Identifying Search Directives on Social Media

Ronald E. Robertson, Amy Dunphy, Shelby Grossman,
Renée DiResta, and David Thiel

Abstract. This study introduces methods for identifying search directives—content that could prompt an online search—and explores their presence on social media. Search directives can be an effective tool for indirect online influence, because instead of guiding people directly to content (e.g., a news article), they indirectly guide people to it through an independent intermediary (e.g., Google Search). By directing viewers to "do their own research" on a known, trusted, or seemingly objective intermediary, search directives have advantages in terms of both persuasion, by making people feel as if they discovered the content on their own, and evasion, by not directly posting the target link(s). Here we present a framework for identifying search directives on social media that includes methods for automated discovery, classification, and query extraction. Data we collected from social media demonstrate the widespread use of search directives, provide construct validity for our definition, and support our framing of search directives as a form of indirect online influence. Last, we report three case studies that highlight the immediate value of identifying search directives to researchers, practitioners, and journalists focused on online trust and safety.

1 Introduction

"To information seekers everywhere: be mindful where the journey leads."
— Francesca Tripodi (2022)

Search engines are a central avenue for discovering and verifying information online. In interview-based studies, interviewees frequently repeat similar phrases when asked about how they seek out information online: "I just Google it" (Toff and Nielsen 2018) and "I Googled it" (Tripodi 2018). These phrases illustrate a level of trust in search engines that is reflected in recent surveys, where people report both obtaining more news from search engines than social media sites (Shearer and Mitchell 2021) and trusting search engines as a source for general news and information more than social media sites (Edelman 2021). Corroborating these qualitative and self-report findings, recent digital trace studies have also often found greater news engagement via search engines than via social media (Allcott and Gentzkow 2017; Guess, Nyhan, and Reifler 2020; Bentley et al. 2019). Together, these studies suggest that web search engines are widely used as trusted independent intermediaries for online content discovery.

© Stanford Internet Observatory, 2023
doi:10.54501/jots.v2i1.133

Given this widespread use and trust in search engines, attempts to influence users' query selection within such intermediaries present a powerful yet understudied form of indirect online influence. While such influence can be exerted in many ways, including broad encouragement to "do your own research" (Tripodi 2022) or strategically repeated *sound bites* in the media (Landtsheer, De Vries, and Vertessen 2008), most relevant here are concepts that describe deliberate attempts to create or amplify short and searchable phrases, such as *strategic new terms* (Golebiewski and boyd 2019), *viral sloganeering* (Donovan and Friedberg 2019), and *keyword signaling* (Tripodi 2019). Of particular concern around such attempts is when these phrases lead to *data voids*: low-quality search results that occur when the results available for a given search query are limited, volatile, or strategically manipulated (Golebiewski and boyd 2019).

In this study, we examine *search directives*—content that could prompt an online search—and investigate their use as a form of indirect online influence on social media. We define search directives within existing linguistic frameworks (Section 3) that allow for a flexible relationship between the intended outcome of a message (e.g., prompting a search) and how it is delivered (e.g., vaguely mentioned or directly suggested) or how effective it was (e.g., causing people to search). For example, to prompt a viewer to search, one can provide a search link (google.com/search?q=chemtrails), suggest in text that viewers conduct search (you could search for "chemtrails"), or provide multimedia content that could prompt a search (e.g., an image that says "google chemtrails"). By directing viewers to "do their own research" on a known, useful, or seemingly objective intermediary (e.g., Google Search), the actors spreading search directives have advantages in terms of both persuasion, by making people feel as if they discovered the content on their own, and evasion, by not directly posting the target link(s).

To examine the use of search directives in real-world settings, we conducted a three-part study that involved a broad exploration, systematic collection and evaluation, and three in-depth case studies. In our exploratory study, we collected and annotated 1.5K search directives from a variety of sources and used these data to identify their lexical boundaries and qualitatively characterize their use in the world (Section 4). Building on our exploratory study, we next developed an approach for automating the discovery, classification, and extraction of linked or text search directives on social media (Section 5). After applying our approach to collect linked and text search directives, we evaluated their use on social media (Section 6), conducted three in-depth case studies on topics relevant to trust and safety teams at both search engines and social media sites (Section 7), and summarized our findings, limitations, and directions for future work (Section 8).

While many search directives are benign or commercially motivated, we show that the tactic is also demonstrably useful for evading platform trust and safety policies. In our case studies, we found data void queries that returned search results containing potential malware risks, financial scams, and misleading featured snippets about alternative medical treatments. For example, we show that one of DuckDuckGo's advanced query operators is being used in search directives posted by an automated network of Twitter accounts to redirect users through DuckDuckGo (skipping the results page) to a video promoting a cryptocurrency scam on YouTube. These findings are likely of particular interest to trust and safety teams at all three platforms involved, and fit within a pattern of findings that suggest search directives can be a broadly valuable signal for proactively detecting indirect online influence campaigns and the networks of actors behind them. Our study provides a foundation for future work on indirect online influence that involves multiple platforms (see Figure 1 on the following page), and raises the need for future work on how search directives may evolve alongside emerging search technologies, such as large language models with interfaces like ChatGPT, Bing, or Bard.

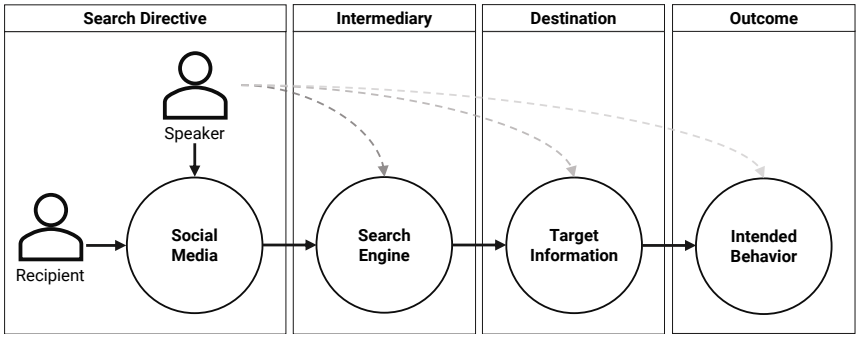


Figure 1: A diagram depicting search directives as a form of indirect online influence. From left to right, a speaker communicates a search directive to a recipient via another platform (e.g., social media), which may prompt the recipient to conduct a search on an online intermediary (e.g., a web search engine), potentially exposing them to content that the speaker intended for them to find (i.e., the target information). Last, the recipient may exhibit the outcome the speaker intended—such as a specific behavior—which could be clicking on a link, making a purchase, or sharing the target information. Though unseen by the recipient, the speaker can influence the information found via the intermediary (by selecting a highly specific search query), and thereby can influence the recipient's destination (e.g., a webpage), and outcome (e.g., attitude change) as well.

2 Background

Here we discuss past work on search engines as trusted intermediaries (Section 2.1), data voids in search results (Section 2.2), and indirect influence via search (Section 2.3).

2.1 Search Engines as Trusted Intermediaries

When search engines filter and rank content from the web in response to a user's query, they not only make that content more visible, but also communicate greater relevance and legitimacy to the perspectives it presents (Gillespie 2017). Some researchers have suggested that search engines are providers of "bent testimony," where "within certain contexts of interactions, users act as if these algorithms provide us with testimony—and acquire or alter beliefs on that basis" (Narayanan and De Cremer 2022, 1). Such findings give cause for concern because people often report a high degree of trust in search engines (Toff and Nielsen 2018; Tripodi 2018; Shearer and Mitchell 2021; Edelman 2021), and lab experiments suggest that search engine rankings can influence users' opinions on important topics, such as politics and health (Epstein and Robertson 2015; Ludolph, Allam, and Schulz 2016; Epstein et al. 2017; Ludolph, Allam, and Schulz 2016).

In recognition of their role as a trusted intermediary, Google solicits feedback on its search results using its Page Quality Rater Guidelines, which includes high standards for topics that carry "a high risk of harm because content about these topics could significantly impact the health, financial stability, or safety of people, or the welfare or well-being of society" (Walls 2022). Such guidelines are needed because many problems have been documented in the quality of search engine results (Sweeney 2013; Noble 2018), and past work suggests that the information discovered while conducting online searches can be persuasive because it gives people the feeling of customizability and control over the information they find (Sundar and Marathe 2010; Fisher, Goddu, and Keil 2015; Fisher, Smiley, and Grillo 2022). Recent work, building on the *IKEA effect*—where people overvalue the things they have built (Norton, Mochon, and Ariely 2012)—suggests an "IKEA effect of misinformation," whereby people are overconfident in information they feel they have discovered themselves, such as via an online search (Tripodi 2022).

2.2 Data Voids in Search Results

A key motivation for identifying search directives is the existence of *data voids*, which are described as "one such way that search users can be led into disinformation or manipulated content," and are said to 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" (Golebiewski and boyd 2019, 2). To combat and raise awareness of such data voids, Google Search places banners at the top of their search results to alert users when the results are (1) rapidly changing (Sullivan 2021), (2) of low-relevance (Tucker 2020), or (3) of low-quality (Nayak 2022).

Golebiewski and boyd (2019) specify five types of data voids—breaking news, strategic new terms, outdated terms, fragmented concepts, and problematic queries—that help inform our study. *Breaking news* data voids are those which "fill up in rapid response to a breaking news incident" (16) and often involve new or uncommon queries that return few results or generic placeholder content (e.g., Accuweather, Yellowpages, Wikipedia). For example, searching "Sutherland Springs Texas" following the 2017 mass shooting in that city led to results that were flooded with content tying it to "Antifa." *Strategic new term* data voids specify "the strategic creation of new terms to divert discourse and search traffic alike" (21) to content that was strategically created to reinforce a narrative. For example, searching "crisis actor" following the 2012 Sandy Hook mass shooting led to established conspiratorial narratives around paid actors in mass shootings.

In contrast, *outdated terms* create data voids that "don't experience spikes of attention like breaking news or strategic terms" but can "emerge when terms stop being regularly used" (26). For example, "social justice" discourse shifted to "racial justice" and "economic justice," leaving a gap in new content for the original term. Next, *fragmented concepts* are data voids occurring for concepts that "are fragmented because the people producing the new content, as well as the people doing the searches, had different ideological commitments" (31). For example, "Vatican pedophiles" returned different results than "Vatican sexual abuse" following the 2018 scandal. Last, *problematic queries* are data voids occurring for queries with problematic language that does not appear on any high-quality and topically relevant websites. For example, "did the Holocaust happen" did not appear on sites with factual content about the Holocaust, leading people who used that search query to Holocaust denial sites (Mulligan and Griffin 2018).

2.3 Indirect Online Influence

Indirect online influence has previously been studied in many other forms, including algorithmically mediated peer effects (Eckles, Kizilcec, and Bakshy 2016), which broadly describe when an individual's behavior is influenced by information about their peers, and distributed distribution, which describes how the spread of links discovered via search engines can play an important role in product adoption (Riedl et al. 2018). Most relevant here are the concepts that describe attempts to guide people to specific content through an independent intermediary, such as the various data void types discussed above (Golebiewski and boyd 2019), and the similar concepts of *viral sloganeering* (Donovan and Friedberg 2019) and *keyword signaling* (Tripodi 2019). However, search directives are distinct from these concepts because they focus on a recognizable yet flexible linguistic strategy (prompting people to search) rather than focusing on queries, search results, or web content. As we will discuss, the distinct format that search directives take offers an opportunity for trust and safety professionals to detect potentially violative content, or for fact-checking or civil society organizations to develop earlier awareness of emerging community concerns.

