Bridging Contextual and Methodological Gaps on the "Misinformation Beat": Insights from Journalist-Researcher Collaborations at Speed

Melinda McClure Haughey
Human Centered Design & Engineering, Center for an Informed Public, University of Washington

Martina Povolo
Journalism & Public Interest Communication, University of Washington

Kate Starbird
Human Centered Design & Engineering, Center for an Informed Public, University of Washington

ABSTRACT

As misinformation, disinformation, and conspiracy theories increase online, so does journalism coverage of these topics. This reporting is challenging, and journalists fill gaps in their expertise by utilizing external resources, including academic researchers. This paper discusses how journalists work with researchers to report on online misinformation. Through an ethnographic study of thirty collaborations, including participant-observation and interviews with journalists and researchers, we identify five types of collaborations and describe what motivates journalists to reach out to researchers — from a lack of access to data to support for understanding misinformation context. We highlight challenges within these collaborations, including misalignment in professional work practices, ethical guidelines, and reward structures. We end with a call to action for CHI researchers to attend to this intersection, develop ethical guidelines around supporting journalists with data at speed, and offer practical approaches for researchers filling a "data mediator" role between social media and journalists.

CCS CONCEPTS

• **Human-centered computing** → Collaborative and social computing; Empirical studies in collaborative and social computing.

KEYWORDS

social media, journalism, misinformation, disinformation

ACM Reference Format:
Melinda McClure Haughey, Martina Povolo, and Kate Starbird. 2022. Bridging Contextual and Methodological Gaps on the "Misinformation Beat": Insights from Journalist-Researcher Collaborations at Speed. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29–May 05, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3491102.3517503

1 INTRODUCTION

In the United States, the year 2020 saw increased levels of misinformation, disinformation, and conspiracy theories — especially online [31, 42, 65]. The Covid-19 pandemic sparked what some referred to as an "infodemic" [4, 66]; a time of widespread medical misinformation leading to adverse health impacts. With more people at home due to the pandemic, Americans spent more time on social media [69], allowing for more encounters with misinformation and disinformation. For example, following the murder of George Floyd, Black Lives Matter protests broke out across the country, spurring disinformation aimed at labeling the organization as deceptive and trying to align it with other movements that provide vectors for right-wing criticism, including Antifa [54]. Later, the U.S. election process was also plagued with disinformation surrounding mail-in-ballots and later wider voter fraud narratives, including false assertions that the voting machine company, Dominion, was changing votes — which played a role in motivating some of those who participated in the January 6th Capitol insurrection attempt [47, 55]. Other debunked conspiracy theories gained traction throughout the year as well, like the QAnon canon [67] and Wayfair conspiracy theory [58, 68].

As public interest in the topic grew, so did the news coverage from those on the so-called "misinformation beat" [60, 70] — an area of journalistic coverage (a 'beat') [37] focusing on 'fake news', misinformation, disinformation, conspiracy theories and harassment [70]. Journalists who work on the 'misinformation beat' aim to inform readers about online activity, investigate and debunk false claims, and hold social media companies accountable for problematic information on their platforms [70]. Those who were already covering the topic grew in popularity (e.g. Donie at CNN — who has been on this beat for years — went viral while reporting on misinformation at the post-election Trump rally that spiraled into the U.S. Capitol Insurrection [40, 63]), and many more, including local journalists (e.g.[43, 45]), ventured onto the beat for the first time.

The rapid increase of this beat came with more journalists facing known challenges of doing this work. Previous work has established that the misinformation beat — which relies heavily on digital trace data due to the outsized role of online platforms — is uniquely challenging due to constantly changing tools, lack of data access, and risks of providing public access to harmful content [11, 71]. Researchers noticed an opportunity to help these journalists get up to speed by bridging gaps in their contextual understanding (of mis- and disinformation) and methodological expertise (for using digital trace data as a "source"). Organizations like First Draft news, the Harvard Shorenstein Center, Poynter, and others offered training and tip sheets for journalists throughout this period, especially focusing on COVID-19 coverage, combatting vaccine misinformation, and election coverage [61, 72–75].

CHI '22, April 29–May 05, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04.
https://doi.org/10.1145/3491102.3517503

Even before this period, the co-production of journalism by journalists and researchers had been increasing, according to a scoping review by MacGregor et al. [36]. This trend continued in 2020. For reasons that are discussed in this paper, journalists new and old to the misinformation beat leaned heavily on researchers for support (e.g. [5, 8, 46, 76]. Researchers took up roles beyond those of a traditional interview source (for background information or an expert quote), increasingly taking on the labor of education, ad-hoc analysis, and helping journalists ethically approach stories. Journalists who regularly turned to misinformation researchers for a quote now turned to these researchers to *collaborate*.

University researchers, too, were eager to get their early misinformation analyses to journalists. They learned during the 2016 presidential election and other elections around the world how viral misinformation can emerge before, during, and after an election; and responding on a typical peer-review timeline can miss the mark on having a timely, positive, and broad impact on the public. A conglomerate of misinformation research teams even set up a rapid-response framework for academic researchers to quickly share analyses of misinformation with journalists leading up to and during the 2020 presidential election [77]. Starting in 2021, these informal collaborations have begun to formalize, as the Harvard Shorenstein center has hired some of the most knowledgeable journalists on this beat [78, 79] and U.S. national newsrooms have started hiring researchers and data scientists to help with misinformation related investigations[1]. All provide evidence of the current journalist-researcher codependence in tackling the challenges of misinformation journalism.

Stepping back — in 2019, we began planning journalist-researcher collaborations to learn how to better serve and support journalists covering this challenging beat. As the unprecedented activities of 2020 and early 2021 unfolded, our research team received more data requests than they could have ever imagined. Our cross-disciplinary team spanned multiple departments at the University of Washington. We both participated in and studied these collaborations simultaneously, taking an ethnographic approach. This piece outlines the findings from thirty researcher-journalist collaborations on misinformation topics and semi-structured debrief interviews with both journalists and researchers.

The findings begin with a typology of five motivations for collaboration on this beat. The frequency of these needs and findings from interviews suggest that researchers fill a necessary gap for journalists on this beat. From education to data access to 'gut-checking', researchers have become an essential part of the ability to do misinformation journalism well, especially for those new to the beat. Yet both journalists and researchers highlight ethical and professional challenges in this newfound closeness, which will be discussed in both the findings and the discussion. Finally, the process of translating journalism questions (like "what impact does misinformation have on the voters?") to ones that can begin to be answered using data inquiry was laborious, and researchers said that the education work required to get newer journalists 'up to speed' was unsustainable.

[1]This information came from the interviews conducted for this paper. Two journalists said their teams were actively hiring for these roles.

These findings help us see these collaborations with more sensitivity and examine some of the socio-technical challenges involved in this work. Our discussion: 1) highlights the value of this situated perspective on the role of academic researchers as data mediators; 2) calls for digital social data researchers to develop ethical guidelines on data sharing and privacy with journalists; and 3) provides practical guidance for researchers managing these relationships.

