Stanford Law School
Law and Policy Lab

Fake News & Misinformation Policy Practicum

Google's Role in Spreading Fake News and Misinformation

RESEARCH REPORT BY:

Danaë Metaxa-Kakavouli, Stanford University, PhD Candidate in Computer Science
Nicolás Torres-Echeverry, Stanford Law School, JSD Candidate

This manuscript is part of a longer report, *Fake News and Misinformation: The Roles of the Nation's Digital Newsstands,* https://www-cdn.law.stanford.edu/wp-content/uploads/2017/10/Fake-News-Misinformation-FINAL-PDF.pdf

This manuscript is a working paper and has not been peer-reviewed.
First version upload date: October 31, 2017
Key Words: fake news; misinformation; search engine optimization; algorithms; news media

INSTRUCTOR AND PROJECT LEADS:

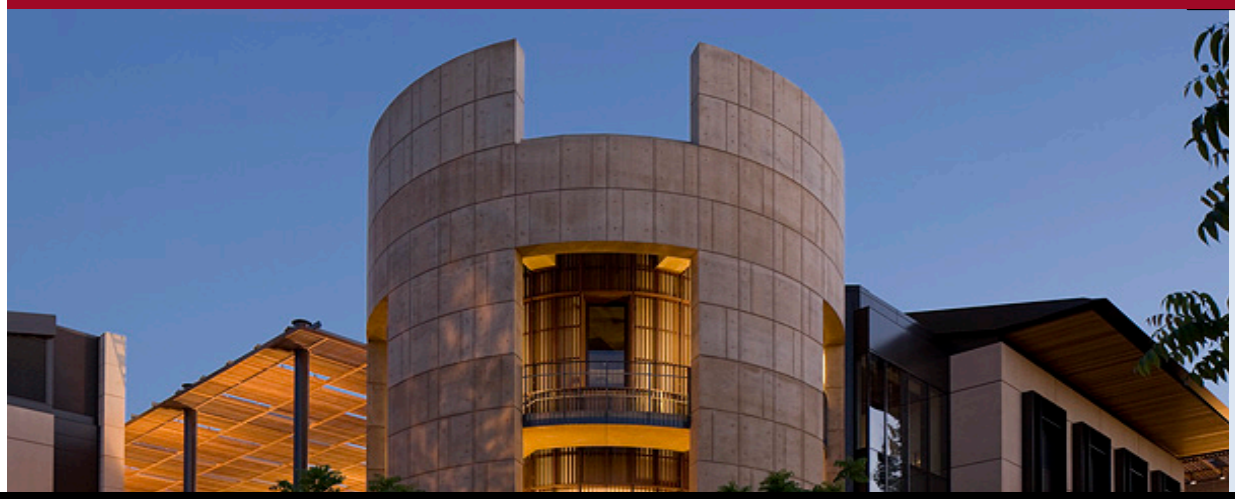
Senator Russ Feingold
Edwin A. Heafey, Jr., Visiting Professor of Law

Luciana Herman, Ph.D.
Lecturer & Program Director, Law and Policy Lab

POLICY CLIENT:

Hewlett Foundation Madison Initiative

Octobe
2017



Contents

Google's Role in Spreading Fake News and Misinformation

Danaë Metaxa-Kakavouli[1,*] and Nicolás Torres-Echeverry[2,*]

Abstract

This paper analyzes Google's role in proliferating fake news and misinformation in the months leading up to and immediately following the U.S. 2016 national election. It is one section of a longer report, *Fake News and Misinformation: The roles of the nation's digital newsstands, Facebook, Google, Twitter and Reddit,* that serves as the first phase of a continuing inquiry over the 2017-18 academic year.[3] This paper reviews the role of Google, and specifically Google Search, in the misinformation landscape. It tracks the problem of misinformation in search engines from the advent of search engine optimization and spam through the present day, focusing on Google's efforts to curb its role in spreading fake news following the 2016 U.S. elections.

Part 1 describes the "arms race" between search engines and spammers exploiting weaknesses in search algorithms, which contributes to Google's role in proliferating fake and/or biased news in the 2016 elections. As part of the continuing accounting of the impact of fake news and misinformation on the 2016 elections, this analysis tracks search results for senate and presidential candidates in that election, revealing that up to 30% of these national candidates had their search results affected by potentially fake or biased content.

Part 2 summarizes Google's recent efforts in 2017 to curb misleading or offensive content through user reporting and human reviewers, along with the opinions of users and experts who are largely supportive of these changes. The section broadly reviews the influence of the Internet on journalism, and then describes Google's recent efforts to invest in initiatives that bolster investigative journalism and news. It concludes with suggestions for policy and research directions, recommending in particular that Google and other companies increase data transparency, in particular for researchers, to better understand misinformation phenomena online. The study concludes that transparency and civilian oversight are the next critical steps towards a society which benefits fully from the ubiquitous and powerful technologies that surround us.

[1] Stanford University, School of Engineering, PhD Candidate in Computer Science
[2] Stanford Law School, JSD Candidate
* Correspondence: metaxa@stanford.edu; nicolast@stanford.edu
[3] *Fake News and Misinformation: The roles of the nation's digital newsstands, Facebook, Google, Twitter and Reddit,* October 2017, https://www-cdn.law.stanford.edu/wp-content/uploads/2017/10/Fake-News-Misinformation-FINAL-PDF.pdf.

I. Introduction

Just as the technologies of radio and television once did, the Internet is reshaping democracy worldwide. The 2016 elections in the United Stated and abroad proved the importance of social media networks like Facebook and Twitter in influencing voting behavior and political views. Fake news and misinformation proliferated on these platforms, posing new challenges to democratic political processes. While we do not yet—and may never fully—know precisely how the spread of fake news and misinformation across these platforms affected political views and outcomes, the events of the recent election, compounded by elections in France and perhaps the U.K. and Colombia, reveal the importance of the platforms on democratic institutions. Although Google did not attract as much attention as did social media platforms on the topic of false information, it remains a crucial actor in this landscape. Google's web search engine is a monolithic intermediary between users and content on Internet, providing information that helps to shape ideas, including political perspectives.

This section answers questions regarding misinformation and modern technology in relation to Google Search: How are democratic and political processes in the United States today affected by Google's search engine? What solutions could be implemented at the platform level and/or though regulatory actions to benefit an informed democracy?[4] This section answers these questions in a six-part analysis. First, it provides an overview of Google, and specifically the relevance of Google Search in today's society, and how the search engine's algorithm has evolved over time to address historical manifestations of spam and misinformation. Second, it discusses the role of Google Search in the 2016 U.S. national elections, and the reasons for the recent wave of misinformation. Third, it assesses the current state of response in the context of Google's interventions with these recent political challenges. Fourth, it provides a brief overview of the relationship between the fake news phenomenon and journalism. Fifth, it describes the legal framework for search engines, focusing on intermediary liability. And finally, it offers public policy recommendations, ranging from the platform level to civil and regulatory actions.

This summary draws on information in three forms: (1) two quantitative analyses, one of a data set of Google search results for U.S. congressional candidates in 2016, and the other of survey data from 475 Google Search users collected in May 2017 through Amazon Mechanical Turk; (2) a qualitative analysis of eight semi-structured interviews with key academic and industry figures; and (3) a review of relevant legal cases and published

[4] Recent reports by the research institute Data and Society, and by Harvard and Northeastern universities, highlight the importance of addressing these questions at the platform level. See Alice Marwick and Rebecca Lewis, "Media Manipulation and Disinformation Online" (Data&Society, 2017), https://datasociety.net/output/media-manipulation-and-disinfo-online/ ; see also, David Lazer, et al., "Combating Fake News: An Agenda for Research and Action," May 2, 2017, https://shorensteincenter.org/combating-fake-news-agenda-for-research/; and Robert Faris, et al., *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election, August 2017,* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3019414.

documents, including academic articles and popular press coverage.[5]

II. Background: Google Search Today and Its Evolution over Time

The Role of Google Search in Today's Information Economy

Nearly half of the world's population – 3.5 billion people – accesses the Internet regularly.[6] According to the International Telecommunication Union 2016 Report, "People no longer *go* online, they *are* online…. Internet users read, shop, bank and date online, thanks to a growing number of websites, services and applications that did not exist a decade ago."[7] The quantity of data available on the Internet is similarly astounding. As emblematized by Wikipedia, the number of articles reached 40 million in 2016.[8] In order to make this information accessible in human scale, it must be preprocessed, indexed, and made searchable; this is the role of a search engine. Legal scholar James Grimmelman aptly points out that "The Internet today is usable because of search engines."[9]

Google's search engine receives between 3.5 and 5.5 billion queries per day.[10] This traffic has grown exponentially. When Google launched in 1998, there were 10,000 daily queries; by 2006, the same amount was searched every second, and by 2017 in less than a tenth of a second.[11] Table 1 presents the demographics of search engines from a 2012 survey made by the Pew Research Center.[12] There is no gender difference in the use of search engines but its use is biased towards white, young, more educated, and higher income populations. These demographics are for all search engines. Pew, however, reports that more than 80% of search engine users use Google Search, suggesting that these demographics describe Google users.

