On the Misinformation Beat: Understanding the Work of Investigative Journalists Reporting on Problematic Information Online

MELINDA MCCLURE HAUGHEY, Human Centered Design & Engineering, University of Washington
MEENA DEVII MURALIKUMAR, Human Centered Design & Engineering, University of Washington
CAMERON WOOD, Human Centered Design & Engineering, University of Washington
KATE STARBIRD, Human Centered Design & Engineering, University of Washington

Journalists are increasingly investigating and reporting on problematic online content such as misinformation, disinformation, and conspiracy theories, leading to the creation of a new misinformation beat. The process of collecting, analyzing, and reporting on this kind of data is complex and nuanced. It is especially challenging as online actors attempt to undermine their work. Through in-depth interviews with twelve journalists, we explore how they investigate and report on online misinformation and disinformation. Our findings reveal some of the unique challenges of reporting on this beat, as well as the ways in which reporters overcome those challenges. We highlight and discuss how journalistic values could be better embedded into the design of tools to support their work, the power dynamics between social media companies and journalists, and the promise of collaborations as a way to support and educate journalists on this beat. This work provides contextual knowledge to researchers looking to better support investigative journalists — on the misinformation beat and beyond — as their work becomes more entangled in sociotechnical systems.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: social media; journalism;misinformation; disinformation; media manipulation

ACM Reference Format:
Melinda McClure Haughey, Meena Devii Muralikumar, Cameron Wood, and Kate Starbird. 2020. On the Misinformation Beat: Understanding the Work of Investigative Journalists Reporting on Problematic Information Online. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 133 (October 2020), 22 pages. https://doi.org/10.1145/3415204

1 INTRODUCTION

Citizens are increasingly turning to online social platforms to seek and share information, including news [83]. This dynamic has enabled information to quickly propagate to large and diverse audiences, spreading within and between online networks in near real-time. In some cases, this information-sharing is problematic — e.g. with the pervasive spread of mis- and disinformation. CSCW researchers have been covering these topics at a growing rate, for example documenting cases where online platforms have facilitated harassment [54, 90], election interference [22], and the spread of conspiracy theories during an epidemic [62].

Authors' addresses: Melinda McClure Haughey, mhaughey@uw.edu, Human Centered Design & Engineering, University of Washington; Meena Devii Muralikumar, mmeena@uw.edu, Human Centered Design & Engineering, University of Washington; Cameron Wood, camw21@uw.edu, Human Centered Design & Engineering, University of Washington; Kate Starbird, kstarbi@uw.edu, Human Centered Design & Engineering, University of Washington.

Journalists, too, are increasingly reporting on problematic information in online environments. Experienced internet, social media, and national security reporters, among others, are starting to cover this new "beat" full time, focusing on issues of "fake news" [84], misinformation (e.g. [78, 80]), disinformation (e.g. [46, 66]), conspiracy theories (e.g. [65, 96]), and harassment (e.g. [38, 63]). In doing this work, journalists aim to inform readers about what is happening online, investigate and debunk false claims, and hold social media companies accountable for the problematic information that spreads on their platforms. Compared to researchers, journalists are often exploring these phenomena on truncated timelines and with fewer resources, yet they rely on some of the same kinds of data and analysis techniques.

Reporting on online activity is complicated, as there are technical, methodological and professional hurdles to navigate. The "online misinformation beat" makes this even more complex. This beat covers both mis- and disinformation. For the sake of definitions here, misinformation is false information that is not necessarily purposefully false, and disinformation is false or misleading information that is intentionally produced and/or spread with an objective [59, 85]. Pervasive mis- and disinformation have been recognized as critical problems by CSCW researchers [60, 85] and journalists alike [25, 93].

Though misinformation, disinformation, and conspiracy theories are not new, journalists reporting on this beat face new challenges in this era of "fake news". For one, the information ecosystems that they exist and compete within contain large numbers of low-quality "news" sites — including clickbait sites and hyper-partisan news sites [57] — that do not attempt to follow professional standards of journalism. Additionally, the "fake news" term has been turned into an attack on professional journalism, attempting to foment distrust in the fourth estate [27]. Stories on misinformation, disinformation, conspiracy theories, radicalization, etc. have an inherent relationship to conflicting claims about "truth", rendering solid analysis and accuracy even more important. In this work, journalists cannot afford to make mistakes. Furthermore, there are people and organizations — e.g. those that manipulate online spaces for political and/or financial gain — who actively endeavor to trip up these reporters. They stand ready to seize upon mistakes made by the journalists trying to hold them to account and use those mistakes to undermine the journalists' work [42].

In this paper, we explore the emerging "online misinformation beat" to understand the challenges that journalists face as they attempt to adapt their work practices to this new kind of work. We document the evolving environment of journalism in studying online social data, the ways in which journalists use their expertise to investigate online social spaces, and the tensions that come from doing this kind of reporting. We interviewed twelve journalists who are writing or contributing to stories that focus on misinformation, disinformation, and other forms of problematic information online. Several of these interviewees are ground-breakers of this new journalistic form. Through one-hour interviews, we asked them to situate themselves within the larger organization or newsroom, and then walk us through the lifecycle of one of their recent reports, among other things. This gave us a glimpse into their world as they navigate this complex work. Using an inductive analytical approach, we surfaced a range of salient themes across the interviews.

The findings suggest that use of computational tools alone in studying online information is insufficient; subject matter expertise, contextual knowledge, and access to data are essential pieces in making sense of these complex topics. Journalists overcome some of these challenges by using unique processes and tools in their investigations and by seeking help from experts. We also found that power dynamics between journalists and social media platforms complicate their efforts. Understanding these dynamics, and how they interplay with traditional journalistic values, is important as we think about the future of this kind of work. Our findings have implications at the intersection of future research and practice — e.g. designing tools to support these emerging

data-based, story-telling practices and exploring collaborations between academic researchers and journalists.

2 BACKGROUND

In this literature review, we first discuss how journalists are adapting their practices to changes in online behavior, economic models, and public trust in journalism. We then note how researchers and newsrooms themselves have been thinking about how to support journalists in light of these changes. We highlight how tools have been seen as one solution, and share areas where researchers are building them for this purpose. Finally, we discuss how online misinformation and disinformation have complicated the work of journalists and how this has led to a new reporting beat — one that CSCW research can contribute to understanding and supporting.

2.1 Changing Journalism Work Practices

Disrupted by the growing reliance on internet technology, traditional journalistic work practices — and the conditions under which that work is done — are changing [37], in part due to the shift in advertising revenues to the internet-based social media platforms [70, 72]. This shift has resulted in a loss of local journalism, under-resourced newsrooms, and the requirement for some professionals to move to remote work [17, 53]. A decrease in newspaper sales and loss in traditional advertising revenue has forced news organizations to adapt and be creative with how to fund their journalism [12]. They have had to reduce staff [53], decide how to implement paywalls [20], and set up social media presence to stay relevant in the attention economy [72].

These pressures are further complicated by an increase in citizen reporting [50] and the related displacement of traditional journalists as gatekeepers. Working outside of the constraints of traditional journalism, citizen reporters are often able to move faster than professional journalists. While this enables coverage of local topics and breaking news [18], it may reduce the quality of information — as these citizen reporters do not have to adhere to traditional journalistic standards like accuracy and verification. Professional journalists must now compete with these citizen journalists for public attention online. This puts pressure on professional journalists to "keep up" or be irrelevant, which can in turn lead to poor reporting or accidental missteps. And these errors can then reduce public trust, especially when the topics covered are controversial.

With the growing salience of online platforms and fewer resources to track down sources in person, journalists are increasingly looking to online social spaces for leads, statements from public officials, and investigations [43, 61, 82]. Additionally, as platforms such as YouTube, Twitter, and Facebook have become modern society's 'keepers of cultural discussion' [48], journalists are now investigating incidents of online participation as "the story" itself. Navigating these online platforms to search for sources and investigate stories requires a new way of working, and that, in turn, requires journalists to learn and adopt new skills. It also impacts the role that they play within the environment. For example, Hermida suggests that in the age of ambient journalism on Twitter, journalists may need to start helping the public process and make sense of the flow of information [58]. Starbird and Dailey examined how journalists are adapting in the context of crisis events, finding that journalists serve the public in these events by stitching social media information together for their readers [35]. Graves wrote about the rise of fact-checking as its own kind of story and how in hyper-partisan times, this role can challenge the traditional role of journalists as objective observers [52]. More recently, Coddington studied the practice of news aggregation stories and noted that journalists must manage a greater amount of uncertainty when aggregating information for readers since they are not in direct contact with sources [32]. He acknowledged that the degree of separation from the original source does, somewhat problematically, "pass the

uncertainty on to a distrustful public." These new roles, prompted by the increased public use of social media, have come with new needs and added pressures for journalists.