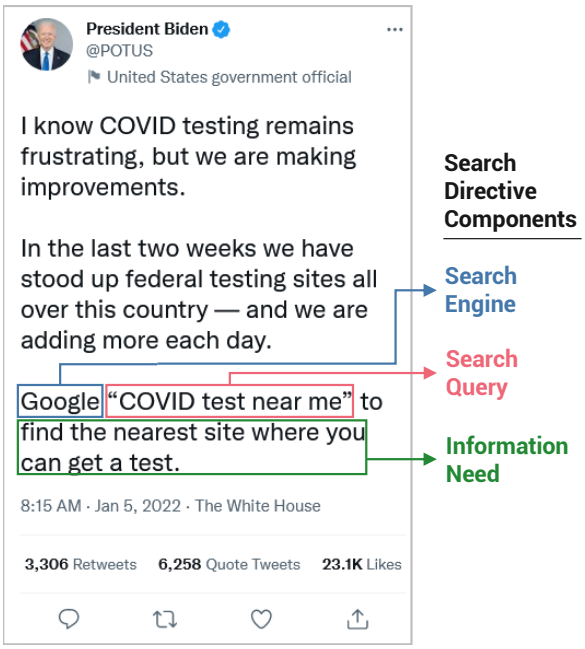


Figure 2: A search directive posted on Twitter in January 2022 by the president of the United States. The post suggests that readers use a specific search engine (Google), to conduct a specific query ("COVID test near me"), to fulfill a specific information need (finding the nearest testing site). This example also shows how other social media factors, such as the account involved or the amount of engagement the post gets, could affect the likelihood of prompting a viewer to conduct an online search.

3 Defining Search Directives

In this section we define search directives (Section 3.1), describe them in the context of established linguistic frameworks (Section 3.2), and develop a taxonomy for distinguishing among different search directive types (Section 3.3).

3.1 Definition and Components

Starting from the dictionary definition of a *directive* as "something that serves to direct, guide, and usually impel toward an action or goal" (Merriam-Webster Dictionary 2023), we define a *search directive* as text or media that could prompt someone to conduct an online search. Under this broad definition, search directives can be said to consist of several sufficient, but not necessary, component parts: a search query, a search engine, and an information need. For example, a Twitter post by the president of the United States on January 2022 includes all three of these components (see Figure 2).

3.2 Linguistic Elements

In linguistics terms, search directives are similar to *illocutionary acts*. Introduced by Austin (1962), an illocutionary act can be defined as "the act performed in making an utterance" (Stiles 1981), which is distinct from both a locutionary act (creating the utterance) and a perlocutionary act (any actual effects of the utterance). That is, an illocutionary act describes the intended outcome of an utterance (e.g., prompting a search), regardless of how it was delivered (e.g., vaguely mentioned or directly suggested) or how effective it was (e.g., causing people to search).

3.3 Search Directive Taxonomy

Using our broad definition, we divide search directives into three primary types based on how their content is delivered, including linked, text, and multimedia search directives. To be coded as a linked search directive, a URL must contain a search query and lead to a page of search results. These digital shortcuts offer the most direct indirect-route towards prompting a specific online search, and by reducing the effort required to conduct the search (a click), they may be especially effective (Fogg 2009). In contrast, text search directives can prompt readers to conduct a search in any grammatical form, including search directives that are suggested ("*Do your research, using #DuckDuckGo, and search for 'dangers of 5G'*"), modeled ("*So weird I can't find this story when I search 'Jackie Gordon'*"), or mentioned ("*'2014 obama coup ukraine' is trending across google search*"). Similarly, multimedia search directives also include a wide range of subtypes based on how they are delivered, including image, video, and audio search directives.

4 Exploring Search Directives

To characterize the usage and lexical boundaries of search directives, we conducted an exploratory study in which we collected and annotated real-world examples. Here we describe our data collection and annotation approach (Section 4.1), with details for linked (Section 4.2), text (Section 4.3), and multimedia (Section 4.4) search directives, as well as for search directives beyond social media (Section 4.5)

4.1 Data Collection and Annotation

Using an iterative approach, we collected a total of 1.5K posts from four social media sites—Gab, Gettr, Twitter, and Parler—using a combination of manual searches on the platforms' interfaces, official APIs, and non-official APIs (e.g., smat-app.com). In each iteration, we annotated the posts we collected according to our definition and taxonomy, noted post-level features that might reveal patterns in usage, and recorded edge cases that were difficult to classify (Appendix A) or extract queries from (Appendix B). Using a convenience sample of 500 Gab posts that contained the word "search," we obtained classifications from two individuals and found they agreed on 90% of posts.

4.2 Linked Search Directives

Data Collection To collect linked search directives (posts containing a URL leading to a page of search results), we conducted searches with popular search engine domain names (e.g., google.com) in our query. One such post provided links to both Brave and Google Search with the same search query prefilled:

> *yes very different results between Brave search and Google*
> *https://search.brave.com/search?q=mass+formation+psychosis&source=web*
> *https://www.google.com/search?q=mass+formation+psychosis*

Query Extraction Linked search directives contain a relatively standardized search engine URL that allows for systematic extraction of exact queries. For example, one post we found prompted viewers to conduct a specific Google Search by providing only a URL ("https://www.google.com/search?q=black+sun"), where the query parameter was "q=<query>" and the extracted query was "black sun," the name for a symbol that originated in Nazi Germany. While we focus here on queries, other parameters present in these URLs could also provide useful context for evaluating search directives.

4.3 Text Search Directives

Data Collection For text search directives, we initially cast a broad net by searching for recent posts that contained the word "search" in their main text. For example,

> *CCP agent and vaccine pusher Leana Wen was a crisis actor for the fake Boston Marathon Bombing. Also, search Leana Wen and planned parenthood.*

These single-keyword searches retrieved a large proportion of posts that did not include search directives, so we began adding phrases ("search for") or punctuation ("search:") that we found in our initial sample to increase the likelihood of obtaining more search directives. One post surfaced by the addition of these cues was:

> *Net Search: Dr. Richard Urso, MD Search for ashley false flag*

We next included the names of popular search engines (e.g., Google, Bing, DuckDuckGo) in our searches, which retrieved examples such as:

> *BING/DDG search "Barack Obama's Smash & Grab"*

Query Extraction In comparison to linked search directives, extracting queries from the text of search directive posts can be a challenging task due to wide variations in language use on social media. Moreover, by our definition, a post can contain a search directive yet provide no search query, an ambiguous search query, or an obfuscated query, as we show in the examples below.

No Query: Search directives can provide little to no guidance on query selection. For example, this post broadly suggests that viewers "search for information":

> *if you genuinely want to be informed, please search for information outside of twitter or carrds. especially when they are most likely outdated. reuters and al jazeera are two great sources for international news and politics. twitter is not your friend when it comes to this.*

Ambiguous Queries: Some search directives do not contain an explicit search query, but instead hint at search queries a viewer might use. For example, one post suggested that viewers search for "doctors by name":

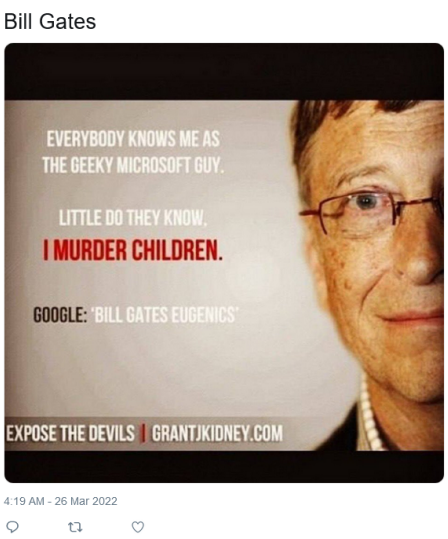
> *Search doctors by name and see how much they've been paid by Big Pharma: https://openpaymentsdata.cms.gov/search#search-results*

Obfuscated Queries: We also found search directive queries that were partially obfuscated or "censored." For example, in a similar vein to the "black sun" query we found earlier, one post replaced the word "Jewish" in a search directive query with asterisks:

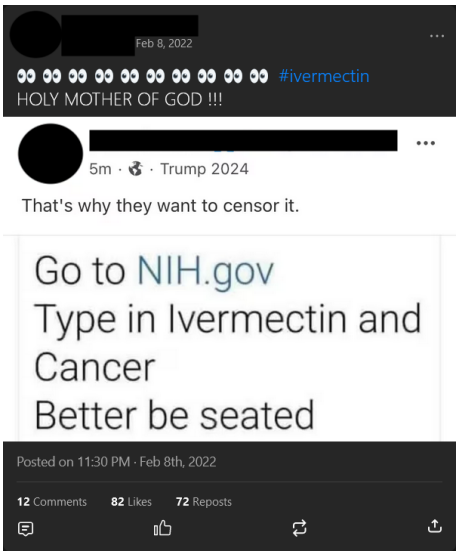
> *cant have people looking up things like "The Complete List Of The 1030 ****** Expulsions In Human History"*

4.4 Multimedia Search Directives

Search directives can also be communicated via multimedia, like images, video, or audio. Among the multimedia search directives that surfaced during our exploratory data collection, we found that their classification as such often required only the multimedia content. For example, some posts prompt viewers to conduct an online search via text present only in their images (Figure 3 on the following page). However, we also found cases that required both the text of the post and the image, such as one whose post text contained a prompt (*"Search this"*) and the image contained a paragraph of text that could potentially be intended as queries.



(a) An example of a search directive communicated solely via an image. We found this search directive on Twitter, and the account that posted it has since been suspended. The top five results for this query on Google Search include two fact-checks (from Reuters and Snopes), and a link to VK (Russian-based social media), which links Bill Gates to "The Eugenics Movement."



(b) Another example of a suggested search directive communicated via text in an image. In this case the viewer is again instructed to use a specific search query ("Ivermectin and Cancer") on a specific search engine (at NIH.gov). This example is one from a larger trend of search directives that pointed people to search on the NIH search engine (Section 7.1)

Figure 3: Search directives communicated via images or other media may be detectable through the use of OCR to automatically extract the text present in the image.

In one video search directive, the text of the post encouraged readers to type "Vaccines causing " with a trailing space into a search engine:

> *Try this on your search engine! Vaccines causing ...*

The video shows how doing this prompts an unknown search engine—all indicators of which one it is have been cropped out of the video—to generate autocomplete suggestions based on the specified prefix. Specifically, the video shows the suggestions updating as they append different letters to the end of that prefix (e.g., "Vaccines causing d," with the first suggestion being "Vaccines causing death"). Another video search directive, which we found on TikTok, shows viewers search results, text passages, and images related to a conspiracy theory about how "13 Illuminati families" control the world.

4.5 Search Directives in TV and News

Beyond social media, search directives can also be shared on TV or news websites. As past research has shown, people in the US often encounter more partisan and fake news via TV than via online websites, suggesting that information spread via TV could have significant impacts (Allen et al. 2020; Muise et al. 2022). One prominent instance of this that we identified from our data was in a text search directive where the author mentioned the "Tucker Carlson Challenge":

> *I took the 'Tucker Carlson Challenge' from last nights show. He advised if you don't believe search engines are suppressing the Hunter Biden story, type "hunter biden weighing crack on scale" into Google. The first image below is the result I received after searching the phrase on Google ; the bottom image is what I received after entering it in DuckDuckGo*

This post recapitulates a search directive that Tucker Carlson (a conservative media personality in the US) made on TV, encouraging viewers to search "Hunter Biden weighing crack on a scale," the day before the author posted it. That search directive is provided in text, video, and audio format on foxnews.com, and states (Carlson 2022),

> *... Type in "Hunter Biden weighing crack on a scale" which is out there. Type it into Google and see what happens. ...*

5 Identifying Search Directives

Following our exploratory examination, we sought to obtain a larger sample of search directives that would allow us to better understand how they are used. Due to the challenges around collecting and processing multimedia search directives, we decided to focus on linked (Section 5.1) and text (Section 5.2) search directives.