[5] For more information on the quantitative and qualitative methodologies see the Appendix.
[6] International Telecommunication Union, "Measuring the Information Society Report," 2016, http://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2016/MISR2016-w4.pdf.
[7] *Ibid.*, 182.
[8] The number of articles was 3.9 million in 2006.
[9] James Grimmelmann, "The Google Dilemma," *NYL Sch. L. Rev.* 53 (2008): 941.
[10] "Google Search Statistics - Internet Live Stats," accessed June 10, 2017, http://www.internetlivestats.com/google-search-statistics/; Danny Sullivan, "Google Now Handles at Least 2 Trillion Searches per Year," *Search Engine Land*, May 24, 2016, http://searchengineland.com/google-now-handles-2-999-trillion-searches-per-year-250247.
[11] Sullivan, "Google Now Handles at Least 2 Trillion Searches per Year."
[12] Kristen Purcell, Joanna Brenner, and Lee Rainie, "Search Engine Use 2012" (Pew Research Center, March 9, 2012), http://www.pewinternet.org/2012/03/09/search-engine-use-2012/.

Who uses search?

% of online adults in each group who use search engines

	% of each group who ever use search	% of each group who used a search engine yesterday
All online adults	91%	59%
Gender		
Male	90	59
Female	92	60
Race/Ethnicity		
White	93*	63*
African American	89*	44
Hispanic	79	44
Age		
18-29	96	66*
30-49	91	65*
50-64	92	52*
65+	80	38
Education		
Some high school	78	34
High school	88*	45*
Some college	94*	65*
College graduate	95*	74*
Household income		
< $30,000	84	45
$30,000 - $49,999	93*	54*
$50,000 - $74,999	97*	66*
$75,000+	95*	76*

* Denotes statistically significant difference with other rows in that category

Source: The Pew Research Center's Internet & American Life Project Winter 2012 Tracking Survey, January 20-February 19, 2012. N=2,253 adults age 18 and older, including 901 cell phone interviews. Interviews conducted in English and Spanish. The margin of error is plus or minus 3 percentage points for internet users.

Table 1. Pew Research Center 2012 report on the demographics of search engine users.[13]

In addition to its widespread use, Google has also appropriated the biggest share of this market. Internationally, 89% of searches occurred using Google in mid-2015;[14] in the US,

[13] *Ibid.*, 6.
[14] Martin Moore, "Tech Giants and Civic Power" (Centre for the Study of Media, Communication and Power, April 2016), https://www.kcl.ac.uk/sspp/policy-institute/CMCP/Tech-Giants-and-Civic-Power.pdf.

65% of searches took place on Google.[15] The search engine owes its success to popular satisfaction with its search results. 88% of the respondents in our survey reported that they are somewhat or extremely satisfied with the information they find using Google Search. (Although the Practicum survey is not representative of the U.S population, the percentage may highlight a trend in users' satisfaction.) Such satisfaction likely reflects Google's agile algorithm which responds automatically to the creation of new webpages and the creativity of its users; indeed, 15% of the queries Google receives each day have never been searched on Google before.[16]

As a tool that enables individual users access to a vast body of knowledge, Google and other search engines are also critical to defining ideas and shaping the success of businesses, events, and the political landscape. Grimmelmann exemplifies this role with five iconic cases.[17] One shows that there is information we can find only because of search engines, for instance the existence of "mongolian gerbils." A second explains how open source content allows information to flow freely, and this available content enables the search engine to organize and display information.[18] But this also implies that humans can affect the rankings being displayed: for example, a third case describes the way the query "jew" caused search engines, using algorithms based on user behavior, to display controversial content. A fourth case presents the importance of search engine rankings to business, recounting how companies have argued that a low ranking would irreparably damage their economic viability. Finally, Grimmelmann's fifth case relates to politics online: a search for "Tiananmen" displayed different results in different countries, based on how the algorithm responded to queries related to those countries' political contexts. Grimmelmann concludes, "Search results matter: to culture, to business, and to society. Every decision about search, and every lawsuit about search, has these inescapably political questions at its heart. That is Google's dilemma, and ours."[19]

Important to an understanding of Google's role in this socio-cultural landscape is an understanding of the market structure in which Google operates. Google receives 89% of its revenues from advertising.[20] The market structure of Google search engine is two-fold. On one side, the search engine provides a service to users in exchange for their user data and who consent to see ads linked to their queries. On the other side, companies pay Google to place the companies' advertisements on relevant search results. The first market between Google and users is closer to a monopoly, the second one between Google and other companies to an oligopoly. In the first case, Google has more freedom to fix the nonmonetary price—the amount of data it extracts from users and the uses it gives to these

[15] "Google Search Statistics - Internet Live Stats."
[16] Ben Gomes, "Our Latest Quality Improvements for Search," *Official Google Blog*, April 25, 2017, http://blog.google:443/products/search/our-latest-quality-improvements-search/.
[17] Grimmelmann, "The Google Dilemma."
[18] In interviews with our research team, Google officials highlighted this aspect of the web as well.
[19] Grimmelmann, "The Google Dilemma," 940-950.
[20] "Google: Distribution of Revenue by Source 2016," *Statista*, 2017, https://www.statista.com/statistics/266471/distribution-of-googles-revenues-by-source/.

data. In the latter, Google must compete with other platforms and media companies for the business of companies seeking advertisement outlets.

Google's Historical Efforts to Combat of Misinformation

The algorithm behind Google's search engine has evolved constantly since its inception. In addition to regular improvements, changes in the algorithm have frequently been triggered in response to spammers attempting to manipulate the system. This process has been termed an "arms race" and has long affected Google and all other search engines. Historical search engines like AltaVista and Lycos were also known to change their algorithms to thwart spammers.[21]

In the early 1990s, first-generation search engine algorithms were based on *vector models* of documents, a theoretical formulation from the field of Information Retrieval. Metaxas writes, "the more rare words two documents share, the more similar they are considered to be."[22] This model soon came under attack from spammers, who began a practice that became known as *keyword stuffing*—creating pages with many rare keywords (sometimes hidden by matching the text to the background color to avoid detection by users on the page), so as to get their pages ranked highly for many different, unrelated user queries.[23]

In response, by 1996, search engines had developed second-generation algorithms with more sophisticated techniques, for instance relying on connections in the network of pages to determine credibility: popularity—where many pages link to any single given page—was taken as an indicator of quality. In response, spammers created *link farms*, clusters of spam pages all linking to each other to help each other appear popular and, therefore, rise in the rankings.

The third generation of search engine algorithms, introduced with Google's famous PageRank algorithm in 1998, built on the idea of a popularity network but weighted the votes of highly reputable pages (*i.e.*, a page with more links to it than pages linked to by it) more than less reputable pages. Unfortunately, this method is also vulnerable to spam: spammers acquire legitimately high rankings on pages about unrelated topics, and then use that reputation to elevate other pages (which might, for instance, be full of ads providing revenue to the spammer).[24] This tactic was known as forming a *mutual admiration society.*

This process of gaming a search engine to move one's webpages higher in a search engine's ranking is called *Search Engine Optimization* (often abbreviated "SEO"). Notably, this kind of spamming is not uncommon; legitimate companies and individuals often employ

[21] Interview with Takis Metaxas, May 1, 2017.
[22] Panagiotis Takis Metaxas, "Web Spam, Social Propaganda and the Evolution of Search Engine Rankings," in *Web Information Systems and Technologies* (International Conference on Web Information Systems and Technologies, Springer, Berlin, Heidelberg, 2009), 176, doi:10.1007/978-3-642-12436-5_13.
[23] Metaxas, "Web Spam, Social Propaganda and the Evolution of Search Engine Rankings."
[24] *Ibid.*

SEO experts to help publicize their content online. In this sense, many of us are small-time propagandists without posing much risk to our communities. The danger lies in those who would use these techniques to spread falsehoods and misinformation, and lead others astray in high-risk and high-stakes contexts: for instance, in matters of politics.

