Data Voids: Where Missing Data Can Easily Be Exploited

Technical Report · October 2019

2 authors, including:

danah boyd
Microsoft

160 PUBLICATIONS **53,714** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Census View project

DATA&
SOCIETY

DATA VOIDS

WHERE MISSING DATA
CAN EASILY BE EXPLOITED

Michael Golebiewski
danah boyd

CONTENTS

Executive Summary

The logic underpinning search engines is akin to a lesson from kindergarten: no question is a bad question. **But what happens when innocuous questions produce very bad results for users?**

Data voids are one such way that search users can be led into disinformation or manipulated content. **These voids occur when obscure search queries have few results associated with them, making them ripe for exploitation by media manipulators with ideological, economic, or political agendas.** Search engines aren't simply grappling with media manipulators using search engine optimization techniques to get their website ranked highly or to get their videos recommended; they're also struggling with conspiracy theorists, white nationalists, and a range of other extremist groups who see search algorithms as a tool for exposing people to problematic content.

Data voids are difficult to detect. Generally speaking, data voids are not a liability until something happens that results in an increase of searches on a term. Some are created by media manipulators, and escape notice for long periods of time. Others are the sudden products of a news spike, as millions are prompted to search names or terms for the first time, and misleading or hateful content is created to meet demand. Search-adjacent recommendation systems, like search bar auto-suggestions, further complicate the data voids problem by providing auto-suggestions that can send people down deeply disturbing paths.

Search engine creators want to provide high quality, relevant, informative, and useful information to their users, but they face an arms race with media manipulators. In this report, we focus on five types of data voids that are currently being corrupted by those spreading conspiracies or hate:

Breaking News: The production of problematic content optimized to terms that are suddenly spiking due to a breaking news situation; these voids will eventually be filled by legitimate news content, but are abused before such content exists.

Strategic New Terms: Manipulators create new terms and build a strategically optimized information ecosystem around them before amplifying those terms into the mainstream, often through news media, in order to introduce newcomers to problematic content and frames.

Outdated Terms: When terms go out of date, content creators stop producing content associated with these terms long before searchers stop seeking out content. This creates an opening for manipulators to produce content that exploits search engines' dependence on freshness.

Fragmented Concepts: By breaking connections between related ideas and creating distinct clusters of information that refer to different political frames, manipulators can segment searchers into different information worlds.

Problematic Queries: Search results for disturbing or fraught terms that have historically returned problematic results continue to do so unless high quality content is introduced to contextualize or outrank such problematic content.

Data voids raise questions about what role search engines can and should play in diverting their users from disturbing search results. **We argue that there is no "fix" for data voids.** Search engines and content creators must work together to identify these vulnerabilities, iteratively respond to attacks, and produce the high-quality content that is needed to fill these data voids.

Introduction

Search engines and recommender systems (a.k.a., "recommendation systems") play a unique role in modern online information systems. Unlike people's use of social media, where they primarily consume algorithmically curated feeds of information shared by those in their social network, people's approaches to search engines typically begin with a query or question in an effort to seek new information. Many recommender systems operate adjacent to search engines and search features, offering recommendations for new searches to query or even allowing content to be streamed based on the result of a search. While these are frequently designed to help increase clarity for the search engine, they may also invite users to traverse a network of information into areas that the searcher never previously considered.

> "There are many search terms for which the available relevant data is limited, nonexistent, or deeply problematic. ...We call these low-quality data situations 'data voids.'"

Not all search queries are equal. Many more people search for "basketball" than "underwater basket weaving." Likewise, a lot more content is created about the sport than the absurdist activity. As a result, when search engines like Bing and Google try to provide users with information about basketball, they have *a lot* more data to work with than they do with underwater basket weaving. The same is true for social media platforms that function as a search engine in many contexts, such as YouTube. Because basketball is more popular with more people than underwater basket weaving, more people produce more content related to and search more often for the former.

There are many search terms for which the available relevant data is limited, nonexistent, or deeply problematic.[1] Recommender systems also struggle when there's little available data to recommend. We call these low-quality data situations "data voids." Data voids lead to low quality or low authority content because that's the only content available. They come about both naturally and through manipulation. When people *do* search for a term that leads to a data void, search engines return results based on limited data. If you type a random set of characters into a search engine – e.g., "aslkfjastowerk;asndf" – you will probably receive no results—simply because no pages contain that random set of letters. But there is a long tail between a term like "basketball," which promises a seemingly infinite number of results, and one with zero results. In that long tail, there are plenty of search queries that can drop people into a data void rife with existing but deeply problematic results. Some of these data voids are intentionally exploited to introduce disturbing content, while others are created to promote political propaganda.[2] Moreover, data voids are difficult to detect. Some are created by obscure search queries that escape notice for long periods of time. Others are the sudden products of a news spike, as millions are prompted to search names or terms for the first time. Generally speaking, data voids are not a liability until something happens that results in an increase of searches on a term.

---

1   "Problematic" is an overarching term attempting to account for a range of content that search engines grapple with. This includes conspiratorial, extremist, hate-oriented, terroristic, graphic, and illicit content. Search engines generally treat this content as acceptable to return when they know that this is what people are intentionally searching for, given a widespread commitment among search engine creators that they should not prevent users from seeking out most information. That said, this category of content is deeply concerning for search engines when they might be exposing people to content that they didn't intend to see.

2   Francesca Tripodi, *Searching for Alternative Facts: Analyzing Scriptural Inference in Conservative News Practices*, (New York: Data & Society, 2018). https://datasociety.net/wp-content/uploads/2018/05/Data_Society_Searching-for-Alternative-Facts.pdf.

The logic underpinning search engines is akin to a lesson from kindergarten: no question is a bad question. Every search teaches the system something about what people are looking for, what they are (or aren't) clicking on. But some search queries can produce very bad results for users, which means search engine companies must be constantly working to improve their systems. Media manipulators have learned to capitalize on missing data, the logics of search engines, and the practices of searchers to help drive attention to a range of problematic content. Sometimes this is simple digital marketing, but these techniques are increasingly being adopted by networks of people invested in spreading hate and polarizing society. Because of this, a new awareness of and approach to data voids is necessary to enable a healthy information ecology.

In this paper, we offer some basic background on search engines before discussing the different types of data voids that appear in search engines and adjacent recommender systems, the challenges that search engines face when they encounter data

> "Media manipulators have learned to capitalize on missing data, the logics of search engines, and the practices of searchers to help drive attention to a range of problematic content."

voids, and the ways data voids can be exploited by media manipulators with ideological, economic, or political agendas. Search engines aren't simply grappling with people who want their favorite team to come up when someone searches for basketball; they're struggling with conspiracy theorists, white nationalists, and a range of other extremist groups who see search as a tool for radicalizing people.

Understanding how these data voids are created and exploited will be crucial for limiting the influence of manipulators. Currently, search engines are locked in a type of arms race with those who wish to twist the landscape of public information, and while traditional efforts to update models and moderate certain problematic queries have long been a key part of search engine operation, this new use of data voids to amplify and fragment content will require new strategies and new collaborations. Content creators themselves will be part of this, by understanding how filling in data voids can create a more secure public sphere. But search engines as well must take additional steps to identify and prevent this type of abuse.

