

# Examining the digital toolsets of journalists reporting on disinformation

**Andrew Beers**

Human-Centered Design and Engineering  
University of Washington  
Seattle, Washington

**Ahmer Arif**

Human-Centered Design and Engineering  
University of Washington  
Seattle, Washington

**Melinda McClure Haughey**

Human-Centered Design and Engineering  
University of Washington  
Seattle, Washington

**Kate Starbird**

Human-Centered Design and Engineering  
University of Washington  
Seattle, Washington

## ABSTRACT

The rise of the socially-connected web has created new reporting challenges for journalists—especially around the spread of mis- and disinformation—and required them to adopt new reporting methods. Increasingly, they have turned to digital tools to verify information, collect leads and story updates, and draw connections between content online. To better understand the range of such tools in current practice, and the processes that journalists use these tools for, we interview 12 national and international journalists who specifically report on mis- or disinformation. We ask them questions about the tools they use in their reporting, and ask them to recount narratives of reporting individual stories. From their responses, we have extracted the range of online tools they use in their journalistic practice, and categorized them according to how they are used in the process of reporting a story. We discuss the implications of our review for both for journalists and for prospective developers of digital tools for journalists.

## CCS CONCEPTS

• **Human-centered computing** → **Social media**; • **Empirical studies in collaborative and social computing**;

## KEYWORDS

journalism, software development, verification, search

## ACM Reference Format:

Andrew Beers, Melinda McClure Haughey, Ahmer Arif, and Kate Starbird. 2020. Examining the digital toolsets of journalists reporting on disinformation. In *Proceedings of Computation + Journalism 2020 (C+J '20)*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/1122445.1122456>

## 1 INTRODUCTION

The rise of social media and, more generally, the socially-connected web has introduced new topics for journalists to report on, and has transformed the methods they use to conduct their investigations. One beat that journalists increasingly report on is the spread and origin of online mis- and disinformation. While the correction of rumors, conspiracy theories, and lies has been a perennial duty of journalists, the increased scope and technologically-supported nature of mis- and disinformation in recent years has required them to adopt new methods. Particularly, many journalists now use digital tools that take advantage of techniques of data science, machine learning, and the software infrastructures of online platforms [5]. Software developers, data scientists, and newsrooms themselves are now working to create more advanced digital tools for journalists, tailored to the practical needs of day-to-day reporting [4].

In this paper, we present the results of structured interviews with twelve journalists who broadly report on mis- and disinformation. While they spoke on a range of subjects, we focus here on the specific technologies that they use in the course of reporting on mis- and disinformation. We categorize tools according to their usage during different tasks within the reporting process, namely content verification, content monitoring, and relationship mapping. We focus on the tools that journalists use themselves, as opposed to tools used by their team members or external collaborators.

We write this paper for two audiences. The first audience is journalists who report on mis- and disinformation or related topics. We hope that other journalists outside of our interview sample can learn from the tools and techniques we

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*C+J '20, March 20-21, 2020, Boston, MA*

© 2020 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/1122445.1122456>

make explicit here, and benchmark their own newsrooms' technical development against those mentioned here. Making these tools visible is especially important for newsrooms that may not have technical development teams of their own. The second audience is software developers and engineers who design tools for journalists. By categorizing and cataloging the range of existing tools, we help surface both what tools have been successful in serving journalists' needs and what reporting processes have had less adoption of digital tools.

## 2 BACKGROUND

### Journalists and Online Misinformation

In her work "Lexicon of Lies," Caroline Jack defines misinformation as "information whose accuracy is unintentional" and disinformation as "information that is deliberately false or misleading" [7]. As public concerns around mis- and disinformation have increased, so too have journalists' interest in reporting on it. For example, journalism education and training resources are being developed to address the growing threat that both the public and journalists' themselves face from disinformation campaigns [e.g. 6]. Journalists also hold a privileged position in the effort to dispel mis- and dis-information. Starbird, Dailey et al. observed journalists on Twitter are more likely than other users to be retweeted when denying rumors [10]. Consequently, it is particularly important for toolmakers to attend to the needs of journalists when they work to reduce the spread of problematically misleading information in online settings.