Acknowledging these pressures of changing conditions and changing practices, researchers — and newsrooms themselves — have been thinking about how to support these journalists by creating tools to help them with their work [39].

2.2 Computational Tools for Modern Journalism

As early as 2000, the HCI community was thinking about how to better empower journalists with computational tools — e.g. mobile tools for reporters on the ground to help them keep up with news while on the go [44]. Since then, researchers have continued to help newsrooms innovate as the internet became more ubiquitous, for example by creating flexible systems that allow them to begin publishing their stories online [69]. As data became more accessible, Hamilton and Turner [33] drew attention to the emerging field of computational journalism and suggested that data tools designed for journalists would enhance and enable new kinds of reporting in the future.

Researchers have continued to design systems to both support journalists as they adapt to technology-driven changes in their work and to study how those systems are used in practice. For example, Diakopolous and collaborators created tools to help journalists pull out important pieces of social media data, such as potential eyewitnesses and newsworthy tweets, from the firehose of social information [40, 41]. Two other tools, GroundTruth and DejaVu, help journalists with image verification. GroundTruth [89] was designed to augment and accelerate journalists' work by connecting them with crowdworkers who could quickly geolocate images for verification. Matatov et al. [68] created DejaVu to help journalists identify manipulated images online and work collaboratively with other journalists to fight misinformation. Matatov et al. presented DejaVu at the annual Computation + Journalism Symposium, where new tools to help journalists navigate social media and other kinds of data are regularly showcased e.g [23, 31, 77].

Researchers have also studied how journalists adopt and use data tools in the newsroom. Though many newsrooms have been able to use tools that aggregate data from social media to understand their audience and track story performance [64], they have been slower to incorporate social media data into the stories themselves. Appelgren and Nygren [21] studied editors' perceptions of data journalism and its corresponding methods, finding that the interest in data journalism was growing but hindered by a lack of time and resources to learn how to handle large data sets. Focusing on social media verification strategies, Brandtzaeg et al. [29] conducted an interview study to understand the practices of journalists in Europe. They concluded that journalists use a combination of traditional methods and online tools, like reverse image searches, to verify images and video on social media. However, they acknowledged that most journalists do not have the skills, knowledge, or access to perform independent and timely assessment of online content. Though there have been meaningful strides in designing for journalists and understanding their needs, the challenges involved in using social media data suggest that there is more to learn. Watkins and Anderson [94] suggests that with the right approach newsrooms can innovate and overcome inherent challenges of technology adoption. Our work examines the tools and processes of a unique and emerging beat in journalism, offering another perspective of modern newsroom challenges and needs.

2.3 The Challenge of Reporting on Misinformation and Disinformation

In particular, this research seeks to understand the emergent practices, skills, and needs of journalists whose work focuses specifically on online misinformation and disinformation — information toxicities that are increasingly recognized by scholars (e.g. [86, 91, 92]) and the broader public (e.g. [14, 25, 71]) as critical societal problems. A small, but growing, group of journalists has begun to

identify with what we are calling the "misinformation beat," though they use a range of different terms for themselves.

As journalists break ground on this new beat, they are learning to recognize and cope with their roles as both reporters and as participants in problematic information flows. These dual roles become particularly salient in the context of disinformation. As mentioned above, misinformation is false information that may or may not be intentional — e.g. a rumor that turns out not to be true. And disinformation is false or misleading information produced and spread for a political, financial, or other objective. Starbird et al. [85] advise that it can be useful to think of disinformation not as a piece of content, but as a campaign. Disinformation campaigns often attempt to obscure the origins and purpose of their actions with the goal of manipulating perceptions of reality in a target population.

Our current understanding of disinformation can trace some of its roots back to Soviet active measures. Bittman [26], a disinformation practitioner who defected and became a researcher, stressed how Soviet "active measures" specifically targeted journalists — e.g. luring them into spreading disinformation through the prospect of getting a "scoop". He described how disinformation operators attempted to manipulate journalists into serving as "unwitting agents" in their campaigns. Relaying a quote attributed to another Soviet military officer, Bittman wrote: "Sometimes I am amazed how easy it is to play these games, if they did not have press freedom, we would have to invent it for them."

Bittman defected in the 1970s and so we can expect that tactics have evolved considerably since his practitioner days, but modern disinformation campaigns have continued to use journalists to achieve their aims. For example, many journalists played a (likely unwitting) role in Russia's efforts to interfere with the 2016 U.S. presidential election by publishing emails that had been hacked by Russian agents and released through Wikileaks [87]. More recently, journalists and media organizations helped to amplify several high-profile, false narratives, including a recorded confrontation between a teenager wearing a "Make America Great Again" hat and a Native American activist [34] and a heavily altered video of Joe Biden appearing to make overtly racist statements [97]. In many cases, individuals and organizations pretended to be someone who they were not — a form of inauthenticity — to seed problematic content that was then picked up and amplified by others, including seemingly well-meaning journalists.

One theory about the long-term consequences (and perhaps objectives) of disinformation on democratic societies is that it functions to confuse and disorient citizens [75] in a way that makes it difficult to come together to govern themselves [45]. In this perspective, disinformation is, in part, an attack on our ability to find information that we can trust. When journalists get "caught" spreading misinformation, this can serve to advance the overall goals of some disinformation actors — i.e. diminishing trust in information providers. Journalists on the misinformation and disinformation beat have begun to understand these dynamics, but they continue to struggle with their own role in these campaigns — and to make decisions about when to publish and what to publish. Even in seemingly cut-and-dry situations, there are concerns that journalists may be amplifying mis- and disinformation through their efforts to debunk it [74].

Researchers have talked about this sort of exploitation of journalists as a form of "media manipulation" and found that it has contributed to decreased trust in mainstream media [67]. And recently, there have also been efforts to develop recommendations [74] and resources [13] for journalists to help them navigate questions of when and what to publish when investigating stories of online harassment, radicalization, and disinformation. However, the "best practices" for journalists on this new beat are still evolving.

This paper explores the work of a unique group of journalists who are trying to investigate and report on the data-intensive and geopolitically-loaded beats of misinformation and disinformation.

We aim to enhance our understanding of their evolving practices, the tools they choose to use, and the collaborations that they seek, with the goal of providing insights into how to support these specialized journalists as they become more entangled in these sociotechnical systems.

3 METHODS

Adopting an interpretive and inductive approach, informed by Charmaz and Strauss[30, 51], we conducted and analyzed twelve semi-structured interviews with journalists from large newspapers focusing on national and international news.

3.1 Semi-Structured Qualitative Interviews

The interviews were designed to gather expert insights on current practices of news reporting using online social data and the difficulty of covering topics like misinformation and disinformation. The interview protocol and practices were reviewed and approved by the University of Washington Institutional Review Board.

We interviewed twelve participants from news media organizations in the United States, Canada, United Kingdom, and Australia. Most were journalists working on the misinformation, disinformation, or conspiracy theory beat for their respective organization. Two worked on peripheral beats but had significant experience using social media data for their journalistic work. Their titles ranged from investigative journalist, to social media reporter, to director for newsroom engineering.

All of the journalists were identified based on recently published work on misinformation, disinformation, or other problematic information online. The journalists we interviewed were all employed by well-known news organizations — including local outlets from large metropolitan cities as well as national and international outlets. To gather as diverse a sample as possible, we initially aimed to have only one participant from any given organization. However, through internal recommendations (an interviewee telling us we also needed to talk to another individual in their organization), we had two participants each from the same two parent organizations, but these individuals worked in different areas of reporting. Three participants were recruited based on having a prior relationship — via social media interactions or as a source for a story — with the emCOMP Lab. To recruit the other participants, we sent interview request emails to twenty journalists who fit the target description — i.e. having authored an article about online misinformation or disinformation. From those requests, we gathered nine additional participants.

We have anonymized those participants in this paper, but we include Table 1 with general information about each one.

Due to the niche nature of this emerging specialization, the potential pool of interviewees was small. This limited the number of participants we were able to recruit. Though additional interviewees may have led to additional insights, the twelve interviews we conducted allowed us to achieve sufficient data saturation [55, 95] for the themes presented here. However, there are likely additional themes that are relevant to this conversation, but which will have to be left for future work.

Nine interviews were conducted via video conferencing and three via phone. Interviews lasted an average of one hour. In most of the interviews, there was one primary interviewer and a note taker. With the permission of the participants, the interviews were recorded.