2 BACKGROUND AND RELATED WORK

In this literature review, we first discuss the process of specialization journalism which — with ubiquitous technology in the profession — requires both technical and contextual expertise. We will discuss this process of specialization, and touch on ways that the Human-Computer Interaction (CHI) community has tended to the technical learning and needs of journalists. Next, we introduce the need — necessitated by the events of 2020 — for journalists to rapidly specialize on misinformation, and the tools, stop-gaps, and collaborations put in place to meet this need. Finally, we introduce collaboration on this beat as a site of learning about the socio-technical challenges that exist between journalists and researchers working to meet evolving public interest needs at speed.

2.1 Beat Specialization in the Online Environment

Online journalism has changed the role of all modern journalists, especially specialists— or beat journalists [38]. A "beat" refers to a thematic specialization in journalism [38]. Reporters cover one or a few beats, allowing them to use their expertise to reliably report on relevant news to the beat. As more information has become available to the masses (the 'new media environment' [44]), expectations from the audience of beat journalists have moved beyond those of a *disseminator or storyteller* to roles that involve even deeper expertise like *analyst, detective, and educator* [17, 22]. These beat reporters are expected to not only have access to the information but to be able to curate and contextualize it for the public in a responsible and meaningful way.

The journalism workforce, according to Marchetti, is structured around the poles of generalists and specialists [39]. The process of specialization — or developing this subject matter knowledge — for technology-heavy beats involves both methodological and contextual learning. To the former, researchers have studied the ways that computational tools and algorithms have impacted and shaped the production and dissemination of the news [14, 29, 80]. For example, one well-established development is through the emergence of Computer-Assisted Reporting (CAR) [13, 26], which requires helpful, reliable tools.

Specifically, within the CHI community, scholars have been eager to prototype and learn from tools that help journalists more quickly understand computational activity and the use of computation-enabled resources like crowdsourcing and big data [56, 81]. CHI researchers have also explored how practitioners support (or impede) journalists through conference workshops [82], exploring innovative news processes [15, 83] and centering needs of journalists in the digital age [19, 53].

Yet tools and data are hardly enough to do this kind of work. CAR and data journalism practices are still in service to professional

journalistic norms and expertise [13, 35]. The second element of the specialization process — contextual, cultural learning — comes through time-consuming news gathering, analysis, and verification related to the area of specialization [34, 38]. Especially for specialization in the online space, journalists must learn over time what kind of information to trust (often posted anonymously or by people who are difficult to interact with) and what meaning should be ascribed to it. One way to do this is by cultivating a network of expertise across disciples that helps them assess — and sometimes outsource [32]— credibility [39, 59]. Doing this specialization at speed, necessitated by the temporal challenges of news-making [9, 84], is challenging and risky. In a recent op-ed, Ivan Oransky discussed the challenges of moving onto the health beat as COVID-19 grew globally [85].

2.2 Rise and Challenges of the Misinformation Beat

The norms and practices of beat reporting vary wildly between media organizations and specializations [50], necessitating analysis of various beats and associated needs individually. Reporters working on the "misinformation beat" must navigate the complexity of the technical infrastructure underpinning misinformation, in addition to the cultural and ethical challenges, this kind of problematic content creates. Research conducted before 2020 highlights a dearth of resources to meet these needs, in particular a struggle to access appropriate, well-maintained technical tools to track and understand their subject matter [71]. Reporting on big data involves technical, methodological, and ethical challenges [7]. There's also deep contextual knowledge needed to understand how and why misinformation spread online and the role of social media platforms — and their algorithmic recommendation systems — in shaping those flows.

To grow this knowledge, journalists have embedded themselves in online spaces and reached out to experts, like industry and university researchers, for assistance [71]. When health, policy, and election misinformation flourished throughout 2020, researchers and tools, like Crowdtangle, became an essential piece of doing this work [86, 87]. Those who were new to the beat — or transferred from peripheral beats — had to get up to speed quickly to serve as trusted knowledge brokers [21] on a beat with such high stakes. Getting it wrong could mean falling victim to 'source hacking' [16], accidentally amplifying harmful information [49], and other forms of media manipulation [74, 88].

Starting in 2016 and increasing in 2020, online tip-sheets and training emerged during this period in an attempt to meet these needs [61, 72–75]. These resources were made publicly available as journalists did work to self-educate both themselves and their newsrooms on how to cover topics like medical misinformation, anti-mask rallies, and claims of voter fraud. However, these stopgap resources alone may not provide the individualized or in-depth contextual knowledge, or the technical expertise needed to report on misinformation as breaking news. Accordingly, reporters look to build relationships with academic and industry experts to build sustainable reporting habits. The relationships themselves contain complexities because of differing publication timelines and working pace, differing incentives, and public orientations [36].

2.3 How Collaborations and Ethnography Set the Stage for Design

This highlights the need to examine the driving forces behind the motivations to initiate and maintain journalist-researcher relationships on this challenging beat. The growing prevalence of these kinds of technical, contextual collaborations around misinformation suggests the existence of socio-technical challenges that hinder journalists from meeting the analysis needs of this beat individually or internally within their newsroom.

Ethnographic approaches help center members' perspectives and needs in collaborations [89] and developing of empathy [90]. Studies using participant observation — an approach to ethnography derived from cultural anthropology [2]— have been used by CHI researchers to understand the socio-technical challenges (e.g. [10, 28, 30]). Ethnographic approaches, like participant observation, are essential first steps in deepening understanding [89]. This kind of work can serve as a foundation for future design provocations and implications.

Collaborations, too, serve a unique role in the Computer Supported Cooperative Work (CSCW) space informing the design of useful systems and solutions [51]. CSCW scholars have recently studied collaborations between fact-checkers and journalists in Bangladesh, highlighting communication challenges between the professions that lead to misaligned understanding toward professional roles in the media ecosystem and urging the further study of these dynamics in developing contexts [23]. Our ethnographic approach to collaborations and subsequent analysis surfaces a better understanding of journalists' and researchers' both computer-supported and collaborative work together to fulfill growing public interest and concerns about misinformation.

3 METHOD

Adopting an interpretive and inductive approach informed by Charmaz [12] and Strauss [20], the authors analyzed 30 collaborations between (the largely U.S. centric) journalists and members of the authors' research team (described in the appendix) on misinformation-related investigations, including 14 interviews with researchers and journalists who participated in these collaborations. There were three parts to the method: 1) the collaboration activities, where the first author took an ethnographic approach to data collection as a participant-observer, 2) semi-structured qualitative interviews of both researchers and journalists who participated in these collaborations, and 3) integrated data analysis of the collaborations and interviews.

