Antisemitic memes in image results highlight vulnerabilities in search

Antisemitic memes appearing in Google Image search results caught the attention of search professionals last week – and broader media this week. Many are urging the company to make tangible moves towards removing such content.

Google isn't the only search engine surfacing these hateful memes – Search Engine Land's Editor-in-Chief, Ginny Marvin, pointed out that they can also be found on Bing and DuckDuckGo.

On its face, the issue may seem simple to resolve by just removing the offending content from the index; however, they and others say, it's not that easy and opens the door to broader decisions on content moderation. The fact that Google has no definitive plans to remove the offensive images speaks to how search engines are designed to work.

How search engines have(n't) responded

In a statement to Search Engine Land, a Google spokesperson said:

"We understand these are disturbing results, and we share the concern about this content. It does not reflect our opinions. When people search for images on Google, our systems largely rely on matching the words in your query to the words that appear next to images on the webpage. For this query, the closest matches are webpages that contain offensive and hateful content. We've done considerable work in improving

instances where we return low quality content, and will continue to improve our systems."

This statement is a reiteration of the initial response the company issued via its Search Liaison Twitter account. In the Twitter thread, Google pointed to "data voids" (more on that below) as the reason this hateful content is able to surface in the results:

"For 'baby strollers', there's lots of helpful content. For this, there's not. That's not surprising. It's not likely a topic normally searched for, nor an actual product that's marketed. There's a "void" of good content to surface that matches what was asked for…"

No plans to remove the offensive content. Google has also said that it does not have a removal policy for this type of content, and that it is working "to see if there are other ways to surface more helpful content." This is not the first time controversial and offensive content has been spotted on Google, of course. In 2009, the company bought a search ad to explain why it wasn't removing an offensive image of Michelle Obama, for example.

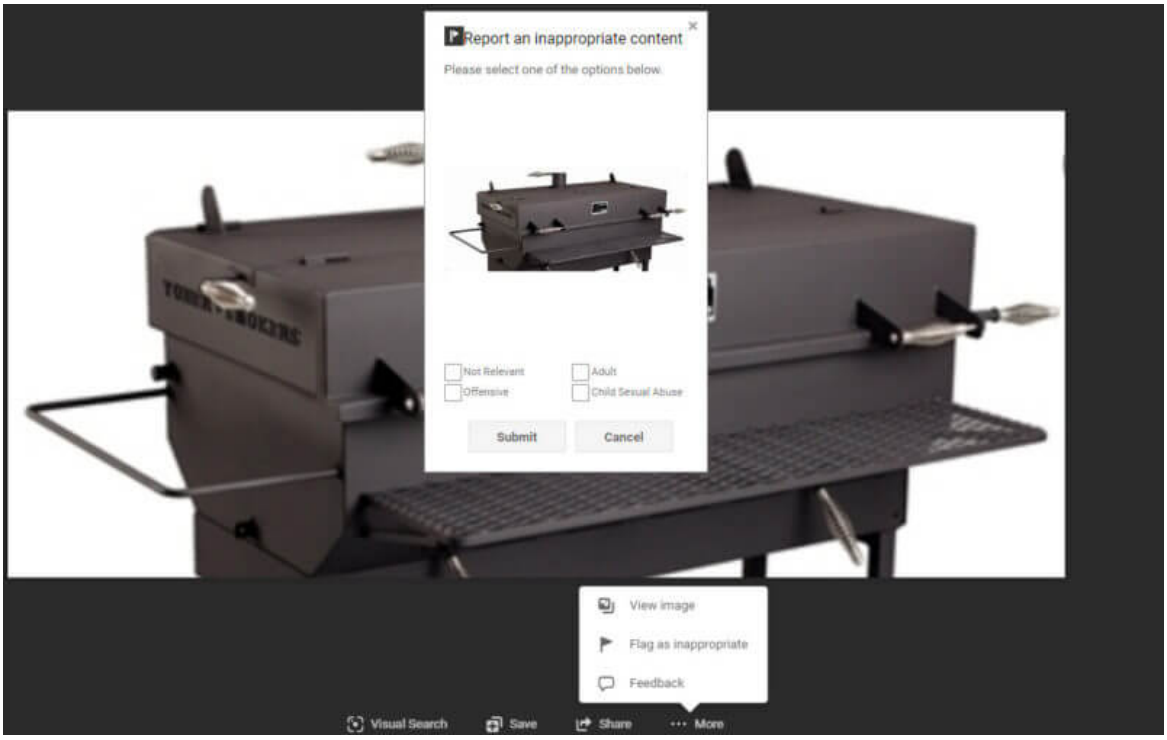
We have also reached out to Bing for comment; however, a response was not received by the time of publication. We will update this article if we receive one.