The interview structure was designed so that we could first get a sense of the role and organization of the journalist. We then asked questions about their process of writing stories using online data, encouraging them to take on a storytelling format and take us through a particular case study in their work. Hearing the process of investigating and writing about online activity afforded us the opportunity to ask follow-up questions about their workflows, tools, pain points, and wishes for

Participant	Role/Title	Type of Organization
P1	Foreign News Editor	Australian National News
P2	U.S. City Bureau Chief	U.S. Popular Culture Magazine
P3	National Security Reporter	Large U.S. City News 1
P4	Reporter	U.S. Internet Media Company 1
P5	Data Visualization Editor	Large U.S. City News 2
P6	Writer	Large U.S. City News 2
P7	Reporter	U.S. Television News
P8	Reporter	U.S. Internet Media Company 2
P9	Reporter	Fact Checking News
P10	Reporter	Canadian National News
P11	Reporter	U.K. National News
P12	Engineering Director	Large U.S. City News 1

Table 1. Participant details

the future. We ended the interviews asking specifically about their perception as a journalist of coping with media manipulation, as well as any safety concerns they had in doing their work.

3.2 Analysis

We conducted a grounded, inductive analysis of the interviews. This approach allowed for insights from the interviews to emerge from the data. Initially, a small team of researchers participated in a collaborative clustering activity. The interviews were transcribed and atomized into individual cards. Over the course of multiple sessions, co-located researchers clustered related comments from the different interviews, identified salient themes from the clusters themselves as well as from connections between clusters, and recorded insights (how themes fit together) that emerged from the analysis. After the initial clustering activity, a subset of researchers returned to the original interviews, continuing to refine and distill the emergent themes. These themes and analysis will be more broadly discussed in the findings section.

4 FINDINGS

Through analysis of the interviews a number of themes emerged. We focus on four main insights. First (1), we discuss the unique challenges to reporting on problematic information online, and to social data more broadly, to highlight the difficulty of reporting on these beats in the evolving journalism environment. Next, we explain how to overcome these challenges, reporters both (2) rely on their own subject matter expertise and close interaction with the data; and (3) reach out to internal and external experts to empower their work. Finally (4), we describe participants' perceptions of what solutions would help them do their job more reliably and efficiently.

4.1 Reporting on problematic information online presents unique challenges in journalism

The participants in this study face unique challenges in their work due to the online context. When reporting on vast and rapidly changing online environments, nine of the participants highlighted their struggle to keep up. And these challenges become even more acute when reporting on misinformation, disinformation, and other related phenomena where perceived "bad actors" are actively manipulating the systems.

> I feel like.. I'm coming with a knife to a gunfight in terms of reporting on bad actors who are trying to spread misinformation and disinformation on social media. I'm coming to an increasing realization of the sophisticated tools and notions of social engineering that these seemingly bad actors have, and I have excel spreadsheets and TweetDeck. - P2

P2 feels overwhelmed and outmatched — that they are ill-equipped in the effort to report on problematic information on social platforms. They feel that the techniques of bad actors are becoming more sophisticated, making their disinformation campaigns harder to unpack. We found that these challenges are exacerbated by a lack of communication with technology companies, tools that do not always meet their needs, and ethical considerations around publishing on these topics, which we discuss here.

4.1.1 Constantly changing environments and structure means that a method that works one day could be gone the next. Though change is not new to journalists, the rapid evolution of online environments and social media platforms challenges reporters to constantly update tools, investigative strategies, and navigation practices across numerous platforms. With the common software development practice of continuous delivery [8] continuously impacting the affordances of platform interfaces, a strategy that participants use to monitor information one day could be unusable the next.

Challenges with changing interfaces extend beyond the platforms themselves to the digital tools that journalists use to analyze online information. One participant pointed out that digital analysis tools — e.g. ones that show trending topics on social media platforms — are often run by third parties. It can be difficult to find these and learn how to use them, but even more vexing when those tools change or disappear underneath them, disrupting their work practices. Additionally, the tools can become obsolete due to a lack of reliable maintenance as platform outputs change or disappear.

> [Access to tools] always changes because the platforms have made it harder and harder for journalists to investigate them, especially in a network capacity. So, for example, we've lost our API access to Instagram that has prohibited research. [...] Because old tools become obsolete really quickly, we've got to find new ones. - P4

Because of this, journalists must continuously update their approaches and tactics. Unexpected changes to third-party tools can profoundly impact ongoing investigations and instantaneously change what is possible in performing this kind of work.

4.1.2 Lack of action and cooperation with data owners hinders journalism practices. This difficulty of working in a continuously updating environment is exacerbated by a lack of cooperation and communication with data owners: the social media companies. Nearly every participant — without prompting— discussed their attitudes about their relationships with social media platforms. Six participants spent a significant amount of interview time talking about this. They feel that they, as reporters, are being increasingly denied access to information. One participant mentioned that it is especially frustrating knowing that in some cases platforms are giving data access to researchers but not to journalists.

> It hurts the journalism field that they're locking us out of their data. [...] I can't stress enough how damaging it is for our reporting. [...] the things that we'd been able to do two years ago, we just cannot fathom doing anymore and that's been really discouraging—and it's not just Facebook. - P4

> [Mentioning our role in academia:] At least y'all get access to the firehose. At least you can conceivably do [investigation] if the need was high. I basically sat in on a conversation at [previous employer] where we realized that even if we wanted to do this in a real way, we couldn't do this. I can't do it. - P12

These quotes demonstrate increasing frustration about limited — and diminishing — access to data from the social media platforms and the real impact that this was having on what kinds of stories the journalists could cover. Our participants repeatedly stressed a desire for more cooperation from the platforms. Because there is no legal requirement for platforms to share data, the reporters feel that they are at the mercy of large social media companies, who have seemingly few motivations to share. From the perspective of the journalists, being denied access to data hinders their ability to do work even when the goal is not to criticize the company who holds the information. However, our interviews suggested that the journalists perceived a significant reason for why they were being denied access was because these companies feared public criticism of their platforms.

> There have been instances in the past where I have approached some of the big social media companies — not even to confront them with any story that would embarrass them or expose any wrongdoing on their behalf — for stories where I just had a tip-off and I wanted to see whether they knew anything about it. I have always found closed doors. There is always the assumption that you are out to get them. - P11

> ...we're just at the mercy of tech companies right now and that is something that frustrates me endlessly... having that lack of visibility. And there's not really a lot we can do about it because we report so aggressively on those companies. [...] I would say this — there is a relatively poisonous incentive there on their behalf — which is the self-preservation in saying that they lock down access, and then the result is that we can't do a broad base of reporting, even though they are only really upset about a small number of things. - P12

These quotes highlight an obvious tension: the news organizations that the participants work for are aggressively reporting on the same social media companies from which the journalists want more data.

Participants also feel that, currently, the social media companies are not adequately identifying and handling problematic information on their platforms, increasing the need for journalists to step in and report. P6, in describing their goals around platform accountability, says "*[the] misinformation ecosystem is very nuanced and the content moderation policies of the social networks are incredibly blunt tools that aren't keeping up.*" This perceived gap motivated several of the journalists we interviewed to continue to do work on this beat.

However, one important area that they continue to struggle with — in part due to lack of communication with platforms — is understanding the magnitude or impact of certain information operations online. This issue is not unique to journalists; researchers are also struggling to measure the impacts of problematic information [85]. Barriers to understanding the reach or impact of problematic information online make it difficult for journalists to make decisions about when to publish or stand down.

4.1.3 Lack of reliable reach data complicates decisions on what — and what not to — report. The issue of media manipulation [67] — i.e. efforts to exploit the media to produce or amplify problematic information — has caused journalists covering this beat to think twice about when and what they publish. Some researchers recommend that journalists sometimes consider not publishing at all. Research and media advocacy organizations [5, 74] are working to produce and update recommendations for professionals working on this beat. P9 discussed how they are trying to encourage their own organization and others to take these publishing considerations seriously.

> I'm super concerned about [media manipulation]. I'm trying to change the way that [my organization] frames things or pursues story ideas. I've also told other news organizations, 'Hey, maybe don't cover Q-Anon so fucking much.' If something doesn't have hundreds of shares — but even more than 700, 800 shares or 1,000 shares — then I'm probably not even going to [publish] it. - P9

This quote reveals some of P9's process for deciding whether or not to cover a controversial topic — weighing the risk of amplifying a marginal and problematic narrative into the mainstream against the danger of ignoring it. In general, the participants we spoke to felt more comfortable publishing a story when they knew that the magnitude and reach of the problematic information was already large, although this standard did vary by organization. In this way, they felt that their role was to share what was happening online alongside doing the work of debunking false or misleading narratives.

However, without a clear understanding of the magnitude and reach of the problematic information, the journalists seemed unsure if they were reporting on a fringe topic that would serve to amplify the intent of malicious actors. P8 discussed how perceived reach can be manipulated and misleading to journalists. They described a case where they were about to publish a story on individuals who appeared to have a massive following on Twitter, but stopped when those accounts were suddenly banned for buying fake followers. They have noticed magnitude measurement issues elsewhere, saying "*In the same way [it happens] on 4chan. I can go in and just open eight different tabs and then interact with myself on 4chan, you know that is the issue you deal with bots of any kind.*" Without reliable information about the reach of the problematic information, decisions on when to publish this unique kind of investigative reporting can be difficult to make.