From a Computation + Journalism perspective, the work of reporters in this space is particularly interesting because it is not only supported by technology but also deeply dependent on it. While many journalists are incorporating digital tools like social media in their labor, finding and telling stories about disinformation campaigns requires journalists to learn how to keep their ear to the internet in new ways — blurring the lines between *technology-supported journalism* and *technology-infused journalism* [9].

### Digital Tools for Misinformation Journalists

Engineers, software developers, and human-computer interaction researchers have taken note of journalists' increasing interaction within online spaces and with online data, and have sought to develop tools to address their needs in this emerging area. However, these tools are not always widely adopted by journalists due to lack of awareness, lack of access, or lack of trust in their abilities [1]. Understanding how we can address this issue is important as the Computation + Journalism community continues to explore tools that help journalists report on mis- and disinformation online [e.g. 8]. Our work contributes to these efforts by describing how

journalists determine when and what technologies support their work.

## 3 METHODS

We conducted an empirical study consisting of twelve semi-structured, hour-long interviews with national and international journalists who report on mis- and disinformation online. The twelve interviews shed light on many topics, including the tools that journalists use in reporting on topics like mis- and disinformation. The journalists were encouraged to tell a story about a recent investigation, discussing the entire process from finding the lead to publishing an article. The interviewers paid attention to tools that the journalists used in their process and asked follow up questions about other tools they may use in this type of work. The interviews were recorded and transcribed, and then analyzed using a grounded theory approach [2]. Though the analysis surfaced a range of emergent themes, for this paper we focus on themes related to the tools that journalists use to do this type of work.

## 4 FINDINGS

Journalists participating in our interviews used different digital tools to address to different needs in their investigations. In the following subsections, we classify the extent of digital tools used into three broad areas of practice.

### Verification and Establishing Provenance

*I personally find it more interesting to go back and do an autopsy or excavation of a conspiracy theory or a particular narrative or meme... How did this start? Who started it? Where did it come from? —Participant*

When encountering a suspect piece of information, most journalists we spoke to first seek to establish its provenance. Typically, this means finding and potentially contacting the person who initially posted that information via social media, and questioning them as to the information's origin. If they are able to find an original poster, they can initiate contact to try to learn more about the information. However, those who post misinformation may not always be contactable; as one journalist summarized, "they hardly ever respond." Furthermore, those who share information may be very socially distant to the original source of that misinformation. In these cases, journalists have to track down the source of the information themselves. The journalists we interviewed used a number of digital tools to support audiovisual, time, location, and identity verification.

*Audiovisual Verification.* The widest variety of tools used by the interviewed journalists pertained to verifying suspicious image and video data. Previous research suggests that audiovisual content is more likely to be shared across several

different social media platforms, and is more likely to be spread in viral social sharing on Twitter [12]. This inherent virality makes images and video a particularly potent vector for misinformation, and a point of specific attention for several of the journalists we interviewed.

Six of the journalists specifically mentioned using a form of reverse image search, in which users can submit an existing image to a service that cross-references it against a large archive of previous images. Four journalists referenced using Google's reverse image search, with the remaining two referencing reverse image search as a generic term without a specific tool name. One journalist also used TinEye and Yandex, two other reverse image search tools, to supplement Google's reverse image search. Three journalists used online tools to verify video content, with two saying specifically that they use InVid, a free tool for video and image verification developed for journalists specifically [11]. These same journalists used InVid and a separate tool, FotoForensics, to assess whether a video or image had been digitally altered.

*Time and Location Verification.* To supplement audiovisual verification, one journalist reported using a wide range of techniques for time and location verification. If this journalist encountered potential misinformation that referenced a specific location at a certain time, he could check Google satellite and street views for discrepancies via Google Maps. The same journalist would check the social media accounts of a person who claimed online to be in a certain location, to see if they had left evidence they were at a different location at the time posting. They also would check online weather archives for the location, to verify both events and imagery presented in a potential piece of misinformation.