3.1 Step 1: Collaboration Initiation and Advertising

The collaborations were positioned as a way of meeting the needs of the journalists working on misinformation and disinformation-related stories. Recruitment started with members of the research team who advertised their availability for collaborations through tweets and blogs. Adjacent team efforts included rapid response research — e.g., publishing blogs and social media posts highlighting election and vaccine misinformation — that also helped make public our willingness to engage with journalists for their investigations. To be considered a collaboration, the request needed to have a data

Table 1: List of Collaborations

Collaboration Number	News Organization Type	Media Organization Number	Journalist Number	Research Team Member	Collaboration Outcome
C101	Academic News Organization	M1	J1	R2, R3	Article
C102	Digital First Outlet	M2	J2	R5	None
C103	International Newspaper	M3	J3	R1	Article
C104	International TV	M4	J4	R1	TV Segment
C105	International News Agency	M5	J5A,J5B	R4	In Progress
C106	Niche Newspaper	M6	J6	R4	Article
C107	Niche Newspaper	M7	J7	R6	Article
C108	Niche, Non-Profit Newsroom	M8	J8	R4, R2	Podcast
C109	Niche, Non-Profit Newsroom	M9	J9	R4	None
C110	U.S. Local Newspaper	M10	J10A	S2, R4	Article
C111	U.S. Local Newspaper	M10	J10B	R2	Article
C112	U.S. Local Newspaper	M11	J11A	R4, R1	Article
C113	U.S. Local Newspaper	M11	J11B	R1	Article
C114	U.S. Local Radio	M12	J12	R3	None
C115	U.S. Local Radio	M13	J13	R6	Radio Segment
C116	U.S. National Newspaper	M14	J14A	R3, R4, S1	Article
C117	U.S. National Newspaper	M14	J14A	R6	Article
C118	U.S. National Newspaper	M14	J14A	R3	Article
C119	U.S. National Newspaper	M14	J14A	R4	Article
C120	U.S. National Newspaper	M14	J14B	R3	Article
C121	U.S. National Newspaper	M15	J15A	R4, R1	Article
C122	U.S. National Newspaper	M15	J15B	R2	Article
C123	U.S. National Newspaper	M15	J15C	R1	Article
C124	U.S. National Newspaper	M16	J16	R7	Article
C125	U.S. National Newspaper	M17	J17A	R4	None
C126	U.S. National Newspaper	M17	J17B	R2,R6	None
C127	U.S. National Radio	M18	J18	R5	Radio Segment
C128	U.S. National Radio	M19	J19	R3,R4	Radio Segment
C129	U.S. National TV	M20	J20A	R4, R7	Article
C130	U.S. National TV	M20	J20B	R2, R4	None

element — quantitative or qualitative — or hint to the need for data or analysis to further develop their story.

The collaborations began when journalists reached out to the team (usually via email) with misinformation beat questions that fell into one or more of the following categories:

- Investigative story about a case, narrative, or spreader of misinformation, disinformation, or conspiracy theories.
- Story covering breaking news events with a misinformation, disinformation, or conspiracy theory element.
- Voting or election-related investigation.
- COVID-19 related investigation.

These requests were sparse at the beginning of 2020 but picked up heavily starting in August 2020 (leading up to the U.S. presidential election), extended through November 2020, and surged again in January 2021 (following the U.S. Capitol Siege). These requests typically went to the inbox of a PI or the team communications director. The request would then be forwarded to an appropriate researcher on the team and the first author would coordinate the start of the ethnographic research component (studying the collaboration).

3.2 Step 2: Participant-Observation and Ethnography of Collaborations

Once a researcher had volunteered to help with the journalist's request, the journalists and researcher(s) would then go back and forth for anywhere from two days to a month exchanging questions and data. These communications and data exchanges happened virtually due to the ongoing COVID-19 pandemic.

The first author performed as both a researcher and ethnographer for most of these collaborations. They took ethnographic notes throughout the collaborations; paying close attention to (1) the communications between the journalists and the researchers, (2) data, and artifacts (like Google Docs) that were exchanged, (3) challenges to data access and sharing, and (4) outcomes of these collaborations. Within the role as a participant-observer — as has been described by Konstan, Chi, and Höök [90] — the first author was able to grow close to the participants' (journalists') experiences to better understand and empathize with their needs. This resulted in rich data about the context from which the journalists were

Table 2: List of researchers and journalists interviewed post-collaboration

Interview Participant Number	Short Participant Description
R1	Researcher is a PhD candidate
R2	Researcher is a PhD student
R3	Researcher is an associate professor
R4	Researcher is a PhD candidate
R5	Researcher is a postdoctoral fellow
R6	Researcher is a postdoctoral fellow
R7	Researcher is a postdoctoral researcher
R8	Research team member is a data engineer[b]
R9	Research team member is a communications director[b]
J10A	Journalist covers misinformation for an international news agency
J14A	Journalist covers business and misinformation for a digital first outlet
J15A	Journalist covers technology for an international TV outlet
J18	Journalist produces general content with a focus on disinformation for a radio show
J20A	Journalist covers misinformation for a U.S. national TV outlet

[b] Two staff members who were involved with the research team were interviewed for this piece

coming and the tensions that they experienced both within their newsrooms and working with collaborators and data along the way.

Table 1 summarizes the 30 collaborations analyzed for these findings. They involved 20 different media organizations, 28 different journalists, and 9 participating research team members.

3.3 Step 3: Semi-Structured Qualitative Interviews

After the collaborations ended, the first author reached out to both the journalist and the researcher team members to set up individual, semi-structured interviews to debrief. The interviews were conducted over video conference. Table 2 lists the journalists and researchers who participated in the interviews.

During the semi-structured interviews, the first author first asked the participants about their experiences working with the researchers or journalists (whichever was appropriate) before 2020. They then asked the participant to walk through — start to finish — the experiences of the collaborations that they participated in. The interviewer asked probing questions about the participants' expectations of the collaborations and their opinions on the outcomes, among other things. The interview protocol and consent practices were reviewed and approved by the authors' University Institutional Review Board.

The participant interviews added color to the ethnographic findings and filled important information gaps about collaborations where the first author was not about to closely observe. These interviews were unique in that they allowed the researchers to learn about a few, in-depth collaborations from both the journalist and source perspective — empirical studies that consider both of these perspectives are rare, according to Magin and Maurer 2020 [38].

3.4 Step 4: Qualitative Analysis

The first, and most significant step of the qualitative analysis was reviewing the ethnographic notes, memos, and artifacts from step

Figure 1: Screenshot of Collaborative Miro Board

one of the collaborations. The first author regularly discussed notes and findings throughout the collaborations with the other authors during weekly 1:1 meetings. The second author, too, reviewed notes and memos before doing collaborative analysis.

Once the final collaboration was complete, the authors gathered and placed the most salient ethnographic notes and memos onto a collaborative Miro board. They talked through the observational notes for each collaboration and the authors were able to identify five key collaboration types and associated motivations.

Next, the interviews were transcribed, atomized, and placed on the same collaborative Miro board (see Figure 1). The first author led an exercise of sorting the 500 atomized quotes and notes from the 14 interviews into themes. As mentioned, the themes from ethnographic notes and memos were also displayed on the board, providing an initial, yet flexible structure for card categories. Many of the interview cards did align to existing ethnographic themes, but other, new findings emerged from the card sort as well. After discussing and refining the groupings, the authors identified the most salient findings to discuss in this paper.

