



A public online resource to track COVID-19 misinfodemic

Thomas Marcoux¹ · Katrin Galeano¹ · Rick Galeano¹ · Karen DiCicco¹ · Hayder Al Rubaye¹ · Esther Mead¹ · Nitin Agarwal¹ · Anna Galeano¹

Received: 15 October 2020 / Revised: 3 February 2021 / Accepted: 6 April 2021 / Published online: 5 May 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2021

Abstract

The COVID-19 pandemic has seen the rise of many unique online narratives through social media and other sources. They can range from theories about the origin of the virus, to misinformation regarding personal health measures. Such narratives have a direct impact on public health and safety. In an effort to inform the general public and with support of the Arkansas Attorney General's office our team sought to keep track of narratives for research purposes and provide near-real-time public documentation via a website with two main goals. The first is to track every unique narrative as curated by our analysts. This allows individuals to fact-check the information they consume. The second goal is to relay recommendations to the State Attorney General on how to detect such misinformation and avoid fraud and scams seeking to profit from this online fear and chaos. This paper showcases our website as well as some research findings from the data we collected. More broadly, this effort showcases a unique collaboration between scientists and policy makers to stem the flow of misinformation during a major public health crisis.

Keywords Misinformation · Disinformation · Social media · COVID-19 · Misinfodemic · Narratives

1 Introduction

As of September 2020, the USA holds the highest number of COVID-19 cases in the world, behind countries like India with over ten times the population density. While a number of factors can explain this, we believe one unique factor is the ongoing stream of misinformation spread through multiple online media, which is strongly politically polarized

in the USA, especially with the upcoming election cycle in November.

In close coordination with the Arkansas Attorney General's office and the Arkansas Research Alliance, we published a website tracking misinformation narratives and listing official tips on how to detect scams and disinformation campaigns.¹ According to Arkansas Attorney General Leslie Rutledge, "This partnership will ensure we are quickly identifying and putting a stop to fake websites that are spreading misinformation with the purpose of stealing from Arkansans."²

False claims, misleading information and the sale of fake cures and treatments undermine Arkansas' efforts to combat this disease and present a hazard to consumers' physical, financial and online safety. The collaboration with COSMOS offers a new avenue to quickly identify the bad actors who want to spread misinformation or steal consumers' money or identity. Through our data collecting, we have seen attempts at preying on the uninformed or elderly with scams such as personal protective equipment marketing schemes, online fraud or identity theft. For these reasons, we find it obvious

✉ Thomas Marcoux
txmarcoux@ualr.edu

Katrin Galeano
kkaniagalea@ualr.edu

Rick Galeano
Ragaleano@ualr.edu

Karen DiCicco
Klcottrill@ualr.edu

Hayder Al Rubaye
hkalrubaye@ualr.edu

Esther Mead
elmead@ualr.edu

Nitin Agarwal
nagarwal@ualr.edu

¹ University of Arkansas At Little Rock, Little Rock, USA

¹ <https://cosmos.ualr.edu/covid-19>.

² <https://arkansasag.gov/media-center/news-releases/rutledge-to-partner-with-ualr-tracking-false-covid-19-websites-and-scams-online/>.

that information has a very real impact on the lives and living conditions of many.

The rest of the paper is structured as follows. In section two, we discuss related research and tools seeking to understand this topic or provide information. In section three, we describe the website, our data collection process and an example of a daily product that was provided to the Attorney General's office. In section four, we expand upon the knowledge from this research. Finally, we conclude the study in section five.

2 Related works

The tool described in this paper constitutes our contribution to the efforts of the information community that has been fighting the spread of misinformation since the beginning of the outbreak. We recognize the importance of providing tools to inform the general public based on the research of Pennycook et al. (2020) finding that misinformation spreads in a viral manner and that consumers of misinformation tend to fail to recognize it as such. The spread of misinformation, even online, constitutes a very real danger to public health and safety (Kouzy et al. 2020), making the identification of such narratives urgently essential. We believe efforts to identify misinformation are helpful in curbing misinformation since research has found that simply recognizing the existence of misinformation and improving our understanding of it can enhance the larger public's ability to recognize misinformation as such (Pennycook et al. 2020). In fact, media outlets have a significant impact in health communications (Liu et al. 2020). Because of the importance of this task, we look into online solutions to curb the spread of misinformation.