4.2 Journalists embed themselves in online environments, rely on their subject matter expertise, and use ubiquitous tools to overcome these challenges

Because the journalists working on social data are often either the only or one of very few assigned to their specific beat in their newsroom, they often work independently to locate leads that they would like to pursue. Ten of the participants talked about embedding themselves within social media environments to look for interesting stories and suspicious activities online.

> And oftentimes, I mean, really a lot of this 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. Then by the fourth time, you know, I realized that seems to be a trend. - P6

The process described by P6 is common to what we call passive monitoring, which often takes the form of simply scrolling through their personal news feeds. Similar processes of embedding themselves in a variety of environments were described by many of our participants.

> That's what I do on a daily basis — basically lurk Facebook and Instagram. - P9

> I literally just go to the websites [4Chan and 8Chan] and read what they say. Yeah it's all manual. Especially because, even our in-house data engineers are like how can we automate this. Unfortunately, there's just no way to automate it. A lot of it is just human judgement. - P4

Along with scrolling through feeds, five participants also use a more organized form of monitoring, in which they specifically search for certain topics or visit particular websites or groups where they know suspicious behavior is likely to take place. This can take the form of visiting problematic Facebook groups, hyper-partisan news sites, Discord chats, and places like 4-Chan and 8-Chan. This is what we refer to as active monitoring.

> I have a bookmark tab of thirty [problematic] websites that once a day I'll open them all up and go through the websites. - P8

Passive and active monitoring were common ways to become familiar with the platforms that they studied. However, the journalists we interviewed suggested that their ability to recognize suspicious activity (that might be worth further investigation) was not easy or innate, but something that they have learned — and are still learning — through their experience working this new beat.

4.2.1 How journalists identify suspicious activity. One reason that disinformation campaigns are effective is that it is difficult for people — including researchers, journalists, and the platforms themselves — to distinguish between authentic and inauthentic activity in online spaces. The journalists we talked to, emerging experts on this new "beat", described how they relied heavily upon their own intuition and accumulating experience to determine whether an account was who it said it was or whether a network of accounts was acting in suspiciously coordinated ways. P1 referred to their reliance on their "human faculties, intuition, expertise." Similarly, others described relying on hunches, which are a kind of information hypothesis. We repeatedly heard journalists describe how these hunches inform their reporting and direction of investigation and that they typically derived not from a specific tool, but from the journalist's embeddedness in the online environment.

> [In talking about why they monitors manually]: I feel like computational solutions ignore the fact that real people fall for it and we gotta look for why the real people fall for it. Unfortunately, computers are not capable of telling you that. - P4

> There's no one tool, there's no magic wand, it was really applying old school strategies of journalists to new media. - P7

This expertise was often operationalized through *rules of thumb* that journalists developed for identifying suspicious behavior. For example, when it comes to verifying online identities, P11 explained that a newly made profile can be an indicator of a fake account. In the case of older accounts, P7 scrolls down their timeline in order to check consistency. And when it comes to the behavior of the accounts in question, P2 uses spreadsheets to gather suspicious interactions and identify patterns.

Many of these heuristics are also used by researchers to study these online activities, prompting us to think about ways to better share these methods across individuals and professions. Because many of these rules exist in the minds of these reporters, "*there's a really high barrier to entry, especially for any older reporter or newer reporter who might not understand the complex environment*" (P4). Some places, like Bellingcat [1] and the Digital Forensics Research Lab [4] have already started sharing tactics through resources and training, and five participants spoke about looking to those groups for assistance.

4.2.2 Journalists primarily rely on ubiquitous tools to do this work. In developing and evaluating their hunches and investigating their stories, participants exhibited a wide range of data processing and analysis skills. For example, P2 and P11 mentioned projects where they spent hours manually copying and pasting content into spreadsheets to create an environment in which they could analyze the content. In contrast, P12 (who runs an engineering group that supports other journalists at their organization) had extremely advanced data collection skills like the ability to query APIs themselves. Somewhere in the middle of these two extremes were participants who did not have advanced coding skills, but had some expertise with specific tools that they used when investigating stories with online social data.

These tools varied somewhat. However, most participants shared that they primarily relied on ubiquitous tools like Google search, TweetDeck, in-platform search functionality, and reverse image search websites. The following two quotes highlight the use of spreadsheets.

> [we] literally just had a big spreadsheet and went through site by site and checked every single one and typed it in the spreadsheet if it was there. So sometimes the old fashioned way works as well. - P10

> We're not usually dealing with massive, massive, massive datasets, so honestly, they'll go with Excel, and Google Drive, and that kind of thing for making rough charts, sketching out shapes... is sufficient for a lot of this work. - P5

These quotes demonstrate reporters on this beat are not heavily relying on advanced analytical tools to do their work. P10 pointed out that they, "*get this feeling that, you know, each sort of type of group that looks at this stuff — there's academics and research, there's journalists, there's law enforcement — everyone thinks that the other group have a special tool that we don't have, and that's usually not the case. [...] a lot of the work we're all doing is the same, just reading and talking to people, chasing down leads, Googling a lot of this stuff.*" P10 acknowledges that although some professionals might think that others covering mis- and disinformation are using special tools to do their work, they are actually relying on their manual labor and ubiquitous tools to investigate these stories. P12, who designs tools for journalists in their newsroom, has recognized this and has pivoted to providing helpful features within tools that journalists already use (like Slack, email, and text messages), instead of creating new tools for their colleagues.

One exception to this is CrowdTangle, a tool acquired and now provided by Facebook that serves as a "content discovery and social monitoring platform tool for publishers and brands"[2]. The tool allows exploration of trends on multiple social media platforms, including data from Facebook that is otherwise inaccessible. Five participants specifically mentioned using CrowdTangle for tasks like monitoring crowd 'signals' and statements by world leaders. However, the platform is not available to everyone. Limited access to the tool brings up questions about who has access to data, and the inherent power relationship between providers and requesters. This idea will be explored in the discussion.

4.3 Journalists utilize experts for help gathering, analyzing, and contextualizing online information

We explored participants' collaborations with others as a method of investigating problematic information. Our research revealed that journalists are adapting to evolving data needs by making use of both external and, when available, internal (within their organization) experts for help gathering evidence for their reporting. The participants seemed eager to supplement their subject matter expertise with data by partnering with academics, industry professionals and researchers, peer journalists, government entities and even amateurs for help collecting, analyzing and contextualizing online information. In some cases, journalists and their organizations also establish formal, direct, and mutually beneficial partnerships with the platforms themselves in order to identify problematic information.

Despite the advantages of these partnerships, mismatches in goals between the journalists and experts (both internal and external) can also present challenges to these collaborations. This section of findings will discuss the nature of these collaborations and what elements, in the opinion of the participants, lead to success or fracture.

4.3.1 External resources are valuable but have shortcomings. Nine participants noted that they reach out to external experts for help — e.g. when they want to discuss online activity that they cannot understand or to solicit help with data crunching when they do not have the skills to do the work themselves. The need to form these partnerships comes from the evolving nature of the field and the unique challenges that come with it. Not surprisingly, P2 noted that this type of external collaboration has "*been happening more lately*".

> I find it very very useful to go to experts outside the organization and get their input on the story that I am working on, either in terms of techniques to crunch data or techniques to get to the bottom line of my investigation. [...] to help me verify and guarantee that I have a solid story on my hands. - P11

For P11 and other journalists, these external experts are a helpful resource for verification, techniques, and help with the technical side of reporting, such as onboarding new tools and dealing

with complex data sets.

Trust in collaborators' methods is important. We also explored who the participants chose to reach out to and found that trust in the external resource's methods was a determining factor. It is important to remember that the public nature and need to maintain their professional reputation makes these journalists sensitive to whom they choose to accept information and insight. P7 felt that some academics and other researchers working in this space have not followed strong methodologies noting that "*there has been some terrible reporting based on some really terrible research over the past 24 months*". P8 felt similarly, sharing that they think those who look at these issues from a data perspective are quick to say that far-right activity is coming from bots, but in their experience it is real people. This makes these journalists cautious to work with outside experts who do not deeply understand this context.

> You're just at the mercy, without a data journalist, to kind of go through [an outside expert's] methodology. [...] And I think that's one of the biggest pitfalls — is having on the research end, somebody you can trust with that. - P6

> And so even in the academic field in very top notch universities we've seen methodologies that would really be up for questioning, if more people were up for questioning in this field. - P4

These quotes reveal significant unease by some of the journalists toward outside methodologies, including those from published research and high-profile researchers in this space.