Table 3: Typology of Collaboration Types

Type Number	Name	Short Description	Associated Collaborations	Details
1	**Follow Up Piece**	Journalists follow up on a piece of public scholarship (like a tweet or blog post) and want to write a story featuring the research team's analysis.	C102 C113 C122 C124	In these cases, public scholarship (e.g., a blog or tweet about a finding from the rapid analysis work) drove interest in the team's research and the establishment of collaborations. These collaborations were typically initiated on Twitter or via email and involved follow-up conversations describing the underlying data of what the researchers had reported and providing additional context (and sometimes data) for the story.
2	**Specific Data Question**	Journalists have a specific data question for a story they are already working on.	C103 C105 C106 C108 C112 C114 C120 C121 C128 C129	In these cases, journalists typically had a story in progress and wanted primarily quantitative data (though occasionally qualitative artifacts were enough) to advance their piece. Answering these questions was possible due to the team's data collection capabilities, access to tools, and robust data storage. Some of the collaborations that began as 2s, also evolved into 3s.
3	**Research "Take"**	Journalists have a story they are working on, and they want to talk about the research team's analysis or "take" on the topic.	C104 C107 C110 C111 C115	Many of these calls included "gut checking" - a need described in the findings. Journalists would either share data, a story, or findings from an interview and get the researchers' take on it. Put simply, asking "am I approaching this right?". Often these ended with quotes from the researcher to help put this finding in perspective for the readers.
4	**Education**	Journalists have a story they are working on, and they want contextual (background) information, data, and misinformation education to help tell the story.	C102 C117 C127	These were important for journalists new to the beat or covering topics that were new to them. Researchers would provide background knowledge about a topic and guide the journalist in how to approach the topic. Sometimes they would provide direction of where to look for specific pieces of information online or help with the framing and approach to the story. These were especially prevalent with stories about vaccine misinformation, especially as many journalists were writing about online anti-vaccination groups online for the first time.
5	**Open Ended Collaborations**	Journalists have an open-ended question or a general topic for an investigative story and want to work with researchers to shape and gather data for the piece.	C109 C116 C118 C123 C125 C126 C130	These collaborations started, often, with an open-ended question. The researcher and journalist worked together to refine the questions and went back and forth with the data many times. These collaborations were the most challenging and time-consuming, yet some of the most rewarding for researchers.

4 FINDINGS

The findings offer four primary insights: First (1), we offer a typology of motivations for journalist-researcher data and information sharing. Next, we expand on the ethnographic findings with interview insights to (2) discuss why journalists are relying more on researchers (compared to other beats) for help getting up to speed, accessing data, and 'gut checking' stories on misinformation; and (3) discuss challenges and professional tensions that emerged from these collaborations. Finally (4), we lean back into the ethnographic findings to describe the challenge of "getting on the same page" — the translational work of turning journalistic questions into data inquiry questions, and the shared understandings needed to get to a productive starting point for the collaborations.

4.1 Typology of Journalist-Researcher Collaborations

Drawing on ethnographic notes and memos from thirty different collaborations throughout one year, the authors identified five categories of motivations for journalists relying on researchers as they developed misinformation journalism. These collaborations involved journalists from a range of organizations: including many well-known, national U.S. outlets, a few international journalists, as well as local print and radio reporters. Because the journalist collaborators were in various cities, all communications were done virtually through e-mail, Slack, phone calls, and video calls.

Some of these collaborations were short — e.g., just one or two emails exchanged to answer a quick question or a few days to investigate the origins of a viral misinformation meme. Some went much longer, spanning dozens of interactions including email exchanges, phone conversations, and the use of collaborative artifacts like Google Docs or data visualizations. These extended collaborations tended to involve the co-creation of journalism between the journalist and the researcher, where the data questions and conversations guided the reporting. Some collaboration structures were innovative and experimental for both the researcher and the journalists. For example, some were done "live", where researchers and journalists co-worked (virtually, in a chat and video environment) during high visibility events like election debates and on election day — giving tips, journalists asking questions, researchers sharing quick analyses.

Many, though not all, of these collaborative efforts, contributed to published pieces. In some cases, the collaboration informed and guided the final piece. In others, the findings — even though they took ample research and journalistic effort — simply did not turn out to be newsworthy enough for publication. In these cases, either a piece wasn't published at all, or the journalists included a piece of the analyses as supporting data for an adjacent story.

Through an ethnographic analysis of these interactions, we identified five distinct types of collaborations — each revealing a different aspect of journalists' needs (generally) or goals for working with researchers on the "misinformation beat". Table 3 enumerates these different collaboration types, providing some details on how they were initiated and unfolded, and serving as a reference point throughout the remaining findings and discussion section.

4.2 Researchers fill an essential — but possibly temporary — expertise gap for journalists working on the misinformation stories

Our interviews revealed that these collaborations were part of a larger trend: journalists are leaning on external researchers for their contextual and methodological expertise in helping develop reporting on the misinformation beat. Journalists, themselves, discussed this dynamic:

> "Academics are such a large part of what I do. without that, I could not do this job" - J20A

> "On this misinformation beat you're heavily dependent on relationships with researchers, I think probably more than most beats" - J14A

This section provides color on *why* journalists are more reliant on the researchers for this beat and *why* researchers struggle to keep up with those needs.

4.2.1 Journalists lean into researchers to fill three main newsroom gaps. J18 said explicitly, "since misinfo is a new topic for newsrooms, there is a lot less internal expertise and infrastructure" on this beat. This statement was confirmed through the ethnographic findings and interviews. The lack of contextual and methodological expertise with social trace data pushes journalists to look externally for needs that are typically met with internal experts — like gut checking and data access and/or analysis. Each of these is explored in this section. Meeting these needs requires that journalists build closer, and more collaborative, relationships with researchers. These close relationships also exist in a grey area where researchers sit somewhere between a source and a collaborator — leading to relationship and independence challenges that will be discussed in future sections.

Gut Checking. Each journalist and researcher interviewed for this piece mentioned the concept of reaching out to researchers to "*gut check*" their stories, data, or framing; something that — according to J18 — is more accessible internally for other beats. The concept of 'gut checking' isn't new to journalists. Schultz writes that the journalistic gut feeling associated with determining newsworthiness is something that accrues with experience [52]. Though journalists were doing newsworthiness gut-checking throughout the collaborations (in some cases they walked us through this process), they were primarily turning to researchers for a different kind of gut checking; one more aligned with the process of beat specialization and ensuring that they were properly discussing the complexities of the online environment. For example, J18 went on to give an example: when covering a story about armed conflict, they can go to a coworker with experience on that beat and talk through their story, approach, framing, and information with them. But on the topic of misinformation, this doesn't yet exist. Researchers, too, noticed this need adding that journalists would reach out "*want[ing] validation that they weren't crazy in the way that they were thinking about this" (R1)* and to "*confirm their suspicion or the hypothesis that they're putting forward or because there's so much uncertainty around like the aspects of the beat" (R9).*

The ethnographic findings revealed that throughout these collaborations, researchers helped check journalists' findings and guide them on approaching challenging stories. After discussing the process of asking a researcher on our team for help with story framing, J18 was quick to add, *"I would be careful to asterisk that we don't ever really run scripts by researchers"*. This practice of looking externally for "gut checks" can contribute to an environment where researchers are contributing to the presentation of the story, moving beyond the traditional role of a source — and creating tension for journalists for whom this may violate professional standards.

Data Access and Analysis. As was shown in the typology, many of the collaborations also involved a specific data or data analysis request from the journalists. Journalists discussed two reasons why existing internal newsroom approaches to data don't currently meet the needs of misinformation journalists: the data is private and the existing data journalists lack experience analyzing social media trace data.

To the former challenge, J14A mentioned the fact that accessing social media data — held exclusively by private companies — is unlike the data access process for other beats they had worked on.