One notable example of the entrance of politics into SEO is the "miserable failure" hack. Third wave search engine algorithms, in addition to PageRank, started using *anchor text* (the text that a website matches with a URL link) to learn something about the linked URL. For instance, if a page wrote something like "Many newspapers [link: nytimes.com] have sports sections," Google would learn that the concept "newspapers" was related to the Times' website. Capitalizing on this feature, a group of activists started using the phrase "miserable failure" to link to President George W. Bush's official webpage, causing Google searches for George W. Bush to yield this phrase.[25] This tactic, termed a *Googlebomb*, gained significant publicity and was deployed against other politicians as well. The enormous potential for political impact of web (and particularly web search) technologies, has been recognized as a significant new form of politicking.[26]

The unspoken fuel behind this phenomenon is the implicit trust placed in search engines like Google by their users. "Users have come to trust search engines as a means of finding information, and spammers have successfully managed to exploit this trust."[27] More often than not, we do not interpret the results of every Google search we make with skepticism. (After all, how many did you make today? That would be exhausting.) Instead, we have learned a heuristic: that Google's results are almost always helpful and well-ordered. This pattern ingrains an unconscious instinct to trust Google's ranking deeply, and to correlate authority with ranking—meaning that companies ranked highly will see their sales and subscriptions rise, politicians ranked highly may be perceived as more credible, and highly ranked spammers' pages will garner more clicks (and, along with that, ad revenue).[28]

In addition to spammers manipulating Google's search rankings, another source of bias is an accidental byproduct of crowdsourced data that Google's algorithms learn from. Like many algorithms powered by machine learning and fed data generated by (flawed, biased) human users, Google Search is vulnerable to displaying the same biases. For example, in 2016 Google had to manually alter its search suggestions (the suggestions that appear when a user begins typing a search query) to remove autocomplete results that appeared when a user started a query with "are jews" or "are women." Before those changes, Google's algorithm, which used machine learning to autocomplete with common phrases from other

[25] *Ibid.*
[26] Grimmelmann, "The Google Dilemma."
[27] Metaxas, "Web Spam, Social Propaganda and the Evolution of Search Engine Rankings," 171.
[28] Bing Pan et al., "In Google We Trust: Users' Decisions on Rank, Position, and Relevance," *Journal of Computer-Mediated Communication* 12, no. 3 (April 1, 2007): 801–23, doi:10.1111/j.1083-6101.2007.00351.x. This article studies the perception of Google's search algorithm by college students. Using an eye tracking experiment, the authors show that students are strongly biased towards links higher in Google's results (even if those pages were less relevant.)

users, would suggest ending either phrase with the word "evil."[29] This example, among others, indicates another potential site of misinformation in Google's Search features—those caused by the accidental effect of machine learning trained with biased data. Such algorithmic bias is the topic of a growing field of study, and points to the need for better data collection and transparent research surrounding search engines. While problematic patterns in autocorrect might become obvious to end users who see them directly, subtler patterns (for instance, a consistent slant in search results for certain political topics) might be harder to catch, and nearly impossible to prove as a systematic problem. Without a thorough research framework for collecting and analyzing data, these sorts of issues risk go unexamined.

To address a third possibility for bias, our research further considered both the possibility of Google intentionally influencing its search results or user search suggestions to achieve some political outcome and the company's occasional explicitly political messages to users. Our external examination of the algorithm did not reveal evidence of manipulation of the algorithm for political ends. In fact, our research suggests that Google strives for transparency when it seeks to achieve particular political outcomes. The company has occasionally leveraged its logo, for example, to communicate a particular political message. In January 2012, "Google blacked out its logo in protest against the Stop Online Piracy Act (SOPA) and Protect IP act (PIPA). It also urged people to sign a petition against the bills. On 22nd May 2015 Google's homepage in Ireland told its users they should #VoteYes in the Irish referendum on gay marriage."[30] These occurrences are akin to any company lobbying for its business values, and not a novelty of the technology era. However, given Google's unprecedented availability to billions of people around the world, this form of influence is worth keeping in mind.

III. Misinformation and the 2016 U.S. National Elections

Related Literature

In the months following the 2016 U.S. national elections, a literature is emerging, studying the impact of fake news the election and its outcome. Most of these studies have focused more on social media than on search engines, making this report one of the few studies in the field.

Imperative to the field is a survey conducted by Allcott & Gentzkow (2017) that tracks online news search behaviors. The survey, which is representative of the U.S population, finds that 28.6% of respondents reported receiving political news primarily online, either

[29] Carole Cadwalladr, "Google, Democracy and the Truth about Internet Search," *The Guardian*, December 4, 2016, sec. Technology, https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook.
[30] Moore, "Tech Giants and Civic Power," 28.

from social media (13.8%) or websites (14.8%).[31] In comparison, a majority (57.2%) report receiving their political news primarily through television. (The survey did not examine respondents' secondary or tertiary sources of political information.) With regard to search engines, specifically, Allcott & Gentzkow find that between October and December of 2016, 30.6% of visits to pages hosted by the 690 top news sites came from search engines in contrast with 10.1% coming from social media sites. Further, 22.0% of visits to a list of 65 fake news sites came from search engines compared to 41.8% from social media sites.[32] These twin findings suggest that social media sites, rather than search engines, may have been primary drivers in the spread of fake news in 2016.

While there may be less cause for concern regarding the magnitude of the role of the Google search engine in amplifying fake news, some vulnerabilities persist. Some psychologists argue that exposure to different balances of favorable and unfavorable search result rankings can sway voters' opinions on candidates, even when those voters do not visit or fully process the information on the pages shown in the ranking. This effect has been termed the Search Engine Manipulation Effect, or "SEME."[33] Further, psychologists have also found that content can be digested unconsciously by users who are unaware of the influence on their beliefs. Such unconscious or biasing effects include peripheral persuasion, confirmation bias, and misinformation correction. We believe that further studies with methodologies able to capture psychological effects may arrive at different conclusions, give a more comprehensive understanding, and find that fake news may, in fact, have a more significant role in political outcomes.

A recent report by Marwick and Lewis (2017) of Data and Society reflects on these psychological factors. The study details media manipulation by various actors including trolls, the "Alt-Right", conspiracy theorists, politicians, and others. The report highlights how distrust and polarization created a feasible environment for the fake news phenomenon and concludes: "we can expect the continuation of some current trends: an increase in misinformation; continued radicalization; and decreased trust in mainstream media."[34] While this report does not address these topics from the internal perspective and management decisions of the platforms, it does examine the role of platforms including Google in perpetuating the spread of misinformation, both directly and indirectly, in terms of the impact the platforms have had on journalism.[35]

Reasons for Misinformation

In our interviews, Google management and experts in the field both were careful to

[31] Hunt Allcott and Matthew Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives* 31, no. 2 (May 2017): 211–36, doi:10.1257/jep.31.2.211.
[32] *Ibid.*
[33] Robert Epstein and Ronald E. Robertson, "The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections," *Proceedings of the National Academy of Sciences* 112, no. 33 (August 18, 2015): E4512–21, doi:10.1073/pnas.1419828112.
[34] Marwick and Lewis, "Media Manipulation and Disinformation Online," 44.
[35] Marwick and Lewis, "Media Manipulation and Disinformation Online."

distinguish between two separate causes of fake news and misinformation: financial incentives and political manipulation.

Evidence of the role of financial incentives in proliferating fake news—news that is patently and purposely false—implicated Google. Well-documented news reveal that Macedonian teens developed websites peddling sensationalist, patently false, stories and hoaxes in order to gain money through a Google advertising algorithm that paid them every time their pages were viewed.[36] Indeed, Google AdSense turns a profit via an algorithm that connects companies advertising their products to websites looking to host ads, and delivers a small portion of that profit to the host sites. Google does place some restrictions on which online publishers are allowed to host ads through AdSense, but those restrictions do not bar "fake news" sites.[37] In this way, people producing fake stories, for instance propagating the so-called Pizzagate conspiracy theory, could bring in revenue (with a cut going to Google) by allowing Google to match advertisements to sites. Google has recently pledged to remove from its network any publishers presenting information under "false or unclear pretenses."[38] Notably, this would still allow the Pizzagate pages to continue working with Google AdSense.

The second motivation for producing misinformation is for political gain, which is perhaps especially dangerous for an informed democracy. In such cases, actors, either foreign or domestic, may pursue Google as a vehicle for spreading their biased or incorrect perspectives with the hope of manipulating American voters. For instance, Stanford Professor Larry Diamond believes there is sufficient evidence to show that Russia intentionally manipulated the 2016 U.S. presidential election for its own political gain. In such cases, curbing ad revenue is insufficient to address the problem, and online sources used by citizens to inform themselves will be a primary target. Unfortunately, it is also harder to quantify this subset of the problem, since many websites promote biased content for political gain, and the metric with which to measure that bias can be subjective.

2016 Election URL Search Analysis Results

To investigate the issue of biased news on Google and its role in the 2016 elections, we accessed an unpublished data set collected by researchers at Wellesley College. The data are comprised of top 10 Google Search result URLs (the first page of results) collected twice per week in the 26 weeks prior to the November 2016 U.S. national elections, a total of 141,313 URLs. The results come from Google searches of all congressional candidate names (incumbents and all other candidates running) in six U.S. states (Arkansas, Arizona, Colorado, California, Alaska, and Alabama). Additionally, the data set contained URLs from searches for four presidential candidates: Clinton, Rubio, Sanders, and Trump.