A few digital tools and websites have been created to help report misinformation and provide reliable resources regarding COVID-19. One site, NewsGuard, provides a list of websites publishing false coronavirus information by providing websites, reports and articles. As of this writing, the NewsGuard Coronavirus Misinformation Tracking Center has over 335 sites from the USA, UK, France, Italy and Germany.³ Users can report misinformation through the misinformation hotline, Signal or WhatsApp, SMS, voice message, email and web form. NewsGuard also provides links to reliable websites such as the World Health Organization and the US Centers for Disease Control and Prevention.

The World Health Organization (WHO) provides a webpage on their website letting users know how to report

misinformation online.⁴ WHO provides links redirecting to the different social media websites on how to report flash news. The social media sites linked are: Facebook, YouTube, Twitter, Instagram, WhatsApp, TikTok and LinkedIn.

The Empirical Studies of Conflict (ESOC) project in collaboration with Microsoft Research provides a catalog of misinformation efforts around the pandemic. The data are provided via a spreadsheet with records of stories on misinformation efforts. ESOC also provides spreadsheets of URLs based on ESOC, AFP and NewsGuard data.⁵

Another website from Johns Hopkins Bloomberg School of Public Health, called COVID-19 Communication Network, is providing articles on many topics including myth-busting, how to report misinformation online, and teaching citizens to spot misinformation.⁶ This website provides a user-friendly card layout that provides the title, buttons to view the article and print and a description of the article that allows users to easily skim through the articles.

3 Methodology

In this section, we describe the thought behind the website and the methodology employed to collect and manage the misinformation narratives we publish and base our research on.

3.1 Website

Our main goals when designing the website (Fig. 1) were education and accessibility. It is especially important for those more vulnerable to misinformation such as elderly or uninformed Internet users. One aspect of the website is to provide a well-designed and trusted source of information where the general public is able to double check the information they come across. Another aspect is educating the general public by sharing official tips on how to detect misinformation and avoid scams. Lastly, we encourage citizen science by providing a "Report" feature. The feature allows users to send suspicious narratives to our team for analysis. We are currently researching the fields of human computer interaction and information visualization and applying our findings to make more compelling, quicker and easier to understand user interfaces. We aim to provide

³ <https://www.newsguardtech.com/covid-19-resources/>.

⁴ <https://www.who.int/campaigns/connecting-the-world-to-combat-coronavirus/how-to-report-misinformation-online>.

⁵ <https://esoc.princeton.edu/publications/esoc-covid-19-disinformation-tracking-report>.

⁶ <https://covid19communicationnetwork.org/featured-resource/rumors-misinformation/>.

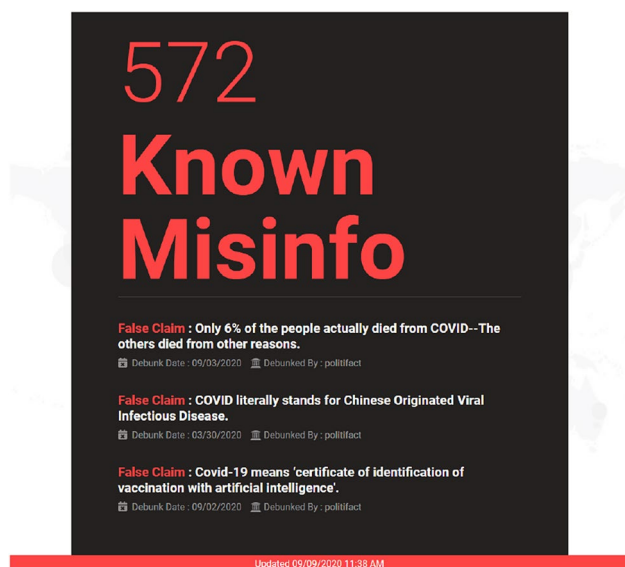
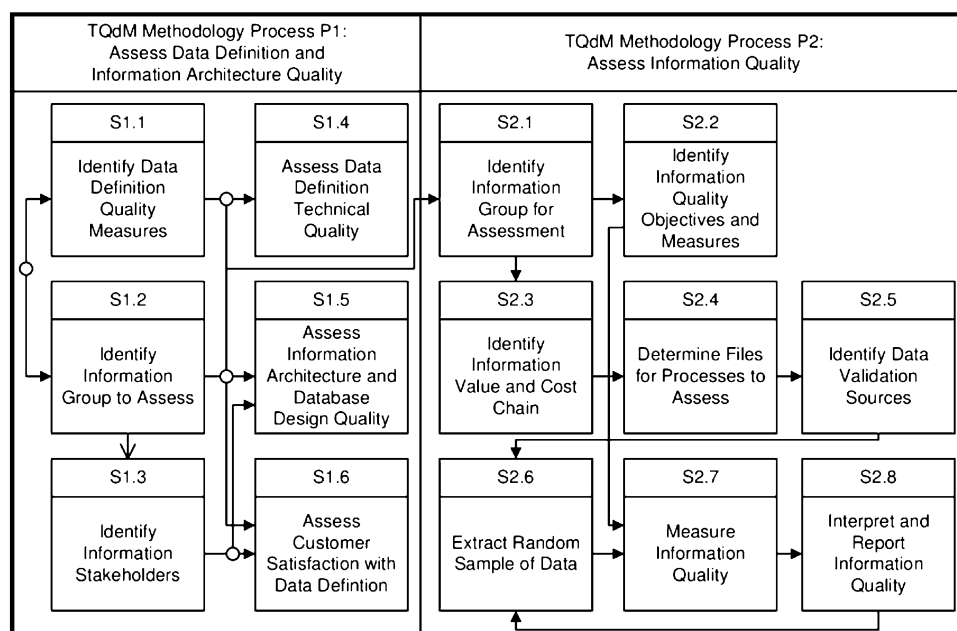


