



Research Article

The weaponization of web archives: Data craft and COVID-19 publics

An unprecedented volume of harmful health misinformation linked to the coronavirus pandemic has led to the appearance of misinformation tactics that leverage web archives in order to evade content moderation on social media platforms. Here we present newly identified manipulation techniques designed to maximize the value, longevity, and spread of harmful and non-factual content across social media using provenance information from web archives and social media analytics. After identifying conspiracy content that has been archived by human actors with the Wayback Machine, we report on user patterns of “screensampling,” where images of archived misinformation are spread via social platforms. We argue that archived web resources from the Internet Archive’s Wayback Machine and subsequent screenshots contribute to the COVID-19 “misinfodemic” in platforms. Understanding these manipulation tactics that use sources from web archives reveals something vexing about information practices during pandemics—the desire to access reliable information even after it has been moderated and fact-checked, for some individuals, will give health misinformation and conspiracy theories more traction because it has been labeled as specious content by platforms.

Authors: Amelia Acker (1), Mitch Chaiet (2)

Affiliations: (1) School of Information, University of Texas, USA, (2) Moody School of Communication, University of Texas, USA

How to cite: Acker, A., & Chaiet, M. (2020). The weaponization of web wrchives: Data craft and COVID-19 publics. *Harvard Kennedy School (HKS) Misinformation Review*, Volume 1(3).

Received: May 25th, 2020. Accepted: September 4th, 2020. Published: September 28th, 2020.

Research questions

- How is health misinformation being archived, weaponized, and propagated online?
- What does provenance information from web archives reveal about the spread and moderation of harmful content in platforms?
- What do these web archiving tactics reveal about information practices during pandemic?
- How does screensampling extend the propagation of health misinformation beyond trackable metrics?

¹ A publication of the Shorenstein Center for Media, Politics and Public Policy, at Harvard University, John F. Kennedy School of Government.

Essay summary

- Using provenance information such as original context, technical specificities, and unique characteristics of online resources from web crawls, and social analytics data from the Crowdtangle API we find that web archives like the Internet Archive's Wayback Machine are being weaponized to propagate and preserve health misinformation circulating on platforms like Facebook and Twitter.
- Here we present two interconnected studies of data craft that leverage archived web resources to intentionally evade automated content moderation efforts and further propagate health misinformation from platforms attempting to combat.
- This research shows how archived URLs of web archived sources of health misinformation found in the Internet Archive's Wayback Machine and screensampling practices of archived content appear to be difficult for automated content moderation systems to identify and ban, and as a result circulate longer and spread on platforms.
- In order to understand these information practices that shape emerging COVID-19 publics convened by private platforms in the aftermath of pandemic, misinformation researchers and platform operators need to carefully consider how the circulation of archival content from the web now appears in platforms and its status in the wake of the coronavirus.

Implications

For many years, computer security researchers and internet researchers have documented the various ways that web archives, such as the Internet Archive's Wayback Machine, have been hacked, hi-jacked, and misused (Caplan-Bricker, 2018; Littman, 2017). High profile cases typically point to zombie content no longer published on the 'live' web, provenance laundering with customized hyperlinks or link shortening, backdating resources, or blocking bots from crawling web pages for search indexing (Madrigal, 2018; Nelson, 2018; Walden, 2012). The coronavirus infodemic (Zarocostas, 2020) has resulted in a slew of data craft techniques propagating health misinformation, which now includes web archives like the Wayback Machine. By data craft we mean "practices that create, rely on, or even play with the proliferation of data on social media by engaging with new computational and algorithmic mechanisms of organization and classification" (Acker, 2018). Here we also discuss screensampling, a data craft technique that extends the propagation of archived misinformation when social media users post screenshots of archived URLs thus removing the ability to click or track these static images of archived online sources. Such data craft often allow misinformation and disinformation campaigns to go undetected, and prove particularly adept at avoiding the automated content moderation algorithms used to increasingly combat fake news and inauthentic behavior. One well-documented data craft technique is mimicking legitimacy by publishing fake content that appears to be credible information (Acker & Donovan, 2019). In their study investigating the misuses of web archives on social media, Zannettou et al. (2018) found that news articles and social media posts were the most common web resources to be saved in Archive.is and the Wayback Machine. They found that these kinds of URLs circulated among Reddit forums when the original web content was assumed to be controversial or ephemeral. The data craft reported here is designed to leverage the legitimacy of the Wayback Machine's archival infrastructure in order to deploy health misinformation into platforms by circumventing moderation efforts.