> "In other beats, you rely a lot on public information, like things you can [FOIA request]. In this beat, everything comes from those private companies" - J14A

Data access, then, is mediated by the social platforms themselves who are hesitant — and in many cases entirely unwilling — to share data with journalists looking into misinformation concerns on their platform. These social media organizations cite privacy concerns in this kind of data sharing, while journalists tended to

suggest that these decisions were more about these organizations' self-preservation interests.

Researchers, though, have unique infrastructure that enables data collection and analysis of social media data. In some cases, this includes privileged access to data from social platforms through partnerships. In other cases, researchers don't have "special" access, per se, but they do have the funding and infrastructure to enable data collection at a meaningful scale. This valuable data was available for the journalists who participated in these collaborations. *Specific Data Requests* (Table 3, Type 2) collaborations, the research team's data collection capabilities filled the newsroom data needs gap.

When it comes to the second need, data analysis, many newsrooms lack resources for internal expertise and tools. Though data journalism is an increasingly important part of newsroom activity, formal data journalism training is largely introductory, often focusing on older, traditional data types [25]. The social media trace data collected in these collaborations is known for having unique challenges requiring both methodological and contextual expertise to properly analyze [27]. The ethnographic notes from these collaborations — and the authors' previous relationships with journalists — suggest that many newsrooms' data journalists do not have the training or experience to collect, maintain, and analyze this kind of data.

4.2.2 Researchers are eager to help journalists on this time-sensitive and impactful beat, but these collaborations feel unwieldy and unsustainable. Universities and funders (including the National Science Foundation) have been encouraging researchers to attend to the broader impacts of their work — including through public scholarship. By working with journalists, university researchers, too, feel like their work can have more of an impact on the public discourse; possibly slowing the spread of misinformation, impacting national and platform policy, and bringing the danger of misinformation more into the public conscience at the right time. Collaborations with journalists on this beat fulfill this public scholarship goal.

> "'So, I kind of see journalists as a potential conduit for our research to the broader public" - R3

> "And I feel incredible whenever I get a [journalist] request. [...] I can help not only inform their story but hopefully better inform the public generally" - R6

Yet these collaborations take a lot of researcher time — not just to answer the questions or do the data queries — but to play a role in educating journalists who are new to the beat about this work. In some cases, the education calls lasted over two hours.

> J10A spoke fondly of these education calls: "every time I talked to [a researcher], I would learn more about this topic. And then I would form a much more nuanced, like vision of it"

> R6 spoke about this process as being a part of the broader impacts of being a researcher: "I want everyone to be really good — at covering this beat. And to do that, it does require a little bit of education. And then that's like one of the responsibilities I guess, being an academic."

Though these efforts feel worthwhile to researchers amid mass online misinformation — educating journalists about the field improves coverage and mitigates amplification risk overall — it also takes away from time that the researcher could have been spending on tasks that are more traditionally rewarded by academia, like writing peer-reviewed papers. This setup felt unsustainable to many of the researchers interviewed. There are also ethical questions surrounding this practice for the long term — e.g., about collaborating too closely with a small number of journalists, feeling pressure to publicize findings before they are peer-reviewed, and taking opportunities from other researchers to receive visibility. Researchers are currently pondering how to maintain the rigor traditionally required of the field while also moving at a pace that can have an impact on this fast-moving phenomenon. Collaborating with journalists brings these tensions to the surface.

4.3 Navigating ethical and professional differences in these collaborations: perspectives on data sharing, newsworthiness, and innovation.

Researchers and journalists both conceptualize themselves as knowledge brokers and in service to the public. Yet the way that these roles are enacted differ — like the speed of publication, sourcing practices, and epistemologies. This section discusses a few of the salient tensions: (1) researchers' concerns about the speed and nature of data sharing and (2) journalists' professional commitment to newsworthiness and meeting the public with stories of interest.

4.3.1 Sharing data at speed: researchers' hesitations. Researchers are keenly aware of how incomplete and unofficial the social media data they collect, and share, are; the data are often grabbed from public APIs, based on certain keywords, or provided directly by platforms and subject to scrutiny. Researchers are still grappling with ethical quandaries around reporting on these public social data. Sharing the data with reporters is even trickier. Researchers cited two primary hesitations to data sharing: (1) not knowing where the journalists' piece was heading and (2) feeling like the speed did not allow for enough "double-checking".

Story Direction. Researchers expressed hesitation in sharing data especially when the researcher didn't know the aim of the piece or what argument the data would be used to support. R2 put this explicitly, stating *"I felt weird about sharing data so openly with [this journalist], on this topic. Especially without knowing what "the person's angle is".* In one case, this concern developed through experience: a researcher reflected on sharing since-deleted tweet data from an inauthentic account with a journalist without much thought to how it would be used:

> "When I was giving him information on these people, I didn't know what the story was about, which is, in hindsight, probably not the best thing. I probably should have been like, why do you need this data?' But I was just like, 'you're [a big national news organization], you're fine.'" - R4

In the interview, this researcher reflected that they felt uneasy about how they handled the data sharing. Though the published

story accurately covered the data, the researcher wasn't aware that the data would be used to confront the account owner — whom the journalist had identified through an external tip — in person. Moving forward, the researcher was set on understanding the aims of the story and the use case for the data before sharing.

Yet some journalists are still developing the direction of the piece before contacting researchers for help (Table 3, Type 5). In a few cases, this created a stalemate, where the researcher wanted to know more about the story before sharing data, but the journalist wanted to see data analysis to help determine the direction of their story.

Data Sharing at Speed. The speed of publication in the newsroom is much faster than that of academia. Though the researchers we interviewed suggested that peer review, too, is a flawed process, they said it does require them to validate their methods in a much more rigorous manner before publication. Researchers spoke about occasionally feeling uncomfortable moving at the journalists' pace during these collaborations.

> "Moving at a rapid scale (esp. <24 hrs.) is challenging and vulnerable for us because we want to make sure things are exactly right. And a lot of things can go wrong, there might be an error in your data you didn't see, or you miscoded something in the script." - R3

> Another researcher said if the same work was for academia: *"I would need to do a lot more work than I did to validate my measures"* - R1

Multiple researchers spoke to the fact that sharing data outside of peer-review — where methods are scrutinized — felt uncomfortable and risky, especially on a topic of great public interest. The risk of putting out an incomplete data set or having a journalist misrepresent information to the public is a fear that brought tensions to some collaborations. One researcher, R2, didn't even think of the fact they had shared an analysis outside of peer review until people started asking for their paper on it when the article came out. R2 felt uncomfortable saying that no paper had yet been written on the topic.

The comfort levels can also vary by the kinds of data or methods required to answer the data query. R2 said they felt comfortable sharing some kinds of data publicly and with journalists because they were confident in it and, in other cases where there was more uncertainty (or were using a new method), they would want it to be reviewed. This was sometimes hard for the team to convey to journalists: why, in one collaboration, the journalists received a straightforward data response and in another, they received a long email explaining the challenges of collecting and sharing a kind of data or analysis.

As researchers have adapted to collaborating, so have journalists. The first author noted in their memos that journalists recognized researchers' discomfort around data sharing at speed. Because it is often in journalists' best interest to establish trust and make researchers feel comfortable sharing data, seasoned journalists on this beat tried to add appropriate context to build up this trust, either by running the text back by the researcher or taking the time to fully understand and contextualize the data themselves.