*Identity Verification.* The second-most significant method was identity verification, in which journalists attempt to verify the legitimacy of a suspect persona posting information online. In addition to attempts to directly contact these personas, four journalists used online databases of identity information to check if these personas matched the biographical details of an already indexed person. LinkedIn, Nexus, and staff directories on potential employers' homepages were all used as databases to check identities against. Two journalists checked suspect personas against online databases of exposed fake accounts. The databases mentioned were Russiatweets.com, which archives tweets created by the Internet Research Agency, and Twitter's published archives of accounts suspended for misinformation and disinformation. One journalist used BotOrNot, an algorithmic tool which determines a given Twitter user's similarity to previously identified bot accounts [3]. One journalist used reverse image tools not only to verify audiovisual misinformation, but to verify the identity of social media users via their profile pictures. Another journalist described tools to verify the

creators of a suspicious website. Specifically, they visited CentralOps, which hosts several different tools for potentially determining the owner or location of registration of a website.

### Tracing the Spread of Content

*I often reverse engineer. I'm very interested in the mechanics of how things spread, right? So I guess a lot of times, what I'm doing is just sort of tracing the the ecosystem of whatever piece of content. —Participant*

Multiple journalists stressed the importance of reporting on not only the veracity and provenance of a piece of information, but also on how that information spread throughout the online information ecosystem.

*Searching.* Ten journalists explicitly mentioned search engines as a key part of their reporting process. Specifically, journalists used the built-in search engines provided in social media platforms such as Twitter, Facebook, Instagram, 4Chan, 8Chan, TikTok, and Reddit. Search tools for these platforms were often used simultaneously in one story, to understand the relevance of a search term in different social media communities. Three specifically mentioned using platforms advanced search features, with two using Twitter Advanced Search and one using Google Advanced Search. Advanced search was used to filter results by date, or to only display results within a certain website domain in the case of Google. Three referenced previously using Facebook's now defunct graph search, which provided similar services to other advanced search tools, but also let users search against Facebook users' tagged photos and location check-ins.

Two journalists used more advanced search tools. One, citing difficulty using Twitter's advanced search, used the Trump Twitter Archive. The Trump Twitter Archive stores all of U.S. President Donald Trump's tweets, categorizes them by topic, and has extensive search functionality, simplifying that reporter's attempts to retrieve information. Another used a marketing toolbox called Khoros (formerly Spreadfast), which let them search for the prevalence of hashtags at different points in time.

*Drawing Connections.* Journalists seek to draw connections, in some cases using the metaphor of a "network" or "ecosystem" to describe the channels by which misinformation is shuttled from person to person, or from platform to platform.

Six journalists used CrowdTangle, a free browser extension and monitoring platform. In its browser extension, CrowdTangle lets users understand where a given link has been shared on Facebook, Twitter, Reddit, and Instagram, and highlights the most popular users who have shared that content. Only one journalist referenced using a function specific to CrowdTangle's monitoring platform, which requires preauthorized access from Facebook, CrowdTangle's owner. This

journalist described compiling lists of accounts to monitor for the reporting practice, and then feeding those lists to CrowdTangle's list monitoring feature.

Few journalists used digital tools for drawing connections other than CrowdTangle. One journalist used BuzzSumo, a marketing platform that can among other use cases track the relative popularity of a search term on different social media platforms. The same journalist used Analyze ID, a marketing platform that searches advertising services like Google AdSense to find websites that share a common owner. Another journalist used Hashtagify, a Twitter marketing tool that displays, for a given hashtag, its most popular users, frequently co-occurring hashtags, its popularity over time, and other metrics.

*Archiving and Analysis.* Eleven journalists mentioned using either Microsoft Excel or Google Sheets to store structured data, with the majority of interviewees favoring Excel. Journalists used these database technologies to store and view the results of content searches, either via the bulk download of searches or, more typically, copying and pasting of search results in spreadsheet rows. Other journalists receive spreadsheets either from collaborators or as outputs from digital tools, such as CrowdTangle. One journalist, whose team focused on data visualization and had more programming experience, mentioned that some members of their team use SQL databases to store information.