Platforms, in their algorithmic sorting and moderation, bring together new online publics, what Gillespie calls "calculated publics" (2014). Finn has shown in her work on information orders before and after disasters that private platforms like Facebook convene groups of people in novel ways (2018, p. 140). Finn argues that after disasters, like earthquakes and pandemics, platforms become public information infrastructures that shape and are shaped by new information practices. Coupled with automated

recommendation algorithms and closely knit, targeted audiences, the subversive propagation of weaponized health misinformation now shapes the calculated publics of the pandemic. How platform algorithms convene these COVID-19 publics are generally unknown, their mechanisms are “black boxed,” providing outsiders with low visibility into their construction, development, and evaluation. Here we show how web archives are being used to mimic legitimacy and spread misinformation to COVID-19 publics through platforms like Facebook, which may in turn provide more insight into apprehending the power of algorithms to label and classify misinformation (Burrell, 2016).

Many content manipulators leverage the “context collapse” afforded in Facebook’s newsfeed and Twitter’s timeline to spread misinformation with free and fast online publishing tools. Because the newsfeed and timeline streams “flatten” all content into one feed or social awareness stream (Kivran-Swaine & Naaman, 2011), it can be hard to distinguish between vetted news articles, targeted advertisements, and other online content. Further, Facebook’s mobile app modifies web articles into their instant article format which they describe as a “buttery smooth” native feature (Facebook, 2020), providing a legitimizing data craft to content that would otherwise be perceived as sketchy and unreliable if viewed outside of the platform, at the content’s original URL. Before we explain how archived content can be weaponized in platforms, understanding how the web is archived is necessary.

Web archives such as the Internet Archive’s Wayback Machine, come from the resource intensive, and purposeful digital preservation of digital material (Brügger, 2018). Web archiving methods and techniques are typically divided into two approaches—micro and macro efforts. Macro web archives usually are managed by large information institutions relying on web crawling, which involves generating “seed lists” and automating routine, repeated crawls to build robust, comprehensive snapshots of the quickly changing web. Web crawling techniques are the most time consuming and resource intensive because they aim to capture whole web pages and online resources by systematically “crawling” each and every embedded hyperlink in a website to capture each part of complex and layered web resources (Milligan, 2016). Crawls can have varying levels of automation, and seed lists are added frequently by web archivists to expand crawlers’ reach. If web crawling is a macro technique that can be automated at scale, micro techniques are more targeted and less routine, such as API extraction, or focused on capturing dynamic features of the user interface with screencasts or screen shots. Micro web archiving projects usually are managed by individuals and small groups of researchers who want to capture particular slices of the web to illustrate an event, social movement, or emerging behavior. Web crawls are not limited to websites indexed by search engines, they also include individual web pages that users save through features like “Save Page Now” or previously the Alexa Toolbar (Rogers, 2017). Once a URL has been added to a seed list, they can also appear in many different collections and be captured by different automated crawlers. Our research has found that both macro tools and micro tools are overlapping to shape COVID-19 publics and spread misinformation across social platforms, which are increasingly used as public information infrastructures.

In their study of the Internet Archives’ preservation of the North Korean Web, Ben-David and Amram found that knowledge generated from Wayback Machine web crawls comes from human and non-human actors, and “includ[ed] proactive human contributions, routine operated web crawls, as well as curated and appraised web crawls of collections, arguing that these archived snapshots are like other algorithmic black boxes (Ben-David & Amram, 2018, p. 195). Despite these routine web-wide crawls, individual human actors are strategically adding to the collections, for a variety of different intentions and memory practices. Many have argued that more studies of archivists’ appraisal decisions and “web archival labour” should be conducted to understand the ways human and non-human actors impact the collections of resources that result in a history of the web (Ogden et al., 2017). Our investigation found that both human and bots were archiving online misinformation, but that more individuals used Save Page Now to archive a resource after Facebook moderated and flagged the live URL as health misinformation.