5.1 Identifying Linked Search Directives

Building on our exploratory study, we start with linked search directives because their consistent structure allows for easier collection, classification, and query extraction.

URL-based Data Collection To obtain a larger sample of social media posts containing linked search directives, we created a search query that contained URL fragments for five popular search engines. The five search engines and their URL fragments were Google Search (google.com/search), Bing (bing.com/search), DuckDuckGo (duckduckgo.com/?), Yahoo (search.yahoo.com/search), and Brave (search.brave.com/search). We also added a filter to our query to limit our collection to posts written in English. Our final query was: "("google.com/search"OR "bing.com/search"OR "duckduckgo.com/?"OR "search.brave.com/search"OR "search.yahoo.com/search") lang:en".

Using the Twitter API, we collected a total of 4.1M URLs from 4M posts that matched our URL-fragment query across a 16-year span (December 19, 2006 to May 27, 2023). These posts were made by 1.5M unique accounts, and the dataset was primarily composed of original posts (54.8%), with the remainder split among reposts (27.2%), replies (16.8%) and quotes (1.2%). Additional details on our collection and cleaning of linked search directives (e.g., expanding short urls) are available in Appendix C.

URL-based Classification and Query Extraction We classified each URL as a search directive if it contained a known query parameter ("&q=" for Google Search, Bing, DuckDuckGo, and Brave, and "&p=" for Yahoo Search). Before checking these parameters, we simplified domains as much as possible (e.g., reducing www2.bing.com to bing.com). We also filtered out URLs that either had a blank query (only spaces and newlines), because these do not return a page of search results (n = 287), or "%s" as a query, which can be used to change the default search engine in a browser, but do not return a page of results (n = 195).

In total, we classified 3.9M (95.1%) URLs as search directives (3.8M posts), with the exceptions largely consisting of similar URLs (e.g., developers.google.com/search...) that did not lead to a page of search results. The frequency of linked search directives leading to each search engine on Twitter was comparable to their search engine market share, with Google Search by far the most popular (85% of URLs), followed by Bing (8.8%), DuckDuckGo (3.6%), Yahoo (2.6%), and Brave (0.1%).

5.2 Identifying Text Search Directives

For text search directives, we used a broad approach to collect more examples, and fine-tuned two models to automate search directive classification and query extraction.

Text-based Data Collection To expand our labeled dataset of search directives, we collected and annotated an additional 1.9K social media posts. Based on the examples we found during our exploratory study (Section 4.3), and to broaden our set of example posts that both were and were not search directives, we primarily collected posts that contained the word "search" in their main text. Prior to labeling, we also removed posts with duplicate text, which were often spam posts or advertisements, and removed longer posts (> 512 characters) from Reddit and Gab. While some of these posts contained search directives, we chose to remove them because they significantly increased the time it took raters to assess each. After combining this new labeled dataset with the data from our exploratory study, we had a total of 3.4K labeled Reddit, Gab, Gettr, and Twitter posts that we could use to train models for classification and extraction. Additional details on the collection and augmentation of this dataset are available in Appendix D.

Text-based Classification To develop our classifier, we split and used our labeled dataset to fine-tune a BERT[1] classifier model that would label the text of each post with either 1 (contains a search directive) or 0 (does not contain a search directive). For our purposes, we used the "bert-base-uncased" model from Hugging Face, which was pretrained on Wikipedia and the book-corpus datasets using masked language modeling and next sentence prediction (Hugging Face 2022). To create a baseline for the classifier, we looked for a set of strings that we qualitatively found to be relatively common indicators of a search directive, including 'search:', 'search for', 'search "', 'search '", 'search term', and 'search bar'. If a post contained any of these, we marked it as a search directive for the baseline. This baseline had an accuracy of 62% across our entire labeled dataset. Evaluating the classifier on the 30% of labeled posts we set aside for testing, we obtained 89% classification accuracy, a substantial improvement compared to the baseline accuracy (+26%), and only slightly lower than the consistency between individual human raters, which was about 90% (Section 4).

Text-based Query Extraction To construct our extractor, we fine-tuned Hugging Face's T5-small model on the subset of 1,670 posts we had labeled as search directives. T5 was introduced by Raffel et al. (2020) to explore the limits of text-to-text transfer learning, a technique where models are trained on massive datasets for self-supervised tasks (e.g., sentence completion), before being fine-tuned for specific purposes on much smaller datasets (e.g., extracting search queries).

To create a baseline query extractor, we defined a set of start tokens (including words and punctuation), which generally occur before directives (e.g., "search:"), and defined a set of stop tokens that could indicate that a query was over.[2] When we took the query to be any text between these start and stop tokens (with whitespace and quotes removed), we found this baseline had an accuracy of 20% for posts with queries, highlighting the difficulty of this task. To evaluate our extractor model's accuracy, we standardized both the human and model-extracted queries—by lowercasing and removing whitespace, quotation marks, and non-ASCII symbols—and counted the model as correct only when its output was an exact match for the human label. By this measure, our extractor's

1. BERT, which stands for Bidirectional Encoder Representation from Transformers, was first introduced in Devlin et al. (2019) and uses a now standard self-attention transformer proposed in Vaswani et al. (2017).

2. Our baseline start tokens were "`search:`", "`search for`", "`search "`, "`search '`", "`search term`", "`"`", and "`'`"; the stop tokens were a period, a single or double quote, and a newline.

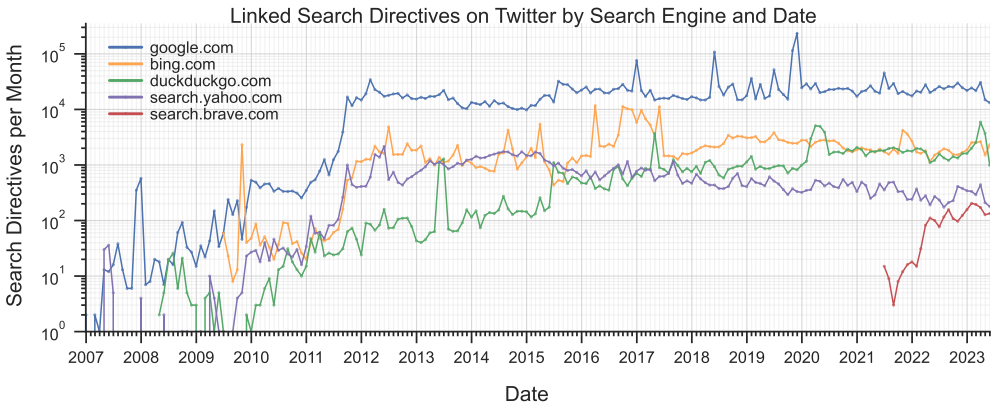


Figure 4: Total number of posts-per-month containing at least one linked search directive by search engine. After a startup period, each search engine reached a relatively steady state of linked search directives over time, with occasional bursts in volume.

accuracy was 75% on the test dataset. Many of the errors we observed occurred on queries that our raters also struggled with, often due to ambiguities that had surfaced in our exploratory study (see Appendix F for examples).

6 Evaluating Search Directives

Here we evaluate the use of search directives by examining the accounts, queries, and search engines involved, over time. We focus our analysis primarily on the linked search directives we collected (Section 6.1), and present exploratory results that we obtained through an application of our text-based methods (Section 6.2).

6.1 Linked Search Directives

Of the 3.9M linked search directives we collected from Twitter, 26.3% were retweets, 16.8% were replies, and 2.1% were quote tweets. We used all of these to count search directives, because each can distribute them into news feeds and contribute to their amplification on social media. To account for differences by search engine, we examined their trends separately and found that each search engine received a relatively steady supply of linked search directives over time on Twitter (Figure 4).

The first linked search directive we found occurred on December 19, 2006—within nine months of the first Tweet—and was about Google indexing Twitter profile pages:

> *@neil, yes google is indexing profile pages... here's the Google SERP for twitter + joy http://www.google.com/search?hl=en&q=twitter+joy*

Highlighting the challenging nature of studying what an online platform might have shown users, the results for that search query on Google today will be very different from what they were in 2006, and Twitter has also substantially evolved as a platform since then. Following that linked search directive at the time of this writing returns results that link to Twitter accounts with "joy" in their screen names, but not the account that posted the first linked search directive. This post also notably contains a mentioned text search directive in its main text ("here's the Google SERP for twitter + joy") that included the same search engine ("Google") and nearly the same query ("twitter + joy") as the link.

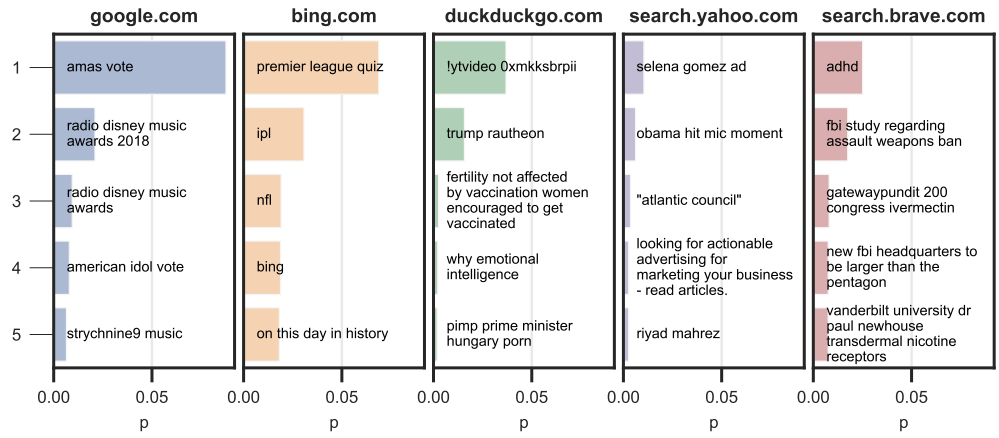


Figure 5: The top 5 queries and the relative proportion of linked search directives they account for by search engine. Overall, the top 20% of queries accounted for 70.4% of all search directives, but this distribution varied substantially by search engine. The more popular search engines, Google and Bing, had substantially more skewed distributions, with the most popular query for either accounting for over 5% of all search directives leading to that search engine.

Accounts A total of 1.4M accounts posted at least one linked search directive on Twitter, but a relatively small number posted the majority. Over two-thirds of the search directives we collected (68.6%, 2.6M) were spread by only 20% (285K) of accounts, and 9.8% came from the top 10 accounts alone. The top two accounts were both automated, exclusively posted Google Search links, and generally used queries containing song names based on a radio station's activity. First, accounting for 4.5% of all linked search directives, was @radio_scrobble, an automated account that tweeted when Beats 1 Radio (a music radio station run by Apple) played new songs and included links to Google Search with queries in an "artist - song" format. The next top account was on6musicnow (1.1% of all search directives), another automated account that followed a radio station and posted Google Search links (BBC Radio 6). These findings highlight the value of considering both the accounts involved and the search engine(s) invoked when evaluating search directives. We highlight additional automated search directive activity in Appendix E.

Queries Similar to the top accounts, the top overall queries were also music-related and exclusively invoked through Google Search links. The top query "amas vote" (referring to the American Music Awards) accounted for 7.4% of all queries, and the next most popular query ("radio disney music awards 2018") accounted for 1.8%. However, both were much less concentrated among a small number of accounts, with the top "amas vote" poster only accounting for 0.2% of all such search directives. In contrast to the search directives we found among the top accounts, which show how a single account can spread many unique search directive queries, these results demonstrate that many unique accounts can also spread a single query.