How Search Engines Work

In order to organize information and respond to queries in a reliable manner, search engines must first obtain a corpus of data to work with. Web search tools like Bing and Google "crawl" the web to map available information (including URLs and their content, links, images, videos, etc.).[3] Search features inside specific services, like YouTube, rely exclusively on content directly uploaded to that service. Search engines do not necessarily include all existing content. Through longstanding technical standards, website owners can signal to search engines that they don't want their content indexed by search engines and, thus, these websites are not included in search results. Likewise, platform companies may choose to exclude certain content from their platforms' search tools, perhaps because users mark that content private or it is illegal in a given jurisdiction.

Once a corpus of data is constructed and organized, engineers design models that allow search engines to quickly identify and prioritize content that most likely matches the desired goals of a searcher. This isn't an easy task, in no small part because what people search for is often vague. Consider what a user is really searching for with a query like "subway." Are they looking for information about the closest transit station and its closures? Information about the fast food restaurant and its hours? A history of subways around the world? Without more context or information, a search engine simply makes a probabilistic guess.

The flip side of vague queries is narrow content. If a website only uses a term like "public transit" – and never references the colloquial concept of "subway" – how does a search engine know that these two concepts are related?

---

3   For more information on how search engines work, see: Jutta Haider and Olof Sundin, *Invisible Search and Online Search Engines: The Ubiquity of Search in Everyday Life* (London: Routledge, 2019); Alex Halavais, *Search Engine Society* (Cambridge: Polity, 2019).

Search engines will need to continue to build improved models that better understand related terms in order to match a user's intention and the content that it might be able to return. At the same time, creating relationships between distinct terms can also connect more than words. For example, a search for "green poop" should probably return similar results as a search for "green stool" because they are commonly used synonymously. Yet, by collapsing these two concepts, a search engine's model creates a situation in which parenting forums discussing children's bowel movements as "poop" are ranked alongside medical and scientific content concerning "stool" (not to mention green IKEA stools for sitting). This may sometimes be desirable, but what if someone is using a narrow phrase like "green poop" in order to get a specific type of content and doesn't want content associated with "green stool"?

The architects of search engines draw on a wide range of available information to help maximize the likelihood that search results give users what they want. This information comes from sources including web pages themselves (e.g., visible text on the page and metadata like anchor text or title of the page), previous searches and interactions from other users, and additional information like the geographical location of the person's computer. Search interfaces are designed to coax the user into offering more information by auto-suggesting additional phrases, which encourage users to narrow the query and increase the likelihood that the search engine returns relevant results.

Search engines aren't human. They're machine learning systems designed by people; they don't understand the *meaning* of words. They focus on the probability that a given page, image, video, or news story will be the result that prompts users to take an action that the search engine registers as positive

feedback (e.g., a click). Overlaps like the one between Sub-way-the-sandwich-shop and subway-the-underground-train aren't identified through manual demarcation, but through statistical probabilities and models derived from the data that suggests a different topological link structure and related word context. For example, very few pages discussing train timetables have detailed descriptions of sandwiches. Thus, within the data model, these two types of content are registered as distinct even if they include the same word. On the other hand, data models must also identify content whose words are distinct even if they are part of the same cluster of knowledge (e.g., "green poop" and "green stool"). While the engineers behind search engines have historically resisted manual human intervention, preferring to improve their algorithmic systems, there are also situations in which human overrides are deemed necessary. For instance, potentially harmful information concerning terrorism and suicide are typically monitored while related searches are contextualized with pointers to crisis hotlines manually placed on those pages. While this can be valuable for certain queries, creating this type of intervention is impractical for all but a tiny proportion of queries.

Search Engine Optimization

Modern search engines all extend early work on "page rank." Page rank involves scraping web pages to determine the link structure and then ranking pages based on the inbound links that a given page receives. When Larry Page, Sergey Brin, Rajeev Motwani, and Terry Winograd first proposed this algorithm in 1999,[4] they imagined that this technique would provide an "objective" measure of what links search users would find desirable. They were responding to earlier search engines that ranked results based on financial offers or editorial considerations. Their page rank technique became the foundation of Google. Google founders (Page and Brin) initially imagined their invention to be resilient to manipulation.[5]

It didn't take long for people to game Google and increase the visibility of their content on the new search engine. Website creators spent significant time and money to increase their page rank with increasingly sophisticated search engine optimization (SEO) efforts. Marketers paid SEO companies to increase the rank of their content through any means possible. SEO shops created "link farms" to artificially inflate the inbound links for a site. They created automated "bots" to click on specific results in search results in order to reinforce the signals that search algorithms use to assess link relevance. They helped website owners alter their webpages to include machine-readable content that was invisible to the user. Google responded to these forms of manipulation by downranking certain signals and prioritizing others. As

---

4   Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd, "The PageRank Citation Ranking: Bringing Order to the Web," (Stanford: Stanford InfoLab Technical Report, 1999). http://ilpubs.stanford.edu:8090/422/.

5   Critics have long argued that there is nothing objective about search engine ranking. See, for example: Lucas D. Introna and Helen Nissenbaum, "Shaping the Web: Why the Politics of Search Engines Matters," *The Information Society* 16 no. 3 (2000) 169-185.; Siva Vaidhyanathan, *The Googlization of Everything (and Why We Should Worry)* (Berkeley: University of California Press, 2012).

a result, the SEO industry de-
veloped new tactics, and Google
responded by preventing results
stemming from behaviors that
they deemed malicious or duplic-
itous in an ever-evolving game of
whack-a-mole. Rather than tak-
ing a direct editorial path (which

"SEO does not require an attack-
er to hack into a search engine
and alter the code; it simply
requires an attacker to alter
the information landscape that
the search engine depends on."

Google was explicitly designed to avoid), Google began
tweaking its page rank algorithm in ways that shaped the
platform as a whole. Moreover, Google chose not to reveal
how their page rank algorithm evolved after the original
academic paper, declaring it a proprietary algorithm and ar-
guing that making this public would allow attackers to gain
new knowledge to exploit. While both security engineers
and marketers consistently seek to reverse-engineer the al-
gorithms underlying search engines, updates and changes to
the system regularly thwart their efforts.

Engineers who work on these systems see themselves as
playing cat-and-mouse games with people seeking to ma-
nipulate the system for their own interests. That said, the rise
of the SEO industry revealed how technical systems could
be exploited through a combination of social and technical
practices. While cybersecurity often focuses on how attack-
ers can achieve unwarranted access or bring down a system,
SEO has long shown that it is essential to also examine how
attackers identify and exploit vulnerabilities in the design
and deployment of a data-driven system without ever pene-
trating the technical architecture. Moreover, what's notable
about SEO and its evolution is that the most powerful
exploits require understanding not just how the system is
designed, but also how human users might try to use the
system in the first place. In the case of search engines, the

ecosystem of human searchers, human-produced content, and algorithmic systems introduces numerous vulnerabilities that attackers – or media manipulators – can exploit. SEO does not require an attacker to hack into a search engine and alter the code; it simply requires an attacker to alter the information landscape that the search engine depends on.

Today, many users approach platform-specific search features as though they are search engines for the entire web. Young people often treat YouTube as their primary search engine when seeking general information.[6] Journalists now take to searching on Twitter and Facebook in response to breaking new topics.[7] As platform-specific search features have grown in significance, platform-specific SEO has also proliferated. Yet the amount of data available on these specific platforms pales in comparison to what sites like Google or Bing can provide, increasing the likelihood that people will encounter problematic or manipulated content.