In this study we sought to discover how online health misinformation is being archived and then weaponized using the data craft tactics of mimicking legitimacy with reliable URLs and a practice we call screensampling. In answering these research questions, we showed how data craft weaponizes web archives and impacts how platforms convene COVID-19 publics, contributing to the coronavirus misinfodemic. Here we argue that there is an opportunity for misinformation researchers to examine this relationship, between passive and active archival agents and their intentions to archive misinformation, as well as the status of weaponized web archives that evade content moderation and removal on platforms because of their trusted URLs.

Findings

In March 2020, our research team began collecting, labeling, and organizing examples of COVID-19 health misinformation spreading through screenshots in posts on social networks. Within the dataset, four screenshots saved from a single Twitter thread, posted by user @narvonocutz on March 12, 2020, revealed health misinformation content that had been captured and archived with the Internet Archive Wayback Machine (Figure 1) (Narvo, 2020).

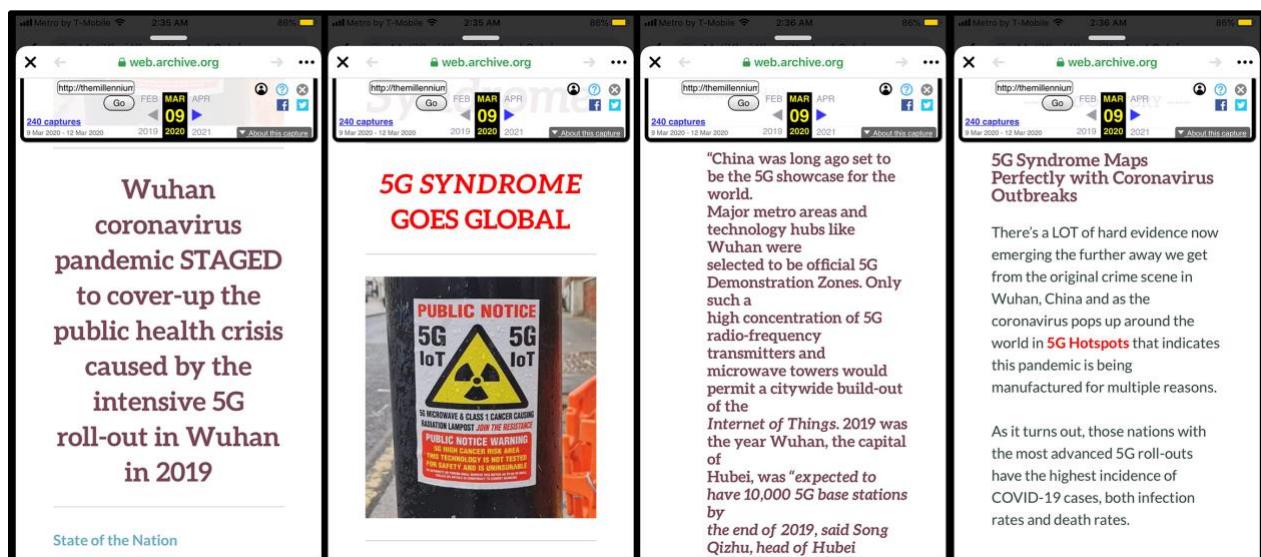


Figure 1. Screensampling: screenshots from web.archive.org taken by @narvonocutz and posted on Twitter.