We also found differences in the type and distribution of queries that led to each search engine (Figure 5). In contrast to the top Google Search queries, which were music-related, the top Bing query was related to sports and involved a Bing-specific feature: searching "premier league quiz" on Bing, even today, returns a quiz at the top of the results that can be completed without leaving the page of search results. The top query for DuckDuckGo we examine in depth for the case studies section of this paper (Section 7.3), while the top query for Yahoo ("selena gomez ad") was largely driven by retweets of a post made

by Selena Gomez's official account about a new fragrance she released at Macy's. Last, the top query for Brave Search was "adhd" a trend that was almost entirely (91.4%) driven by a single individual, often in replies to their own threads, but constituted a small absolute number of search directives relative to the broader dataset (n = 53).

In addition to evaluating the entire text of each query, we also examined 13 advanced search query operators[3] that users can add to their queries in order to filter the results to specific websites (e.g., "site:bbc.com"), filetypes (e.g., "filetype:pdf"), or time ranges (e.g., "before:2023-01-01"). Overall, 1.5% of the linked search directive queries we collected contained one of these operators. The vast majority of these queries (90.5%) used the "site:" operator.

Among the search directive queries containing any operator, the most frequently used one filtered the results using a site specific search operator ("site:theconsumerhealth.com") to limit the search results to webpages from that domain ("theconsumerhealth.com"). We found 8,873 posts containing this specific operator in their search directive queries between June 6, 2013 and May 6, 2023, with the rest of the query typically involving health-related topics, as the domain name suggests. Accounting for 36.4% of all such queries, the top query was: "tropical+fruit+burns+site:theconsumerhealth.com." When we examined the posts prompting this search, we found that most were via replies (92.5%), most of the accounts involved posted a link with this query at least twice (72.4%), and most of the posts occurred on a small number of days (91.8% occurred on 5 days in June and August 2013). The text of these posts was unique 99.6% of the time, but their differences were often fairly minimal. For example, after removing mentions and URLs, many of these posts used the same two words at the start and end of their text, such as these two examples:

omg already down 7 pounds this web-site i came across on google

omg already burned 7 lbs that site i discovered on google

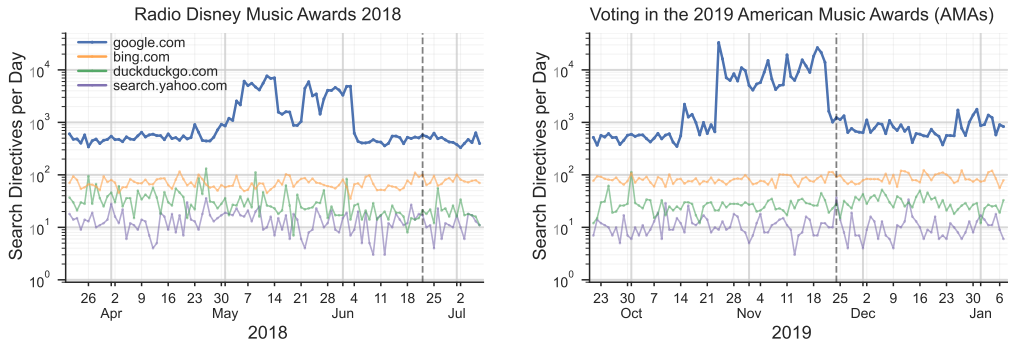
Together, these signals suggest that these posts were part of an indirect influence campaign leveraging Google's "site:" search operator. More importantly, when examining the full URLs used in these linked search directives, we found that nearly all of them (99.3%) also included a parameter "btnI" that can be used to enable Google's "I'm Feeling Lucky" functionality. This function, also available as a button on Google's homepage, takes users directly to the website of the first search result, skipping the results page itself. These examples highlight the importance of considering URL parameters when evaluating search directives, and show how an indirect influence campaign can combine both a query operator ("site:") and a URL parameter ("btnI="I'm Feeling Lucky") to send users who click on it directly to the first search result from a specific web domain.

The domain in this case ("theconsumerhealth.com") is now for sale, but Internet Archive records from June 2013 show that the website was selling "Garcinia Cambogia Formula," a weight loss pill. When used on Google Search today, the "I'm Feeling Lucky" URL parameter still works, but Google Search now provides users with a notice prior to completing the redirect that states: "Redirect Notice. The previous page is sending you to <url>. If you do not want to visit that page, you can return to the previous page." This change appears to have occurred in October 2019.[4]

Bursts in Usage The use of linked search directives on Twitter is relatively consistent over time (Figure 4) but we observed several bursts in activity that corresponded to

3. The operators we examined were: `site:`, `intitle:`, `allintitle:`, `inurl:`, `allinurl:`, `intext:`, `allintext:`, `source:`, `before:`, `after:`, `related:`, `filetype:`, and `ext:`.

4. https://support.google.com/websearch/thread/15794018?hl=en



(a) A burst in search directives that led to Google Search and were related to the 2018 Radio Disney Music Awards. This burst started nearly two months prior to the event (June 22, 2018; see dashed vertical line) and ceased three weeks before the event. The top three queries in May 2018 were: "radio disney music awards 2018" (54.5%), "radio disney music awards" (23.5%), and "rdma voting" (3.1%).

(b) The burst in search directives that led to Google Search during October and November 2019 was due to the American Music Awards (AMA), with the search queries "amas vote" and "ama voting" accounting for 69% of all queries in those months. This burst started about a month prior to the event (November 24, 2019; see dashed vertical line) and ceased only a few days prior to the event.

Figure 6: High-volume bursts in the amount of linked search directives on Twitter. We also found query bursts for a variety of other events, including a sports tournament, and featuring other search engines, including DuckDuckGo, Bing, and Yahoo (see Appendix E).

real world events (Figure 6). For example, the burst in Google search directives that occurred in October and November of 2019 (Figure 6b) largely accounts for the most popular query in our dataset—"amas vote"—and occurred in the weeks leading up to the 2019 American Music Awards (AMA). During this time period, the daily number of search directives increased by over an order of magnitude for Google Search, and "amas vote" accounted for 80.6% of all Google search directive queries from October through December 2019. We highlight additional bursts in activity in Appendix E.

6.2 Text Search Directives

After training our classifier and extractor models, we ran them in sequence on a dataset of posts that we had collected from Gab. This dataset contained 373K posts on Gab that contained the word "search" from between April 2021 and March 2023. For consistency with our labeled dataset, we filtered out longer posts (> 512 characters), leaving us with 225K posts. In total, the model classified 24.7% of the posts in this dataset as search directives, and extracted a search query from 14.7% of posts.

These data differ from the linked search directives due to the differences in the social media platform they came from, limiting our ability to compare across types. Moreover, our exploratory dataset composed nearly half of our labeled dataset, and the exploratory nature of its collection could have biased the types of examples we found due to the time range (2021-2023) or other factors (see Appendix D). However, these models were still a useful proof-of-concept for surfacing text search directives with potential data void queries, and we limit our use of the results from this dataset for that purpose.

The most frequent text search directive query we found in the Gab dataset was "died suddenly," which refers to a conspiracy theory centering the COVID-19 vaccine that we examine in the case studies in the next section. We also found a number of queries related to other politicized topics or phrases, including abortion-related queries (e.g., "abortion clinics" and "crisis pregnancy centers") and race-related queries (e.g., "white lives matter"), both of which have been found to be areas of concern for data voids in past work (Mejova, Gracyk, and Robertson 2022; Noble 2018).

7 Case Studies

To illustrate the value of identifying search directives for trust and safety teams, we report three case studies involving queries that led to data voids and fell under the "your money or your life" categories that Google Search specifies carry a "high risk of harm" (Walls 2022). First, we examine search directives that promoted alternative medical treatments and encouraged the use of scientific sources and intermediaries (Section 7.1). Second, we examine queries about a conspiracy theory that the COVID-19 vaccine was causing people to "die suddenly" (Section 7.2). Last, we examine a set of linked search directives that used an advanced search operator to lead people who clicked on them directly to a YouTube video promoting a cryptocurrency scam (Section 7.3).

7.1 Scientific Sources and Alternative Medical Treatments

One of the recurring themes we found in our exploratory study involved both specific queries, often about alternative medical treatments, and specific intermediaries, which were often scientific sources. Many of these queries mentioned Ivermectin, which is an antiparasitic drug that is largely used in veterinary medicine yet has received substantial attention as an alternative medical treatment since the COVID-19 pandemic. In terms of intermediaries, the search directives we collected on this topic often encouraged viewers to specifically use the NIH and other scientific sources as their intermediaries or included advanced query operators to automatically filter the results. This case highlights a category of search directives that could impact users' health, and shows how the use of source names and advanced operators in search directive queries can be used to further narrow the results that such searches return, raising questions about how such settings could be better communicated to searchers on the results page.

We found search directives with queries mentioning Ivermectin in text, linked, and multimedia formats, and these queries also often mentioned "cancer," implying that Ivermectin could be used as an alternative treatment. For example, an image search directive we found in our exploratory study suggested that people search for "Ivermectin and cancer," and that they specifically conduct that search on the NIH website (Figure 3b). Similarly, a text search directive suggested the same query and site-specific search engine:

> *Hey people try this search for yourself @NIH.GOV Ivermectin and cancer in the search bar!*

The linked search directives we found on this topic often led to DuckDuckGo (57%)[5] and often had queries that were permutations of the same two terms (e.g., "ivermectin cancer," "cancer ivermectin"). For example:

> *Not that Joe had anything to do with it, but since you mentioned a cure for cancer, if you search: Ivermectin Tablets for Cancer, you will see how it has been found out that it is useful for cancer. https://duckduckgo.com/?q=Ivermectin+tablets+for+cancer*

When these queries deviated beyond these two terms, they often did so by including the name of an intended intermediary or search results source: the NIH. For example, we found linked search directives with queries that mentioned the NIH (e.g., "nih ivermectin cancer" and "nih, ivermectin and inhibiting cancer cell growth."), which could increase the likelihood that a search engine returns results from the NIH website. In the same vein, but more effectively, we also found DuckDuckGo queries that used advanced search operators to filter for results only from the NIH website ("ivermectin

5. Followed by Google (34%), Bing (6.3%), and Brave (2.5%).

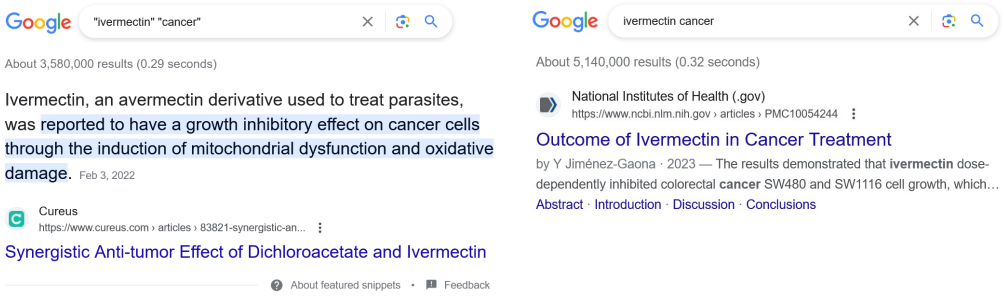


Figure 7: The top search result for “ivermectin cancer” on Google Search with (left) and without (right) quotation marks around the key terms. Both the featured snippet (left) and the snippet extracted from the standard result (right) suggest that Ivermectin can cure cancer, but it has yet to be tested on humans.

cancer site:nih.gov”) and Google Search queries using quotations around both terms: *“ivermectin” “cancer”*

When we tested the latter query on Google Search, we found that the addition of the quotation marks generated a misleading featured snippet (Figure 7). Despite claims that Ivermectin cures or protects against cancer, research suggests that using Ivermectin to treat cancer may cause harm in the amounts needed to benefit from any potential antitumor effects, and it has not yet been tested on humans (Jiménez-Gaona et al. 2023). As of this writing, Google Search does not provide any indication that a site-specific or site-limiting query has been used aside from what appears in the search bar, which the average user might not be aware of or pay close attention to when clicking on a linked search directive.