The same data used for platform-specific search is also used for platform-specific recommendation systems, including YouTube's Auto-Play and Twitter's Trending Topics features. Platform-specific SEO focuses not solely on optimizing terms for search, but also terms that can shape the adjacent recommendation systems. Paralleling the field of SEO, attempts to influence recommender systems are sometimes referred to as recommendation engine optimization

---

6   Pew Research Center has examined aspects of YouTube use. See: Aaron Smith, Skye Toor, and Patrick Van Kessel, "Many Turn to YouTube for Children's Content, News, How-To Lessons." November 7, 2018. https://www.pewinternet.org/2018/11/07/many-turn-to-youtube-for-childrens-content-news-how-to-lessons/

7   Muck Rack conducts research into the state of journalism. As of their 2019 report, they find that Twitter is the most popular social media site and that a large number of journalists turn to Twitter first. See: *Muck Rack, The State of Journalism 2019,* July 2019, https://info.muckrack.com/stateofjournalism.

(REO). For example, a media manipulator might try to create connections between popular videos, like movie trailers or music videos, so that someone who is watching a popular video might be recommended to watch the media manipulator's video. Anywhere that a search or recommender system makes decisions based on public data, there is an opportunity for determined, data-literate manipulators to influence other users' exposure to content.

There is a fine – and often fuzzy – line between appropriate and abusive SEO and REO. Search engines want the most relevant content to appear first. Thus, if Delta Airlines produced a website with no meaningful metadata and no other site linked to the airline's website, a query for "delta" might be more likely to return mathematical concepts and references to sororities. While this may be what some users are looking for, it's more likely that they're looking for the airline. As a result, search companies expect content creators to produce metadata and other SEO signals that are well-structured for search engines to use and return information that users want. Yet the same techniques that search engines expect of companies like Delta are also deployed – and often in much more sophisticated ways – by people determined to get their content in front of someone, regardless of the searcher's interest or intention. Being the top result on a search engine can be profitable and influential, which is important to both Delta and media manipulators. As a result, SEO and REO have long been a site of contestation. What matters most to search engine creators, however, is that users value the results they get – both organic results and advertising content. The metrics they use to assess this often align with business interests. Users grow frustrated when SEO manipulation reduces the quality of results. Social media platforms, including those with robust search

features, are in a more precarious place; they are often trying to create a more delicate balance between users' interests and the interests of content creators, including advertisers. Their economic and organizational interest is often to get people to spend more time on the platform. These different logics and priorities create different vulnerabilities that can be exploited.

From Voids to Vulnerabilities

Data voids exist because of an assumption baked into the design of search engines: that for any given query, there exists *some* relevant content. But this is simply not true. When search engines have little available content to return for a particular query, the "most relevant" content is likely to be low quality or problematic or both. With low-data queries, not only are search engines less likely to statistically predict what the user is looking for, but there might simply be no high-quality content to return. Data voids can easily be manipulated, offering problematic content in a range of situations that can cause serious harm. In this report, we focus on five types of data voids that are currently being manipulated by those spreading conspiracies or hate:

1. Breaking News
2. Strategic New Terms
3. Outdated Terms
4. Fragmented Concepts
5. Problematic Queries

There are ways for search engines to work against these forms of manipulation, which we discuss in the final section below. But to understand the possible solutions, we first need to understand the specific type of vulnerabilities that allow data voids to be exploited in the first place.

DATA VOID TYPE #1: BREAKING NEWS

One of the most obvious and high-impact types of data void are those that fill up in rapid response to a breaking news incident. When news is breaking, journalists and other content creators produce new material that must be integrated into search engines. At the same time, a wave of new searches that have not been previously conducted appears, as people

use names, hashtags, or other pieces of information to seek new information. Given the weight of journalistic content and the flood of search queries, media manipulators often seek to capitalize on breaking news in order to influence public perception.

Consider what happened on November 4, 2017. On that Sunday, many people in the United States picked up their phone to a notification that an active shooting was underway in Sutherland Springs, Texas. As the day unfolded, the public learned that a disgruntled white man had walked into a Baptist church and opened fire on worshippers. But the notification that people received on their phones did not provide additional information; the only specific information was the location. With no additional information, people began searching in unprecedented numbers for "Sutherland Springs Texas."
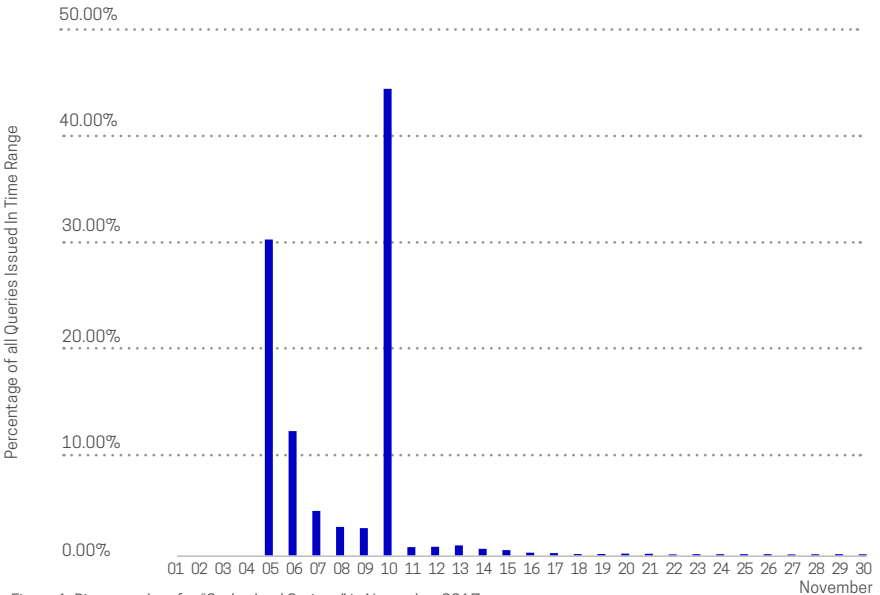

Figure 1: Bing searches for "Sutherland Springs" in November 2017.

As best we can tell, no one had searched for this town for years. Moreover, there was almost no information on the web about this town. To understand the dearth of information at that point, consider the search results for "Harrold-Oklaunion" (another small town in Texas) in summer 2019. Both Bing and Google promise you tens of thousands of results, but the information you see is primarily algorithmically generated content provided by services like Accuweather. com, City-Data.com, Yellowpages.com, and Acrevalue.com. There is a smattering of links to Wikipedia entries, news stories, and court records, but the "most relevant" content provides little information in a breaking news context. Data voids like this are not vulnerable until something happens.
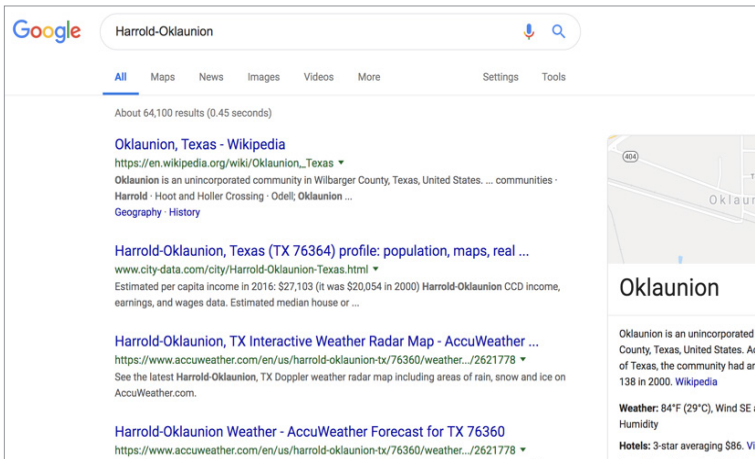

Figure 2: Until the shooting, a search for "Sutherland Springs Texas," returned results that mirror Harrold-Oklaunion's. This was a data void, but one with no consequence until the attack.