These screenshots were taken of a conspiracy article, “CORONAVIRUS HOAX: Fake Virus Pandemic Fabricated to Cover-Up Global Outbreak of 5G Syndrome,” which had been archived by the Internet Archive’s Wayback Machine web crawlers on March 9, 2020 (The Millennium Report, 2020). The original article, which appeared on The Millennium Report website on March 2, 2020, was first crawled and preserved by the Wayback Machine on March 2, 2020 (hereafter we use “the original URL” and “the archived URL” to refer to these two sources). By examining the screenshots, we were able to locate the original URL as well as the archived URL hosted by the Internet Archive’s Wayback Machine at web.archive.org. Using provenance information from the Internet Archive’s multiple web crawlers, we found that individual human actors had archived and crawled the web page, which then seeded bots for automated routine crawls of the website.

By using these archived snapshots as a kind of proxy rather than the original URL, web.archive.org links can easily bypass existing content moderation systems used by platforms. As previously described by Donovan, the “hidden virality” of the article was not in its original URL form, but instead in the archived version stored in the Wayback Machine (Donovan, 2020). The Wayback Machine web archive allows for a public, relatively anonymous (with no profile or login necessary) means of spreading disinformation from the web and then hosting it—even when the original URL has been taken down or unpublished on the live web.² This tactic of storing misinformation and highly ephemeral content enables manipulators to use the web archive as a distribution mechanism, allowing it to evade content moderation and live longer on platforms.

The hidden virality of Wayback archived URLs can be further compounded by the practice of screensampling, where digitally extracting an archived snapshot creates a new digital asset that can easily increase the spread of dubious content (as seen in Figure 1). Posting images of text allows human readers to view the content while bypassing content moderation mechanisms because image formats with text are not easily machine-readable. Screensampling excerpted sources from web archives allows users like @narvonocutz to recontextualize and propagate content from a trusted source (the Wayback Machine), while constructing a new post or thread of recontextualized content made of images that both evade moderation but also obfuscate the archived URL by disabling the hyperlink and shortening the original URL. Abstracted from their original source, these screenshots sever attribution to the original URL and spreading the content in an untraceable manner, and creating “memetic abstraction” (Chaiet, 2019). The archived URL of the Millennium Report’s “CORONAVIRUS HOAX” article, as of this writing, has been captured with web crawls through various Internet Archive web archiving tools approximately 448 times since the beginning of March 2020 (Figure 2).

² Since the publication of Donovan’s article, the Internet Archive has since labeled the two archived webpages with a notice explaining that they were removed from the live web for violating Medium.com’s content policies and that their appearance in the Wayback Machine is not an endorsement.