Fig. 1 COVID-19 website

Fig. 2 Process for measuring data quality



easy to understand user interfaces that reduce user friction and encourage education.

3.2 Database construction

The information displayed on the website was provided through a daily-updated database. The fields of the database were determined by the requirements of the project.

Each item, may it be a piece of misinformation or scam, received a brief title to provide the end user with a claim. It was then placed into one of four categories: false claim, mislead, scam or sensational headline. In order for the end user to learn more about the claim, a summary was added to the line item, followed by a debunking link from a reputable source, and the publication date of the debunking source.

To track the platforms on which the misinformation and scams were spread, the database includes fields to indicate the name of the platform(s), which included websites, social media, text, email, direct mail and verbal. If a website was involved, we included the domain name. This was done to determine which websites were spreading misinformation and scams. Oftentimes, social media platforms only allow a limited amount of characters to be published. Misinformation and scam spreaders are utilizing social media as a tool to entice an audience to click on an attention-grabbing headline that will lead to a website sharing disinformation or employing scam tactics.

Because numerous database users added input on a daily basis, we also included an internal comment field that did not feed into the public-facing website. Our data scientists were able to collaborate through this field adding additional information that did not fit into one of the aforementioned data fields. We shaped our data entry

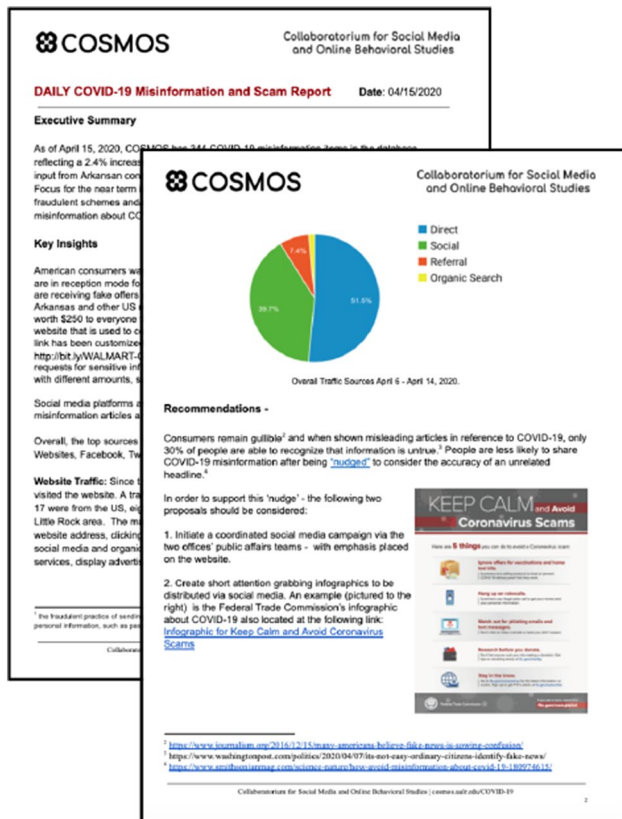


Fig. 3 Sample of daily COVID-19 misinformation and scam report

methods by increasing the number of data entry personnel and created a template for data input. Ensuring that we had a standard we followed Larry English's definition of data quality, "consistently meeting all knowledge worker and end-customer expectations," thus enabling them to do their jobs effectively.⁷ This is highlighted in Fig. 2 showing the methodology for the measurement of data quality. To facilitate this, our initial template for database entry was hastily done as initial with the intent of identifying what data existed and whom the messaging was targeting. Now, our process more closely resembles Total Quality Data Management (TQDM) as English described.