[36] Samanth Subramanian, "Inside the Macedonian Fake-News Complex," *WIRED*, February 15, 2017, https://www.wired.com/2017/02/veles-macedonia-fake-news/.
[37] Ginny Marvin, "Google Isn't Actually Tackling 'Fake News' Content on Its Ad Network," *Marketing Land*, February 28, 2017, http://marketingland.com/google-fake-news-ad-network-revenues-207509.
[38] *Ibid.*

Building on the Wellesley data set, we compare these URLs against the 150 URLs in PolitiFact's "guide to fake news websites," which flags sites that have in the past produced disputed content.[39] (We also considered other available lists, such as that curated by Melissa Zimdars,[40] but settled on PolitiFact because it is not only carefully vetted but also relatively short, focusing on popular sites.) That comparison enabled us to check how many, if any, of the top ten results in the data set are from fake or questionable news sites.

	Overall number	Number flagged by PolitiFact's list	Percentage flagged of total
All search result URLs	141,313	2,152	1.52%
Unique search result URLs	9,573	283	2.95%
Politicians	356	103	28.9%

Table 2. Findings in the URL analysis show that over 1.5% of all results shown for the politicians in our data set were from disputed sites, and that this amounted to nearly a third of all politicians in the data set having their Google Search results affected by the presence of these sites.

As summarized in Table 2, our analysis shows that over 1.5% of all URLs in our data set belonged to websites disputed by PolitiFact. When filtering for only unique sites, this rises to 2.95%. This indicates that the non-flagged sites were more likely to repeat (appearing consistently from week to week) in our data set than the flagged sites—in other words, flagged sites were more likely to appear and disappear, rather than stay consistently in the top 10, a fact which speaks to the ephemeral, quick-reaction, non-authoritative nature of these sites. Furthermore, we find that when measuring the 356 politicians in our data set, nearly a third had their results affected by these flagged sites in the time period over which data was collected.

[39] Joshua Gillin, "PolitiFact's Guide to Fake News Websites and What They Peddle," *PunditFact*, April 20, 2017, http://www.politifact.com/punditfact/article/2017/apr/20/politifacts-guide-fake-news-websites-and-what-they/.
[40] Melissa Zimdars, "Resource-False-Misleading-Clickbait-Y-and-Satirical-'News'-Sources-1.pdf," November 2016, http://d279m997dpfwgl.cloudfront.net/wp/2016/11/Resource-False-Misleading-Clickbait-y-and-Satirical-%E2%80%9CNews%E2%80%9D-Sources-1.pdf.

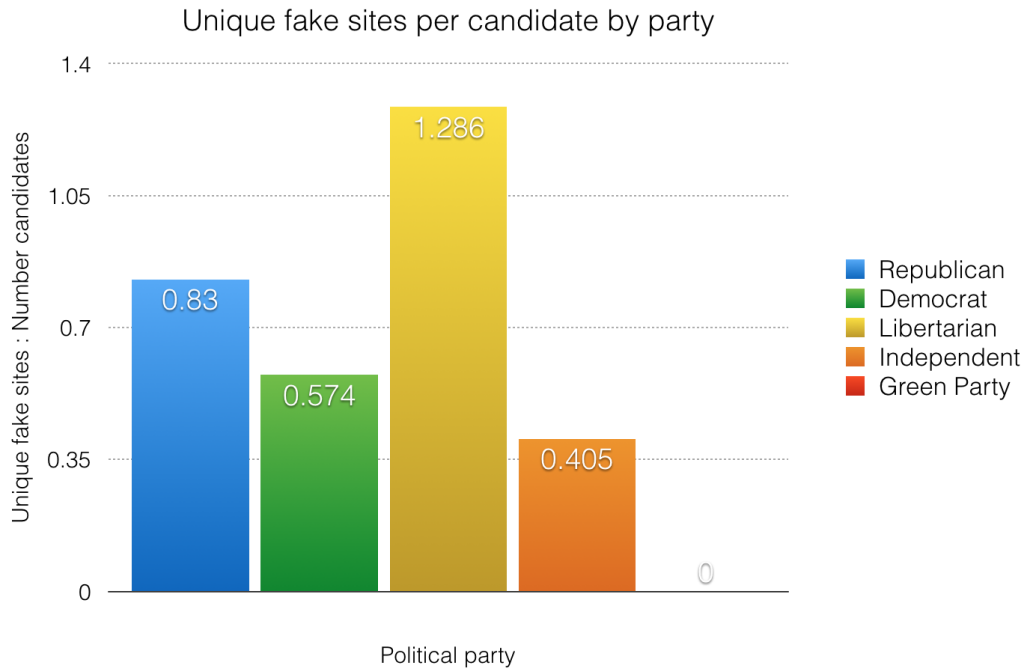


Figure 1. Considering the presence of disputed sites in search results by candidate, we find that Libertarian candidates were much more likely to have their results display URLs from disputed sites than were other candidates.

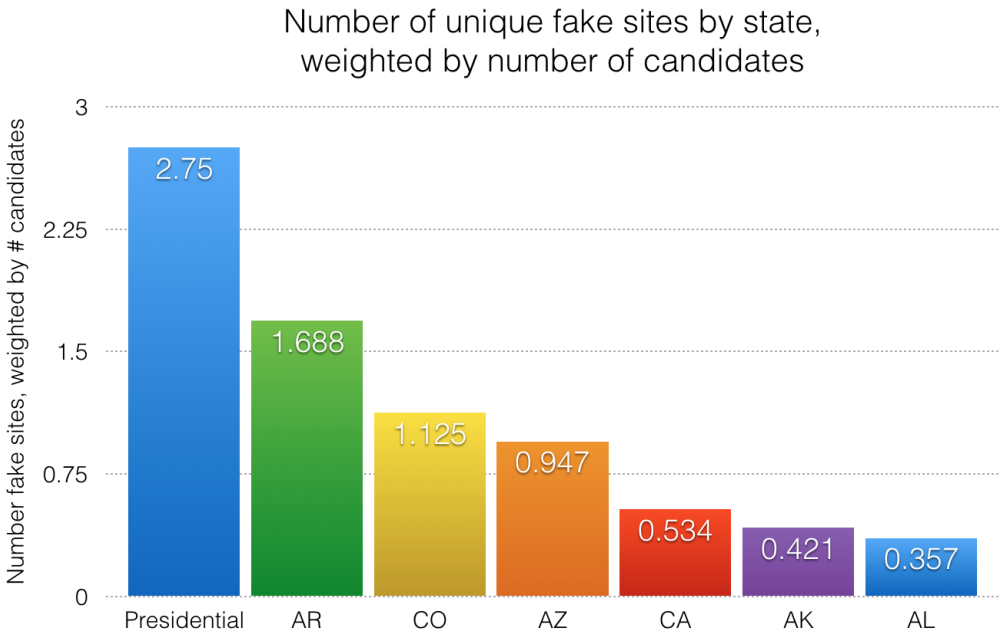


Figure 2. When grouping candidates by position, we find that the candidates in the presidential race were more likely than congressional candidates to see disputed sites in their search results, confirming our hypothesis.

When analyzing by political party, we find that Libertarian candidates were much more likely than others (and Republicans more likely than Democrats) to have their results affected by disputed sites (Figure 1). This may be in part because these less-mainstream candidates inspire coverage from less-mainstream sites whose journalistic standards are sometimes lacking, or because the political positions of those candidates or the journalistic standards of their constituents inspire more editorializing. We also observe that the four presidential candidates were more likely to have their search results affected than were congressional candidates, confirming our hypothesis that the highest-stakes race would attract more spammers and others peddling misinformation or biased content (Figure 2).

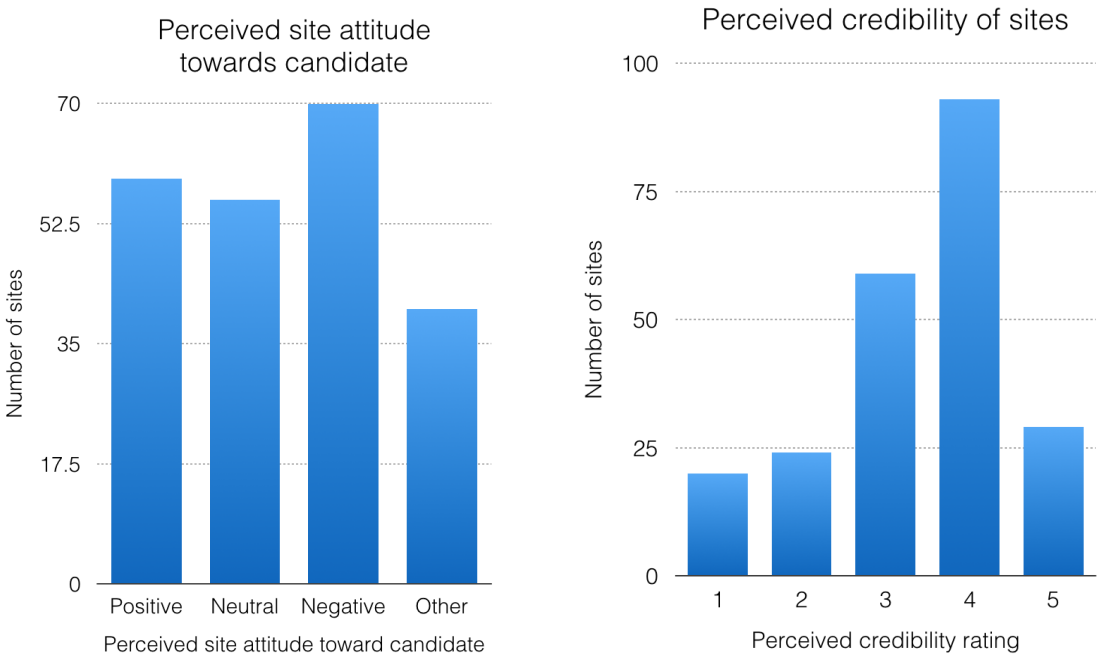


Figure 3. When web users on Mechanical Turk were asked to evaluate the disputed webpages in our data set, they reported that those pages were not overwhelmingly slanted in one ideological position relative to the candidate, nor were the sites perceived as low in credibility.