Reporting offensive images. At one point, Google enabled users to report offensive images in its Image results. This feature no longer exists, at least not in the form it once did.

Now, users can send feedback by clicking on an image in the Image results and selecting "Send feedback" from the three-dot menu in the preview pane. Users are then presented with a window to highlight the content they're submitting feedback on and a text field to provide more details.

Users can report specific images in Google web search, but

this functionality is only available when image previews show up on the main search results page, and there are no options to provide details on why the image is being reported.



Bing's Image search results allow users to flag inappropriate content.

Bing's image reporting functionality is more robust, enabling users to flag an image as inappropriate by marking off "Not Relevant," "Adult," "Offensive," and/or "Child Sexual Abuse" as the reason. Users that want to leave more specific feedback can do so by selecting "Feedback," which provides a pop-up window similar to Google's.

What are data voids?

Data voids refer to search terms for which available relevant data is "limited, non-existent, or deeply problematic," according to Microsoft's Michael Golebiewski and Danah Boyd in their report *Data Voids: Where Missing Data Can Easily Be Exploited*.

Data voids are a consequence of search engine design, the report asserts. The assumption "that for any given query, there exists some relevant content," is at the heart of the

issue. When there is little available content to return to users for a particular search term, as was the case with "Jewish baby strollers," the most relevant content has a higher potential for problematic, low quality content to rise to the top.

Why are data voids a problem?

When there is little competing content for a query, its easier to get content to rank for that query. When there is a data void, it's even easier for low quality or malicious actors to do so because they aren't competing with volumes of content from more reputable and/or relevant sources.

It's unrealistic to think that appropriate content will always outweigh hateful content, especially for problematic queries. "Without new content being created, there are certain data voids that cannot be easily cleaned up," Golebiewski and Boyd wrote – a notion that falls in line with Google's statement that it is evaluating other ways to surface more helpful content.

When no one is searching for these terms, the issue can go unnoticed – and an unnoticed issue is often a nonexistent issue as far as most companies responsible for fixing them are concerned. However, when problems like this do gain attention in the media, the problem gets compounded.

Case in point: By writing about this particular issue and featuring screenshots of the offensive memes, Google indexes those image instances, increasing the visibility of the antisemitic memes in Image results (and why I have opted to embed tweets instead of screenshots of the images themselves).

Problematic queries aren't the only place data voids exist. In their report, Golebiewski and Boyd bucketed data voids into five categories:

- Breaking news: Users may be searching for terms related to a breaking news story before accurate, high quality content about it becomes available.
- Strategic new terms: New terminology may be coined by malicious actors, and trustworthy content relevant to that terminology may not yet exist. The example of "crisis actors" was given, referring to trained actors playing mock victims; conspiracy theorists then applied these terms to content about the 2012 Sandy Hook Elementary School shooting to spread their false narrative.
- Outdated terms: Words or phrases that are no longer in use, such as terminology that is now considered racist, may not appear in content published by reputable sources. Additionally, seasonality (such as searches for the Winter Olympics between games) may also result in an exploitable data void.
- Fragmented concepts: Users searching seemingly synonymous terms may be presented with entirely different results; for example "undocumented aliens" and "illegal aliens." While these terms may converge over time, "media manipulators can also exploit search engines' resistance to addressing politically contested distinctions to help fragment knowledge and rhetoric."
- Problematic queries: "Jewish baby strollers," while not inherently a hateful term, is one example of a problematic query. This category includes conspiratorial and other unpredictable search queries.

In each of the scenarios above, data voids present malicious actors with an opportunity to use search engine algorithms to distract and misinform users.

The search community's response

Numerous search professionals, and the public at large as this story begins attracting their attention, have urged Google to

take swifter action. In the tweet below, Mordy Oberstein, liaison to the SEO community at Wix, pointed to similarly hateful memes appearing for the query “Jewish bunk beds.”

