made," said Cernovich.

The insider says that unlike many actions that Facebook content moderators can take against pages, the "deboost" action, which appears to occur algorithmically, does not notify the page's owner. "[W]ith these 'deboost live stream' things, there was no warning sent to the user... These were actions that were being taken without the users knowing."

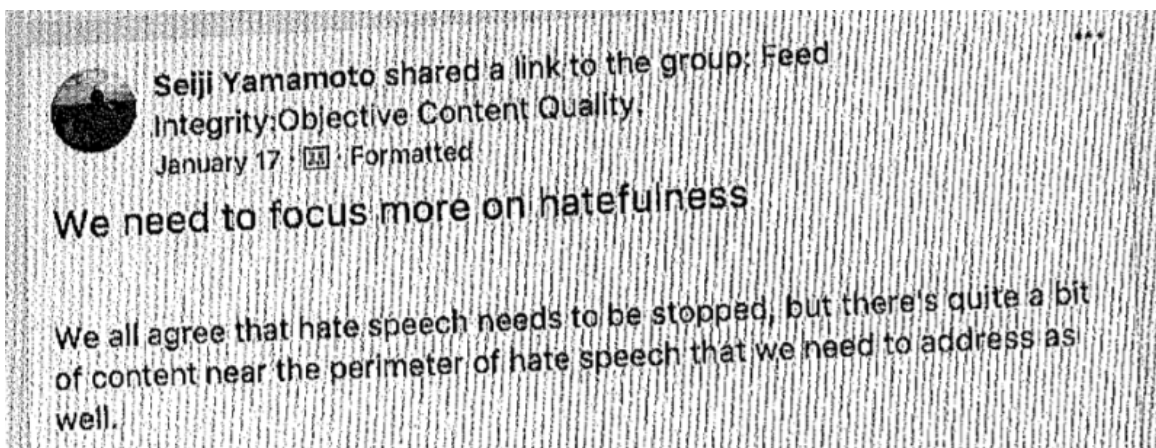


Upon further review, the insider says she did not notice the tag on any left-wing pages. "I looked at the Young Turks' page, I looked at Colin Kaepernick's page, none of them had received the same *deboost* comment."

The "deboost" tag appears after the word "Sigma," which Project Veritas has learned is an artificial intelligence system used to block potential suicide and self-harm posts. Both Mike Cernovich and Steven Crowder cannot recall having ever produced any videos on Facebook that promote suicide or self-harm. Mike Cernovich told Project Veritas that in fact he has long spoken out against suicide and self-harm, and provided tweets of his and a blog post as evidence.

"They're shifting the goal post"

Also in the in the documents was a presentation, authored by Facebook engineers Seiji Yamamoto and Eduardo Arino de la Rubia, titled "Coordinating Trolling on FB." Yamamoto is a Data Science Manager, and de la Rubia is a Chief Data Scientist at Facebook. The presentation appears to describe the current actions, as well as potential future actions, Facebook can take to combat alleged abusive behavior on the platform.



Yamamoto, who is responsible for "News Feed Reduction Strategy," also authored a post where he said Facebook should address "...quite a bit of content near the perimeter of hate speech." Said the Facebook insider, the "perimeter of hate speech" means "things that aren't actually hate speech but that might offend somebody. Anything that is perceived as hateful but no court would define it as hate speech."

The insider believes Yamamoto's plans appears to be political in nature, rather than in response to abusive behaviors, "[i]t was clearly kind of designed... aimed to be the right wing meme culture that's become extremely prevalent in the past few years. And some of the words that appeared on there were, using words like *SJW*... *MSM*... the New York Times doesn't talk about the *MSM*. The independent conservative outlets are using that language."

Also in Yamamoto's report was a line appearing to say that online Facebook trolls are involved in "destructive behaviors" such as "[r]ed-pilling normies to convert them to their worldview."

In online circles the term "red-pilling" refers to bluntly showing the truth, and "normies" refers generally to apolitical or uninformed people. Directly below the line in the document is hyperlink labeled "example video."

Yes, their meme content is seriously bad, but let's examine Troll behavior more broadly rather than just focus on content. Trolls are involved in many destructive behaviors on Facebook, including:

- Raids
  - Coordinated attacks on groups, pages, and individuals
- Doxing and Harrassement
  - Targeting specific individuals, e.g. Sophie Labelle
    - <https://io9.gizmodo.com/transgender-creator-of-assigned-male-webcomic-facing-de-1795442512>
- False reporting of groups, pages, and individuals to Facebook Ops
- Toxic meme creation and templating for widespread dissemination to various FB groups and pages
- Black Propaganda
- Red-pilling normies to convert them to their worldview
  - example video: <https://www.youtube.com/watch?v=c31NAbNF7PM>

5

The video linked in the presentation was made by Lauren Chen, a conservative commentator who now hosts a program on BlazeTV. "If you actually watch the video you can see that it clearly isn't abusive or promoting harassment, the video was a criticism of social justice," said Chen when asked for comment on this story. She added that "the video actually promotes equality and individualism."