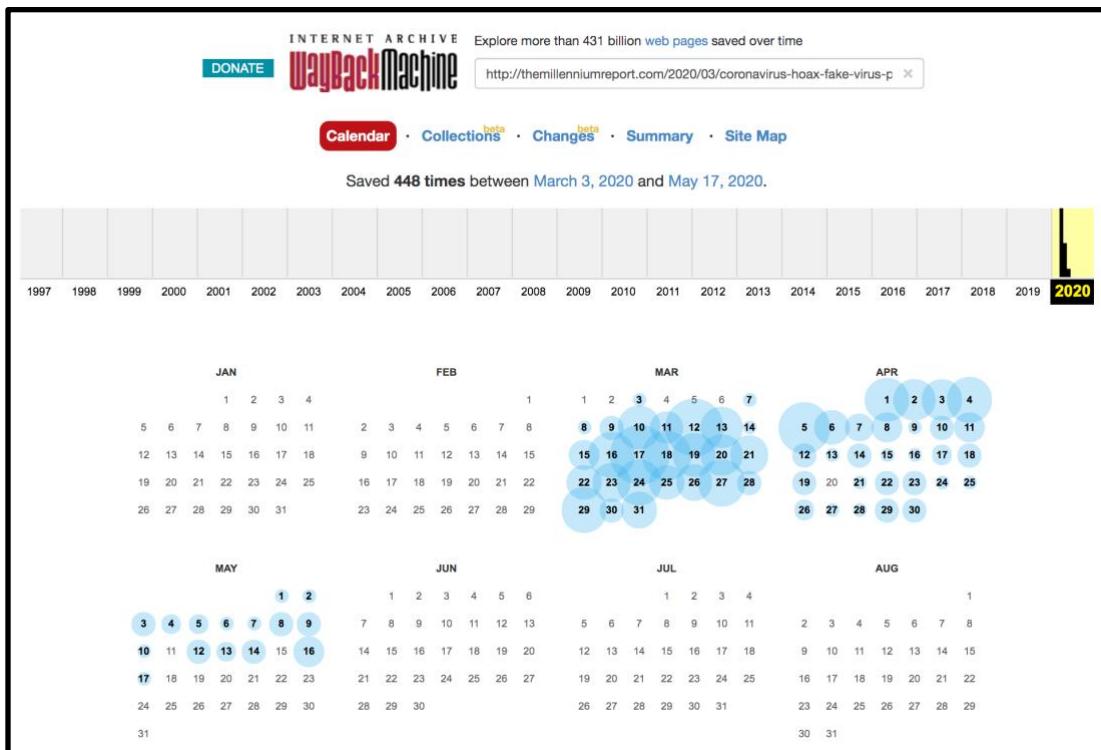


Figure 2. Calendar view from the Internet Archive Wayback Machine displaying 448 web crawl captures of The Millennium Report “CORONAVIRUS HOAX” article, from March 3, 2020 to May 17, 2020.

Recently the Internet Archive Wayback Machine has begun to publish provenance information about the source and type of web crawls (Internet Archive, 2018). This feature allows us to see more information about the purpose or context of the capture, as well as associated collections to which the crawl is saved. This provenance information allows researchers more insight into when human agents purposefully archived a snapshot with the Save Page Now feature, and whether the snapshots were then automated as part of routine archive collections carried out by web crawlers and bots. By applying this provenance feature at scale and collecting information about all types of web crawls of a particular web resource, we can compare proactive human contributions to the Wayback Machine to automated, routine crawls that feed particular Internet Archive collections. Most notably, different types of crawls that have appeared in the Wayback Machine of the original URL first published on 2 March 2020 follow a pacing that matches external fact checking and moderation of the original URL on Facebook. We have verified N=17 different kinds of web crawls that indicate a broad ecology of web archive agents—both non-human and human actors. These web crawls seed a number of specific collections at the Internet Archive, including collections of outlink URLs posted to Twitter, collections of Fake News, Archive-It partner collections that subscribe to Internet Archive’s web services, as well as wide crawls of the whole web (Internet Archive, 2018). In collections like “Fake News II” web archivists direct the seeding of web crawls but collections like “Live Web” proxy crawls are mostly fed by people using Save Page Now. Both broad and content-specific seeds play important roles in appraising what will (and won’t) be accessible in the future web (Summers & Punzalan, 2017), as well as when individual human actors choose to use Save Page Now.

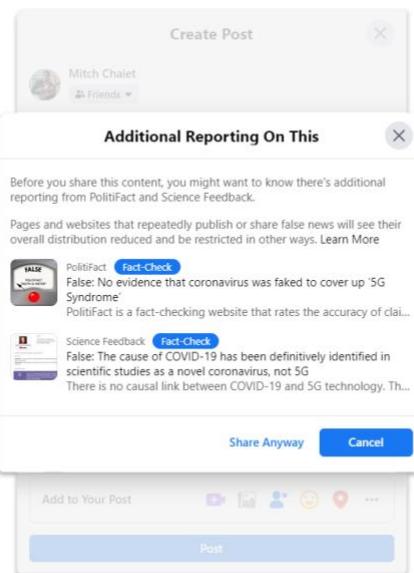


Figure 3. Facebook Fact-Check dialog that appears when posting the “CORONA HOAX” article.