In order to better understand the URLs in our data set and the perception of those webpages by web users, we posted the URLs on Amazon's Mechanical Turk crowdwork marketplace, along with instructions to read and evaluate each webpage on its perceived attitude towards the candidate in whose search results it appeared, as well as the respondent's perceived credibility of the site. Our results, shown in Figure 3, found that the webpages were not perceived as overwhelmingly positive, negative, or neutral with regards to the candidate and, even more surprisingly, that the sites were perceived as

relatively credible, scoring an average of 3.38 out of 5. These results have a couple possible implications. From an optimistic perspective, this could indicate that although sites that have been known to distribute misinformation were flagged in our data collection, the majority of their content is reasonably credible and not overtly biased propaganda. This could point to a tactic of those engaging in misinformation: to sprinkle the misinformation in throughout a larger body of quality content, as was done by early spammers in their mutual admiration societies. On the other hand, more serious issue may be that those pages, while not actually credible, were perceived as such by users, who are unable to effectively evaluate the information with which Google presents them. This is supported by the fact that users did report pages as having a slant either for or against the candidate in the majority of cases, but still found most pages moderately to highly credible.

IV. Current Efforts to Combat Misinformation

To address the problem of misinformation, Google, like other platforms, has included mechanisms to fact-check content and engage users in reporting on low quality search queries and "Featured Snippets" ("which shows a highlight of the information relevant to what you're looking for at the top of your search results"). The company has avoided taking content down unilaterally. Instead, it has tried to improve its algorithm by considering user feedback and hiring quality evaluators. Both Facebook and Google have reacted quickly to public pressure generated by the fake news phenomenon in 2016. They have an incentive to act quickly and avoid social pressure for regulation of their markets; some of our interviewees alluded to this: 2016 was a year that put many eyes on these platforms, and, as Daphne Keller pointed out, "Google is being smart." These platforms know that the government, think tanks, and the public in general are looking at them, and therefore now is an opportune time to push them to address this issue pro-actively.[41]

Google has claimed that the fake news problem is rather minimal, but still worth addressing. As rationale for their recent adjustments, Google claims that only a fraction of a percent (around 0.25%) of user queries "have been returning offensive or clearly misleading content, which is not what people are looking for,"[42] however the company is doing things to address the problem. In April 2017, Google publicly announced three solutions to combat misinformation, as summarized in a recent post by Danny Sullivan.[43] Two solutions include feedback forms with which users can send feedback about search suggestions and "Featured Snippets" answers.[44] A third solution tries to enhance

[41] Interviews with Jonathan Zittrain and Daphne Keller, May 10, 2017.
[42] Gomes, "Our Latest Quality Improvements for Search."
[43] Danny Sullivan, "Google's 'Project Owl' -- a Three-Pronged Attack on Fake News & Problematic Content," *Search Engine Land*, April 25, 2017, http://searchengineland.com/googles-project-owl-attack-fake-news-273700; Gomes, "Our Latest Quality Improvements for Search."
[44] Gomes, "Our Latest Quality Improvements for Search."

authoritative content by hiring 10,000 “search quality evaluators” to give feedback on search results.

Figures 4 and 5 display the forms with which users can provide feedback to the platform. The form in Figure 4 is for search suggestions, which allows users to report which of the search suggestions seem inappropriate and why. The form in Figure 5 is for Snippets, and includes more predetermined options for reporting as well as a box for comments or suggestions.

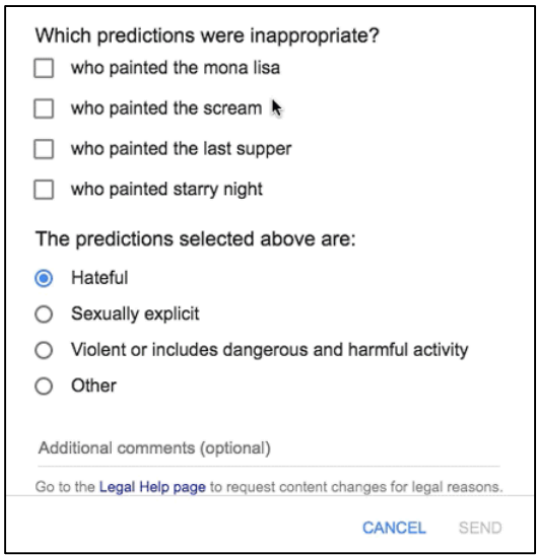


Figure 4. Form for users to give feedback on search suggestions.

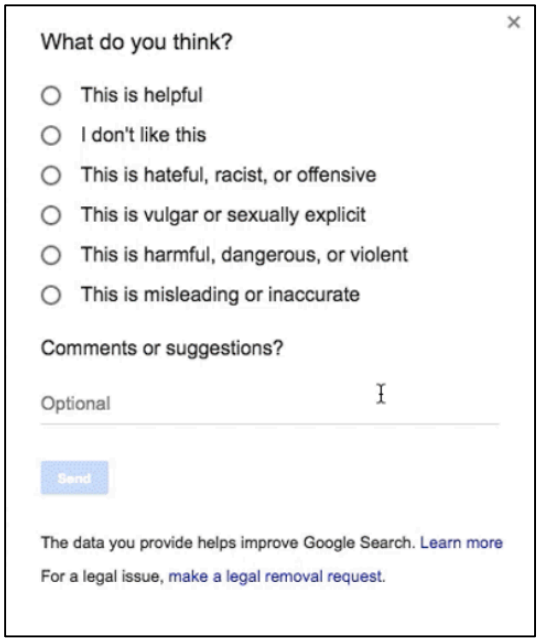


Figure 5. Form for users to give feedback on Snippets.

Notably, at present, Google does not allow individual users to give feedback on search results. Google attempted to allow user-level feedback of search results in 2008 but discontinued this feature, claiming that it was unused.[45] Instead, the company has hired professional evaluators to play this role, a decision that may be influenced by the way Google has scoped the problem and restricted it to a very small proportion of searches. These quality evaluators follow a protocol established by Google to rate search results according to specific and very detailed criteria; the protocol is 157 pages long.[46] Evaluators' provided rankings, weighted into the Search algorithm, combining with the existing infrastructure. This means that in addition to such other features of the algorithm as key words, user historical queries and place, among others, the evaluators' rankings help determine the search results to specific queries.

The guidelines employed by the quality-control evaluators describe upsetting or offensive content as including the following:[47]

- "Content that promotes hate or violence against a group of people based on criteria including (but not limited to) race or ethnicity, religion, gender, nationality or citizenship, disability, age, sexual orientation, or veteran status."
- "Content with racial slurs or extremely offensive terminology."
- "Graphic violence, including animal cruelty or child abuse."
- "Explicit how-to information about harmful activities (e.g., how-tos on human trafficking or violent assault)."
- "Other types of content which users in your locale would find extremely upsetting or offensive."

Researchers cannot yet assess the impact of these efforts, but no doubt Google is carefully collecting data in order to do so. However, the new features do not seem to have been widely publicized. In our survey of 467 Google users, conducted in May 2017, 72% of respondents did not know Google had made changes to curb fake news, and less than 4% reported having been informed by Google's own communications (see Figure 6). When informed of the details of these features, however, 63% of those surveyed expressed support for the changes.

[45] "SearchWiki: Make Search Your Own," *Official Google Blog*, November 20, 2008, https://googleblog.blogspot.com/2008/11/searchwiki-make-search-your-own.html.
[46] Google, "General Guidelines," May 11, 2017, https://static.googleusercontent.com/media/www.google.com/en//insidesearch/howsearchworks/assets/searchqualityevaluatorguidelines.pdf.
[47] Danny Sullivan, "Google Launches New Effort to Flag Upsetting or Offensive Content in Search," *Search Engine Land*, March 14, 2017, http://searchengineland.com/google-flag-upsetting-offensive-content-271119.