Challenges in academic partnerships. Through our interview protocol, we specifically discussed partnerships with academic researchers at universities. In seeing our participants' perspectives on these collaborations, academics looking to work with journalists on data intensive topics can carefully consider their methods of interacting with and empowering such professionals.

Participants who have worked with researchers at universities on investigating problematic content online described a wide variety of benefits for collaboration; such as lead generation for their beat, data collection and analysis of information trajectories online, advice on where to look for problematic pieces of information, and contextualization of problematic activity. P6 notes that their "*thinking changes as a result of speaking to people like yourself who are coming from an academic perspective on a lot of this stuff. There's just historical perspective that I don't have as much of, and I think that that's really helpful.*" Three journalists also listed a desire to have more collaboration with academic researchers in the future, something we discuss further in the final findings section.

Those who had experience working with academic researchers on these topics did highlight significant challenges, primarily around timing and motive. Though they recognized the value of in-depth research and peer review, the participants felt like the timelines of academia are generally too slow in helping to meet reporting deadlines and make a meaningful impact in exposing problematic information online. P12 notes that journalists at their organization may be at fault here for this too, noting that journalists typically reach out to academics too late in the story building process for it to be helpful. They say, "*it is better [to get them] involved in those stories early.*"

Motive differences also appeared in some interviews as an issue when working with academics. P6 spoke about a desire to work more with academia but suggested that it can be difficult when academics and journalists have different priorities. In their opinion, while academics may want to watch a bad actor or suspicious narrative online to better understand its tactics, journalists may want to publish information quicker to put public pressure on platforms to make a change. In addition, journalists may not be aware of the kinds of privacy concerns that limit how academics can share data about individuals. These tensions suggest challenges for future collaborations between journalists and academic researchers.

4.3.2 Internal resources empower data journalism but are limited. Because the participants in this study worked at large reporting organizations with national and international audiences, some had access to internal resources to help them gather and analyze data. However, it is important to stress that not all organizations have these kinds of resources. P3, who came from an organization with many resources, noted that it "*is unusual in the resources that [P3's Organization] devotes to research staff within the newsroom. So, I'm surrounded — and sit next to — a full-time dedicated researcher who exists there to do deep research work on the reporting targets that we generate*". This impacts the generalizability of this piece of the findings, because journalists from smaller organizations who may be working on this same beat are likely to not have the same kinds of resources. However, our research provides insight into the kinds of resources that may be needed to do this kind of journalism well.

Reporters working at organizations that have internal resources like data journalists and data engineers tend to turn to them for data collection. Instances included scraping data from pdfs online (P5), collecting tweets of politicians (P12), automating online processes like clicking on links to look for dark patterns (P6), and writing short programs to automate looking at Facebook ads spending of politicians (P7).

Beyond providing technical assistance, these research staff members can also provide important, immediate feedback to journalists on what is possible to do with data, which can empower journalists and greatly impact the direction of their stories:

> In talking to our data engineer, I'll ask if it is possible to pull or see what time everybody tweeted this hashtag just by pulling from the API or something like that, he can tell me, because I don't know what's available in there, and how easy or hard it is to process that data. - P10

> We have in-house data engineers and data reporters who we can sort of pick up the phone and call and say, "Hey, we have this idea... is it doable or not?" And then we team up across newsroom to do this. [...] My coding skills are very barebones, but for example my [colleagues] can help us do this type of work. They understand the environment a lot better. - P4

Not all internal data teams serve the needs of the participants we spoke to. Some teams are more experienced with traditional data sets, such as census and financial disclosures, rather than social media data (P7, P12). This means that they may not be skilled at collecting and interpreting the social media data needed for investigations into problematic information online. For example, P7 notes how nuanced analyzing social media data can be by saying, "*We have a data team that is very good on sort of the more traditional data sets [...] But we don't really have a team in place that can creatively query social APIs and then also run analysis on those.*"

There are clear benefits to having access to internal and external resources so long as they are trustworthy, accessible, and can move on a deadline favorable to journalists. The interview discussions about tools, collaborations, and training naturally led to conversations around what journalists want, but do not yet have. The final section of the findings outlines potential solutions that the participants think would positively impact their ability to report on online social data.

4.4 Journalists desire better tools, collaborations, and training to enhance their journalistic skills and avoid media manipulation

Across all twelve interviews, we repeatedly heard about the increasing need for participants to access and analyze social data for their work. The journalists without strong data science skill sets, internal support, and effective tools for analyzing those data felt disadvantaged in achieving their reporting goals. This section highlights specific wants and needs as presented by the participants.

4.4.1 Data Journalism skills could bring independence. The desire for self improvement around data collection and data analysis repeatedly came up in the interviews.

> Every time I have worked with people who knew how to use those tools, I felt very jealous because I wish I could also map those networks of disinformation, misinformation, and to really find a way to easily understand some stories circulate online and where they come from, and where they go and who are the people pushing them out. - P11

The data collection desires tended to focus on the need to create or find the right data set to use for their investigation. P9 discussed a case in which they attempted to fact check a spreading claim online about crime rates for Black Americans but struggled to even find a dataset to recreate the claim, making it difficult to know if the information was problematic. They discussed their desire for more data journalism skills because it would allow them to "*quickly and efficiently locat[e] data that [they] need for a story.*" P10 also discussed the desire to be able to scrape data themselves to create datasets to look at. Others mentioned that they wanted to be able to interact with social media APIs themselves.

We also saw a desire for data analysis skills, which would allow the participants to look for interesting trends in existing data sets or on websites in an automated way. In some cases, a lack of these analysis skills and experiences made it hard for them to understand the value in published work. P7 said, "*We see a lot of graphs where data scientists will group accounts together in networks. [...] It can be very difficult to see what is going on, or in my case, to build a narrative to turn into a story. But that might just be something I need to work more on.*"

In the same vein, P12, who had experience building tools for journalists in their newsroom, discussed the importance of contextualizing leads and suspicious activities for journalists. It is not enough to simply provide a set of anomalous activity. There needs to be contextualization and analysis to understand why specific activity may be an anomaly and what they would expect to see in a normal scenario. They explained that the data-plus-context approach has been successful in their own newsroom and has the potential to lower the barriers of entry to studying this type of information, allowing new journalists to investigate these stories and learn analysis skills through doing.

4.4.2 Training on media manipulation can benefit journalism industry. The journalists we talked to expressed concern about media manipulation in their field writ large. They felt that they, as experts working on this beat, are more aware — compared to those not working on this beat or working for less resourced news organizations — of the risks and ethical considerations associated with reporting on problematic information.

> We know that one of the biggest sharers [of misinformation] is local news and radio stations, mostly radio actually. And we know from a recent study that only 18 percent of newsrooms have been trained for looking for media manipulation which is a pretty low number considering the shitstorm that we're in right now . - P4

Because of these concerns, they expressed a want for more training around mis- and disinformation, as well as general online activity, for their peers and for other newsrooms around the nation. In talking about a local newsroom who shared an inauthentic picture, P11 felt like the local journalist could have "*debunk[ed information] very simply, by just doing a bit of Googling, and having a little bit of a critical spirit.*" This suggests that journalists on this beat desire for their peers to not only be trained on tools but also the methods and mindset of exploring online information with a critical eye.

5 DISCUSSION

In this paper, we have explored how journalists whose work focuses on online mis- and disinformation investigate and report on these challenging topics. We described how their efforts are enabled and shaped by skills, computational tools, intuitions, and collaborations. We found that the

constantly changing nature of social data and the lack of cooperation with social media companies are unique challenges for journalists trying to make sense of this space. We also learned that journalists primarily rely on their own expertise and ubiquitous tools (such as spreadsheets and platform search capabilities) to identify and investigate suspicious online behavior. Increasingly, they are reaching out to internal (within their organization) and external experts for help with data collection and analysis. Lastly, they shared a desire to improve their own data skills and provide more training to journalists writ large to prevent media manipulation.

Through this work, we have come to see important elements to effectively supporting the practices of this type of journalist. In the process of investigating how reporters on the problematic content beat do what they do, we uncovered more about the data ecosystem in which they work, including their professional values, the challenging power dynamics with platforms, and their desires for the future. The following sections explore these elements in more depth.

5.1 Supporting Journalistic Values with Computational Tools

Researchers often think about supporting journalists with data-intensive beats by giving them advanced analytical tools (e.g. [40, 68, 76]). Our interviews reinforce that effectively supporting these professionals requires aligning with their current practices while being sensitive to their values. Our findings highlight values specific to journalists investigating and reporting on misinformation, disinformation and other kinds of problematic content online. And we suggest that HCI and CSCW designers be sensitive to these values — for example, by following processes suggested by Friedman and colleagues [47] — as we design tools for these unique "expert" users.