What's notable about queries like these is that there *is* content, but it is placeholder content. There is nothing but upside for companies like Accuweather and Mapquest in producing a page for every town in the country. Moreover, many users appreciate the data-centric information provided by these sites as well as those that are generated on Wikipedia, where many US towns have an associated stub article that was initially created by algorithmically combining census data along with other sources. Given the long-standing and well-received status of sites like Accuweather, Mapquest, and Wikipedia by users, it is unsurprising that these algorithmically generated pages rank highly in a context where there is no other content. Yet these individual pages are not so highly ranked so as to not be overtaken by newly created content that appears to be less formulaic during a breaking news situation.

A Sutherland Springs-style data void is easy to manipulate because automated map and weather data pages are easy to surpass in relevance. Without a breaking news situation, such manipulations would have an audience of none; after all, the void exists because no one is searching for these terms in the first place. But this can all change when a news event occurs. During a breaking news situation, journalists know that they need to produce content quickly in order to get information in front of all the sudden searchers.[8] In the process, their articles, blog posts, social media updates—all of these also generate new data for search engines to use to update their results. The void begins to fill up. Unfortunately, the time between the first report and the creation of massive news content is when manipulators have the largest opportunity to capture attention.

8   Robyn Caplan and danah boyd discuss the interplay between search engines and news media in "Isomorphism Through Algorithms: Institutional Dependencies in the Case of Facebook," Big Data & Society 5, no. 1 (February 2018): https://doi.org/10.1177/2053951718757253.

Shortly after the announcement of the shooting in Sutherland Springs, a distributed network of people began coordinating on various forums in an effort to shape media coverage and search engine results. They were driven by a political agenda to influence public perception about this shooting. They first targeted Twitter and Reddit, knowing that search engines like Google and Bing elevate content from these sites when no other material is available. To increase the likelihood of visibility of their content on search engines, they tweeted and posted content that includes words and phrases related to the incident in the early moments before higher-authority content (news media) appeared.

At the same time, these manipulators attempted to influence journalists by using a series of "sock puppet" (inauthentic) accounts on Twitter to ask journalists about the shooter or drop hints to send journalists in the wrong direction. In this case, they tried to encourage journalists to consider whether the shooter might be associated with left-leaning groups by asking questions or pointing to misleading social media posts. While their primary goal was to influence news coverage, this tactic also helps waste journalists' time.

Even when journalists are aware of manipulation, the limitations of search engines can create powerful pockets of low-quality information. Quickly after the Sutherland Springs shooting, a reporter at *Newsweek* uncovered the coordinated attempts to manipulate the story. He wrote a scathing article detailing the manipulation with an unfortunate headline: "'Antifa' Responsible for Sutherland Springs Murders, According to Far-Right Media." This headline causes harm because both Google and Bing chop long news headlines when displaying them at the top of search. Thus, in the first critical hours, most searchers were shown the truncated phrase: "'Antifa' Responsible for Sutherland

Springs Murders…" Eventually, *VICE* ran a headline "No, the Sutherland Springs Shooter Wasn't Antifa" and Snopes created a report on the topic. However, in a breaking news context, the earliest frames can significantly influence public perception.[9] Furthermore, even headlines intended to negate rumors can help spread them.[10]

Breaking news situations are typically time bound as a news story runs its course. While search engines saw a flood of queries related to Sutherland Springs in the hours and days following the shooting, few people search for that term today. Because of the wave of high-quality content produced after the shooting, this kind of data void is cleaned up naturally, but the damage created by the data void was done in the hours after the shooting when search engines amplified content designed to manipulate public perception. Media manipulators' ability to sow chaos by amplifying problematic frames during a breaking news situation is a vulnerability that challenges search engines to this day.

DATA VOID TYPE #2: STRATEGIC NEW TERMS

Manipulation attempts like the Sutherland Springs–Antifa effort are focused on short-term disinformation and increasing chaos at the time of breaking news, but there are other techniques for exploiting data voids that attempt to establish a longer-running narrative. One such approach involves the strategic creation of new terms to divert discourse

---

9   In the field of mass communication, there is a notion of "frame theory," which refers to the agenda-setting power of media frames. Much of this work is rooted in Erving Goffman's 1974 book *Frame Analysis: An Essay on the Organization of Experience* (Boston: Northeastern University Press, 1986).

10   In psychology, this is referred to as the "boomerang effect." This term was coined in 1953 by Carl Hovland, Irving Janis, and Harold Kelley in *Communication and Persuasion: Psychological Studies of Opinion Change* (Westport, Conn: Greenwood Press, 1953).

and search traffic alike into areas full of disinformation. This technique's focus on specific terms not only preys on the infrastructures of hashtags and keywords that exist on social media, but it echoes a longstanding political PR strategy for reshaping public debate. When combined with a breaking news situation, this type of data void can be especially damaging.

To understand this type of data void, consider what unfolded after the horrific 2012 shooting at Sandy Hook Elementary School in Connecticut that took the lives of 20 children and six educators. Long before the Sandy Hook massacre, many conspiracy theorists had propagated false narratives whenever mass shootings occurred, typically implying that the shooting hadn't occurred or that it was a state-sponsored activity. In the hours and days following this tragedy, members of well-established conspiracy forums produced hundreds of posts attempting to debunk the shooting. They believed, falsely, that no shooting occurred, that the news story was manufactured by the Obama administration to justify restrictions on guns, and that distraught parents were faking their emotions on national TV. In their conversations, they settled on the conspiratorial ideal that the parents and kids appearing in news footage of the shooting's aftermath were actually paid actors. They labeled them "crisis actors" and began mobilizing around this term.

Prior to the creation of this conspiracy, the term "crisis actor" referred to a job in which trained actors or volunteers would play mock victims during disaster simulations to help train first responders. While there was some web content referring to those simulations, few people searched for this term and few people created web content referencing this job. Thus, this term was ripe for manipulation.

Immediately following Sandy Hook, conspiracy-minded media manipulators began creating websites dedicated to "crisis actors." They began commenting on news articles to label people as "crisis actors." A highly visible conspiracy theorist began using this term in podcasts and YouTube videos. Others created websites about this issue. After the news cycle faded, these conspiracy theorists continued to produce content talking about "crisis actors" in order to create a network of information associated with the term and its conspiratorial logic. Some edited Wikipedia entries about new shootings in an attempt to legitimize these conspiracies, engaging in edit wars when Wikipedia's editors attempted to debunk this conspiracy theory.

Although there was a spike in searches connected to this term associated with every shooting, this concept did not break through into national news coverage until the Marjory Stoneman Douglas High School massacre that took place in Parkland, Florida, in 2018. At the height of news coverage, CNN anchor Anderson Cooper asked survivor David Hogg on national TV if he was a "crisis actor." This was intended to allow Hogg to deny the conspiracy theory, but it ended up breathing life into it. More news outlets – and news comedy shows – started using the term. And the more that the term was used in the media, the more people searched for it.