On March 9, 2020, a week after The Millennium report published the article and the original URL began to appear on Facebook and Twitter, PolitiFact fact-checked the claims and sources cited in the article (Kertscher, 2020). Reporting had found that there was no credible evidence confirming the claims in the “CORONA HOAX” article. Shortly thereafter, Facebook began issuing warnings to users intending to share the original URL (Figure 3.). On the same day, Wayback Machine web crawlers “LiveWeb” and “WebWideCrawl” began to archive the original URL for snapshots. Both collections are fed mostly by the Save Page Now feature, which only saves a single page (Internet Archive, 2018). While the original URL had previously been crawled by automated collections crawlers from March 3 to March 8, it was only after the article had been fact-checked and flagged by Facebook that individual human agents began to proactively archive it with Save Page Now and Save Page Now proxies as compared to previously automated Wayback Machine web crawlers feeding collections like Twitter outlinks or Archive-It partners.

The spread of the original URL and the archived URL on Facebook can also be compared using social analytics data from CrowdTangle, which is owned by Facebook. They provide public analytics for how far the links spread on Facebook in the beginning of March, and the number of total interactions with the post indicating the popularity and reach of each URL. Table 1 shows that the archived URL circulated on Facebook outperformed the original URL in reach, engagement, views, and shares (Fraser, 2020). Although the original URL has now been fact-checked, flagged, and moderated by Facebook, users still are able to post the health misinformation today. However, the archived URL of the same misinformation, as yet, has not been flagged or identified by the platform as violating platform policies.

Table 1. Engagement data from CrowdTangle comparing original URL and archived URL. The Wayback Machine URL outperformed on every Facebook engagement metric, thus spreading misinformation farther than the original URL.

	Original URL	Archived URL
Facebook Reactions (Likes+)	7,606	16,583
Facebook Comments	6,050	14,318

Facebook Shares	5,770	10,113
Facebook Interactions (Total)	19,426	41,014

Weaponizing web archives and screensampling to evade misinformation moderation efforts are not only data craft for platforms trying to detoxify their networks of disinformation and harmful misinformation, but also a new challenge for misinformation researchers, tracking tactics and developing new methods for studying online behavior. As data craft, screensampling becomes another hurdle for users, researchers, and platforms trying to determine the original source of the content—and social analytics tools like CrowdTangle will never be able to quantify the number of users who screenshot an article and then post portions of it with their own framing commentary. As researchers confront the pandemic, there have been many calls for closer attention to information practices, digital archiving efforts, data management, and the importance of preserving this moment (Xie et al., 2020). Here we challenge our research communities to consider the information practices of emerging COVID-19 publics and the circulation of misinformation stored in web archives because it reveals both a mistrust and awareness of platforms' current automated moderation and fact-checking efforts. As scholars of misinformation continue to examine the information practices found in private platforms that convene COVID-19 publics, we need to expand our scope of to consider the circulation of dubious information found in web archives and examine their status as they become weaponized in platforms with data craft.

Methods

Following on Ben-David and Amram's innovative method of collecting provenance information from IAWM web crawls (2018), we used forensic analysis to learn when human agents directed web crawlers to archive health misinformation related to the 5G coronavirus conspiracy (as compared to automated seed lists and bots that archive the web). The availability of web crawl provenance information data provided readily available data and descriptive metadata for us to analyze. Once we identified the original URL from screenshots posted on Twitter, we scraped the provenance information from web crawls of the Millennium Report's archived URL (beginning March 3, 2020 and ending May 17, 2020). Then we compared archived snapshots to the original URL using CrowdTangle, Facebook's social analytics dashboard, to measure the engagement between the original URL and the web archived URL to compare their spread and reception. By using CrowdTangle analytics to parse the engagement data and confirm greater spread of the archived URL than the original URL, we were able to identify hidden virality of a web archive URL that evades platforms' swift, automated moderation of harmful misinformation because it is hosted by a trusted web archive domain. In observing screensampling methods of archived URLs circulated amongst COVID-19 publics on platforms, the same web archives may be appropriated for different uses to increase doubt and spread dangerous and unreliable health misinformation.