Highlighting the use of such query operators in the search interface—for example, by including a banner at the top of the results when one is activated—may make such filters more transparent to users who click on linked search directives, thereby helping them better evaluate the results. Indeed, we found that DuckDuckGo does offer such a banner for queries including the “site:” operator, which we discovered when we checked the search results for the query “ivermectin cancer site:nih.gov” (Figure 8). This banner appears at the top of the results and informs the searcher that the search engine is “Only showing results from nih.gov,” notes that the user can “Clear filter to show all search results” and offers a button that the user can click to do exactly that.

This case illustrates the proactive value of monitoring search directives: while the earliest search directive we found on this topic was posted in May 2021, the earliest fact check we found (addressing a claim that Ivermectin is a government-withheld cure for cancer) did not occur until February 2023 (Fichera 2023), nearly two years later.

7.2 Harms and Origins of the COVID-19 Vaccine

Related, but distinct in focus from the Ivermectin search directives, are those that focused on the origins and effects of the COVID-19 vaccine. Some of the queries we found explicitly mentioned the vaccine, including the video example from our exploratory study, which showed viewers related queries (e.g., “vaccines causing death”; see Section 4.4). At one points in the video, the dropdown list of autocomplete suggestions disappears and reveals a result titled “COVID-19 mRNA vaccination leading to CNS inflammation: a case series.” Similar to the alternative medical treatments case study, that search result appears to leads to an NIH website (ncbi.nim.nih.gov), again leveraging scientific sources as the target content.

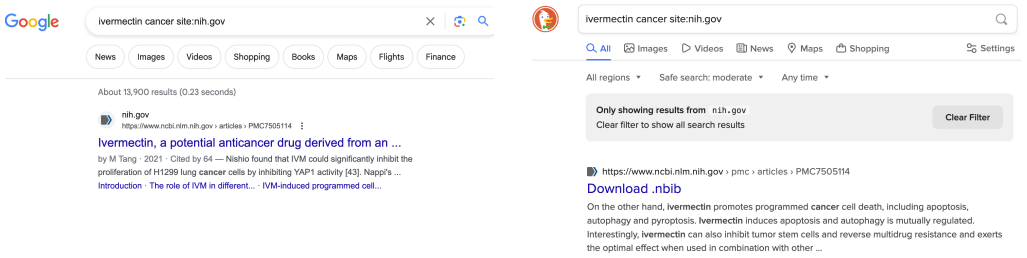


Figure 8: An example of a search for the query "ivermectin cancer site:nih.gov" on Google Search (left) and DuckDuckGo (right). As shown in these partial screenshots, DuckDuckGo, but not Google Search, provides a banner at the top of the results that informs users that the filter is active and provides a button for them to clear it and reconduct the search.

However, not all the queries we found on this topic were as explicit, and many were more subtle. Among these were queries that mentioned sudden deaths, including the top query ("died suddenly") that we found in the text search directives we identified on Gab (Section 6.2). For example, a text search directive we found on this topic suggested that readers conduct a search on Twitter:

> *These stories are tragic. Search twit for "died suddenly" or search the news for "cardiac arrest" and see hundreds of young athletes who've died. Look at the meteoric rise of heart attacks under 30. It's mind boggling.*

One of the queries containing a "suddenly died" reference that we found also used an advanced query operator to filters the results by time: "died suddenly before:2020-12-1." Advanced query operators such as this one (and "after:<date>") can be used to modify the time range of the search results that searchers receive, with the usage in this case ("before:2020-12-1") filtering the results for webpages published before December 1, 2020. This query was used in three replies to three separate threads, and in each thread, the original post was by a prominent account (followed by about 26K, 300K, and 500K) and suggested a link between recent sudden deaths and the COVID-19 vaccine. Similar to the examples we found in our exploratory study, these examples highlight how the threads that search directives appear in could also be useful in prioritizing closer evaluation. Here the value for social media sites comes from this thread discovery, and the value for search engines comes from identifying how advanced operators are being used to filter their results in links.

In this case, all of these search directives were posted by a single account that was using this advanced operator to try and dissuade people that the COVID-19 vaccine was the cause of recent sudden deaths. For example, using the "before:" operator to show results prior to the existence of the vaccine:

> *Did these people die from the vaccine too? Does it travel through time? https://www.google.com/search?q=died+suddenly+before:2020-12-1*

However, in one text search directive we found during our exploratory study, we found the same tactic being suggested for the opposite purpose:

> *Run a Google News search for "suddenly died" and set it for 2022*

We began observing linked search directives around this topic on October 3, 2021, and the largest burst in activity occurred in July 2022. This case also highlights the potential proactive value of identifying search directives, as a documentary that promoted this conspiracy theory (titled "Died Suddenly") was not released until November 21, 2022.

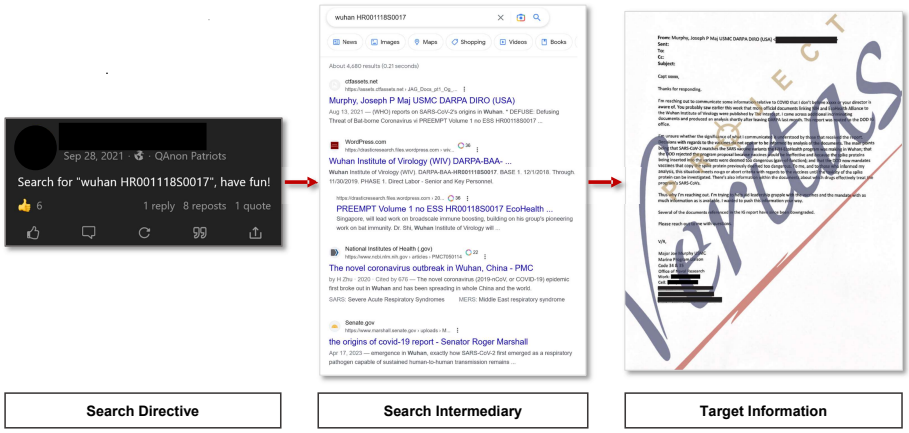


Figure 9: A diagram depicting an example of a search directive leading to a data void. In this example, a post we found on the alternative social media site Gab tells viewers to conduct an online search using an extremely specific query ("wuhan HR001118S0017"). At the time of this writing, searching that query on Google Search returns PDFs with the "Project Veritas" watermark in the top results. These PDFs appear to be leaked US Department of Defense emails and documents, and push a narrative that SARS-CoV-2 is "an American recombinant bat vaccine, or its precursor virus."

By surfacing new claims before they reach a large audience, search directives may also be useful to journalists and fact checkers, as early detection would allow them a greater window of time to assess new but spreading claims.[6] Indeed, search directives may often occur alongside claims because claims are more persuasive when evidence is offered, and search results from a trusted intermediary can fulfill that role.

One data void query on this topic stood out for its use of a highly specific keyword ("Wuhan HR001118S0017").[7] When we checked the results for this query on Google Search, we found a data void that largely consisted of PDFs claiming that COVID-19 is a human-made biological weapon (Figure 9). Another query we found containing a highly specific keyword also leveraged the use of quotations to get results that had an exact match:

"c9h9no3" "wuhan"

The first part of this query appears to be the molecular formula ($C_9H_9NO_3$) for adrenochrome, a chemical compound that is mentioned in several conspiracy theories, including those around QAnon and PizzaGate. While the first two Google Search results for this query point to listings on the NIH website that provide technical details about this chemical compound, the results that follow appear to be a data void. The first of those results, ranked third overall, redirects to a website with an IP address domain that is blocked by a commonly used browser extension (uBlock) because it appears on a list of domains that present a malware risk. When examining the cached version of the webpages these results led to, many appear to have a similarly formatted wall-of-text promoting various conspiracies around adrenochrome.

6. In this respect, search directives may be especially useful in combination with existing tools for evaluating data voids based on some researcher-selected seed content (Flores-Saviaga, Feng, and Savage 2022).

7. We found this query while examining queries containing the word "wuhan," which we qualitatively found to appear in the search directive queries and posts around the topic of COVID-19 vaccines.

7.3 Cryptocurrency Scams: I'm Feeling Lucky

This case study highlights how advanced search operators—additions to a query that alter a search engine's results—can be used in combination with search directives to skip the results page and send people directly to specific content. Among our linked search directives, we found that this combination appeared in the most popular search query for DuckDuckGo: "!ytvideo 0xMKksbrPiI." This query uses an advanced search operator that is specific to DuckDuckGo ("!ytvideo")[8], is meant to be paired with a YouTube video ID ("0xMKksbrPiI"), and takes searchers directly to that video on YouTube (youtube.com/watch?v=0xMKksbrPiI).[9] In this case, that search directive had the effect of sending people who clicked on it directly to a YouTube video titled "How I LEGALLY Made $200k Passive Income Daily Farming On Uniswap (Arbitrage bot+MEV) Sept 2022." The channel that posted this video has since been terminated "due to multiple or severe violations of YouTube's policy against spam, deceptive practices, and misleading content or other Terms of Service violations."

When we dug into the posts that contained this linked search directive, we found several characteristics that suggest a coordinated campaign to spread this search directive via automated accounts. Among the 5,225 search directive posts that contained the top DuckDuckGo query ("!ytvideo 0xmkksbrpii"), about half were retweets (50.9%) and half were replies (49.1%). The retweets were always of two original tweets that contained almost identical post text, one of which was:

> *More people deserve to know about this new Method called Mev+Arb Bots. RT please! Basically free money at this point, it's easy if you follow this tutorial. I'm pulling about $2,000 per day with this The only way you can profit, Bear market or not. https://duckduckgo.com/?q=!ytvideo+0xMKksbrPiI*

The remaining search directives with this query were replies that often occurred in threads related to cryptocurrency. For example:

> *@TygeneCrypto How is it- possible that this guy has made over $250K by. exploiting Uniswap? It's outrageou_s.(, https://duckduckgo.com/?q=!ytvideo+0xMKksbrPiI*

These replies in particular had several characteristics—including time of posting and text similarity—that suggested a coordinated campaign to spread this specific linked search directive. For example, all 2,657 of these replies occurred on a single day (April 14, 2023), and involved only 98 unique accounts. While the text of these replies was always unique, they also frequently contained seemingly random distortions involving punctuation (e.g., *"It's outrageuo_s.(,"* in the example above). These distortions appear designed to intentionally obfuscate the similarity of these replies, as when we removed all punctuation from their text, we found that nearly 95% had at least one duplicate, and the top 10 most commonly used replies accounted for 24.8% of all replies. We also found that many of the accounts spreading this linked search directive used the exact same de-punctuated reply text, with a median of 16.5 unique accounts posting each unique reply. The most common de-punctuated reply text was "this is an interesting and inspiring story," accounted for 3.4% (n=91) of all such replies, and was posted by 64 out of the 98 unique accounts that posted this linked search directive in a reply.

While this YouTube video has since been taken down, we found several other queries using the "!ytvideo" operator, the next most popular of which is still up. When we searched for mentions of its video ID ("Vhkb9G2SyVw") across the web, we found several posts on cryptocurrency forums in which people either asked if the video is promoting a scam, or

8. https://duckduckgo.com/bang_lite.html
9. Similar to the "I'm Feeling Lucky" functionality that Google Search provides, but activated via a query operator rather than a URL parameter, and without the redirect notice we documented in Section 6.1.

stated that they had fallen victim to the scam. We reported this video to YouTube upon finding it, but this case study demonstrates a proactive approach that Twitter or YouTube could have used to surface this scam as early as March 31, 2023. Similarly, for the top video ID ("0xMKksbrPiI"), the account that posted it was active on YouTube until at least May 30, 2023 (when it was last captured in the Wayback Machine), but we observed linked search directives leading to it as early as March 21, 2023.