On a page from the presentation titled "Strategies we use today," Yamamoto and de la Rubia list "demote bad content." They add, "... we should still of course delete and demote, but we can do even more..."

*(Similar documents from big tech companies can be sent to Project Veritas here.)*

Other actions that could be interpreted as "bad content" could be posting words such as "zucced," "REEE," and "normie." Said the insider, Facebook is "shifting the goal post. It's one thing, if you're dropping the n-word, or things like that, using some kind of homophobic or racial slur, by all means that's something that a platform should not want on it. But now you're moving it to things like, jokes that conservatives tend to make."

"Special features" triggered "leading up to important elections"



## Are you a Facebook troll?

Do you...

Use words like "REE" or "normie?"

Share popular memes?

Join or "like" many different Facebook groups?

---

### Tactic 5: Build a troll classifier

- Build classifier that predicts if a user is a Troll
- Trolls have distinct vocabulary with rare words
  - cuck, zucked, REEE, normie, IRL, lulz, Shadilay, etc.
- Accounts that post anything with hits to the meme cache or toxic comment cache are more likely to be trolls
  - Troll memes and toxic comments are like marked bills which we can use to identify trolls
- Membership in certain groups and associations with other troll accounts would likely help
- This could be a stand-alone classifier, or be integrated into Deep Entity Classification and the Fake Account index

Two of the "tactics" outlined in the presentation that the Facebook engineers propose for dealing with "troll operations" involve the introduction of a "Troll Twilight Zone."

Yamamoto and de la Rubia's presentation says that "troll accounts," can have their internet bandwidth limited and experience forced glitches like frequent "auto-logout[s]" and the failed upload of comments. These "special features" would be triggered "leading up to important elections."

Facebook could identify trolls by their vocabulary, friend network, and behavior, according to the presentation. "Facebook has what's called a Fake Account Index," explained the insider, "where they assign a score which helps them determine whether the account is a real person or just a dummy spam account. And rightfully so, they want to delete those accounts, that's okay. They created the *troll score* so they could help identify, using words they would post, pictures, if they were friends with other trolls and then using that to determine whether this person should be on the platform or not."

The insider thinks that Facebook's system to score trolls is problematic because, "there's no accountability and especially when they are using machine learning to do this, whenever an individual actions an account, there is a process where at least you can send a message to Facebook... However, this is all being done without the user's knowledge, there is no recourse for them."

Another proposed tactic in the presentation would apparently alert a "troll's" friends list when they have been banned. The presentation reads:

"When a user does something egregious, warranting an account suspension or deletion, we should notify the friend network

"John Smith's account has been suspended for 7 days because he shared hate speech in the group Kekistani Special Forces"



The presentation says that notifying a “troll’s” friend list would “strike fear in the hearts of trolls...” and “[n]otified users who accidentally befriended the offender might be more mindful of suspicious accounts, increasing overall herd immunity.”

The insider now works for Project Veritas



Said the insider:

*“I think that the biggest thing, that getting the documents, getting video or still pictures of what was going on that shows that it is actually happening. This isn’t rumors, they talk about how right-wingers, they come up with all these crazy theories, and that’s not actually happening at these social media companies. They pooh pooh it. But here it is and it’s in your face.”*

The insider was terminated by Facebook shortly after Project Veritas published [undercover video of Twitter employees discussing “shadow banning”](#) and other data privacy abuses. When asked if supplying documents and testimony to Project Veritas was “worth it,” the insider said “Yes... I knew what I had seen... this is something they were trying to keep in the shadows, that they did not want to public to know and yet the public has a right to know.”

Project Veritas founder James O’Keefe believes that “our collective future depends on those who are willing to give up everything for what they believe.” He believes that if more insiders from large companies step forward and expose similar dishonesty and wrongdoing, that the country will be better educated. O’Keefe said:

*“What are you willing to give up? How many of you will step forward? While they may be able to stop one of us, they won’t be able to stop an army. Be brave. Do something.”*

([Big tech insiders can reach out to Project Veritas here to help expose similar newsworthy wrongdoing.](#))

## HELP EXPOSE CORRUPTION

SEND A TIP

DONATE NOW

1214 Boston Post Road No. 148  
Mamaroneck, NY 10543

(914) 908-2300

© 2020 Project Veritas. All rights reserved.

[Privacy Policy](#)