Screensampling creates a memetic abstraction from the original source by converting a web resource into an image, resulting in a transmedia transformation of the existing content. In addition to converting text into a rasterized image, the screenshot may simultaneously encapsulate more layers of contextual information for researchers to examine, such as device's mobile network, timestamps, domain names, and other revealing diegetic user interface elements (Chaiet, 2019). While such identifying features are akin to traditional metadata found in digital formats, they are captured in the amber of an image instead of the object's metadata that could otherwise be extracted programmatically. Screensamples (like Figure 1) are screenshots of text so subsequent contextual "metadata" are not machine-readable, yet these image-based messages are human-readable and can subvert text-based content moderation systems.

Like other researchers who have studied web archivists and their crawling decisions (Maemura et al., 2018; Ogden et al., 2017), we find that individual human contributions played a role in the spread of this misinformation on platforms like Facebook and Twitter, as well as in its appearance in a number of Wayback collections seeded by Save Page Now. Social media analytics allows us to examine performance trends of URLs in the platform by comparing engagement metrics within the dataset to those of earlier or different versions of the URLs and their derivatives. Even so, more research needs to survey users, creators, and those that leverage web archive tools like Save Page Now about their perceptions and understandings of web archived URLs as they relate to trust, content moderation, and the spread of misinformation in platforms.

Bibliography

- Acker, A. (2018). *Data Craft: The manipulation of social media metadata*. Data & Society.
<https://datasociety.net/output/data-craft/>
- Acker, A., & Donovan, J. (2019). Data craft: A theory/methods package for critical internet studies. *Information, Communication & Society*, 22(11), 1–20.
<https://doi.org/10.1080/1369118X.2019.1645194>
- Ben-David, A., & Amram, A. (2018). The Internet Archive and the socio-technical construction of historical facts. *Internet Histories*, 2(1–2), 179–201.
<https://doi.org/10.1080/24701475.2018.1455412>
- Brügger, N. (2018). Web history and social media. *Sage Handbook of Social Media*, 196–212.
- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1–12. <https://doi.org/10.1177/2053951715622512>
- Caplan-Bricker, N. (2018, December). Preservation acts. *Harper’s Magazine*.
<https://harpers.org/archive/2018/12/preservation-acts-archiving-twitter-social-media-movements/>
- Chaiet, M. (2019). *Engineering Inflammatory Content* [University of Texas].
<https://mitchaiet.github.io/RussianAds/>
- Donovan, J. (2020, April 30). *Covid hoaxes are using a loophole to stay alive—Even after content is deleted*. MIT Technology Review.
<https://www.technologyreview.com/2020/04/30/1000881/covid-hoaxes-zombie-content-wayback-machine-disinformation/>
- Facebook. (2020). *Example articles—Instant articles—Documentation*. Facebook for Developers.
<https://developers.facebook.com/docs/instant-articles/example-articles/>
- Finn, M. (2018). *Documenting aftermath: Information infrastructures in the wake of disasters*. The MIT Press.
- Fraser, L. (2020). *What data is CrowdTangle tracking?* CrowdTangle.
<http://help.crowdtangle.com/en/articles/1140930-what-data-is-crowdtangle-tracking>
- Gillespie, T. (2014). The relevance of algorithms. In P. J. Boczkowski, K. A. Foot, & T. Gillespie (Eds.), *Media technologies: Essays on communication, materiality, and society* (p. 167). The MIT Press.
- Internet Archive. (2018). *Save pages in the wayback machine*. Internet Archive Help Center.
<http://help.archive.org/hc/en-us/articles/360001513491>
- Kertscher, T. (2020, March 9). *PolitiFact—No evidence that coronavirus was faked to cover up ‘5G Syndrome.’* Politifact. <https://www.politifact.com/factchecks/2020/mar/09/facebook-posts/no-evidence-coronavirus-was-faked-cover-5g-syndrom/>