> "I think we reporters have all been in a sort of situation where [a researcher] has rightly been mad [at us] getting something wrong. We want it to be right. But sometimes there's something lost in translation. I always try, especially with academics, to say, 'this is what I think I hear you saying'. Sometimes they reply with 'not quite', then we'll like to massage it to a place that's true" -J20A

The journalists that help researchers feel comfortable sharing data — by working together to frame findings and provide ample context — are the ones that benefited more frequently from the knowledge of the research team in these collaborations. These trust-building practices brought certain, privileged journalists and researchers even closer together.

4.3.2 Newsworthiness and what is interesting. When collaborating with researchers, the journalist participants also had their professional intentions and priorities to consider. Among other things, journalists are eager to write engaging content that will attract readers, educate them, and that subscribers would find valuable. The journalists interviewed talked about how they would break down the dense misinformation research and make it more digestible and relatable for the public.

> "I have to balance the vegetables [info about misinformation] with what they're interested in. If I can't get them to like, read it, then it feels a little bit less impactful." - J10A

J20A, too, talked about doing this by putting more data substance in the middle of the article and making the piece interesting by having the beginning and ends be more engaging. This skill, of course, is what makes sharing research with journalists beneficial to the public — there's little value if nobody reads the piece.

In some of the collaborations, however, journalists and researchers had a hard time finding harmony between meeting the needs of the business of journalism and using the data output from the collaboration. Sometimes, the analysis findings just simply weren't that interesting or newsworthy to the journalists (though the researcher sometimes disagreed). Choosing not to run a story or include a quote or finding from the researcher who worked on gathering the data was common (e.g., C103, C109, C114, C118).

> "The stuff that we ended up finding ended up being not that newsworthy. And at the end of the day, I have to use my news judgment." - J14A

Sometimes this led researchers to feel like they had wasted time.

> "So, I did all this work. And I did data, I even roped in another researcher to do all of this. And then I gave [the journalist] feedback. Like, why I think this is happening. Here's a deeper analysis of my expertise in the space. and nothing happened" - R6

This highlights an interesting tension for the journalist — balancing the need to meet the public with interesting, informative content with the need to maintain a positive relationship with researchers — especially as the misinformation beat continues to change and comes with ample journalistic capital [38].

4.3.3 Despite challenges, these collaborations can lead to innovative research and journalism practices. For the reasons described above, not all collaboration efforts resulted in publications — despite many of them consuming researchers' and journalists' time. However, in the post-collaboration interviews, researchers and journalists reflected on ways that even unpublished efforts ended up contributing to their work, though many of them didn't realize the contributions until later. In their interviews, researchers noted that these collaborations led to innovative methods development (R1, R2, R8) and powerful case studies for presentations and forthcoming papers (R2, R4). Journalists who participated in Open-Ended *Collaborations (3, Type 5),* noted how working closely with researchers on co-production activities furthered their understanding of how data analysis works and the difficulty of gathering, cleaning, and making sense of social media information. Even when these collaborations were seemingly unsuccessful (i.e., no story was published) they had the potential to inspire thoughtful conversations and new ways of thinking about misinformation online.

4.4 Getting on the same page: translational work from journalism questions to data questions

So far, we've discussed the necessity of these kinds of work on rapid timelines and the ethical and normative roles of both reporting and academia's role in this beat. This section centers the practical; the *how*. How do these challenges manifest and what design provocations might this open for researchers and designers in the CHI community?

One major procedural challenge to the '*how*' of academic-journalist collaborations on this beat was persistent: getting on the same page. Specifically, aligning journalism questions with data questions that were actionable with the infrastructure and expertise that the research organization had. Two elements of this "getting on the same page" process are discussed: 1) scoping the journalism question to one that could be answered using the data and capabilities at hand (i.e., addressing the challenge of broad questions) and 2) having the right contextual expertise to accurately guide and understand the implications of those findings (the challenge of making sense of the data).

4.4.1 Question translation: from journalism question to data inquiry. One of the most salient and challenging issues the researchers discussed was the process of tackling big, bold questions from journalists. An example might be: "how much misinformation is on Twitter now compared to before the 2020 election?". From the perspective of the researchers, properly answering this question would require a minimum: a definition of what constitutes misinformation, access to all tweets in a period before and after the election, and the ability to process and find tweets that meet the misinformation criteria.

Receiving questions like this, over time, became burdensome for researchers for a few reasons: These kinds of questions elicited conversations about data access, data analysis, and data assumptions — as well as education on why these kinds of questions can't easily be answered.

> R9 reflected that, "[Some journalists have] a misconception that we have this big thing of data, [and] we can just press a couple of buttons and have all your answers like, 'here's your disinformation campaigns'. It was hard to explain to those journalists that's not necessarily how it works"

These kinds of questions also prompted educational conversations about the definition of misinformation, the role that public accounts play in the ecosystem of mis/disinformation, and more. In short, these broad questions often necessitated a great deal of methodological and contextual "onboarding".

As we've discussed in previous findings, researchers were eager to help with this, but over time the education work started feeling unsustainable.

> "It felt a little bit like helping to educate [them] on a bunch of different fronts, which you know, has value, I guess, if [they're] gonna go on and write about these, about these things, but it did feel a little bit scattered" - R3

The researcher and journalist would then enter a scoping phase, where those broad questions were refined into ones that could be answered using our data infrastructure and expertise. For example, the "how much misinformation is on Twitter now compared to before the 2020 election?" question might become "how much voting misinformation did these specific accounts share in the months leading up to the 2016 election vs the 2020 election?" In this case, scoping to a topic, accounts of interest and a time frame helped shape a data question that was answerable to our researchers. They could then use qualitative and/or quantitative methods to begin an analysis.

Over time, some journalists improved the way they asked questions, making them more actionable for the research team. Researchers, then, were more eager to tend to these well-scoped questions and prioritize these requests — hinting at challenges of equity in doing this work. Through this back and forth, both researchers and journalists learned how to translate these bold journalistic questions into actionable data queries.

4.4.2 Contextual, cultural expertise is necessary. The activities of 2020 reveal that the "misinformation beat", in addition to a beat of its own, spans across numerous other beats. And expertise on those intersecting beats — like health, armed conflict, vaccine science, politics, and elections — was essential for success in analyzing and covering the misinformation emerging from those topics. In many cases, researchers — unfamiliar with the context in question — probed the journalists for example artifacts, hashtags, or other evidence of misinformation that could serve as a starting point for a data-led investigation (as suggested in investigative digital ethnography [62]). It became clear, in a two difficult cases, that journalists knew very little about the communities they were investigating, which made data analysis harder, less valuable, and put them at risk of manipulation [88].