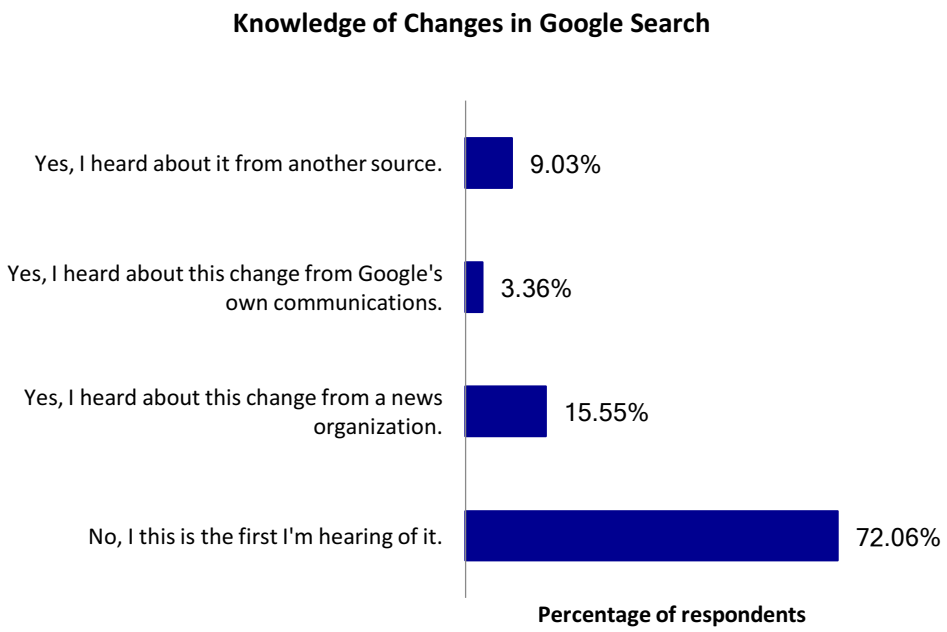


Figure 6. When informed about Google's changes to Search intended to curb fake news and hate speech, users reported being unaware of the changes; only a small minority reported being informed by Google itself of these changes.

Our interviewees, experts in this area, agreed with users in their approval for Google's changes. However, they also pointed out the need for complementary efforts.[48] These experts focused on the need to protect the First Amendment and avoid making Google the "arbiter of truth," instead suggesting that additional efforts be pursued both internally at the platform level, as collaborations between platforms and external groups, and purely externally.

Regarding such collaborative solutions, two of our interviewees, Takis Metaxas and Jonathan Zittrain, suggested that librarians could play a central role in the fact checking mechanisms. Zittrain additionally suggested that schools could be included, involving students in a multifaceted collaboration with platforms by engaging them in evaluating information on these platforms. The value of a critical civic education and critical thinking skills came up consistently in our interviews, and should not be dismissed. However, such solutions would likely be most effective in the longer than the short term.

Regarding complementary solutions outside the platforms, Robert Epstein and Dennis Allison referred us to their own efforts to create a civic organization to conduct research and advise on regulation, for instance by collecting and analyzing data from across the country in a manner modeled after the Nielsen Ratings system. Their proposed organization is called the Sunlight Society. Zittrain echoed this point, saying we need "a model of research that is collaborative and requires sharing of data early and often" between researchers and the platforms. He noted that in

[48] Metaxas, Keller, Sullivan.

the ideal scenario this effort might have some cooperation from platforms to aid researchers in collecting data and other information, though Epstein and Allison suggest, instead, that this data could be passively collected from a random sample of volunteers in order to decrease dependency on the platform.

Another relevant point that several respondents stressed in interviews was that opportunities for platforms to deal with the issue of misinformation, though important, are relatively limited. While many aspects of the issue are directly related to platforms, the underlying issues—distrust in journalism, political polarization, civic education, international interference in domestic elections, among others—are much bigger issues that do not fall cleanly under the purview of companies like Google.

V. The Impact of Google Search on Journalism

It is impossible to reflect on the phenomenon of fake news without reflecting on journalism. The fake news crisis has been predicated on distrust in traditional media sources. People are seeing news from traditional sources less frequently and, increasingly, question whether traditional news sources provide reliable information.[49] This was not a main focus on our research, yet the subject of the current state of journalism came up repeatedly throughout our research, leading us to conclude that the subject is essential to this analysis.

Current John S. Knight Journalism Fellows at Stanford rehearsed with us the crisis in the advertising business model that has traditionally supported print and television journalism. In "The Platform Press: How Silicon Valley Reengineered Journalism," Emily Bell and Taylor Owen of Columbia University, agree, arguing: "The 'fake news' revelations of the 2016 election have forced social platforms to take greater responsibility for publishing decisions. However, this is a distraction from the larger issue that the structure and the economics of social platforms incentivize the spread of low-quality content over high-quality material. Journalism with high civic value—journalism that investigates power, or reaches underserved and local communities—is discriminated against by a system that favors scale and shareability."[50]

The Knight Fellows see the very function of social networking sites as promoting the diffusion of low-quality emotional content over higher quality content. Platforms benefit from so-called clickbait articles that entice users to interact and share those materials on the platform, which results in higher revenue advertising. In other words, short, controversial content engages more users than do longer, well-researched, analytical or investigative articles. The Knight journalists point out that the platform companies have little business incentive to diminish emotional content. Moreover, without sufficient demand, quality journalism will not maintain the necessary funding

[49] The DataSociety report highlights distrust as one of the major factors to why the media is vulnerable. See Marwick and Lewis, "Media Manipulation and Disinformation Online."
[50] Emily Bell and Owen Taylor, "The Platform Press: How Silicon Valley Reengineered Journalism," *Tow Center for Digital Journalism Blog*, March 29, 2017, http://towcenter.org/research/the-platform-press-how-silicon-valley-reengineered-journalism/.

to survive.[51]

On the other hand, representatives from Google strongly refute this account. For one thing, Google claims that emotional content has always been an issue for journalism. A high-level Google official, for example, referred to the new market ecosystem as yet another in a string of adjustments to which journalism will adapt, as it has in the past. In fact, the official argues, the vast amount of data available provides opportunities for journalism to grow in new directions.[52] He pointed to The Trust Project, a Santa Clara University venture in collaboration with Google News, as an example of this emerging transformation. The Trust Project "crafts tangible digital strategies to fulfill journalism's basic pledge: to serve society with a truthful, intelligent and comprehensive account of ideas and event," seeking to "bake the evidence of trustworthy reporting—accuracy, transparency and inclusion—plainly into news practices, tools and platforms."[53]

This rings true in comments received from Google users in response to our survey. Several comments underscore the decline in trust in the media. One respondent lamented: "BBC, CNN, NY Times - FOX news may be considered traditional, but they are the most egregious offenders when it comes to misleading news, they are absolutely not trustworthy!" Another criticized the integrity of mainstream media, musing, "Yeah, hmm, CNN, MSNC, CNBC, New York Times, Washington Post - these are FAKE NEWS media outlets. Better yet, I should say they are public relations agencies for the Democratic Party and the political left in general." Further, we received comments suggesting that news outlets are not effectively communicating their norms and standards. As one respondent put it: "Many liberal sites are publishing content from unverified anonymous sources as if they ARE news, thus creating fake news because it is more important to them to harm Trump than tell the truth." Another expressed concern about a liberal feedback loop: "Fact checking in and of itself is problematic because liberal sources [fact-]check liberal sources."

While it is not yet possible to conclude whether platforms like Google are helping or hindering journalism, we see an opportunity for collaboration. Companies like Google are not only thriving in the current ecosystem, but also helping to create it. If they can be more fully engaged as strong allies with journalism, they can help to advance high quality news and promote the journalistic integrity necessary to a vibrant democracy.

---

51 Interviews with the Stanford Knight Fellows centered on the negative impact on journalism of changes in technology and the expansion of internet. These interviews also highlighted diminishing local journalism. The fellows discussed the impact of the shifting revenue model for local news as cutting into local coverage and generating "news deserts" where there are no daily local news outlets. News deserts leave large sectors of the U.S. with no local news and they create distance and distrust between news outlets and citizens. For a graphic description of "news deserts," by *Columbia Journalism Review*, see: "America's Growing News Deserts," *Columbia Journalism Review*, accessed June 10, 2017, https://www.cjr.org/local_news/american-news-deserts-donuts-local.php.

[52] Interview with Google officials, May 12, 2017.

[53] Taken from "The Trust Project," http://thetrustproject.org/.

VI. Legal Framework

In the U.S, under the First Amendment and the Communications Decency Act of 1996, 47 U.S.C. § 230(c), search engines are not liable for the content they provide nor for the ranking they give to webpages on search results.[54] Opinions issued in *Search King, Inc. v. Google Technology, Inc.*[55] and *Langdon v. Google Inc.*[56] uphold these protections.[57] Therefore, if false information shows up in Google's search results, the company is not liable. This is a straightforward answer. Most lawsuits against search engines pose legal questions that focus on rankings. Namely, what is the algorithmic reasoning behind search results? What constitutes an adequate or effective ranking in response to a search query? These questions comprise the baseline for developing case law.