When they searched, they found the conspiratorial content that had been staged over multiple years. Even though there was some content designed to debunk the conspiracy, conspiratorial content was highly ranked in web searches and in searches on platforms because it had been there for years and because the network of content surrounding it was highly optimized for search engines and recommender

systems. With no major news coverage or other authentic sources using this term, the conspiratorial content came up first in nearly every search context until debunking videos started overtaking the results. But even debunking videos helped spread this particular conspiracy.

As of August 2019, searches for information about parents whose children were murdered in the Sandy Hook shooting returned conspiratorial content; the top hit for "Robbie Parker" on YouTube offers a video that claims he's a crisis actor because he smiled at one point. The comments are filled with conspiratorial narratives. In response to all that has unfolded, some parents whose children were murdered have sued the most well-known conspiracy theorist; at least one parent died, in part because of the harassment experienced after their tragic loss.

In the online communities that organize these types of manipulation efforts, there is frequent discussion of the political strategy work of Frank Luntz. In the 1990s, Republican pollster and "public opinion guru" Luntz became famous for using focus groups and polls to develop pithy phrases that would reframe political concepts. Many of his terms are part of the contemporary lexicon, including "climate change," "death tax," and "partial-birth abortion." What made him successful as a political operative was his ability to create catchy phrases and convince elected officials to repeat these terms until the news media helped spread them across the country. In effect, he created a linguistic drumbeat that

"... media manipulators who create strategic phrases are not necessarily looking to get a new term to stick; they are more interested in getting people to search for these phrases and encounter the web of information that they have produced by exploiting data voids before those data voids are cleaned up."

pushed these terms into public life, relying heavily on the news media to amplify his terms.

While Luntz expected his phrases to do political work directly, media manipulators who create strategic phrases are not necessarily looking to get a new term to stick; they are more interested in getting people to search for these phrases and encounter the web of information that they have produced by exploiting data voids before those data voids are cleaned up.

Measuring the impact of strategic terms is difficult. What we do know is that those who engage in this tactic celebrate in online forums when journalists pick their terms up. We also know that participants in various hate-oriented and conspiratorial forums reference these terms in describing how they found the forums in the first place. One of the clearest descriptions of this came from a white supremacist mass murderer in 2015. Before this young white man attacked and killed nine people (injuring one other) at the Emanuel African Methodist Episcopal Church in Charleston, South Carolina, he produced a manifesto describing how a targeted data void introduced him to extremist content. This manifesto included the following passage (*emphasis ours*):

> The event that truly awakened me was the Trayvon Martin case. I read the Wikipedia article and right away I was unable to understand what the big deal was. It was obvious that Zimmerman was in the right. But more importantly *this prompted me to type in the words "black on white crime" into Google, and I have never been the same since that day*. The first website I came to was [hate site]. There were pages upon pages of these brutal black on white murders. I was in disbelief. At this moment I realized

that something was very wrong. How could the news be blowing up the Trayvon Martin case while hundreds of these black on white murders got ignored?

In other words, one search query of a strategic term led him to a data void shaped by white nationalists.[11] Media manipulators who embraced white nationalism and white supremacy took credit for intentionally placing the phrase "black on white crimes" on various websites, including on Wikipedia articles, to affect SEO. Yet, unlike "crisis actors," the term "black on white crimes" never made it to mainstream news sources. Even without that more common source of amplification, this media manipulation campaign influenced this particular terrorist enough to encourage him to spend time learning more about white nationalism.

DATA VOID TYPE #3: OUTDATED TERMS

Strategic terms can be created to fill data voids, but data voids can also emerge when terms stop being regularly used. These data voids don't experience spikes of attention like breaking news or strategic terms. Instead, they have long lives in the gaps of search engines. As search engines respond to new trends and new words, old terms can be left behind, associated only with outdated content, leaving more and more room for manipulators.

---

11   In her 2018 book *Algorithms of Oppression*, Safiya Noble also discusses this case. While she centers her argument on how algorithms directed this terrorist to this content, our concern is that this data void was strategically manipulated by adversarial forces and then undetected as a vulnerability before this atrocity. Safiya Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press, 2018).

Search engines are programmed in an effort to balance between content that is recent and content that is "authoritative."[12] All search engines – including Bing and Google as well as those features on YouTube, Twitter, and Facebook – respond to breaking news events, and will deliver a greater proportion of recent content for certain terms if they experience an algorithmically recognizable surge in content and search attempts. They do so because they are designed to presume this is what a searcher wants. Meanwhile, when users search for terms that are more consistent and associated with much older but highly authoritative content or more established content creators, search engines are more likely to return authoritative – "evergreen" – content alongside news content. Moreover, even during a flu outbreak where there may be a surge in flu news content and an increase in people conducting flu-related queries, authoritative medical content is still likely to be among the top results because of the cyclical nature of these types of queries.

There are many cyclical patterns in both new content and search queries. For example, new content associated with the Academy Awards (a.k.a., "the Oscars") tends to start emerging in January of each year, with more and more content coming online until late February, at which point there is a peak in both new content creation (e.g., news stories, tweets, etc.) and search queries. And then interest in that topic dies off for another year. Many search-relevant topics have seasonality, ranging from elections to major sporting events. Generally speaking, both content creation and search queries rise alongside each other for rhythmic search topics. Thus,

---

12   Within search engine discussions, "authoritative" content refers to content that comes from verified sources, has a long history, and/or has significant linkage patterns from other known sources. This measure was adopted to prevent people from creating a new website with spam-style linkage patterns from auto-generated "spam" sites as part of early SEO attempts.

while it might be possible to coordinate a large-scale search engine optimization campaign around "the Oscars" in July, the number of people then searching for this is small.

Other terms are more faddish in nature. That is, they are used heavily for months or years as they catch on and then quickly fade from use as they fall out of fashion. Consider a term like "social justice." This term still appears in textbooks and is used by many companies and organizations. But many of those who initially created content around this term – including the activists, institutions, and community groups who are working toward social justice – currently prefer to use more narrow and specific terms to describe their work, such as "racial justice" or "economic justice." As a result, there is a decline in new content that is created among people seeking equality that explicitly talks about "social justice." The lack of new content with this term allows a data void to form.

Google and Bing have less difficulty with a query like "social justice" because of the sheer quantity of content they have related to this term from authoritative sources. Both return Wikipedia content, dictionary content, and a range of results from sites focused on defining the term. Of course, as of summer 2019, Google also returns a politically oriented result from an organization that is hostile to "social justice" efforts, but that result is a well-researched 2009 article outlining a different way of understanding the term.

Bing and Google have significantly different results for a query like "social justice warriors" versus "social justice." "Social justice warriors" has become a pejorative label for progressive activists regularly used by political conservatives and other reactionary activists to undermine the meaning of social justice and the people committed to it. Given the very distinct patterns of search and content creation associated

with these two terms, they do not regularly collide on Bing and Google. The closest overlap between these two distinct branches of content is that "social justice warrior" is an auto-complete option for those typing in "social justice."

YouTube faces a different challenge. Searches on its platform only return links to YouTube videos. It has a limited back catalogue of content and it prioritizes new content. Furthermore, the people who currently produce new content associated with "social justice" on YouTube are primarily doing so in an antagonistic fashion, often systematically associating social justice with social justice warriors to undermine the legitimacy of the social justice concept. In other words, those who oppose feminism, racial justice, and economic justice have exploited this data void on YouTube to reclaim and twist the term. Additionally, they have created significantly more content associated with "social justice warriors," and the structural factors that limit the collision of these terms on Bing and Google do not exist on YouTube. Starting around 2015, we began noticing that those seeking to alter the meaning of social justice were successful in shaping the results for queries like "social justice" on YouTube.