Among those journalists who stored structured data in their reporting, few reported using any additional tools to analyze that data. One journalist specifically mentioned Excel's pivot tables functions, which can create aggregated metrics and visualizations from spreadsheet data. Another used InfoGram, a data visualization and report generation tool. Two journalists working in data visualization and engineering respectively had members of their team who used the programming languages Python and R. Notably, only one journalist interviewed outside of these roles mentioned using programming languages, although many had newsroom team members or collaborators who could use such languages for them. Most commonly, analysis methods were ad-hoc and performed within the native capabilities of Excel and Google Sheets.

### Monitoring New and Developing Stories

*A lot of the story ideas that I get either come from something that pops across the feed, and then I go in and sort of, you know, dive deeper. Or it's something where I can sort of see something happen over a period of weeks or months - like one or two or three or four times. And then by the fourth time, you know, I realized that seems to be a trend. —Participant*

While the previous reporting processes discussed are retrospective, analyzing misinformation that has already been

posted, all journalists interviewed used tools and strategies to note the real-time emergence of *new* information.

*Social Media Platforms.* All journalists engage in informal monitoring on social media platforms, whose built-in features are reconstituted into journalistic tools. Using their personal accounts or, in rare cases pseudonymous accounts, journalists watch live updates provided by the platform, such as the Twitter Timeline or Facebook's News Feed. Others will join groups within these platforms, such as Facebook's groups or private chatrooms on Discord or Telegram, and monitor their responses. Online forums such as 4Chan and 8Chan were also monitored by journalists for discussions of new disinformation campaigns or of how to sustain existing disinformation campaigns. One journalist periodically checks a stored list of news websites known to distribute mis- and disinformation, and two other journalists use browser's built-in bookmarking features to do the same.

*Advanced Tools.* Few journalists used advanced digital tools to enhance this process. Two journalists used TweetDeck, a tool built that allows for more customized monitoring of one's Twitter timeline and the ability to manage multiple accounts simultaneously. One used CrowdTangle's monitoring platform to watch for updates from predetermined lists of users. One journalist with programming experience took advantage of 4Chan's API to search for all links to Discord chatrooms, which have a standardized URL format, and then joined and monitored those chatrooms.

## 5 DISCUSSION

In this paper, we have discussed the digital tools used by a set of journalists reporting on mis- and disinformation, as conveyed to us via semi-structured interviews. We categorize tool usage by misinformation journalists into three different areas of practice: verification, drawing connections, and real-time monitoring. We show how journalists assemble sets of diverse tools — based on their needs, their expertise, their resources, their collaborators — into their own, personalized digital work environment.

While digital tools have to some extent been integrated into each of these areas of practice, some areas had a wider variety and adoption of tools than others. The journalists we interviewed had access to many resources for verifying and establishing the provenance of data, but relatively fewer resources for monitoring content and drawing connections to understand information trajectories. Whether these differences reflect a greater need in these areas, or a higher level of satisfaction with the tools already available, should be researched further. Supporting the former explanation, however, is that many journalists self-deprecated about what they perceived as their primitive monitoring efforts. As one

journalist said when describing their workflow: *That's what I do on a daily basis — basically lurk on Facebook and Instagram.*

Database usage, and subsequent methods of structured data analysis, were a relative technological bottleneck in these reporters' processes. Most journalists use only Excel or Google Sheets as their primary database service, and did not have experience with programming languages that could manipulate spreadsheet data contained in these tools. Accordingly, their analyses were mostly contained to the native features of Excel and Google Sheets. Those looking to update or create tools for journalists would do well to note this software convergence point in the reporting process, and understand their tools' interoperability with spreadsheet manipulation resources.

A common theme across all areas of practice is that, except in rare cases, these digital tools are not developed for journalists. Most tools are developed either for general or specifically for marketers. Only the video verification tool InVid and the link monitoring tool CrowdTangle are marketed to journalists specifically. This could alternately indicate that there is a need for additional tool development aimed towards journalists, or conversely that journalists benefit from the affordances of tools that are built for and sustained by a larger audience.