Scholar James Grimmelmann offers the latest and most comprehensive revision of case law relevant to address these questions.[58] He reviews three theories: the conduit, the editor, and the advisor. Conduit theory argues that search engines should simply connect users with webpages and provide a diverse view of possible search results. The results should be a representative survey or "map" of the universe of web pages related to a specific query. In contrast, the editor theory claims that search engines should edit the results and prioritize that which is most relevant according to the search engine criteria. Where the editor theory presumes an underlying "truth", advisor theory assumes that search engines should behave like a friend and provide results that best fit users' queries. It asks: what does the user want?

Grimmelman argues that the courts should adopt advisor theory as best-suited to user's search habits and expectations. He points out that "the conduit theory goes wrong because it treats relevance as a fact about websites and ignores users' normative opinions of websites. The editor theory goes wrong because it conflates users' normative opinions about websites with search engines' descriptive opinions about which websites users will find relevant. In contrast, the advisor theory of relevance—a descriptive opinion about users' normative opinions of websites— yields a straightforward test based on loyalty to the user."[59]

---

[54] Other countries are not so protective. "German law prohibits Holocaust denial; Thai law prohibits insulting the king. Google frequently removes links to these and many other kinds of content when ordered to do so by local authorities. These deletions directly inhibit users' ability to seek out the information they seek. The German government doesn't let users make up their own minds about the Holocaust; the Thai government doesn't let them decide whether its monarchy is worthy of respect." James Grimmelmann, "Speech Engines," *Minn. L. Rev.* 98 (2013): 901.

[55] No. CIV-02-1457-M, 2003 WL 21464568, (W.D. Okla. May 27, 2003).

[56] 474 F. Supp. 2d. 622, 626 (D. Del. 2007)

[57] "Supreme Court precedents compel the conclusion reached by these two courts, for seven related reasons. First, Internet speech is fully constitutionally protected. Second, choices about how to select and arrange the material in one's speech product are likewise fully protected. Third, this full protection remains when the choices are implemented with the help of computerized algorithms. Fourth, facts and opinions embodied in search results are fully protected whether they are on nonpolitical subjects or political ones. Fifth, interactive media are fully protected. Sixth, the aggregation of links to material authored by others is fully protected. Seventh, none of this constitutional protection is lost on the theory that search engine output is somehow "functional" and thus not sufficiently expressive. And, eighth, Google has never waived its rights to choose how to select and arrange its material." Eugene Volokh and Donald M. Falk, "Google: First Amendment Protection for Search Engine Search Results," *JL Econ. & Pol'y* 8 (2011): 886.

[58] Grimmelmann, "Speech Engines."

[59] *Ibid*., 931.

Grimmelmann claims that the Federal Trade Commission (FTC) has adopted a view that agrees generally with his view about the validity of the advisor theory. An antitrust investigation reached the FTC arguing that Google was gaming its search engine to benefit its own products in an abuse of a dominant market position. By employing a very close approach to advisor theory, however, the FTC reached a different conclusion and dismissed the search-bias portion of the investigation. The FTC considered Google to be providing the information the users wanted and therefore did not engage in an antitrust violation—under this theory Google is likely to be safe, as people are very satisfied with Google Search results.

Therefore, under U.S. law, Google is not liable for search results, nor is there currently the legislative appetite to make it so. As Daphne Keller, a scholar on issues of intermediary liability, states: "As a lawyer with long experience handling takedowns from Google web searches, I believe that there are responsible ways to remove illegal content from platforms. A good start is to have courts decide what violates the law — not machines and not company employees operating under the threat of huge fines." To make Google the censor is to threaten freedom of speech.[60]

The legal questions will remain but more immediate benefit comes from focusing on the criteria underlying search result rankings, rather than on intermediary liability for content. False, fake and misleading news will continue to appear, and the solution is not, as Keller argues, to make "Google the censor."[61]

VII. Policy Options

The 2016 election serves as a case study for escalating concern with the proliferation of misinformation. This study focusing on the role of the platforms in that proliferation reveals opportunities for partnerships with journalism and civic institutions to enhance civic engagement through online forums. We present three forums where such opportunity may be most impactful in strengthening democratic institutions and governance: The platforms, civil society organizations and journalism, and government.

1) The Platform: Increase Transparency and Collaboration

Google should engage in research partnerships with universities and think tanks, sharing data collected from Google Search that pertains to democratic institutions. A first stage of this project should address two questions: (1) what data can and should Google share? And (2) with whom can these different data be shared with optimal effect on strengthening democratic institutions?

Fundamental to considering what data to share is the balance between the potential benefits of transparency with privacy and spamming. Privacy law limits the type of data than can be shared, barring, for instance, personally identifiable information from becoming public. In addition to this

[60] Daphne Keller, "Making Google the Censor," *The New York Times*, June 12, 2017, sec. Opinion, https://www.nytimes.com/2017/06/12/opinion/making-google-the-censor.html.
[61] Ibid.

concern, Google is restrained in what it can make public about the workings of its algorithms, as spammers will take advantage of any openness to game the system more effectively for purposes ranging from ads and other forms of revenue to political manipulation.

A second stage of the project might encourage Google to share information more widely among research scholars and the public. Protocols could be set in place dictating which data is shared with the general public, which with academics, and, potentially, which with third-party auditors to analyze impartially and confidentially. These considerations could lead not only to greater insight into Google's effect on its users, but could also inform users themselves in innovative ways, much like the Google Trends tool does.

2) Civilian Oversight: Civil Society Organizations and Journalism

In addition to lobbying for Google's explicit collaboration in these efforts, this report encourages further independent research collection and analysis. Existing groups, such as Stanford's Center for Internet and Society and Harvard's Berkman-Klein Center for Internet and Society, for example, have a long history of industry-parallel work. It is also worth considering novel proposals such as Robert Epstein's idea for a Nielsen-like system for collecting data from volunteers across the country to examine the workings of Google search independently of the company, or Takis Metaxas's and Jonathan Zittrain's suggestions to develop a corpus of fact-checked knowledge run by impartial, trusted individuals such as librarians, in which the discourse and debate around every check is documented and publicly visible. Projects that enhance collaboration between journalists and Google and other platforms, with special attention to the value of local and public news outlets, may help bolster journalism and foster public trust in media.

3) Public Accountability

There is widespread agreement in the United States that the First Amendment should be upheld and that platforms should not act as arbiters of truth, deciding what content people can and cannot produce or consume online. This forestalls efforts that might seek to force platforms to perform such functions in the U.S. (though such efforts may be undertaken elsewhere). There may, however, be an opportunity to regulate the information that these companies must make public or allow to be audited by third parties. The amount of power that platforms gain in collecting data on their users should result in some level of responsibility and accountability to the public interest.

4) Further Research

To better understand Google's impact on democratic institutions, this study recommends four further lines of research:

1. Research on specific elections tracking Google Search results with political outcomes. This could be done in different states or countries and at different political election levels, for instance, elections of local mayors or city councils, or state legislators, or U.S. representatives. Midterm elections could be a good time to execute these studies, allowing ample time to gather information

before the elections and to track search results on specific issues or candidates. This line of research could also examine how people who use Google may change their political behavior in response to different kinds of interventions. The studies do not need to be restricted to Google and in fact could be more insightful if they include other platforms.

2. Studies on the psychological and civic impacts of Google search results. This line of study could address the following questions: What subconscious effects do Google results have on citizens' political views? Do citizens gain a more diverse view of politics by using Google? If so, under what conditions? What is the impact on deliberative democracy of new features that engage users in assessing the quality of content on the platform? What platform features produce more civic engagement in political debates of high quality and which ones of low quality? Which features cause or increase political polarization?
3. Further investigation of the legal and regulatory framework for search engines. This line of study could address such questions as: What is the regulatory framework for digital footprints? How is that market regulated? What alternative frameworks exist, as for example in other countries? How might these various regulatory models attach to intermediary liability in the spread of misinformation?
4. Surveys to provide not only demographics but also descriptive statistics on how different users engage with content on the platform, with an eye to misinformation.

VIII. Conclusion

As evidenced by the explosion of interest in fake news in late 2016, the issues of misinformation, propaganda, and bias, and their propagation through online tools, are paramount. In our research, we examined a long "arms race" history of how search engines deal with spammers' and other actors' intentional efforts to mislead. Google's deep attention to these problems may partially explain why Google Search was not as hard hit by fake news gaming its algorithm as were social networking sites (SNS), and, thus, why it continues to maintain a high level of trust among users.

We also identified the misinformation risk on Google search, which can be subtler (e.g. a biased collection of search results) than on SNS, and influence users unconsciously. This search result risk for unconscious, peripheral persuasion warrants serious attention, since it can only be uncovered by systematic, aggregated analysis that considers the experiences of large swaths of users over time. Using the 2016 election as a case study, our research leads us to conclude that a precise estimate of the effect of fake news is out of reach without such broad data collection and analysis, but that misinformation did spread on Google in political web search results at that time.