The journalists we interviewed revealed that proximity to social data — and embeddedness within the online communities that create that data — is critical to identifying problematic information and suspicious activity. Bad actors often portray real people, and they can evade being noticed in aggregation tools. Journalists tend to rely upon their own expertise and manual labor, closely reading and following the activity in these groups, to detect irregularity and identify and understand problematic information. In practice, this kind of work involved scrolling through Twitter feeds, becoming part of Discord chats, routinely visiting problematic sites, and even manually investigating suspicious Facebook accounts. Tools built to serve those who investigate this type of online activity should enhance journalistic inquiry methods, such as being close to the content, instead of solely providing aggregate numerical data and visualizations. These suggestions align with work by Rotman et al. [79] who encouraged researchers of online environments to go beyond high level views of the data, which were inherently limited. When thinking about supporting journalists, we likely need to be designing to support what Rotman and colleagues termed "extreme ethnography."

Our findings also suggest that effective tools need to understand and account for journalistic values and practices in order for journalists to trust them and rely on them for reporting. One important dimension of this centers around verifying sources. Traditional investigative journalists question, verify, and cross-examine human sources themselves, using well-established processes [36]. These practices are disrupted when journalists shift from using people as their sources to using social media data as their source. Instead of relying on traditional practices of verification, journalists looking at online data have to develop means of establishing the credibility of their data — how it was collected, processed, and analyzed — before they can be comfortable using it in reporting. But vetting social media data, which is known to be fallible [24], as a source can be difficult, especially when the data was collected by others or through systems that are not fully transparent. Doing this verification also requires a level of methodological expertise. Software tools and other solutions designed for journalists investigating online data should support their efforts to establish the credibility of the source data, for example by enabling them to explore the data in ways that reveal the data's contours and by making data collection and data transformations

transparent. Though all journalists data intensive beats need to trust their data, this is especially important in the context of reporting on disinformation and other types of problematic information, because a misstep or misinterpretation can be seized upon by the groups they are studying and exploited to undermine their work [67].

Another set of values to consider are magnitude and relevance, noted by Harcup and O'Neal [56] as guiding journalists' decisions about what is newsworthy. These are especially relevant in the context of disinformation reporting, because they can be key to deciding what to investigate and when to publish. To guide those decisions, researchers and media advocacy groups have produced recommendations for journalists on this beat [13, 74]. For example, Phillips suggested that, to avoid becoming unwitting agents in disinformation campaigns and doing the bad actors' work of amplification for them, journalists should publish stories about problematic information only when it has already reached widespread audiences [74]. However, our interviews revealed that journalists struggle to assess the magnitude or reach of an existing piece of problematic information, and consequently to make decisions about when to publish — and when not to.

5.2 Platforms and Power Dynamics

Another challenge for journalists on this beat was navigating relationships with the social media platforms that they often needed to rely upon for data. These relationships were inherently imbalanced, with the social media platforms having considerable power over who gets to report on — and perhaps what gets reported from — their data. Gillespie [48, 49] suggests that even the use of the word "platform" is political. The word is used to blend the line between commercially created and user-generated content and this blending is one factor in why problematic content thrives in online spaces. For example, a disinformation campaign designed to garner support for alternative medical treatments may be commercial in nature, but they can claim to be user-generated or "free speech" when their content is threatened or taken down. Because these platforms are the "keepers of cultural discussion" [48], yet provide little transparency to those on the outside, they have become a powerful gatekeeper. And journalists currently have to work with them, or within their terms of service, to do their reporting.

The journalists we spoke to noted that it was challenging to get representatives of the companies to speak to them about content or activity taking place on the platforms — e.g. a disinformation campaign or harassment incident. They reflected upon a somewhat antagonistic relationship between their broader news organizations and the platforms, where as they are trying to gain information about a particular story of problematic information, a colleague might be developing a story critiquing the platforms about privacy infringement or content censorship. Many of our interviewees communicated that they recognized the platforms were navigating challenging terrain. Yet, they still felt frustrated that there was no clear way to move beyond the gate and have their questions answered, even when their goal was not to specifically criticize the company they were reaching out to. The participants feel that there is a critical societal need to understand, analyze, and share what is happening on social media platforms, especially as they increasingly serve as a centerpiece of culture, but that much of the information and the power to do that lie behind the platform walls.

In some cases, platforms are starting to slightly open their gates by providing certain individuals — including journalists and academic researchers — with data. For example, Twitter has created new opportunities to access certain data, like a massive sample of tweets related to Covid-19, through an application process [88]. Similarly, in partnership with the Social Science One initiation [16], Facebook has made data available — after a rough and at times contentious process [73]— to a select group of researchers through data grants awarded through a competitive grant process. Facebook has also opened up tools like CrowdTangle for use by researchers and journalists. CrowdTangle was

used by a number of participants in our study, and follow-up conversations with some interviewees reveal that they are relying more and more on this platform for their reporting.

However, in each of these cases, individual platforms are opening up the gate to specific people, not all people, based on personal connections or application processes. In our own experience, the platforms' rationale for this selectivity is to ensure that the data are being analyzed in methodologically sound ways, by experts who understand the contextual and ethical complexities. Their stated aim is to limit misreporting which could contribute to the misinformation problem. However, these selective access programs highlight the persistent position of power that social media companies maintain. Just as boyd and Crawford [28] wrote about a data divide for academics, we see this extending to journalists who are trying to investigate problematic content online. Those who have the relationships are able to investigate (with the permission of the platforms) and those who do not are left out. This power imbalance can be coercive, as journalists with data privileges may feel compelled — in order to retain their access — to portray the platforms themselves in a more positive light.

5.3 Future Work: Collaborations between Researchers and Journalists

Practices of understanding and reporting on online misinformation, disinformation, and other forms of problematic information continue to evolve. Researchers and journalists have acknowledged how complex this work is and are starting to work together to understand and combat the spread of this type of content. These collaborations include education (e.g. [7, 15]), skillsharing (e.g. [6]), and collaborative investigation (e.g. [19, 81]). We now discuss the benefits and challenges of this type of collaboration, and suggest future work.

When the journalists we spoke to first started covering this beat, they primarily had to learn how to navigate and investigate online spaces themselves. Recently, training provided by organizations and researchers on the topics of content verification [10, 11] and data analysis [3, 9] are available to them. The growing availability of these types of training for journalists mirrors what we found; that these journalists are eager to learn more about how to investigate problematic information online themselves. They also noted that taking time away from their work for these training sessions is difficult and suggested that they have filled their own knowledge gaps by collaborating with data experts inside their organizations and researchers studying the same topics.

Participants mentioned that they are increasingly reaching out to researchers, primarily for help collecting and analyzing online activity data. By working together, journalists can receive meaningful data to inform their work and researchers can help combat problematic information on a timeline that is faster than the typical peer review process. We can currently see this happening in coverage of mis- and disinformation related to Covid-19 [19, 81].

Though these informal collaborations can be mutually beneficial, our findings suggest that mismatches in methods, timelines, and policies can strain these partnerships. Journalists working on this beat described needing more transparency from their collaborators, especially when it comes to methods. They do not see all researchers as reliable on this topic, so they want to understand researchers' processes before attaching their reputation to the findings. Researchers should enable journalists to be closer to the instruments and the data so they can not only use the findings of their investigations, but also to improve their own skills by learning to do the collection and analysis themselves. The journalists we spoke to want to become the analysts and be able to more independently explore and report on this beat. Researchers can help them with this in the short term, both through collaborations and training them to do this work on their own.

There is also opportunity to study these collaborations as they take shape and evolve. Future work should explore these collaborations in action and ask critical questions about the nature, structure, and particularly the ethical challenges they entail. The CSCW community is well-positioned to do

this kind of research, and this future work could further inform how we can design to support such collaborations — addressing journalists' current challenges and navigating inherent infrastructural and organizational tensions.

6 CONCLUSION

Online platforms have become increasingly salient in breaking news events, culture, and politics. Consequently, journalists are looking to these spaces for stories and investigations. This research provides a window into the practices of journalists who are investigating their stories on and through social media. In particular, we focus on journalists working an emergent new beat of online mis- and disinformation. We find that journalists overcome the unique technical, methodological, ethical challenges of this work by embedding themselves in online communities, working with experts, and improving their own data skills. Situating the findings within literature about changing journalism practices, computational tools designed to support modern journalists, and the complexities of online disinformation, we contribute insight into how to better support these kinds of journalists in the future. We suggest that designers of data tools for journalists pay special attention to the traditional values held by those in the journalism field. We discuss the power dynamics that exist between journalists and platforms, and we suggest that future work involves studying the informal collaborations between journalists and researchers that are already taking place. We foresee this "beat" will grow to include more journalists, including those looking specifically at online mis- and disinformation as well as adjacent topics such as bullying, harassment, and radicalization, and we see supporting these emerging experts as an opportunity for future CSCW research.