A data void produced by outdated terms doesn't have to be strategically manipulated to be dangerous. Consider the term "Negro" in English. Although some Black individuals use this term with pride, much new content funneled into search engines using this term is pejorative. Moreover, a lot of historic content that contains this term is racist in nature. For example, the top result on YouTube in summer 2019 is a video of President Reagan talking about how his administration has addressed discrimination, in part by helping "Negro" private colleges and universities. The comments on the video are a mix of explicitly racist comments and a "debate" about whether Reagan's use of the term was racist.

In effect, the video itself stands as a historic record, but the debate in the comments is where the introduction to more racist ideologies occurs.

Data voids in this category emerge when a disconnect occurs between content creation and search queries. Content creators typically leave behind words faster than people stop searching for them. They typically do not produce content that connects the old terms to the new terms in ways that allow a searcher to understand the development of language. Moreover, if people are searching for a term where no new content is being produced, media manipulators can easily fill in the gap and use this disconnect to their advantage.

> "Content creators typically leave behind words faster than people stop searching for them. They typically do not produce content that connects the old terms to the new terms in ways that allow a searcher to understand the development of language. "

DATA VOID TYPE #4: FRAGMENTED CONCEPTS

Media manipulators can exploit data voids by developing new terms or reclaiming forgotten terms. In doing so, they look to channel people's attention to specific terms through other popular channels. But manipulators can also work to intentionally separate manipulated content from more popular content through the creation of distinct terms that are too fraught to connect. The terms people search for matter.[13] Because search engines regularly seek to connect synonyms (e.g., "poop" and "stool"), many manipulated narratives can be overwhelmed by a sufficient amount of high-quality

---

13   Francesca Tripodi documents a range of different fragmentations in search terminology in her 2018 report. Tripodi, *Searching for Alternative Facts.*

mainstream information. In such cases, manipulators might work to prevent connective bridges from being created in a search engine's model. Media manipulators can also exploit search engines' resistance to addressing politically contested distinctions to help fragment knowledge and rhetoric.

When large communities of people approach a news item with different political frames, the result can be a "naturally" fragmented model of search results—two largely unconnected sets of results. In many cases, however, changes in coverage and search behaviors lead to bridging this fragmentation. For instance, in late summer 2018, a scandal concerning the Vatican and sexual misconduct was front-page news. As the news was breaking, a search for "Vatican sexual abuse" returned entirely different results than a search for "Vatican pedophiles," especially on YouTube. These results were fragmented because the people producing the new content, as well as the people doing the searches, had different ideological commitments. But over time, the results converged on web search engines. This occurred partly because journalists began referencing both phrases in their news coverage and partly because the term "Vatican" became the more dominant anchor.

This convergence occurred over time with no intervention because distinct linguistic communities began using each other's terms enough that the system detected them as being of the same kind. Yet, in the process, content associated with "Vatican sexual abuse" ranked higher, because those sites received more interactions and had other signals that suggested they were of higher quality. In the process, the "pedophiles" frame was drowned out.

Both manual and algorithmic efforts by search engines to bridge concepts and information can be controversial. For example, when someone searches for "illegals caravan," search engines might return results that include references to both "undocumented" people and "illegal aliens." Algorithmic systems may associate this concept with the terms "migrant" and "immigration" more broadly. If searchers do not intentionally restrict the system to exact phrase matching, it is quite likely that content using other phrasings might appear in the top results given the high quantity and quality of sources using other terms.

For politically charged content, any decision by search engine designers becomes political itself. Collapsing ideologically split terms might be the automatic response of a search model to the news content produced, but this technique might also be critiqued for mixing together such diametric terms. At the same time, if search engines intervene manually to shape the model for a specific set of terms, they risk accusations of bias by other political communities. Most recently, there has been sustained scrutiny of search engines and social media for alleged "anti-conservative bias" in the wake of numerous controversial news topics.

Because of charges of "anti-conservative bias" by US political figures, search engine designers are especially loathe to help associate synonymous terms that cross political lines. At the same time, media manipulators can systematically work to make certain that new content does not connect concepts, to make sure that fragmented concepts remain fragmented. What is emerging as a result are distinct clusters of information that are neither manually nor algorithmically bridged. Depending on the terms users put in a search query, they can end up in an entirely different sphere of

information than others who seek information on similar topics using different terms.

DATA VOID TYPE #5: PROBLEMATIC QUERIES

Those who design and run search engines know that users enter countless messy, emotional, and unpredictable search queries: "Who should I ask out on a date?" "fever rash black tongue." The queries people construct often reveal anxieties, fears, and intimate thoughts. Some of what people seek to know is deeply disturbing or taboo. Even more concerning are those who produce information for people who might conduct such disturbing or taboo queries.

There are many search queries with disturbing language that have historically returned problematic results. Some, such as "did the Holocaust happen," have been addressed by a concerted effort by both content creators and search engines.[14] Yet, these queries are often tricky. Initially, those who produced factual content about the Holocaust did not produce content that included the kind of language contained in this query. When content creators were informed of the concerted effort by Holocaust denial conspiracy theorists, they contributed to addressing the problem by creating new content. But conspiracy theorists consistently produce newer content and seek to optimize their content in new ways, creating a challenge for those who are producing factual information. While Google and Bing have been

---

14   Deirdre K. Mulligan and Daniel S. Griffin argue that Google's response to "did the Holocaust happen" fails to get to the root of the problem. They suggest that search engines need to make more fundamental changes to address these kinds of problematic queries. Deirdre K. Mulligan and Daniel S. Griffin, "Rescripting Search to Respect the Right to Truth," *2 Georgetown Law Tech Review* 557 (2018). https://georgetownlaw-techreview.org/rescripting-search-to-respect-the-right-to-truth/GLTR-07-2018/.

somewhat successful at addressing this particular query, YouTube consistently has too little data in this arena, which gives conspiracy theorists an easier path to outrank factual creators. Furthermore, as previously discussed, many conspiracy theorists, hate groups, and media manipulators attempt to push searchers to use specific, problematic search queries they know will lead to these voids. Unfortunately, many Holocaust deniers produce videos and talk on the radio, encouraging listeners to search for specific new phrases in order to circumvent the content optimized to address the original problematic query.

Fact-checking websites have traditionally been more concerned with establishing a center of facts than debunking every specific conspiracy theory. And those groups focused on debunking, like Snopes, have only been able to take on those pieces of misinformation that rise to certain popularity. Most other content producers do not think to devote time to producing content designed to debunk potential conspiracies. Given the absence of inoculation content, conspiracy theorists used SEO to fill up data voids

> "Only in recent years have historians, archivists, and other web producers started recognizing that they must make certain that their content is optimized for problematic queries as well as more common ones."

so that conspiratorial content would appear when someone articulated a disturbing query. Historians simply posted evidence, not even thinking to optimize their content for someone who might have considered a conspiratorial frame or to bridge their content to be viable for problematic queries. This created a significant distinction between a query like "Holocaust" and a query like "did the Holocaust happen?" Only in recent years have historians, archivists, and

other web producers started recognizing that they must make certain that their content is optimized for problematic queries as well as more common ones. Through the systematic creation of new content and efforts by search engines, this particular query has become less problematic because new, high-quality data is available. Yet, there are countless other problematic queries and new conspiracy theories that search engines struggle to combat.