We find that Google is taking steps to mitigate the issue of spreading fake news. Our data shows that although users are largely unaware of Google's efforts, once they are informed, they strongly support the company's efforts. Here we identify a tension: systematic, rigorous, third-party analysis is necessary to ensure protection of free speech, but Google has widespread consumer

trust and may be less inclined to engage in changes that risk public misunderstanding or bad publicity. We encourage the company to develop infrastructure for rigorous review of search results across users and over time. Creating transparency and review out of a highly personalized, complex, black box algorithm is the next step towards fully understanding the ubiquitous technological environment in which we are all steeped, and its tremendous potential impact on our national civic beliefs and culture.

Acknowledgements

The authors would like to thank the Hewlett Foundation's Madison Initiative for sparking this collaboration; Senator Russ Feingold, Dr. Luciana Herman, and Ashwin Aravind for their guidance throughout the project; Professor Takis Metaxas of Wellesley College for generously providing access to 2016 Google Search data; and all the interviewees for taking the time to participate in this work.

Bibliography

Allcott, Hunt, and Matthew Gentzkow. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31, no. 2 (May 2017): 211–36. doi:10.1257/jep.31.2.211.

"America's Growing News Deserts." *Columbia Journalism Review*. Accessed June 10, 2017. https://www.cjr.org/local_news/american-news-deserts-donuts-local.php.

Bell, Emily, and Owen Taylor. "The Platform Press: How Silicon Valley Reengineered Journalism." *Tow Center for Digital Journalism Blog*, March 29, 2017. http://towcenter.org/research/the-platform-press-how-silicon-valley-reengineered-journalism/.

Cadwalladr, Carole. "Google, Democracy and the Truth about Internet Search." *The Guardian*, December 4, 2016, sec. Technology. https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook.

Epstein, Robert, and Ronald E. Robertson. "The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections." *Proceedings of the National Academy of Sciences* 112, no. 33 (August 18, 2015): E4512–21. doi:10.1073/pnas.1419828112.

Gillin, Joshua. "PolitiFact's Guide to Fake News Websites and What They Peddle." *PunditFact*, April 20, 2017. http://www.politifact.com/punditfact/article/2017/apr/20/politifacts-guide-fake-news-websites-and-what-they/.

Gomes, Ben. "Our Latest Quality Improvements for Search." *Official Google Blog*, April 25, 2017. http://blog.google:443/products/search/our-latest-quality-improvements-search/.

Google. "General Guidelines," May 11, 2017. https://static.googleusercontent.com/media/www.google.com/en//insidesearch/howsearchworks/assets/searchqualityevaluatorguidelines.pdf.

"Google: Distribution of Revenue by Source 2016." *Statista*, 2017. https://www.statista.com/statistics/266471/distribution-of-googles-revenues-by-source/.

"Google Search Statistics - Internet Live Stats." Accessed June 10, 2017. http://www.internetlivestats.com/google-search-statistics/.

Grimmelmann, James. "Speech Engines." *Minn. L. Rev.* 98 (2013): 868.

———. "The Google Dilemma." *NYL Sch. L. Rev.* 53 (2008): 939.

International Telecommunication Union. "Measuring the Information Society Report," 2016. http://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2016/MISR2016-w4.pdf.

Keller, Daphne. "Making Google the Censor." *The New York Times*, June 12, 2017, sec. Opinion. https://www.nytimes.com/2017/06/12/opinion/making-google-the-censor.html.

Lazer, David, Matthew Baum, Nir Grinberg, Friedland, Kenneth, Hobbs, and Mattsson. "Combating Fake News: An Agenda for Research and Action," May 2, 2017. https://shorensteincenter.org/combating-fake-news-agenda-for-research/.

Marvin, Ginny. "Google Isn't Actually Tackling 'Fake News' Content on Its Ad Network." *Marketing Land*, February 28, 2017. http://marketingland.com/google-fake-news-ad-network-revenues-207509.

Marwick, Alice, and Rebecca Lewis. "Media Manipulation and Disinformation Online." Data&Society, 2017. https://datasociety.net/output/media-manipulation-and-disinfo-online/.

Metaxas, Panagiotis Takis. "Web Spam, Social Propaganda and the Evolution of Search Engine Rankings." In *Web Information Systems and Technologies*, 170–82. Springer, Berlin, Heidelberg, 2009. doi:10.1007/978-3-642-12436-5_13.

Moore, Martin. "Tech Giants and Civic Power." Centre for the Study of Media, Communication and Power, April 2016. https://www.kcl.ac.uk/sspp/policy-institute/CMCP/Tech-Giants-and-Civic-Power.pdf.

Pan, Bing, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Geri Gay, and Laura Granka. "In Google We Trust: Users' Decisions on Rank, Position, and Relevance." *Journal of Computer-Mediated Communication* 12, no. 3 (April 1, 2007): 801–23. doi:10.1111/j.1083-6101.2007.00351.x.

Purcell, Kristen, Joanna Brenner, and Lee Rainie. "Search Engine Use 2012." Pew Research Center, March 9, 2012. http://www.pewinternet.org/2012/03/09/search-engine-use-2012/.

"SearchWiki: Make Search Your Own." *Official Google Blog*, November 20, 2008. https://googleblog.blogspot.com/2008/11/searchwiki-make-search-your-own.html.

Subramanian, Samanth. "Inside the Macedonian Fake-News Complex." *WIRED*, February 15, 2017. https://www.wired.com/2017/02/veles-macedonia-fake-news/.

Sullivan, Danny. "Google Launches New Effort to Flag Upsetting or Offensive Content in Search." *Search Engine Land*, March 14, 2017. http://searchengineland.com/google-flag-upsetting-offensive-content-271119.

———. "Google Now Handles at Least 2 Trillion Searches per Year." *Search Engine Land*, May 24, 2016. http://searchengineland.com/google-now-handles-2-999-trillion-searches-per-year-250247.

———. "Google's 'Project Owl' — a Three-Pronged Attack on Fake News & Problematic Content." *Search Engine Land*, April 25, 2017. http://searchengineland.com/googles-project-owl-attack-fake-news-273700.

"The Trust Project." Accessed June 10, 2017. http://thetrustproject.org/.

Volokh, Eugene, and Donald M. Falk. "Google: First Amendment Protection for Search Engine Search Results." *JL Econ. & Pol'y* 8 (2011): 883.

Zimdars, Melissa. “Resource-False-Misleading-Clickbait-Y-and-Satirical-‘News’-Sources-1.pdf,” November 2016. http://d279m997dpfwgl.cloudfront.net/wp/2016/11/Resource-False-Misleading-Clickbait-y-and-Satirical-%E2%80%9CNews%E2%80%9D-Sources-1.pdf.

Appendix: Methodology

Quantitative

We have access to data collected twice per week in the 26 weeks prior to the November 2016 U.S. national elections. The data include the top ten URLs returned when searching the names of all Senate and House representatives, as well as both of the presidential candidates. We also have access to lists of fake news sites from various sources, including Professor Melissa Zimdars and PolitiFact. As a first analysis step, we are using available lists of verified fake news URLs to check how many, if any, of the top ten results for these representatives are from sites suspected of spreading misinformation. The precise methodology here is to use computational methods to scan our data for any URLs matching the list of fake news sites.

This analysis, however, cannot allow us to definitively conclude that misinformation or fake news appeared with certainty in these search results, as some organizations publishing fake news may also publish legitimate information. In other words, the appearance of an article from www.breitbart.com in our data set does not guarantee that fake news was being spread, as some might argue that not all of Breitbart's articles are illegitimate. In order to examine this issue more closely, we have asked human annotators to annotate a subset of the data on dimensions including whether the content on that specific page perceived as credible, or is favorable or unfavorable to the candidate. These annotators are online crowd workers hired from Amazon Mechanical Turk and paid the equivalent of minimum wage to evaluate one URL at a time.

Qualitative

We interviewed experts in search engine manipulation, particularly with regard to its political implications, as well as scholars at the forefront of law, policy, and democracy studies:

- Dennis Allison, Lecturer in the Computer Systems Laboratory, Stanford University. Sunlight Society scholar studying the presence of search bias in algorithms
- Larry Diamond, Senior Fellow, Freeman Spogli Institute for International Studies and the Hoover Institution, Stanford University is a scholar of democratic institutions.
- Robert Epstein, American psychologist studying the search engine manipulation effect.
- High-level officials at Google who are at the forefront of the debates about Google's roles in news and technology. These officials requested anonymity in the public version of this report.
- Daphne Keller, Director of Intermediary Liability, Stanford University Center for Internet and Society; former Associate General Counsel for Intermediary Liability and Free Speech issues at Google.
- Takis Metaxas, Professor of Computer Science, Wellesley College.
  Scholar studying crowdsourcing and social networks in the context of political events and news literacy.
- Danny Sullivan, Search Engine Land and Third Door Media
  Analyst, journalist, and expert on search engines studying changes in Google's search

engine and algorithmic interventions for fake news and misinformation.

- Jonathan Zittrain, George Bemis Professor of International Law, Harvard Law School; Faculty Director, Berkman-Klein Center for Internet and Society
  Scholar in digital property and content and the roles of intermediary platforms.