ACKNOWLEDGMENTS

This research was done at the University of Washington emCOMP lab and Center for an Informed Public with support from NSF Grant 1749815. Thank you to the journalists who spoke with us for this work. Lastly, thank you to the University of Washington students who provided transcription and analysis help as part of a Directed Research Group on the topic.

REFERENCES

[1] [n.d.]. Bellingcat. https://www.bellingcat.com/
[2] [n.d.]. CrowdTangle. https://www.crowdtangle.com/
[3] [n.d.]. Data Journalism Bootcamp. https://www.ire.org/events-and-training/boot-camps
[4] [n.d.]. Digital Forensic Research Lab. https://www.atlanticcouncil.org/programs/digital-forensic-research-lab/
[5] [n.d.]. First Draft. https://firstdraftnews.org/
[6] [n.d.]. The International Fact-Checking Network. https://www.poynter.org/ifcn/
[7] [n.d.]. MediaWell: teaching and learning. https://mediawell.ssrc.org/teaching-and-learning/
[8] [n.d.]. Rapid release at massive scale. https://engineering.fb.com/web/rapid-release-at-massive-scale/
[9] [n.d.]. Training: Data Journalism. https://newsinitiative.withgoogle.com/training/course/data-journalism
[10] [n.d.]. Training: Verification Course. https://newsinitiative.withgoogle.com/training/course/verification
[11] [n.d.]. Verification Training. https://firstdraftnews.org/training/verification/
[12] 2019. Newspapers Fact Sheet. https://www.journalism.org/fact-sheet/newspapers/
[13] 2020. Coronavirus: Responsible reporting and ethics. https://firstdraftnews.org/long-form-article/coronavirus-resources-for-reporters/
[14] 2020. Novel Coronavirus(2019-nCoV): Situation Report - 13. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf?sfvrsn=195f4010_6
[15] 2020. Reporter Resource Hub. https://firstdraftnews.org/training/
[16] 2020. Social Science One: Our Facebook Partnership. https://socialscience.one/our-facebook-partnership
[17] Penelope Muse Abernathy. 2018. *The expanding news desert.* Center for Innovation and Sustainability in Local Media, School of Media and ....
[18] Elena Agapie, Jaime Teevan, and Andrés Monroy-Hernández. 2015. Crowdsourcing in the field: A case study using local crowds for event reporting. In *Third AAAI Conference on Human Computation and Crowdsourcing.*

[19] Davey Alba. 2020. Virus Conspiracists Elevate a New Champion. https://www.nytimes.com/2020/05/09/technology/plandemic-judy-mikovitz-coronavirus-disinformation.html
[20] Mike Ananny and Leila Bighash. 2016. Why drop a paywall? Mapping industry accounts of online news decommodification. *International Journal of Communication* 10 (2016), 22.
[21] 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.
[22] Ahmer Arif, Leo Graiden Stewart, and Kate Starbird. 2018. Acting the part: Examining information operations within# BlackLivesMatter discourse. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–27.
[23] Aleszu Bajak, John Wihbey, Gibson Free, and Paxten Merten. 2019. Newsroom Textual Analysis and Visualization Tools Built With R Shiny. In *Proceedings of the Computation + Journalism Conference, University of Miami, FL*.
[24] Nancy K Baym. 2013. Data not seen: The uses and shortcomings of social media metrics. *First Monday* 18, 10 (2013).
[25] Susan Benkelman. 2019. Getting it Right: Strategies for truth-telling in a time of misinformation and polarization. https://www.americanpressinstitute.org/publications/reports/white-papers/truth-telling-in-a-time-of-misinformation-and-polarization/
[26] Ladislav Bittman. 1985. *The KGB and Soviet disinformation: an insider's view*. Washington: Pergamon-Brassey's.
[27] danah boyd. 2018. The Messy Fourth Estate. https://gen.medium.com/the-messy-fourth-estate-a42c1586b657
[28] Danah Boyd and Kate Crawford. 2011. Six provocations for big data. In *A decade in internet time: Symposium on the dynamics of the internet and society*.
[29] Bert Jan Brands, Todd Graham, and Marcel Broersma. 2018. Social media sourcing practices: How Dutch newspapers use tweets in political news coverage. In *Managing democracy in the digital age*. Springer, 159–178.
[30] Kathy Charmaz, Liska Belgrave, et al. 2012. Qualitative interviewing and grounded theory analysis. *The SAGE handbook of interview research: The complexity of the craft* 2 (2012), 347–365.
[31] Euirim Choi, Gillian Brassil, Katie Keller, Jessica Ouyang, and Kate Wang. 2020. Bankruptcy Map: A System for Searching and Analyzing US Bankruptcy Cases at Scale. (2020).
[32] Mark Coddington. 2020. Gathering evidence of evidence: News aggregation as an epistemological practice. *Journalism* 21, 3 (2020), 365–380.
[33] Sarah Cohen, James T Hamilton, and Fred Turner. 2011. Computational journalism. *Commun. ACM* 54, 10 (2011), 66–71.
[34] Nick Corasaniti. 2020. How a Misleading Biden Video Spread. https://www.nytimes.com/2020/01/07/us/politics/biden-video-disinformation-spread.html
[35] Dharma Dailey and Kate Starbird. 2014. Journalists as crowdsourcerers: Responding to crisis by reporting with a crowd. *Computer Supported Cooperative Work (CSCW)* 23, 4-6 (2014), 445–481.
[36] Walter Dean. [n.d.]. Journalism as a discipline of verification. https://www.americanpressinstitute.org/journalism-essentials/verification-accuracy/journalism-discipline-verification/
[37] Mark Deuze and Jo Bardoel. 2001. Network journalism: Converging competences of media professionals and professionalism. (2001).
[38] Caitlin Dewey. 2015. 48 hours inside the Internet's 'most toxic' community. https://www.washingtonpost.com/news/the-intersect/wp/2015/03/26/48-hours-inside-the-internets-most-toxic-community/
[39] Nicholas Diakopoulos. 2016. Computational journalism and the emergence of news platforms. *The Routledge Companion to Digital Journalism Studies. London: Routledge* (2016).
[40] Nicholas Diakopoulos, Munmun De Choudhury, and Mor Naaman. 2012. Finding and assessing social media information sources in the context of journalism. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2451–2460.
[41] Nicholas Diakopoulos, Mor Naaman, and Funda Kivran-Swaine. 2010. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *2010 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, 115–122.
[42] Joan Donovan and Brian Friedberg. 2019. Source hacking: Media manipulation in practice. *Retrieved from Data&Society website: https://datasociety. net/output/source-hacking-media-manipulation-in-practice* (2019).
[43] J. Doubek. 2017. Conspiracy Theorist Alex Jones Apologizes For Promoting "Pizzagate.". https://www.npr.org/sections/thetwo-way/2017/03/26/521545788/conspiracy-theorist-alex-jones-apologizes-for-promoting-pizzagate
[44] Henrik Fagrell, Kerstin Forsberg, Erik Johannesson, and Fredrik Ljungberg. 2000. NewsMate: Providing mobile and distributed news journalists with timely information. In *CHI'00 Extended Abstracts on Human Factors in Computing Systems*. 121–122.
[45] Henry Farrell and Bruce Schneier. 2018. Common-knowledge attacks on democracy. *Berkman Klein Center Research Publication* 2018-7 (2018).
[46] Sheera Frenkel. 2018. Facebook Tackles Rising Threat: Americans Aping Russian Schemes to Deceive. https://www.nytimes.com/2018/10/11/technology/fake-news-online-disinformation.html
[47] Batya Friedman. 1996. Value-sensitive design. *interactions* 3, 6 (1996), 16–23.