While some media manipulators produce original content to optimize for problematic queries, others focus on increasing the visibility of content produced for other reasons. Take a query like "female pedophiles." Search results for this query swing from highlighting academic articles on gender and sexual predation to conspiratorial content suggesting that women are the "real" sexual predators. Yet the bulk of the results come from individual news articles which, when presented in aggregate, give the impression of an epidemic. The effect of this effort is particularly notable in an image search, which reveals countless mugshots of white women. This data void was intentionally exploited by those seeking to associate sexual misconduct with women. To achieve this goal, they focused on getting individual journalists to cover individual cases so as to build up a corpus of content, rather than a single story. To those perpetuating this conspiracy, search offers evidence. In forums, where this topic comes up, nonbelievers are told to "just do a Google image search" as a way of showing that the conspiracy is true. As with strategic terms, people are also encouraged to conduct problematic searches like "female pedophiles" in order to encounter content that was staged. Those who search for this and dig in deeper can find information that debunks this frame, but a quick search does not provide the context necessary to undo a conspiratorial frame that is staged elsewhere.

Problematic queries can come in many forms and reveal that the content creators interested in countering misinformation must not only think about the paths people take to reach positive content, but also the paths people take to reach problematic content. By associating positive content with problematic queries, content creators can help limit the isolating quality of certain data voids. In order to combat data voids that exist because of problematic queries, both content creators and search engines need to work together to identify disconcerting paths and produce content that is valuable for those who are not explicitly seeking out problematic content.

Data Voids in Search-Adjacent Recommender Systems

The data voids that emerge through search engines are sometimes reinforced by adjacent features that function more like recommender systems. Recommender systems are the various features that encourage users to consume additional content by making recommendations or pushing new content to consume. While the most well-known recommender systems are on media consumption platforms like Netflix or Spotify and e-commerce sites like Amazon, recommender systems are also integrated into search engines and social media platforms in ways that support and reinforce search functionality. Because search-adjacent recommender systems often rely on many of the same sets of data that undergird search engines and features more generally, they can often compound existing vulnerabilities and open up new vulnerabilities for media manipulators to exploit.

In order to better understand how these search-adjacent recommender systems further complicate the data voids problem, we are going to explore two contexts in which this problem emerges:

- Search bar auto-suggestions
- YouTube's next/auto-play

SEARCH BAR AUTO-SUGGESTIONS

On both Bing and Google (as well as most platform-specific search engines), when a user starts to type a query, the system attempts to complete their thought by recommending a range of possible queries that begin with the letters already typed. Theoretically designed to help users with spelling or minimize how much they must type (which is especially valuable on mobile), this feature encourages users to add

additional words to their query which, in turn, helps search engines produce more accurate results.

Auto-suggestions are generated based on previous queries from users and help reveal fragmentations in language patterns. Thus, if a user starts a query with "subway," it's not surprising that auto-suggest includes both "subway menu" and "subway near me." Choosing one of those helps the search engine narrow the result. While this may be helpful, auto-suggest also introduces new phrases users might never have considered. For example, amidst the auto-suggestions about food and transit, seekers may be given "subway surfers game." Some number of prior users were looking for this particular mobile phone game. Yet, in effect, a query that might begin with seeking information about the sandwich shop may lead a user to take up a new game.

> "Before search engines started overhauling their auto-suggestions around phrases like 'women are' the results ranged from offensive to terrifying. The fundamental problem was a data void."

Unfortunately, auto-suggestions can also send people down deeply disturbing paths. Before search engines started overhauling their auto-suggestions around phrases like "women are" the results ranged from offensive to terrifying. The fundamental problem was a data void. The amount of non-problematic user data that began with this phrase was almost nonexistent. As a result, toxic auto-suggestions popped up, prompting people to click on them and increasing the signals that search engines received, which suggested that these were the desired auto-suggestions. Search engine companies have responded by trying to minimize such dangerous queries, but media manipulators often look to exploit these data voids to encourage certain auto-suggest results. They may attempt to get a distributed group of

people to search for a specific term or even create automated systems ("bots") to do this for them. Unlike SEO, they aren't focused on encouraging people to search new phrases. Instead, they are working to extend commonly searched phrases with additional words that could lead users to intentionally produced problematic content.

Auto-suggest data voids require a different intervention than data voids in search engines. Rather than needing higher-quality content to fill in the data void in search, auto-suggest data voids are typically addressed by understanding and limiting the possibility of auto-suggest on more problematic topics. While this is effective in some contexts, it is a constantly evolving problem because of iterations in language and the efforts of media manipulators.

YOUTUBE'S "UP-NEXT" AND AUTO-PLAY FEATURES

While YouTube is best known as a social media platform, it also functions for many of its users as a search engine that is enhanced through a recommendation engine. When a user searches on YouTube, they are given results for videos on the platform. Once a user clicks on one of these links, they are then shown a page for that individual video. On the side of the page is a list of thumbnails under the label "Up Next," which users can click at any time. And once a video ends, users are automatically shown the next video as a part of their "auto-play" feature.

As previously discussed, search results on YouTube reveal and perpetuate numerous data voids, in no small part because the amount of content available on YouTube pales in comparison to Bing and Google, which can return anything on the public web. Yet YouTube's auto-play and

recommendation features make visible another genre of data voids. Recommendations are not a result of a user-centered search. Rather, they functionally serve as a machine-driven search for content that is similar to the content currently being consumed or that may otherwise appeal to the particular user. The designers at YouTube are incentivized to keep people on the system for as long as possible because of their advertising model and internal metrics.[15] As a result, YouTube's system creates both recommended videos and auto-plays after every video on their site. They accomplish this not by extrapolating from user search queries, but from a whole set of signals: what previous users watched after they watched a given video, what videos are from the same creators, what videos include comments from or are liked by the same people, and what other videos include similar metadata or description text. They also draw on the historical viewing patterns of a user to deeply personalize the recommendations.

YouTube allows creators to monetize their videos, and the advertising revenue that results is shared between the company and the creator. Users who are seeking broad audiences are incentivized to optimize information surrounding their video so that it might be recommended to users watching something else. For example, a lesser-known musical artist might want to make sure that their music video is recommended after someone watches a more famous band's music video. To encourage these connections, those invested in recommendation engine optimization might ask their viewers to help them purposefully engage in activity on YouTube that could influence its systems. In other words, recommendation engine optimization parallels search engine optimization.

---

15   To better understand YouTube's history, business, and role within the ecosystem of content creation, see: Jean Burgess and Joshua Green, *YouTube: Online Video and Participatory Culture,* 2nd ed. (Cambridge: Polity Press, 2018).

The specifics may differ, but the fundamental goal is the same: make certain the algorithm ranks your content well.