One specific advantage of this strategy—advanced query operators in search directive queries—may be the *evasion* of content moderators. For social media, links to specific YouTube videos are easy to identify, but are harder to detect when only the video ID is included as part of a longer link (i.e., in the search query of a linked search directive). Moreover, for the target destination (YouTube in this case study), people who click on these linked search directives will be routed through the search engine intermediary (DuckDuckGo in this case study), potentially making the search engine appear as the referring website (i.e., where the user came from) and obfuscating the actual referrer, which can be a useful signal for trust and safety teams.

8 Conclusion

In this paper we introduced the concept of *search directives* (content that could prompt an online search), documented their widespread use on social media, and highlighted their value in identifying search queries that lead to data voids of low-quality results. Specifically, our case studies show that our methods for identifying search directives can be used to proactively surface data void queries that touch on topics—alternative medical treatments, vaccine disinformation, and active cryptocurrency scams—that carry a high risk of harm to users. As such, our methods for identifying search directives could be of immediate use to wide range of professionals, including system designers, trust and safety teams at social media sites, web search engines, and other online search intermediaries, and journalists or fact checkers that would benefit from early detection of emerging conspiracies.

Our results suggest that in order to protect their users from scams and misinformation, search engines, social media platforms, and other websites (e.g., NIH.gov) would benefit from being aware of the search directives that are being shared on, or about, their platforms. In cases where a search directive has clear potential to cause harm to users, such as the cryptocurrency scam our method surfaced (Section 7.3), our approach could also be used by social media companies or consumer protection agencies to identify and potentially take action on the actors involved. Our results also offer immediate practical recommendations for developing interventions: Google Search would benefit from alerting users to the use of advanced query operators as DuckDuckGo does (Figure 8), and DuckDuckGo would benefit from alerting users who are being redirected through their search engine to a specific search result as Google Search does (Section 6.1). Similarly, both search engines would benefit from informing users when quotations are used in a search query, but neither appears to currently do so.

For external investigators of online platforms, our framework can be extended to inform an algorithm auditing component, in which the queries discovered via search directives are funneled into various search engines. While approaches to algorithm audits—and targeted content discovery more broadly—often rely on researcher-selected seed content (e.g., specific keywords, URLs, or accounts) as a starting point, search directives specify a flexible linguistic strategy that can be used to surface completely unknown content. The quality of the results returned for those queries could then be evaluated through qualitative, quantitative, or mixed-methods measures, and a quality threshold could

be established to identify data void queries for further manual review. With that data, the extent to which the search engines producing those results are aware of and are taking action on those data voids could be measured using the presence of the various warning banners placed at the top of the results, such as the warning banners that Google Search uses for rapidly changing (Sullivan 2021), low-relevance (Tucker 2020), and low-quality (Nayak 2022) data voids. Building such pipelines may be especially of interest to trust and safety professionals working at a platform whose search engine is a frequent target of search directives, as well as to those working on new search technologies (e.g., ChatGPT).

Although we focused on identifying search directives that appeared in the text of social media posts written in English, they can also occur in a variety of media (such as TV broadcasts, newspapers, or radio), and can be delivered via a variety of formats (including images, audio, or videos). This focus on search directives in English may mean that our study provides lower bound estimates, as there is evidence that content moderation on major platforms is limited for languages other than English (Borge et al. 2021). For multimedia search directives, future work should explore methods for systematically collecting them, and investigate their potential to be doctored (e.g., showing fake search results in an image or video). Future work should also evaluate other search engine features that can be triggered in a URL beyond its parameters, such as the "Goggles" feature offered by Brave, or in a query, including those that filter results based on file type (e.g., PDFs). Last, our model for indirect online influence provides a foundation for future work (Figure 1), and studies examining real user behavior could potentially be used to estimate how often search directives are followed.

References

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31 (2): 211–36. https://doi.org/10.1257/jep.31.2.211.

Allen, Jennifer, Baird Howland, Markus Mobius, David Rothschild, and Duncan J. Watts. 2020. "Evaluating the Fake News Problem at the Scale of the Information Ecosystem." *Science Advances* 6 (14). https://doi.org/10.1126/sciadv.aay3539.

Austin, J. L. 1962. *How to Do Things with Words.* The William James Lectures 1955. Oxford, England: Clarendon Press.

Bentley, Frank, Katie Quehl, Jordan Wirfs-Brock, and Melissa Bica. 2019. "Understanding Online News Behaviors." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* Glasgow, Scotland: ACM. https://doi.org/10.1145/3290605.3300820.

Borge, Olivia, Victoria Cosgrove, Elena Cryst, Shelby Grossman, Shelby Perkins, and Anna Van Meter. 2021. "How Search Engines Handle Suicide Queries." *Journal of Online Trust and Safety* 1 (1). https://doi.org/10.54501/jots.v1i1.16.

Carlson, Tucker. 2022. "Tucker Carlson reacts to Hunter Biden's business dealings: Whatever helps China, Joe Biden has dutifully done." *Fox News* (July 11, 2022). https://www.foxnews.com/opinion/tucker-carlson-hunter-biden-business-dealings-china-joe-biden.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.* arXiv: 1810.04805 [cs.CL].

Donovan, Joan, and Brian Friedberg. 2019. *Source Hacking: Media Manipulation in Practice.* Technical report. Data & Society. https://datasociety.net/library/source-hacking-media-manipulation-in-practice/.

Eckles, Dean, René F. Kizilcec, and Eytan Bakshy. 2016. "Estimating Peer Effects in Networks with Peer Encouragement Designs." *Proceedings of the National Academy of Sciences* 113 (27): 7316–22. https://doi.org/10.1073/pnas.1511201113.

Edelman. 2021. *Edelman Trust Barometer 2021.* Technical report. Edelman. https://www.edelman.com/trust/2021-trust-barometer.

Epstein, Robert, and Ronald E. Robertson. 2015. "The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections." *Proceedings of the National Academy of Sciences* 112 (33). https://doi.org/10.1073/pnas.1419828112.

Epstein, Robert, Ronald E. Robertson, David Lazer, and Christo Wilson. 2017. "Suppressing the Search Engine Manipulation Effect (SEME)." In *Proceedings of the ACM on Human-Computer Interaction,* 1:1–22. https://doi.org/10.1145/3134677.

Fichera, Angelo. 2023. "No, ivermectin isn't being withheld as cancer 'cure'." *AP News* (February 16, 2023). https://apnews.com/article/fact-check-ivermectin-nih-cancer-cure-629592291079.

Fisher, Matthew, Mariel K. Goddu, and Frank C. Keil. 2015. "Searching for Explanations: How the Internet Inflates Estimates of Internal Knowledge." *Journal of Experimental Psychology: General* 144 (3): 674–87. https://doi.org/10.1037/xge0000070.

Fisher, Matthew, Adam H. Smiley, and Tito L. H. Grillo. 2022. "Information without Knowledge: The Effects of Internet Search on Learning." *Memory* 30 (4): 375–87. https://doi.org/10.1080/09658211.2021.1882501.

Flores-Saviaga, Claudia, Shangbin Feng, and Saiph Savage. 2022. "Datavoidant: An AI System for Addressing Political Data Voids on Social Media." *Proceedings of the ACM on Human-Computer Interaction* 6 (CSCW2): 1–29. https://doi.org/10.1145/3555616.

Fogg, BJ. 2009. "A Behavior Model for Persuasive Design." In *Proceedings of the 4th International Conference on Persuasive Technology,* 1–7. Persuasive '09. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/1541948.1541999.

Gillespie, Tarleton. 2017. "Algorithmically Recognizable: Santorum's Google Problem, and Google's Santorum Problem." *Information, Communication & Society* 20 (1): 63–80. https://doi.org/10.1080/1369118X.2016.1199721.

Golebiewski, Michael, and danah boyd. 2019. *Data Voids: Where Missing Data Can Easily Be Exploited.* Technical report. Data & Society. https://datasociety.net/library/data-voids.

Guess, Andrew M., Brendan Nyhan, and Jason Reifler. 2020. "Exposure to Untrustworthy Websites in the 2016 US Election." *Nature Human Behaviour* 4 (5): 472–80. https://doi.org/10.1038/s41562-020-0833-x.

Hugging Face. 2022. *Text Classification.* https://huggingface.co/docs/transformers/tasks/sequence_classification.

Jiménez-Gaona, Yuliana, Oscar Vivanco-Galván, Gonzalo Morales-Larreategui, Andrea Cabrera-Bejarano, and Vasudevan Lakshminarayanan. 2023. "Outcome of Ivermectin in Cancer Treatment: An Experience in Loja-Ecuador." *Nursing Reports* 13 (1): 315–26. https://doi.org/10.3390/nursrep13010030.

Landtsheer, Christ'l De, Philippe De Vries, and Dieter Vertessen. 2008. "Political Impression Management: How Metaphors, Sound Bites, Appearance Effectiveness, and Personality Traits Can Win Elections." *Journal of Political Marketing* 7 (3-4): 217–38. https://doi.org/10.1080/15377850802005083.

Ludolph, Ramona, Ahmed Allam, and Peter J. Schulz. 2016. "Manipulating Google's Knowledge Graph Box to Counter Biased Information Processing During an Online Search on Vaccination: Application of a Technological Debiasing Strategy." *Journal of Medical Internet Research* 18 (6): e137. https://doi.org/10.2196/jmir.5430.

Mejova, Yelena, Tatiana Gracyk, and Ronald E. Robertson. 2022. "Googling for Abortion: Search Engine Mediation of Abortion Accessibility in the United States." *Journal of Quantitative Description: Digital Media* 2. https://doi.org/10.51685/jqd.2022.007.

Merriam-Webster Dictionary. 2023. *Directive.* https://www.merriam-webster.com/dictionary/directive.

Muise, Daniel, Homa Hosseinmardi, Baird Howland, Markus Mobius, David Rothschild, and Duncan J. Watts. 2022. "Quantifying Partisan News Diets in Web and TV Audiences." *Science Advances* 8 (28). https://doi.org/10.1126/sciadv.abn0083.

Mulligan, Deirdre K., and Daniel S. Griffin. 2018. "Rescripting Search to Respect the Right to Truth." *Geo. L. Tech. Rev.* 2:28. https://ssrn.com/abstract=3228671.

Narayanan, Devesh, and David De Cremer. 2022. "'Google Told Me So!' On the Bent Testimony of Search Engine Algorithms." *Philosophy & Technology* 35 (2): 22. https://doi.org/10.1007/s13347-022-00521-7.

Nayak, Pandu. 2022. "New Ways We're Helping You Find High-Quality Information." *Google,* https://blog.google/products/search/information-literacy/.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism.* New York: New York University Press.

Norton, Michael I., Daniel Mochon, and Dan Ariely. 2012. "The IKEA Effect: When Labor Leads to Love." *Journal of Consumer Psychology* 22 (3): 453–60. https://doi.org/10.1016/j.jcps.2011.08.002.

PRAW. 2022. *PRAW: The Python Reddit API Wrapper.* https://praw.readthedocs.io/en/stable/.

Raffel, Colin, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer." *Journal of Machine Learning Research* 21 (140): 1–67. http://jmlr.org/papers/v21/20-074.html.