> "We were hoping that we could work together to get better questions. And if they don't understand the context well enough, then that that doesn't really work."- R3

Even when questions *were* scoped to actionable data requests, we found that without contextual, cultural expertise — either from the journalist or the researcher — it was challenging to make sense of the data. For example, when a journalist who formerly covered education began asking questions about voter fraud narratives, the researcher team encountered two challenges: 1) There were hours of contextual and methodological onboarding work to get them up to speed on misinformation and data collection capabilities; and 2) the researcher working with this journalist felt like the journalist couldn't assess the value of the data to a potential story. After going through a long scoping phase and then doing an in-depth data analysis, the researcher — who did have contextual expertise in the topic — felt like their insights had produced meaningful content for the journalist. The researcher later reflected that other journalists *with* contextual expertise would have considered these findings be interesting and newsworthy. Yet, P4 said, *"They kept wanting more and more. And from our perspective, we were like 'the story is right here!'" - R4*

Working with researchers without deep specialization also made some researchers feel 'cagey' — or reluctant to share information — especially when a historically marginalized group was involved. In reflecting on one collaboration, researchers spoke about hesitations working on a data collaboration with a journalist investigating possible inauthentic activity on discourse about a minoritized community in the US. Both the researchers and the journalists had very little knowledge about the community and didn't consider themselves to be part of it.

> "The stakes are high, you're dealing with the marginalized community, I didn't have background in it. All those things, I think made me cagey early on and just make me cagey re-reading [the conversation with the journalist]" - R2

When dealing with marginalized communities, there are added stakes to the conversation — getting it wrong can mean further marginalization or discounting authentic community organizing.

In contrast, the researchers felt that successful collaborations were ones where journalists did understand the context and did come with specific data questions. For example, a reporter on war and government asked about a piece of misinformation spreading about President Biden related to international conflict. The journalist provided a few online artifacts as a starting point. The researcher was able to use internal and external data tools to trace the provenance of that data and then share the findings with the journalist. Because the journalist was well versed in reporting on armed conflict, they were able to find value in the shared data and incorporate it into their story.

5 DISCUSSION

In our role as participant-observer researchers in these collaborations, we assist the broader research community in seeing these types of collaborations on the 'misinformation beat' with more sensitivity, taking a "this is what happens" (Dourish 2006, 547) [89] approach. We share insights and a situated perspective on a unique type of collaboration happening at the intersection of social media data and journalism. This all serves to highlight issues at this boundary and give a rich sense of this site of phenomenological interaction [24].

In this discussion, we (1) highlight the most salient tensions in these collaborations and provide practical guidance for social data researchers engaging with journalists; (2) discuss the growing role of researchers as data mediators; and finally, (3) call for digital social data researchers to develop ethical guidelines on data sharing and privacy with journalists.

5.1 A Situated Perspective on the Role of Academic Researchers as Data Mediators

As discussed in the background, the HCI community has explored tactical ways to equip journalists with tools, data skills, and online crowds to support their work [15, 56, 57, 82]. In this paper, we build upon those insights to offer a situated, ethnographic perspective on the collaboration needs of journalists who are reporting on the "misinformation beat" — a role uniquely challenging due to the accessibility of data, analysis, and the possibility of amplifying the very information that they are trying to mitigate.

Drawing attention to this intersection is important as the role of academic researchers as data mediators increases. This role is becoming more prominent, both as it is enacted through these collaborations, but also as it becomes institutionalized through data transparency initiatives. For example, in November 2021, the Aspen Commission on Information Disorder [64] called for (1) legal protections for both journalists and researchers who violate platforms terms of service in efforts to write about matters of public concern; and (2) platforms to disclose public interest data to qualified researchers, creating a "safe harbor" to analyze how information spreads on social media. Similarly, Senators Coons, Portman, and Klobuchar recently announced legislation that would grant qualified researchers access to social media data if/when their proposals were approved by the U.S. National Science Foundation [48, 91]. Each of these recommendations positions academic researchers in positions of specialized access to social media data for the public interest — reifying their role as data mediators. However, as we demonstrate here, for researchers of online mis- and disinformation, that role is unstable and often fraught, due to pressure to move at the pace of these unfolding phenomena.

5.2 Call to Action: Developing Ethical Guidelines for Data Sharing with Journalists

As requests for collaborations around these data will persist, researchers (and the research community at large) will continually be put in the position of making rapid ethical judgments on what data and insights to share with journalists. We know from boyd and Crawford's work that researchers have a role in the culture of big data — what it means, who gets access, how it is shared, and to what ends [7]. The researchers participating in these collaborations were faced with these culture decisions largely alone and at speed: e.g. churning out analysis; determining which journalists to partner with; making quick user data privacy decisions; and worrying if their work would be properly contextualized in the final piece. The interviews revealed that the weight of this work and the need for quick decision-making was heavy — especially since many of these

researchers were keenly aware of the harm that can come from revealing personal data or from getting a misinformation analysis wrong in the public eye.

Our work finds that the researcher's role as data mediator on misinformation work comes without guidance, like tactical and ethical frameworks for data sharing, which can lead to uncertainty, vulnerability, and potentially high-profile mistakes at speed. The most common gray areas that our researchers faced centered on the speed of publishing, the robustness of analysis, and user anonymization. **We call for social computing researchers, within the CHI community and beyond, to develop ethical frameworks that consider these grey areas of misinformation work, including the need to work at speed and nuances around user privacy.**

5.2.1 Ensuring methodological soundness. We know that moving too fast and making methodological or analysis mistakes can hurt the public too. Disinformation purveyors actively look for these cases to undermine trust in trusted knowledge brokers [49]. Thus, working on a framework to help researchers decide what kinds of questions can be done at speed and what kind of analysis they should feel comfortable publishing without traditional peer review is important.

These determinations will likely involve several interacting dimensions, including complexity, researcher expertise, and the potential damage of being wrong. For example, simply sharing recent tweets from a public figure or a list of the most-engaged-with Facebook pages may be straightforward and uncomplicated, assuming the methods for calculating or finding the data is clear. More complicated analyses — for example, network graphs that can both reveal and mislead — may need to be treated with more care. A central question could be: Has this research team previously published a peer-reviewed paper featuring a similar type of analysis? And even in those cases, we likely want to develop other practices to ensure validity.

5.2.2 Alternatives to peer review at speed. When it comes to reporting on misinformation, researchers and journalists feel pressure to move quickly — moving too slow to analyze and report on a viral misinformation campaign can allow it to thrive and harm the public. The peer-review process plays an important role in gatekeeping and validating researchers before it goes public, yet the pace is too slow to keep up with evolving threats.

One potential route is to build up a network of trusted researchers and to encourage parallel analysis of specific data or online phenomena that warrant a fast-paced response. In the aftermath of the 2020 U.S. Election, the Center for an Informed Public (CIP) researchers were able to point to similar findings from the Social Technologies Lab at Cornell Tech, who were also publishing not-yet-peer reviewed data and findings on election misinformation [1]. The CIP researchers used the work of the Social Technologies Lab to support a decision to publish a list of "repeat spreaders" of misleading claims. Though redundancy can feel like a waste of resources and may reduce the rewards (of seeing one's research featured in an article), the benefits of independent research teams converging on similar conclusions are likely worth the trade-off. Some journalists from the cohort discussed in this paper pursued this kind of triangulation in their data collaborations — reaching out to multiple research teams to confirm findings. Aligning our recommendations with that practice, we encourage journalists, researchers, and even funders to support this kind of triangulation.

5.2.3 Developing nuanced guidelines around anonymization. Existing research provides some ethical guidance on using social media user data in research and protecting user privacy. Fiesler and Proferes [92], for example, present considerations in using social data in published research, focusing on considering users' preferences in the absence of traditional academic processes of consent. They make recommendations such as anonymizing tweets and avoiding using deleted content, among others. Yet in the context of online misinformation, the operators of accounts that play a large — and even intentional role — in the spread of harmful misinformation may not want their activities analyzed and made public, even though it may serve the public interest to do so. Determining if/when it is ethical or not to anonymize their data can be challenging.