Manipulators exploit these same structures to associate problematic content with popular content. They use metadata, tagging, commenting, and other tools in an attempt to connect videos together. Powerful networks of influencers co-host one another on their YouTube channels, which creates the signals for YouTube's algorithms to recommend someone watching one influencer to consume content from another. For those seeking to introduce someone to an extreme viewpoint, being hosted for a debate by a less-extreme influencer is a boon.[16]

In order to better understand how media manipulators have effectively exploited YouTube's recommender system, consider the effort by anti-vaccination conspiracy groups. Although the tenets of this conspiracy have been systematically debunked and their activism has resulted in measles outbreaks and deaths, many who espouse anti-vaccination views believe either that vaccinations are dangerous and/or that the scientific evidence is questionable. Coordinated networks of believers of this conspiracy have actively targeted social media to convert new parents and seed doubt in public health efforts. On YouTube, they actively seek to associate their videos with health videos. While much of their factually inaccurate content has been removed as disinformation, they continue to create content that asks questions or conveys stories of distraught parents. Because government and health professional content is more highly weighted for search queries related to vaccination, content

---

16   Becca Lewis describes the formation of one of these networks in *Alternative Influence: Broadcasting the Reactionary Right on YouTube* (New York: Data & Society, 2018). https://datasociety.net/output/alternative-influence/.

produced by the Center for Disease Control and other similar organizations is almost always at the top of the search results on YouTube. Yet, because of the concerted efforts by anti-vaccination groups (and the ineffective SEO and REO efforts by medical professionals), anti-vaccination videos often appear as recommended videos that follow scientific videos. The SEO/REO strength of a group with an ideological agenda, combined with the weakness of groups who are providing factual information, makes YouTube's system easy to exploit.

Managing Data Voids

Any ranking, rating, or recommender system on the internet is being and will be exploited if media manipulators can benefit from doing so. Search engine optimization is over 20 years old. The PageRank algorithm was an early technical intervention to resist SEO, but those invested in SEO simply evolved to manipulate PageRank and its technical progeny. Major search engines consistently struggle to alter their systems so that they return high quality results under a constant barrage of manipulation attempts. While the sheer quantity of content available on the internet has made Bing and Google fairly resilient to SEO for many queries and in many languages, there are still significant vulnerabilities in this information ecosystem. One notable type of vulnerability – data voids – is regularly exploited by media manipulators determined to shape information in their interests.

While Bing and Google can – and must – work to identify and remedy data voids, many of the vulnerabilities lie at the heart of what these search engines do. Bing and Google do not produce new websites; they bring to the surface content that other people produce and publish elsewhere on third party platforms. Without new content being created, there are certain data voids that cannot be easily cleaned up. The type of data void also matters: Search engines are able to address issues with problematic queries and forgotten terms much more easily than strategic terms or breaking news. Fragmented concepts raise a myriad of more challenging questions for search companies.

Media manipulators also regularly exploit the interplay between news organizations and search engines, between social media sites and search engines, and between large collaborative projects like Wikipedia and search engines. Finding the manipulated signals amidst the large quantity of good signals is a never-ending challenge. Moreover, even

if they can be identified, the content availability problem presents another hurdle. Factually inaccurate information, hyper-partisan content, scams, conspiracy theories, hate speech, and other forms of problematic content are harmful to individuals and societies, but media manipulators have a higher incentive to create such content than those who seek to combat it. Search engine creators want to provide high quality, relevant, informative, and useful information to their users, but they face an arms race with media manipulators. While this report focuses on the dynamics occurring in English on these sites, these problems are likely to be of even greater concern in non-English settings where there is even less data.

Recommendation data voids benefit more from identification because one type of remedy is simpler and less fraught. In effect, some query stems should simply not be recommended to users. Data voids are especially problematic for YouTube, which is working with far less data and using a recommendation engine to encourage users to stay on their site. While they have begun restricting certain kinds of problematic content from their site through their Terms of Service, media manipulators are almost certain to find new ways to undermine their efforts so long as YouTube provides them access to desired audiences. Up-next and auto-play are nowhere near as robust as the contemporary search ranking systems. Given this, it may be imperative for YouTube to simply turn off this feature in certain contexts.

> "… companies have a responsibility to identify these vulnerabilities, iteratively respond to attacks, and support content creators who can produce the high-quality content that is needed to fill these data voids. "

Search engines like Google and Bing face a more difficult set of challenges, in no small part because there is no "fix." This issue must be treated with the same level of seriousness as any security issue. These companies have a responsibility to identify these vulnerabilities, iteratively respond to attacks, and support content creators who can produce the high-quality content that is needed to fill these data voids. This will be especially daunting in contexts where facts and evidence are contested.

Fundamentally, to address data voids, search engines must directly grapple with both the desires of their users and also the practices of media manipulators. Searchers want to get the information they're seeking. Media manipulators want to leverage search engines and search-adjacent recommendation engines to amplify content and get it into the hands of as many searchers as possible, regardless of whether the content is actually what a searcher seeks. What differentiates media manipulators from other content creators is the tactics they use and the goals they have. While search engines wish to be neutral platforms, they are going to increasingly face governance challenges that reveal just how political the project of providing information can be.

Debates over search bias and content moderation are increasingly attracting public attention. These discussions raise significant questions about what role search engines can and should play in making problematic content accessible. By and large, these discussions center on restricting the visibility of certain kinds of content or limiting the amplification of certain types of results. Yet, data voids present a different set of challenges: How can search engines more effectively detect vulnerabilities in search? Who is going to provide viable content for search engine users seeking

information where the current quality of content ranges from mediocre to atrocious? What role should search engines play when someone searches for a data void? And who is responsible for addressing the vulnerabilities at the intersection of different websites, services, and user practices? Even as technology companies increasingly seek solutions to this challenge, the practices of media manipulators reveal that this is not a problem to "solve." Instead, data voids are a security vulnerability that must be systematically, intentionally, and thoughtfully managed.

Notes on Methodology

In conducting this research, both authors have extensively examined a range of specific data voids. We have observed media manipulators in action, tracked the impact of their actions, and supported engineers who are trying to tackle this problem. For the purposes of this paper, we have chosen examples that are well-known or reasonably addressed so as to not contribute to the problem. We have also chosen not to cite specific manipulators by name or amplify known hate sites. The examples we provide are, intentionally, fairly innocuous and simplistic so as to help readers understand the problem; much more work is needed to identify and address much more harmful data voids without aiding and abetting media manipulators.

Author Biographies

Michael Golebiewski

Michael Golebiewski is presently a Principal Program Manager on Bing at Microsoft. In this role he is responsible for content moderation and working to build more responsible and ethical AI. In doing this work he has played a significant role in working to address some of the worst aspects of the internet including online Child Sexual Abuse, Revenge Porn, online censorship, and misinformation. In addition to his role on Bing he consults on cross-company issues including fairness and bias in AI. Michael came to Microsoft in 2008 and since that time has worked on various aspects of Bing. Before joining Microsoft he has a long career in online technology with experience from Real Networks, Parametric Technology Corporation, and Lexis-Nexis. Michael is a graduate of Case Western Reserve University in Cleveland and holds a Masters in Computer Science.

danah boyd

danah boyd is the founder and president of Data & Society, a partner researcher at Microsoft, and a visiting professor at New York University. Her research is focused on making certain that society has a nuanced understanding of the relationship between technology and society, especially as issues of inequity and bias emerge. She is the author of "It's Complicated: The Social Lives of Networked Teens" and has authored or co-authored numerous books, articles, and essays. She is a 2011 Young Global Leader of the World Economic Forum, a Trustee of the National Museum of the American Indian, and a Director of Crisis Text Line. Originally trained in computer science before retraining under anthropologists, danah has a Ph.D. from the University of California at Berkeley's School of Information.

Acknowledgements

DATA & SOCIETY

Data & Society is an independent nonprofit
research institute that advances new frames
for understanding the implications of data-
centric and automated technology. We conduct
research and build the field of actors to
ensure that knowledge guides debate,
decision-making, and technical choices.

www.datasociety.net
@datasociety

DATA&
SOCIETY