Riedl, Christoph, Johannes Bjelland, Geoffrey Canright, Asif Iqbal, Kenth Engø-Monsen, Taimur Qureshi, Pål Roe Sundsøy, and David Lazer. 2018. "Product Diffusion through On-Demand Information-Seeking Behaviour." *Journal of The Royal Society Interface* 15 (139). https://doi.org/10.1098/rsif.2017.0751.

Shearer, Elisa, and Amy Mitchell. 2021. *News Use Acrosss Social Media Platforms in 2020.* Technical report. Pew Research Center. https://www.pewresearch.org/journalism/wp-content/uploads/sites/8/2021/01/PJ_2021.01.12_News-and-Social-Media_FINAL.pdf.

SMAT. 2022. *Social Media Analysis Toolkit.* https://www.smat-app.com.

Stiles, William B. 1981. "Classification of Intersubjective Illocutionary Acts." *Language in Society* 10 (2): 227–49. https://doi.org/10.1017/S0047404500008654.

Sullivan, Danny. 2021. "A New Notice in Search for Rapidly Evolving Results." *Google,* https://blog.google/products/search/new-notice-search-rapidly-evolving-results/.

Sundar, S. Shyam, and Sampada S. Marathe. 2010. "Personalization versus Customization: The Importance of Agency, Privacy, and Power Usage." *Human Communication Research* 36 (3): 298–322. https://doi.org/10.1111/j.1468-2958.2010.01377.x.

Sweeney, Latanya. 2013. "Discrimination in Online Ad Delivery." *Queue* 11 (3): 10. https://doi.org/10.1145/2460276.2460278.

Thompson, Stuart A. 2022. "Fed Up With Google, Conspiracy Theorists Turn to DuckDuckGo." *The New York Times,* https://www.nytimes.com/2022/02/23/technology/duckduckgo-conspiracy-theories.html.

Toff, Benjamin, and Rasmus Kleis Nielsen. 2018. "'I Just Google It': Folk Theories of Distributed Discovery." *Journal of Communication* 68 (3): 636–57. https://doi.org/10.1093/joc/jqy009.

Tripodi, Francesca. 2018. *Searching for Alternative Facts: Analyzing Scriptural Inference in Conservative News Practices.* Technical report. Data & Society. https://datasociety.net/output/searching-for-alternative-facts/.

———. 2019. "Devin Nunes and the Power of Keyword Signaling." *WIRED,* https://www.wired.com/story/devin-nunes-and-the-dark-power-of-keyword-signaling/.

———. 2022. *The Propagandists' Playbook: How Conservative Elites Manipulate Search and Threaten Democracy.* New Haven: Yale University Press. https://yalebooks.yale.edu/book/9780300248944/the-propagandists-playbook.

Tucker, Elizabeth. 2020. "Getting to Great Matches in Google Search." *Google,* https://blog.google/products/search/getting-great-matches-google-search/.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. *Attention Is All You Need.* arXiv: 1706.03762 [cs.CL].

Walls, Amanda. 2022. *The New YMYL Guidelines and What This Means for Marketers.* https://www.searchenginewatch.com/2022/09/08/the-new-ymyl-guidelines-and-what-this-means-for-marketers/.

Authors

Ronald E. Robertson is a Postdoc at the Stanford Internet Observatory (SIO).

(rer@acm.org)

Amy Dunphy is a Masters student at Stanford University.

(adunphy@stanford.edu)

Shelby Grossman is a Research Scholar at SIO.

(shelbygrossman@stanford.edu)

Renée DiResta is Research Manager at SIO.

(rdiresta@stanford.edu)

David Thiel is Chief Technologist at SIO.

(dthiel@stanford.edu)

Acknowledgements

We thank the anonymous reviewers for their helpful comments.

Data Availability Statement

Replication files will be made available after publication.

Funding Statement

Not applicable.

Conflict of Interest

Not applicable.

Ethical Standards

Throughout our study, we reported harmful content we found, such as the YouTube cryptocurrency scam (Section 7.3), through standard channels.

Keywords

Search directives; search engines; social media; indirect online influence; content discovery

Appendices

Appendix A: Collecting and Labeling Text Search Directives

The text of social media posts was often challenging to both to collect and to definitively classify as containing a search directive or not. Throughout our exploratory study (Section 4), we tracked groups of non-search-directive posts that arose during collection, and tracked edge cases that arose during annotation. In this section, we list some of the categories that emerged during that process, and provide examples from each.

Appendix A.1 Single Keyword Collection

Single-keyword searches (e.g., all posts containing "search") retrieved a large proportion of the posts that did not include a search directive. Among the posts we found containing the word "search" but not counting as a search directive were to entertainment, poems, news stories, and banal platitudes:

> *"The search to find the next #AmericanIdol continues TONIGHT! Raise your hand if you're ready for Idol Sunday!"*
>
> *"Every night, I search for you in the stars. Every night, I find you there... ..."*
>
> *"Detectives Search for Discarded Newborn's Mother as Angry Feminist Unleashes Rant https://thenewamerican.com/detectives-search-for-discarded-newborns-mother-as-angry-feminist-unleashes-rant/"*
>
> *"Individuals search for truth, groups search for consensus ~ Naval"*

The second to last example above highlights a frequent type of post returned when only filtering for the use of the word "search"—the reposting of a news story with a headline that mentions a (often non-online) search. Given these relationships to news, the frequency of posts containing "search" can substantively change over time.

Appendix A.2 Offline Searches

Posts that include the word "search" sometimes are in reference to a physical search of a real-world location. For example, we found a number of posts about Mar-a-Lago property searches, largely due to the overlap of our exploratory data collection with news of the FBI's real-world search of Mar-a-Lago. Similarly, we found a number of posts calling for Hunter Biden's laptop to be "searched." We did not code cases like these—real-world searches or digital searches without an independent intermediary—as search directives.

Appendix A.3 Search Engine Optimization

Another common type of post that we did not code as a search directive was those that included the word "search" in the context of search engine optimization (SEO):

> *England SEO Search Engine Optimization (SEO) Services Company UK #seoservices #digitalmarketing englandseo.co.uk/*

Appendix A.4 Missing or Ambiguous Context

Without additional context, some posts can initially appear to be a search directive, such as this example: *"omg.. i need to search inazuma now"*. While this appears to be a

mentioned search directive with an extractable query, "Inazuma" is actually a region in a video game, and the user is stating that they need to explore the region in the video game, not that they are going to conduct an online search for the region's name. In other examples we found, there simply was not enough context to tell if the poster was recommending an online search, or another type of search, such as the physical search of a real-world location.

Appendix A.5 Incoherent Posts

Some posts were incoherent or otherwise difficult to understand, but could potentially be classified as search directives. One short example was *"Chairman run go en ex go search hit"*. A longer example was:

> *top 10 best foods for cockroaches google search top 10 mcdonalds diners mcrib yummy food yum fries hungry cockroach chicken sandwich burger yes hungry yes dessert dessert dessert dessert*

Appendix A.6 Ads

We also found posts containing what looked like SEO-motivated advertisements that could potentially be classified as search directives. These frequently contained the phrase "Are you searching for...," and it was unclear whether they were referring to an online search or a metaphorical search. For example:

> *Perfect Aspen Therapy Centre Near Me Many variables influence how ultrasonography is used in physical therapy, including frequency, intensity, and duration of treatment. You can search for aspen therapy near me option to find the good one.*

Appendix A.7 Discussions About Search Engines

With our single keyword approach, we also found discussions about specific search engines. Especially in alternative or conspiratorial communities, there is widespread discussion about the relative qualities of search engines, with Google often considered to be censoring results, and the use of other search engines, such as DuckDuckGo, is often encouraged (Thompson 2022). For example:

> *Zero search results on DuckDuckGO: It's a second rate search engine. Zero search results on Bing: Figures, Microsoft always sucked. Zero search results on Google: They're hiding something.*

Appendix A.8 Posts in Non-English Languages

We used language filters during data collection to limit this initial investigation to English-language posts. However, we still encountered several posts which were in a language other than English, or which contained a mix of one or more languages. For example, the following post was written in Krio, and bears enough resemblance to English to complicate manual and automated classification.

> *@officials_dc @officials_dc E don tey wey I see your tweet abi na me no dey updated If I search am na the acct wey I dey see be dis @Agba_dc or na the same acct boss*

Appendix A.9 Humor and Satire

We found that search directives are often used in a humorous or satirical fashion. This can make classification challenging, because the goal of such posts is typically to entertain,

not to prompt an online search. For example, the text of one search directive we found reads *"Oh No!!!"* and its image depicts the top of a Google Search for "how long do weed effects last" on a mobile device, with a visible knowledge box that says "22 billion years in the future is the earliest possible end." Another post we found took a more satirical approach to describe what technically meets our definition for a suggested search directive with no query: *"why don't you just googoo it. wook it up on googoo."*

Appendix B: Query Extraction Challenges

Throughout our exploratory study (Section 4) and automated extraction approach (Section 5.2), we encountered a number of cases in which an exact query was especially difficult to extract. Aside from the cases discussed in the main text—no query, ambiguous queries, and obfuscated queries—we also encountered disconnected queries, queries that required the context of a longer thread, queries involving platform-specific search features, and queries in other languages.

Appendix B.1 Disconnected Queries

In some cases a query is provided, but comes in disconnected pieces meant to be combined by the recipient. For example, one post with a disconnected query suggested that readers search for "Lincoln newspapers":

> *Suggest you Google "Lincoln" and "newspapers" under the same search...*

Appendix B.2 Threaded Context

Others ambiguities can be due to the broader context of the thread or conversation a search directive appears in. For example, a post replying to a thread might tell recipients to search a query mentioned in a prior post, or might tell someone to search a modified version of a query from a prior post.

Appendix B.3 Platform Context

Some search directives tell recipients to use a platform-specific search feature, such as one post which stated: *"@[redacted] Search Groups"*. In the context of the platform and thread this post appeared on, it seems intended to tell viewers to conduct a search specifically on the Groups page within that platform rather than to conduct a general online search using "groups" as a query.

Appendix B.4 Translated Queries

Another case we found was search directives where the poster recommended search terms in non-English languages. Searching in another language can have a large impact on the search results returned (Borge et al. 2021), and could potentially be used to lead viewers to a data void without the use of a hyper-specific query. For example, one post prompted viewers to search for "western media" in Russian:

> *Western Media translated to Russian - западные СМИ (copy and paste into google image search, latest or Twitter, latest (this has a translate option). Same bots and propaganda though, but some truths.*

Table 1: Number of posts of each directive type collected and labeled from each platform

	Not Directive	Mentioned	Modeled	Suggested	Total
Gab	727	323	167	250	1,467
Twitter	723	149	98	274	1,244
Reddit	271	35	31	20	357
Parler	88	21	4	19	132
Gettr	6	3	2	19	30

Appendix C: Collecting Linked Search Directives

Seeking out linked search directives with URL fragments casts a wide net that can result in false positives due to the use of similar URLs with different subdomains. For example, the Twitter API returned posts with URLs (e.g., https://developers.google.com/search/docs) that matched one of our URL fragments (e.g., "google.com/search") but did not lead to a page of search results. However, this approach also offers greater coverage of unknown subdomains (e.g., search directive URLs for Bing included: bing.com, www.bing.com, www2.bing.com, and www4.bing.com) or query parameters (q=<query> and query=<query> both work on Google Search), that would have been difficult to predict at the start of data collection, and could change over time. Posts can contain multiple search directive URLs, and in our final dataset, 86.2% of Tweets contained only one search directive URL, and 99% of Tweets contained three or fewer. Across all posts, we found 916K unique URLs.