We also used Deming's research for the foundation to ensure TQDM standards for the data warehouse were continually improving. This process allowed for the overall management of the multiple sources, the continuous training of the team and the changing needs of the customer. Blending both these data scientists for the database construction was paramount for the end product which is discussed in the next section.

⁷ English, L. P., Improving Data Warehouse and Business Information Quality, Wiley, New York et al., 1999.

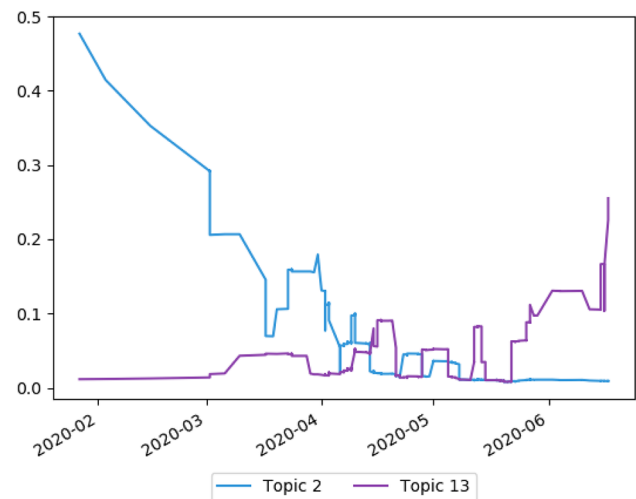


Fig. 4 COVID-19 misinformation topics stream

3.3 Database quality assessment framework

The data quality now was focused on input, evaluation and daily assessment reports to the Attorney General's (AG) office. Figure 3 displays one of the many assessment reports submitted to the AG's office. These reports are publicly available on the website.⁸ Our transition from longer reports to short and concise one to two page reports resulted in quicker turnaround times, more regionally focused analysis and increased feedback from within the partnership. The refined report included an executive summary, key insights and recommendations. These three topic areas also included any significant graphics or images that had been observed during that reporting period.

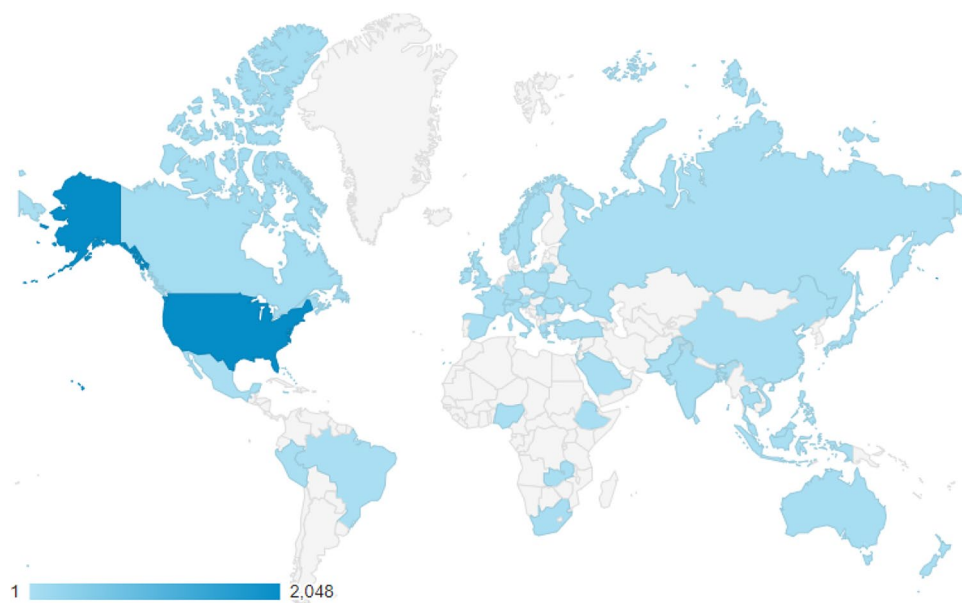
This was relevant to ensure the data input, evaluation, etc., was templated for evaluation by the Attorney General's office. This was done in the form of a daily report. Summarizing this was TQDM which allowed the overall management of the multiple sources, subjective judgment of data, limited resources, volume of data and the changing needs of the customer.

4 Results

In this section, we discuss our observations from the curated misinformation dataset, the performance of the website, as well as other research stemming from this effort.