The journalists we interviewed for this paper are relatively well-resourced, which is a limitation of this study. Higher budgets, a broader reporting beat, and historical prestige can give journalists additional time for learning how to use new tools, additional contact with the developers and users of advanced tools, additional funds to purchase marketing tools, and a greater mandate to report on datasets of national or global scope. Thus, the tool usage of this group of journalists likely does not reflect the usage of all journalists nationwide. Indeed, tool usage varied within this study, with some journalists interviewed barely using any technology outside of the native features of social media platforms, and others using upward of ten individual services for their reporting.

These interviews contained more insights than an explicit survey of digital tools can provide. Journalists also discussed their reporting goals, their pain points with reporting on misinformation, the ethical considerations in amplifying misinformation, and more. These findings will be explored in future work. We also hope to expand this research into local and regional newsrooms, and to use findings from these interviews to create mechanisms of support for journalists writ large who are investigating information online.

## 6 CONCLUSION

We have written this review of digital journalistic tools-in-practice to serve two audiences. By providing journalists a list and categorization of existing tools, we normalize the ways in which journalists report on mis- and disinformation

and potentially lower the barrier of entry into this space. In examining the tools used in practice, tool developers can investigate where there are gaps in providing for journalists and why certain tools may be more used than others in practice, which can inform future design.

## ACKNOWLEDGMENTS

This research was done at the University of Washington em-COMP lab with support from NSF Grant 1749815. We thank the journalists who were willing to interview with us for this work as well as the many students who provided significant assistance to this project through their participation in a Directed Research Group on the topic.

## REFERENCES

- [1] Ester Appelgren and Gunnar Nygren. 2014. Data Journalism in Sweden: Introducing new methods and genres of journalism into "old" organizations. *Digital Journalism* 2, 3 (2014), 394–405.
- [2] Kathy Charmaz. 2014. *Constructing rounded theory*. Sage.
- [3] Clayton Allen Davis, Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2016. Botornot: A system to evaluate social bots. In *Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web Conferences Steering Committee, 273–274.
- [4] Nicholas Diakopoulos. 2016. Computational journalism and the emergence of news platforms. *The Routledge Companion to Digital Journalism Studies*, London: Routledge, Taylor and Francis group (2016).
- [5] James T Hamilton and Fred Turner. 2009. Accountability through algorithm: Developing the field of computational journalism. In *Report from the Center for Advanced Study in the Behavioral Sciences, Summer Workshop*. 27–41.
- [6] Cherylyn Ireton and Julie Posetti. 2018. *Journalism, fake news & disinformation: handbook for journalism education and training*. UNESCO Publishing.
- [7] Caroline Jack. 2017. Lexicon of lies: Terms for problematic information. *Data & Society* 3 (2017).
- [8] Hana Matatov, Adina Bechhofer, Lora Aroyo, Ofra Amir, and Mor Naaman. 2019. DeJaVu: A System for Journalists to Collaboratively Address Visual Misinformation. In *Proceedings of the Computation + Journalism Symposium*.
- [9] C Seth. 2016. Mapping the human-machine divide in journalism. *The SAGE handbook of digital journalism* (2016), 341.
- [10] Kate Starbird, Dharma Dailey, Owla Mohamed, Gina Lee, and Emma S Spiro. 2018. Engage early, correct more: How journalists participate in false rumors online during crisis events. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 105.
- [11] Denis Teyssou, Jean-Michel Leung, Evlampios Apostolidis, Konstantinos Apostolidis, Symeon Papadopoulos, Markos Zampoglou, Olga Papadopoulou, and Vasileios Mezaris. 2017. The Invid plug-in: web video verification on the browser. In *Proceedings of the First International Workshop on Multimedia Verification*. ACM, 23–30.
- [12] Joshua A Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. [n.d.]. *Social media, political polarization, and political disinformation: A review of the scientific literature*.