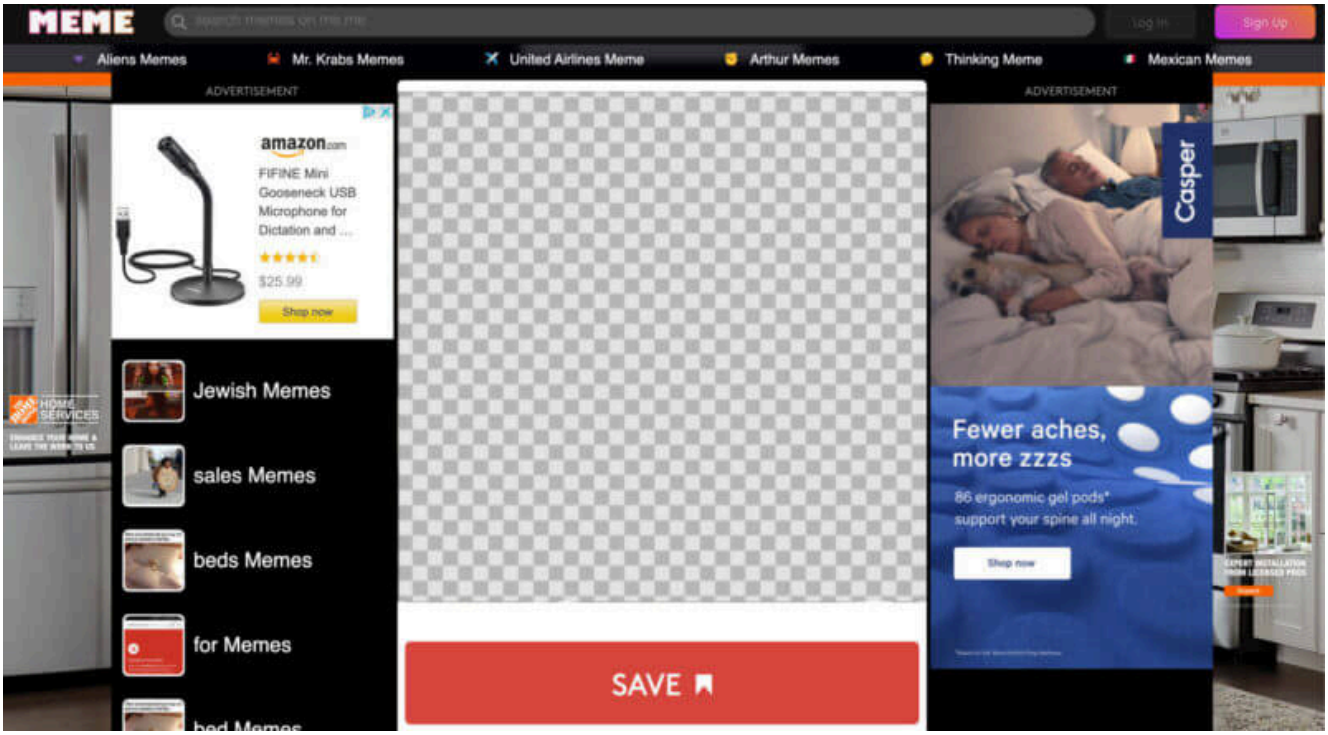
As Google suggested in its official responses, the solution may not be as simple as manually removing the offending content from the index, an unfortunate reality that Victor Pan, head of technical SEO at Hubspot, and former Googler Pedro Dias alluded to.

> *I’ve tried to explain people why you can’t make linear decisions like this, as much as it sucks and pains. It’s not about protecting Google, it’s about accounting for when the axe swings in the opposite direction.*
> *Unfortunately, people don’t get this, unless it aligns w them.*
>
> *— Pedro Dias (@pedrodias) September 28, 2020*

Brand safety implications

Google’s ad systems don’t appear to be monetizing these results — we have not seen ads on the pages. However, the offensive memes found in Image search live on actual web pages — and those web pages typically carry display advertising — a lot of display advertising.

The censored screenshot below is an example of a situation that many marketers would hope to steer their clients away from: Ads for the Home Depot, Casper and a product on Amazon were seen running alongside the offensive content. In this case, the ads are being served via Amazon’s interest-based ad targeting network, not the Google Display Network.



Navah Hopkins, director of paid media at Hennessey Digital, said that she would encourage her clients to pull their Google Ads spend in the fourth quarter if Google doesn't fix this problem.

"Where my frustration and hesitation to encourage brands to continue to invest in respective ad platforms comes from is the reluctance to put safeguards in place against these bad acts, as well as the refusal to take down hateful content," Hopkins told Search Engine Land, "We already filter out placements that we consider racist and hateful – if ad networks can't take a stand on hate-based content, we will need to apply the same standards."

About The Author



George Nguyen is an editor for Search Engine Land, covering organic search, podcasting and e-commerce. His background is in journalism and content marketing. Prior to entering the industry, he worked as a radio personality, writer, podcast host and public school teacher.

http://feeds.searchengineland.com/~r/searchengineland/~3/33u-Pjzgn-0/antisemitic-memes-in-image-results-highlight-vulnerabilities-in-search-341267