⁸ <https://cosmos.ualr.edu/reports>.

Fig. 5 COVID-19 misinformation tracker website visitor world map (as observed during March 26–September 26, 2020)



4.1 Misinformation data

In a separate publication (Marcoux et al. 2020), we discuss the particular themes of misinformation we observed when aggregating these narratives. We used topic modeling to visualize the distribution and evolution of these themes over time. Topic modeling reveals latent topics by identifying common keywords occurring in similar documents. For example, “Topic 2” includes the words “purposeful,” “creators,” “bill gates,” leading us to label Topic 2 as narratives related to conspiracy theories, suggesting the virus is a man-made attack. The evolution of this narrative can be seen in Fig. 4. As speculations on the origin of the virus dwindle, we see a rise in various attempts at taking advantage of vulnerable citizens through scams or online identity theft. Figure 4 shows the rise of “Topic 13” which includes keywords like “scam,” “phishing,” “giveaways.” One other dominant topic is “Topic 15” with the words “government,” “control,” “citizens,” and “predicted.” This leads us to believe Topic 15 reveals conspiracies, suggesting the virus stems from a government effort. The main takeaway is that misinformation items attempting to spread fear about a potential COVID-19 vaccine and phishing scams remained prominent during June. During the month of July, the main themes of the misinformation items shifted back to attempts to downplay the deadliness of the novel coronavirus. Another prominent theme in July was attempts to convince the public that COVID-19 testing is inflating the results.

4.2 Website traffic

From March 26, 2020, to September 26, 2020, the website registered 3,776 page views and 2,857 unique page views

from users that accessed the website. Figure 5 shows the world map of all the users’ location. The average time users spent on the home page were 2 min and 55 s, and 198 users interacted with the “Report Scam” feature on the website. The age range of users that visited the website were: 25–34 (21.44%), 55–64 (19.8%), 65+ (17.78%), 45–54 (16.77%), 35–44 (15.26%) and 18–24 (8.95%). Out of the users that visited the site 54.7% were female and 45.3% were male. 58.61% of the users accessed the website via a mobile device, 37.13% via desktop and 4.26% via tablet.

5 Conclusions and future directions

Through our efforts in making this tool available and accessible to the general public, we hope to do our part in curbing the spread of misinformation and subsequent dangers to public health by informing the more vulnerable. To ensure this, we have worked closely with the Arkansas Attorney General and made frequent reports on our findings, which can be found on the website along with other press releases.⁹

While developing this tool, our efforts have led to a thorough data collection methodology used in multiple research projects. This includes research on epidemiological information propagation, computational detection of misinformation themes and narratives through topic modeling, etc. Future research with this tool may include the utilization of this growing dataset to analyze the spread of misinfodemic

⁹ <https://cosmos.ualr.edu/press>.

themes over time or determining how these themes fit in the context of political space.

Funding This research is funded in part by the US National Science Foundation (OIA-1946391, OIA-1920920, IIS-1636933, ACI-1429160 and IIS-1110868), US Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2675, N00014-17-1-2605, N68335-19-C-0359, N00014-19-1-2336, N68335-20-C-0540), US Air Force Research Lab, US Army Research Office (W911NF-17-S-0002, W911NF-16-1-0189), US Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock and the Australian Department of Defense Strategic Policy Grants Program (SPGP) (award number: 2020-106-094).

Code availability The website presented in the current study uses custom code.

Declarations

Conflict of interest There are no conflicts of interest.

Availability of data and material The datasets generated and analyzed during the current study are available at <https://cosmos.ualr.edu/covid-19>.

References

- Kouzy R et al. (2020) Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter. eng. In: Cureus12.3 Publisher: Cureus, e7255–e7255. issn: 2168–8184. doi:<https://doi.org/10.7759/cureus.7255>
- Liu Q et al (2020) Health communication through news media during the early stage of the COVID-19 outbreak in China: digital topic modeling approach. *J Med Internet Res* 22(4):e19118
- Marcoux T, Mead E, Agarwal N (2020) The Ebb and Flow of the COVID-19 Misinformation Themes. 5th international workshop on mining actionable insights from social networks special edition on dis/misinformation mining from social media (MAISON 2020). Co-located with 29th ACM international conference on information and knowledge management (CIKM 2020), October 20, 2020
- Pennycook G et al (2020) Fighting COVID-19 misinformation on social media experimental evidence for a scalable accuracy-nudge intervention. *Psychol Sci*. <https://doi.org/10.1177/0956797620939054>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.