[48] Tarleton Gillespie. 2010. The politics of 'platforms'. *New Media & Society* 12, 3 (2010), 347–364. https://doi.org/10.1177/1461444809342738
[49] Tarleton Gillespie. 2018. Platforms are not intermediaries. *Georgetown Law Technology Review* 2, 2 (2018), 198–216.
[50] Dan Gillmor. 2004. We the media: The rise of citizen journalists. *National Civic Review* 93, 3 (2004), 58–63.
[51] Barney Glaser. 1967. 8: Strauss, A.(1967). The discovery of grounded theory. *Alsine de Gruyter, New york* (1967).
[52] Lucas Graves. 2016. *Deciding what's true: The rise of political fact-checking in American journalism.* Columbia University Press.
[53] Elizabeth Grieco. 2020. U.S. newspapers have shed half of their newsroom employees since 2008. https://www.pewresearch.org/fact-tank/2020/04/20/u-s-newsroom-employment-has-dropped-by-a-quarter-since-2008/
[54] Joshua Guberman, Carol Schmitz, and Libby Hemphill. 2016. Quantifying Toxicity and Verbal Violence on Twitter. (2016), 277 – 280. https://doi.org/10.1145/2818052.2869107
[55] Greg Guest, Arwen Bunce, and Laura Johnson. 2006. How many interviews are enough? An experiment with data saturation and variability. *Field methods* 18, 1 (2006), 59–82.
[56] Tony Harcup and Deirdre O'neill. 2001. What is news? Galtung and Ruge revisited. *Journalism studies* 2, 2 (2001), 261–280.
[57] Ariel Hasell and Brian E Weeks. 2016. Partisan provocation: The role of partisan news use and emotional responses in political information sharing in social media. *Human Communication Research* 42, 4 (2016), 641–661.
[58] Alfred Hermida. 2010. Twittering the news: The emergence of ambient journalism. *Journalism practice* 4, 3 (2010), 297–308.
[59] Caroline Jack. 2017. Lexicon of lies: Terms for problematic information. *Data & Society* 3 (2017).
[60] Shan Jiang and Christo Wilson. 2018. Linguistic Signals under Misinformation and Fact-Checking: Evidence from User Comments on Social Media. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 82 (Nov. 2018), 23 pages. https://doi.org/10.1145/3274351
[61] Anna Kata. 2010. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. *Vaccine* 28, 7 (2010), 1709–1716.
[62] Yubo Kou, Xinning Gui, Yunan Chen, and Kathleen Pine. 2017. Conspiracy talk on social media: collective sensemaking during a public health crisis. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–21.
[63] Rachael Krishna. 2018. TikTok Creators Say They Are Being Bullied And The Company Isn't Helping. https://www.buzzfeednews.com/article/krishrach/tik-tok-users-bullying-abuse-complaints
[64] Kenza Lamot and Steve Paulussen. 2020. Six Uses of Analytics: Digital Editors' Perceptions of Audience Analytics in the Newsroom. *Journalism Practice* 14, 3 (2020), 358–373.
[65] Jane Lytvynenko, Ryan Broderick, and Craig Silverman. 2020. These Are The Fake Experts Pushing Pseudoscience And Conspiracy Theories About The Coronavirus Pandemic. https://www.buzzfeednews.com/article/janelytvynenko/coronavirus-spin-doctors
[66] Jane Lytvynenko and Craig Silverman. 2020. We're Keeping A Running List Of Hoaxes And Misleading Posts About The Nationwide Police Brutality Protests. https://www.buzzfeednews.com/article/janelytvynenko/hoax-misleading-claims-george-floyd-protests
[67] Alice Marwick and Rebecca Lewis. 2017. Media manipulation and disinformation online. *New York: Data Society Research Institute* (2017).
[68] Hana Matatov, Adina Bechhofer, Lora Aroyo, Ofra Amir, and Mor Naaman. [n.d.]. DejaVu: A System for Journalists to Collaboratively Address Visual Misinformation. ([n. d.]).
[69] Min-Che Ho and Tsai-Yen Li. 2005. An experimental news production system with flexible role-based access control. (2005), 287–291.
[70] Amy Mitchell. 2016. *State of news media 2016.* Pew Research Center.
[71] Robert S Mueller and Man With A. Cat. 2019. *Report on the investigation into Russian interference in the 2016 presidential election.* Vol. 1. US Department of Justice Washington, DC.
[72] Merja Myllylahti. 2018. An attention economy trap? An empirical investigation into four news companies' Facebook traffic and social media revenue. *Journal of Media Business Studies* 15, 4 (2018), 237–253.
[73] Alex Pasternack. 2019. Frustrated funders exit Facebook's election transparency project. https://www.fastcompany.com/90412518/facebooks-plan-for-radical-transparency-was-too-radical%22
[74] Whitney Phillips. 2018. The oxygen of amplification. *Data & Society* (2018).
[75] Peter Pomerantsev and Michael Weiss. 2014. *The menace of unreality: How the Kremlin weaponizes information, culture and money.* Vol. 14. Institute of Modern Russia New York.
[76] Susan Reilly. 2017. The need to help journalists with data and information visualization. *IEEE computer graphics and applications* 37, 2 (2017), 8–10.
[77] Ronald E Robertson and Christo Wilson. 2020. WebSearcher: Tools for Auditing Web Search. (2020).

[78] Kevin Roose. 2018. We Asked for Examples of Election Misinformation. You Delivered. https://www.nytimes.com/2018/11/04/us/politics/election-misinformation-facebook.html
[79] Dana Rotman, Jennifer Preece, Yurong He, and Allison Druin. 2012. Extreme ethnography: challenges for research in large scale online environments. (2012), 207–214.
[80] Elise Samuels. 2020. How misinformation on WhatsApp led to a mob killing in India. https://www.washingtonpost.com/politics/2020/02/21/how-misinformation-whatsapp-led-deathly-mob-lynching-india/
[81] Elyse Samuels and Meg Kelly. 2020. How false hope spread about hydroxychloroquine to treat covid-19 — and the consequences that followed. https://www.washingtonpost.com/politics/2020/04/13/how-false-hope-spread-about-hydroxychloroquine-its-consequences/
[82] Scott Shane. 2017. The fake Americans Russia created to influence the election. *The New York Times* 7, 09 (2017).
[83] Elisa Shearer. 2018. Social media outpaces print newspapers in the U.S. as a news source. https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/
[84] Craig Silverman. 2016. Hyperpartisan Facebook Pages Are Publishing False And Misleading Information At An Alarming Rate. https://www.buzzfeednews.com/article/craigsilverman/partisan-fb-pages-analysis#.iudGgWZ1Zv
[85] Kate Starbird, Ahmer Arif, and Tom Wilson. 2019. Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–26.
[86] Kate Starbird, Ahmer Arif, and Tom Wilson. 2019. Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 127 (Nov. 2019), 26 pages. https://doi.org/10.1145/3359229
[87] M. Sullivan. 2020. Journalists can't ignore hacked data meant to disrupt elections. https://www.washingtonpost.com/lifestyle/style/journalists-cant-ignore-hacked-data-meant-to-disrupt-elections-but-heres-what-they-can-do/2019/04/26/4ff6a0fa-6785-11e9-a1b6-b29b90efa879_story.html
[88] Adam Tornes. 2020. Enabling study of the public conversation in a time of crisis. https://blog.twitter.com/developer/en_us/topics/tools/2020/covid19_public_conversation_data.html
[89] Sukrit Venkatagiri, Jacob Thebault-Spieker, Rachel Kohler, John Purviance, Rifat Sabbir Mansur, and Kurt Luther. 2019. GroundTruth: Augmenting Expert Image Geolocation with Crowdsourcing and Shared Representations. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–30.
[90] Jessica Vitak, Kalyani Chadha, Linda Steiner, and Zahra Ashktorab. 2017. Identifying Women's Experiences With and Strategies for Mitigating Negative Effects of Online Harassment. (2017), 15. https://doi.org/10.1145/2998181.2998337
[91] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151. https://doi.org/10.1126/science.aap9559 arXiv:https://science.sciencemag.org/content/359/6380/1146.full.pdf
[92] Claire Wardle. 2019. Misinformation Has Created a New World Disorder. https://www.scientificamerican.com/article/misinformation-has-created-a-new-world-disorder/
[93] Charlie Warzel. 2019. Epstein Suicide Conspiracies Show How Our Information System Is Poisoned. https://www.nytimes.com/2019/08/11/opinion/jeffrey-epstein-suicide-conspiracies.html
[94] EA Watkins and CW Anderson. [n.d.]. Managing Journalistic Innovation and Source Security in the Age of the Weaponized Internet. I A. Bygdås, S. Clegg & AL Hagen (Red.). *Media management and Digital transformation* ([n. d.]).
[95] Susan C Weller, Ben Vickers, H Russell Bernard, Alyssa M Blackburn, Stephen Borgatti, Clarence C Gravlee, and Jeffrey C Johnson. 2018. Open-ended interview questions and saturation. *PloS one* 13, 6 (2018).
[96] Brandy Zadronzny and Ben Collins. 2018. How three conspiracy theorists took 'Q' and sparked Qanon. https://www.nbcnews.com/tech/tech-news/how-three-conspiracy-theorists-took-q-sparked-qanon-n900531
[97] Brandy Zadrozny. 2019. Twitter account that amplified Covington Catholic D.C. march video appears linked to California teacher. https://www.nbcnews.com/news/us-news/twitter-account-amplified-covington-catholic-d-c-march-video-appears-n961981

Received January 2020; revised June 2020; accepted July 2020