Appendix C.1 Short URLs

The Twitter API outputs we used provided each post's text and any URLs present, with the URLs already extracted from the text, in the majority of cases. However, the Twitter API did not extract URLs from every post it retrieved, and resolved some but not all shortened URLs (e.g., bit.ly URLs). We resolved 1M of the 1.3M short URLs we identified in our URL dataset, increasing the percentage of full URLs relative to all URLs from 80.8% to 95.6%.

Appendix D: Social Media Datasets

Appendix D.1 Reddit

The 357 Reddit posts we annotated were collected using the Reddit API with the PRAW python wrapper (PRAW 2022). We initially gathered posts from across the site, including original posts and comments, by collecting posts from the last year that contained the words "search," "look" (as in "look up"), or "google". However, to increase the number of positive examples in our dataset, we next began using only "search" to collect data. Since we are interested in trust and safety applications, we also focused on conspiracy-oriented content by limiting our collection to the subreddit r/conspiracy (1.9 million member subreddit centered around conspiracies) from the last year. While this introduces possible biases in our dataset, these biases may also be useful in surfacing potential data void queries for further examination.

Table 2: Examples of the translated search directives that we used to augment our labeled dataset.

Original Post	Translate-Augmented Text
@Paprwiz Search Disney “Clone Lab” Very frightening	@Paprwiz Searching for Disney’s Clone Lab is scary
Do an internet search for “PANDA EYES”.....	Search for “PANDA EYES” on the internet...
@JaneDoe1976 Do an internet search for ‘Dov Zakheim’.	@JaneDoe1976 Search the internet for “Dov Zakheim”.

Appendix D.2 Gab, Gettr, and Parler

Gab, Gettr, and Parler are alternative social media platforms that function similar to Twitter. However, both have a more right-wing audience and less content moderation than Twitter. To collect posts from these sites, we used the Social Media Analysis Toolkit (SMAT) third-party API (SMAT 2022), which allows users to collect all posts containing a certain word. We used SMAT to collect all of the Gettr and Parler posts, and 833 of the Gab posts, in our labeled dataset.

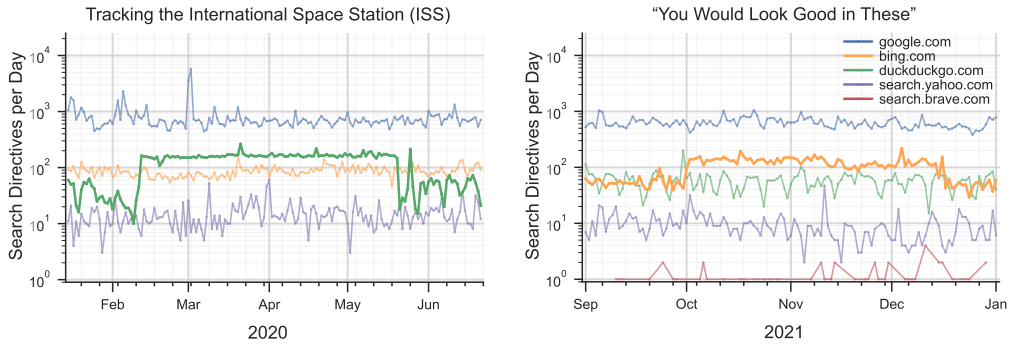
Appendix D.3 Twitter and Gab

After obtaining and annotating the posts from from SMAT, we built an early version of our search directive classifier and found it correctly classified 84% of the labeled posts we had. To expand our set of labeled examples we collected two larger datasets of of social media posts, from Twitter and Gab that contained the word “search,” using the DGAP, Stanford Internet Observatory’s data gathering tool. Using an active learning approach, we applied this early classifier to larger Gab and Twitter datasets and relabeled a random subset of the posts that the classifier marked as directives. At the end of this process, we annotated an additional 1,244 Twitter posts and 830 Gab posts.

Appendix D.4 Translation-Augmented Data

In order to augment our labeled dataset, we used the Google Cloud Translate API to translate examples with queries into Chinese and then back into English. Chinese was chosen as a language very different than English, which would therefore perturb the grammar and structure of the post enough for the translation to be distinct from the original. Through this process, we added a total of 100 posts to our dataset. We manually relabeled the augmented data to account for the small changes in the search terms themselves (especially cases of singular vs. plural, which translate poorly between English and Chinese) that also occurred during this process.

In Table 2 we provide examples of original posts and their translate-augmented versions, which were created by translating into Chinese and back using Google translate. The translation perturbs the sentence structure, while keeping the directive itself. 100 pieces of augmented data were created and then hand-relabeled, in order to account for changes in queries. Almost all of the 100 augmented posts were suggested directives. Both the augmented data and the posts that we augmented data from were included exclusively in the training set, in order to prevent situations where near-identical posts were in the training and test sets.



(a) The sustained spike in DuckDuckGo search directives that spanned February through May 2020 was due almost entirely to a single automated account that was posting links where the query was a set of GPS coordinates. This account's GPS posts constitute 87.8% of all DuckDuckGo search directives in February, 91% in March, 89% in April, and 83.8% in May.

(b) The query "leggings" was a major driver of this spike, accounting for 68.1% (n = 2,558) of all Bing search directive URLs in October, 63.8% (n = 2,086) in November, and 45.4% (n = 985) in December. These search directives often had the same text, with one ("hey you would look good in these! <url>") accounting for 93% (n = 5,242) of all such links.

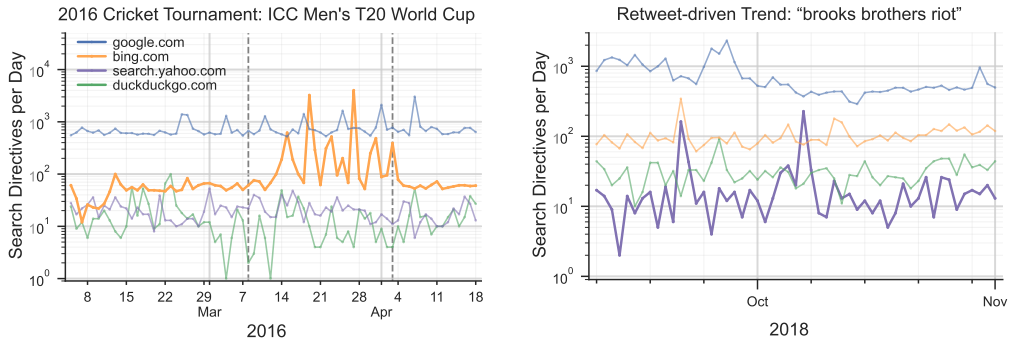
Figure 10: Sustained bursts in linked search directives on Twitter that lead to DuckDuckGo (10a), and Bing (10b). Both of these bursts appear to be the result of automated activity, but the burst leading to DuckDuckGo was due to a single account, while the burst leading to Bing was due to many accounts.

Appendix E: Additional Results

Appendix E.1 Exogenous Shocks

In addition to the exogenous shocks we highlighted in Figure 6 of the main text, we investigated other bursts in activity around a variety of event types. Among these were a wide variety of events, including a sports tournament (Figure 11a). However, we also found bursts in activity that were not due to a single event, but rather due to a single post that was retweeted many times (Figure 11b), a single automated account posting frequently for a short period (Figure 10a), or a collection of potentially automated accounts posting similar versions of the same text (Figure 10b).

One burst in search directives leading to DuckDuckGo was sustained for two months, starting in February 2020 (Figure 10a). During this time, the daily number of search directives linking to DuckDuckGo rose by an order of magnitude, from about 20 per day to 150 per day. However, the spike in DuckDuckGo search directives was not due to a query rapidly increasing in popularity, but due to a single account named "COMP680_bot," which began tweeting out template text ("The International Space Station is currently here:") followed by a linked search directive with a unique set of GPS coordinates as the query (e.g., "-0.2182, 174.1660"). When we examined the time between posts for this account, we found a fairly consistent average time of 10.9 minutes. Given these signals indicating account automation—the use of templates and consistent posting schedule—and that the account's screen name appears to contain a course listing ("COMP680"), it is likely that this spike in DuckDuckGo search directives on Twitter was due to a computer science course project. The account abruptly stopped posting linked search directives on 2020-05-26, at which point it appears to have switched to posting Google Maps screenshots showing the location of the International Space Station. This case also demonstrates that value of documenting the URL arguments commonly used by popular search engines to categorize the type of search results they are returning, as these URLs contained a parameter ("&iaxm=maps") that lead to DuckDuckGo's map search results. In this search context, GPS coordinates are a special query case where only one result is possible.



(a) A spike in activity around the ICC Men's T20 World Cup 2016, a biannual cricket tournament that took place in India from March 8 to April 3, 2016 (marked by the dashed vertical lines). During this window of time, most of the search directive queries for Bing were team name abbreviations. For example "ind vs aus" (27.7% of all Bing queries) and "ind vs pak" (22.5%).

(b) The query "brooks brothers riot" drove this spike and accounted for 45.3% (n = 240) of all Yahoo search directives in October 2018. These were almost all retweets of a single post containing a link to a page of Yahoo image search results for "brooks brothers riot," the event which ended the U.S. Presidential Election recount. Over half of these retweets occurred on 2018-10-07, the same day as the original post.

Figure 11: Temporal spikes in linked search directives on Twitter that lead to Bing (11a) and Yahoo Search (11b). A legend for the search engines appearing in these time periods is available in Figure 11a, and in both plots, activity for the search engine of focus is distinguished by a bolder line.

Similarly, we found a sustained spike in search directives for Bing between 2021-09-01 and 2022-01-01 that was also almost entirely driven by templated queries (Figure 10b). This trend was also driven by a small number of users, with 10 users accounting for 75.8% of these search directive posts. Last, for Yahoo Search we found a search directive trend that was largely driven by retweeting of a single post containing a search directive leading to Yahoo image search results for "brooks brothers riot" (Figure 11b).

Appendix F: Automated Query Extraction

The model we fine-tuned to automate the extraction of queries from search directives often produced errors on posts that were also difficult for human raters to decisively extract queries from. Among such challenges were ambiguities in terms of the start and end of a query, multiple queries in a single post, and instances in which the extractor shortened or changed the spelling of a word (Table 3). For search directive posts that contained a search query, the accuracy for exact matches for all queries was 67% (true positives). For posts with multiple queries, the model was able to correctly retrieve at least one of the queries in the post for 71% cases, and among the cases where the extracted query didn't match, 36% were cases where it incorrectly reported there was no query (false negatives).

Table 3: Paired examples of human and machine-extracted search queries. The first two rows show cases where it can be ambiguous where a query starts and ends, a condition which our human raters also struggled with. The third row shows a case where a search is mentioned, and the model correctly extracted a query that our human raters missed. In the fourth row, we show a case where the model failed to retrieve all queries, but did correctly retrieve one. Last, the fifth row show a rarer cases where the classifier changed the spelling of a word, or changed a specific word.

Original Post	Human Query	Extractor Query
just search up like uwu aesthetic 1d headers for twitter :3	uwu aesthetic 1d headers for twitter	uwu aesthetic 1d headers
You can search physician deaths by year. 2018 has one doctor's death listed. He was 90. 2022 already has 150 physician deaths. What has changed?	physician deaths by year	physician deaths
An immediate search finds this article from March 2, 2021 And another April 15th article from same year, 2021 wsj.com/articles/almost-half-of-federal-cases Does anyone have the update of these "79 ANTIFA rioters" having their charges dropped?	None	ANTIFA rioters
Hmm.. strange (they are right, I tried searching - there is one result for pericarditis and many for heart attack) Well isn't this just a little Suspicious...When searching Myocarditis on the NHS Website, it says No Results are found....	pericarditis, heart attack, Myocarditis	Myocarditis
No Person search results found for "Johnny Massacre"	Johnny Massacre	John Massacre