


LEAK: Huge Teams Engaged in Manual Interventions on Google Search Results



 Google wants to "Police Tone"
Adrian Dennis/AFP/Getty

Google has “huge teams” working on manual interventions in search results, an apparent contradiction of sworn testimony made to Congress by CEO Sundar Pichai, according to an internal post [leaked](#) to Breitbart News.

“There are subjects that are prone to hyperbolic content, misleading information, and offensive content,” said Daniel Aaronson, a member of Google’s Trust & Safety team.

“Now, these words are highly subjective and no one denies that. But we can all agree generally, lines exist in many cultures about what is clearly okay vs. what is not okay.”

Breitbart TV



[Google Searches For ‘Recession’ Hit Highest Point Since Late 2009](#) 

“In extreme cases where we need to act quickly on something that is so obviously not okay, the reactive/manual approach is sometimes necessary.”

The comments came to light in a leaked internal discussion thread, started by a Google employee who noticed that the company had recently changed search results for “abortion” on its YouTube video platform, a change which caused pro-life videos to largely disappear from the top ten results.

In addition to the “manual approach,” Aaronson explained that Google also trained automated “classifiers” – algorithms or “scalable solutions” that corrects “problems” in search results.

Aaronson listed three areas where either manual interventions or classifier changes might take place: organic search (“The bar for changing classifiers or manual actions on span in organic search is *extremely* high”), YouTube, Google Home, and Google Assistant.

Aaronson’s post also reveals that there is very little transparency around decisions to adjust classifiers or manually correct controversial search results, even internally. Aaronson compared Google’s decision-making process in this regard to a closely-guarded “Pepsi Formula.”

These comments, part of a longer post copied below, seem to contradict Google CEO Sundar Pichai’s sworn congressional testimony that his company [does not “manually intervene on any particular search result.”](#)

According to an [internal discussion thread](#) leaked to Breitbart News by a source within the company, a Google employee took issue with Pichai’s remarks, stating that it “seems like we are pretty eager to cater our search results to the social and political agenda of left-wing journalists.”

According to the posts leaked by the source, revealed that YouTube, a Google subsidiary, manually intervened on search results related to “abortion” and “abortions.” The intervention caused pro-life videos to disappear from the top ten search results for those terms, where they had previously been featured prominently. The posts also show YouTube intervened on search results related to progressive activist David Hogg and Democrat politician Maxine Waters.

In a comment to Breitbart News, a Google spokeswoman also insisted that “Google has never manipulated or modified the search results or content in any of its products to promote a particular political ideology.”

Pichai might claim that he was just talking about Google, not YouTube, which was the focus of the leaked discussion thread. But Aaronson’s post extends to Google’s other products: organic search, Google Home, and Google Assistant.

Aaronson is also clear that the manipulation of the search results that are “prone to abuse/controversial content” is not a small affair, but are the responsibility of “huge teams” within Google.

“These lines are very difficult and can be very blurry, we are all well aware of this. So we’ve got huge teams that stay cognizant of these facts when we’re crafting policies considering classifier changes, or reacting with manual actions”

If Google has “huge teams” that sometimes manually intervene on search results, it’s scarcely plausible to argue that Pichai might not know about them.

Aaronson’s full post is copied below:

I work in Trust and Safety and while I have no particular input as to exactly what’s happening for YT I can try to explain why you’d have this kind of list and why people are finding lists like these on Code Search.

When dealing with abuse/controversial content on various mediums you have several levers to deal with problems. Two prominent levers are “Proactive” and “Reactive”:

- Proactive: Usually refers to some type of algorithm/scalable solution to a general problem
 - E.g.: We don't allow straight up porn on YouTube so we create a classifier that detects porn and automatically remove or flag for review the videos the porn classifier is most certain of
- Reactive: Usually refers to a manual fix to something that has been brought to our attention that our proactive solutions don't/didn't work on and something that is clearly in the realm of bad enough to warrant a quick targeted solution (determined by pages and pages of policies worked on over many years and many teams to be fair and cover necessary scope)
 - E.g.: A website that used to be a good blog had it's domain expire and was purchased/repurposed to spam Search results with autogenerated pages full of gibberish text, scraped images, and links to boost traffic to other spammy sites. It is manually actioned for violating policy
 - [These Organic Search policies and the consequences to violating them are public](#)

Manually reacting to things is not very scalable, and is not an ideal solution to most problems, so the proactive lever is really the one we all like to lean on. Ideally, our classifiers/algorithm are good at providing useful and rich results to our users while ignoring things that are not useful or not relevant. But we all know, this isn't exactly the case all the time (especially on YouTube).

From a user perspective, there are subjects that are prone to hyperbolic content, misleading information, and offensive

content. Now, these words are highly subjective and no one denies that. But we can all agree generally, lines exist in many cultures about what is clearly okay vs. what is not okay. E.g. a video of a puppy playing with a toy is probably okay in almost every culture or context, even if it's not relevant to the query. But a video of someone committing suicide and begging others to follow in his/her footsteps is probably on the other side of the line for many folks.

While my second example is technically relevant to the generic query of "suicide", that doesn't mean that this is a very useful or good video to promote on the top of results for that query. So imagine a classifier that says, for any queries on a particular text file, let's pull videos using signals that we historically understand to be strong indicators of quality (I won't go into specifics here, but those signals do exist). We're not manually curating these results, we're just saying "hey, be extra careful with results for this query because many times really bad stuff can appear and lead to a bad experience for most users". Ideally the proactive lever did this for us, but in extreme cases where we need to act quickly on something that is so obviously not okay, the reactive/manual approach is sometimes necessary. And also keep in mind, that this is different for every product. The bar for changing classifiers or manual actions on span in organic search is *extremely* high. However, the bar for things we let our Google Assistant say out loud might be a lot lower. If I search for "Jews run the banks" - I'll likely find anti-semitic stuff in organic search. As a Jew, I might find some of these results offensive, but they are there for people to research and view, and I understand that this is not a reflection of Google feels about this issue. But if I ask Google assistant

“Why do Jews run the banks” we wouldn’t be similarly accepting if it repeated and promoted conspiracy theories that likely pop up in organic search in her soothing voice.

Whether we agree or not, user perception of our responses, results, and answers of different products and mediums can change. And I think many people are used to the fact that organic search is a place where content should be accessible no matter how offensive it might be, however, the expectation is very different on a Google Home, a Knowledge Panel, or even YouTube.

These lines are very difficult and can be very blurry, we are all well aware of this. So we’ve got huge teams that stay cognizant of these facts when we’re crafting policies considering classifier changes, or reacting with manual actions – these decisions are not made in a vacuum, but admittedly are also not made in a highly public forum like TGIF or IndustryInfo (as you can imagine, decisions/agreement would be hard to get in such a wide list – imagine if all your CL’s were reviewed by every engineer across Google all the time). I hope that answers some questions and gives a better layer of transparency without going into details about our “Pepsi formula”.

Best,

Daniel

Breitbart Tech will continue to investigate Google’s manipulation of search results on both its search engine and the YouTube video platform.