- Kivran-Swaine, F., & Naaman, M. (2011). Network properties and social sharing of emotions in social awareness streams. *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, 379–382. <https://doi.org/10.1145/1958824.1958882>
- Littman, J. (2017, November 6). *Vulnerabilities in the U.S. Digital Registry, Twitter, and the Internet Archive*. Social Feed Manager. <https://gwu-libraries.github.io/sfm-ui/posts/2017-11-06-vulnerabilities>
- Madrigal, A. C. (2018, April 25). The evidence is not with Joy Reid. *The Atlantic*.
<https://www.theatlantic.com/technology/archive/2018/04/the-evidence-is-not-with-joy-reid/558935/>
- Maemura, E., Worby, N., Milligan, I., & Becker, C. (2018). If these crawls could talk: Studying and documenting web archives provenance. *Journal of the Association for Information Science and Technology*, 69(10), 1223–1233. <https://doi.org/10.1002/asi.24048>
- Milligan, I. (2016). Lost in the infinite archive: The promise and pitfalls of web archives. *International Journal of Humanities and Arts Computing*, 10(1), 78–94.
<https://doi.org/10.3366/ijhac.2016.0161>
- Narvo. (2020, March 12). *They are saying this is a cover up for the 5G roll out in Wuhan??* Twitter.
<https://web.archive.org/web/20200312132404/https://twitter.com/narvonocutz/status/1238036968635658241>
- Nelson, M. L. (2018, March 23). *Weaponized web archives: Provenance laundering of short order evidence*. National Forum on Ethics and Archiving the Web, New York, NY, USA.
<https://www.slideshare.net/phonedude/weaponized-web-archives-provenance-laundering-of-short-order-evidence>
- Ogden, J., Halford, S., & Carr, L. (2017). Observing web archives: The case for an ethnographic study of web archiving. *Proceedings of the 2017 ACM on Web Science Conference*, 299–308.
<https://doi.org/10.1145/3091478.3091506>
- Rogers, R. (2017). Doing Web history with the Internet Archive: Screencast documentaries. *Internet Histories*, 1(1–2), 160–172. <https://doi.org/10.1080/24701475.2017.1307542>
- The Millennium Report. (2020, March 9). *CORONAVIRUS HOAX: Fake Virus Pandemic Fabricated to Cover-Up Global Outbreak of 5G Syndrome* [Web archive]. Internet Archive Wayback Machine.
<https://web.archive.org/web/20200309221231/http://themillenniumreport.com/2020/03/coronavirus-hoax-fake-virus-pandemic-fabricated-to-cover-up-global-outbreak-of-5g-syndrome/>
- Walden, K. (2012). *Digital zombies, cultural cryogenics and troubles with malware: An archaeological adventure in the archives*. <http://uhra.herts.ac.uk/handle/2299/13988>
- Xie, B., He, D., Mercer, T., Wang, Y., Wu, D., Fleischmann, K. R., Zhang, Y., Yoder, L. H., Stephens, K. K., Mackert, M., & Lee, M. K. (2020). Global health crises are also information crises: A call to action. *Journal of the Association for Information Science and Technology*.
<https://doi.org/10.1002/asi.24357>
- Zannettou, S., Blackburn, J., Cristofaro, E. D., Sirivianos, M., & Stringhini, G. (2018, June 15). Understanding web archiving services and their (mis)use on social media. *Twelfth International AAAI Conference on Web and Social Media*.
<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17860>
- Zarocostas, J. (2020). How to fight an infodemic. *The Lancet*, 395(10225), 676.
[https://doi.org/10.1016/S0140-6736\(20\)30461-X](https://doi.org/10.1016/S0140-6736(20)30461-X)

Acknowledgements

The authors wish to thank Susan Shafer and the anonymous peer reviewers for their detailed feedback. This research collaboration was made possible in part by Good Systems, a UT Grand Challenge and Disinformation and Misinformation Studies at UT Austin.

Funding

This research was made possible by a grant from the Institute of Museum and Library Services. The project's grant number is RE-07-18-0008-18.

Competing interests

The authors confirm that we have no conflicts of interests to declare.

Ethics

This research involves no human subjects and is therefore exempt from institutional review board approval. All data gathered and used in this research is publicly available online.

Copyright

This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

Data Availability

All materials needed to replicate this study are available via the Texas Data Repository Dataverse: <https://dataverse.tdl.org/dataset.xhtml?persistentId=doi:10.18738/T8/LUZATR>