Academic researchers make careful commitments to mitigate potential harms to participants (and social media users become participants in our work, often without consenting). Journalists have a different set of commitments, often working — ostensibly in the public interest — to expose 'bad actors'. The tensions between these two sets of competing commitments are especially salient in the domain of online misinformation, where there are recognized harms to public health, democracy, and vulnerable groups [64], but also where, even for intentional disinformation campaigns, the majority of participating accounts are "unwitting agents" [6]. Unfortunately, it is often difficult to distinguish between a "bad actor" and a "sincere believer".

Our research community will need to come together to surface and navigate through these tensions, building off existing privacy research (e.g. [3, 18, 41]) to develop nuanced guidelines for how/when to protect the identities of accounts and account owners — and how/when to specifically call out bad actors. Relatedly, as we build tools for exporting data (including network visualizations), we may want to develop (and enact as default display options) criteria for anonymizing accounts — for example, that are private, unverified, or under a certain threshold of friends or followers.

5.3 Practical Guidance for Researchers Working with Journalists on Social Data

Finally, this work reveals some of the reasons why journalists are increasingly seeking support from academic researchers, describing how researchers are creating a (perhaps temporary) bridge for those journalists, helping to fill institutional gaps in contextual and methodological know-how for investigations into an online phenomenon. Our collaborations and subsequent analysis also surface some of the professional challenges that journalists and researchers face in working together to fulfill public interest goals at speed.

Collaborations with the most tension usually exhibited one or more of the following: (1) ample back-and-forth time spent molding journalism questions into actionable entries for data investigations; (2) lack of contextual expertise; and (3) lack of clarity around how the work would be credited. These were discussed in findings 4.2.2, 4.3.1, and 4.4.

Drawing inspiration from the process of Investigative Digital Ethnography [62], we suggest a light framework for researchers

to keep in mind during the initiation of these kinds of collaborations. These can help to save time, set expectations, and mitigate professional tensions.

Establish a Focus: Upon receiving a data request, the researcher would benefit from asking the journalists to clarify the focus of the inquiry, which can be done by asking for 1-2 explicit research questions. This task can help researchers ideate the possible methods available to answer the questions at hand and establish a focus for the investigation. The questions will likely be tweaked through the life of the collaboration, yet clarity upfront can help with getting on the same page quicker.

Probe for Prior Knowledge: Finding 4.4.2 highlighted tensions where journalists or researchers lacked contextual expertise about the topic being researched, making data analysis time-consuming and challenging. Asking questions of both the requesting journalist and the research team can help researchers assess if this investigation has the potential to be fruitful and ethical — e.g. doing data research without knowing about the community can lead to marginalization and put both parties at risk of manipulation [33]. Questions may include: What do you already know about the impacted parties? Where are you in your current story/research?

Roles and Credit: Finally asking explicitly: 'How do you envision us helping with this investigation?' can force articulation of expectations and help researchers identify the collaboration type. Researchers may also probe for what kind of credit the journalist usually gives for this kind of contribution and provide journalists with a sense of how much time this work takes.

6 CONCLUSION AND LIMITATIONS

In this paper, we have framed the intersection of social data researchers and journalism as a site of interest. We have discussed tensions, mismatches in expectations, and ethical challenges that can occur during collaborations between these professionals. Better understanding this work will help construct frameworks for more productive collaborations as researchers increasingly serve as data mediators between social platforms and journalists. There is a pressing need for the development of an ethical framework to guide this work. This research is limited by the fact that most of the journalist collaborations were U.S.-centric and took place during a time of civil and political unrest. We acknowledge that journalism is contextual and cultural. More research is needed to understand how collaborations may differ in non-U.S.-centric contexts.

ACKNOWLEDGMENTS

This research was conducted at the UW Center for an Informed Public (CIP) with support from NSF Grant 1749815 and CIP funders. We want to acknowledge the CIP funders that supported aspects of this work, including the John S. and James L. Knight Foundation, William and Flora Hewlett Foundation, Craig Newmark Philanthropies, Gates Ventures, and the Omidyar Network. We also want to express appreciation to the journalists and the CIP researchers and staff who participated in this research: Andrew Beers, Joe Bak-Coleman, Ian Kennedy, Kolina Koltai, Rachel Moran, Joseph Schafer, Paul Lockaby and Michael Grass. Finally, we would like to acknowledge and thank the students in the Fall 2020 UW directed research group and Ahmer Arif — who helped evolve this work by supporting the authors.

REFERENCES

[1] Anton Abilov, Yiqing Hua, Hana Mataov, and Mor Naaman. VoterFraud2020. Retrieved from https://github.com/sTechLab/VoterFraud2020
[2] Paul Atkinson, Amanda Coffey, Sara Delamont, John Lofland, and Lyn Lofland. 2001. *Handbook of Ethnography*. SAGE Publications Ltd, 1 Oliver's Yard, 55 City Road, London England EC1Y 1SP United Kingdom. DOI:https://doi.org/10.4135/9781848608337
[3] Karla Badillo. 2019. A Review of the CSCW 2018 "Privacy in Context" Workshop. *ACM CSCW*. Retrieved January 8, 2022 from https://medium.com/acm-cscw/a-review-of-the-cscw-2018-privacy-in-context-workshop-ae380040f1ec
[4] Philip Ball and Amy Maxmen. 2020. The epic battle against coronavirus misinformation and conspiracy theories. *Nature* 581, 7809 (May 2020), 371–374. DOI:https://doi.org/10.1038/d41586-020-01452-z
[5] Julian E. Barnes. 2021. Russian Disinformation Targets Vaccines and the Biden Administration. *The New York Times*. Retrieved September 5, 2021 from https://www.nytimes.com/2021/08/05/us/politics/covid-vaccines-russian-disinformation.html
[6] Ladislav Bittman. 1985. *The KGB and Soviet disinformation: an insider's view*. Washington: Pergamon-Brassey's.
[7] 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*.
[8] Br, y Zadrozny Br, y Zadrozny is a senior reporter for NBC News She covers misinformation, extremism, and The Internet. YouTube's recommendations still push harmful videos, crowdsourced study finds. *NBC News*. Retrieved September 5, 2021 from https://www.nbcnews.com/tech/tech-news/youtubes-recommendations-still-push-harmful-videos-crowdsourced-study-rcna1355
[9] Matt Carlson and Seth C Lewis. 2019. Temporal reflexivity in journalism studies: Making sense of change in a more timely fashion. *Journalism* 20, 5 (May 2019), 642–650. DOI:https://doi.org/10.1177/1464884918760675
[10] Stefan Parry Carmien and Gerhard Fischer. 2008. Design, adoption, and assessment of a socio-technical environment supporting independence for persons with cognitive disabilities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '08), Association for Computing Machinery, New York, NY, USA, 597–606. DOI:https://doi.org/10.1145/1357054.1357151
[11] Regina Kayla Catipon. Watching Them, Watching Us: How the Misinformation Beat Redefines Journalism's Relationship with Conspiracy Theories. University of Chicago. DOI:https://doi.org/10.6082/uchicago.2922
[12] Kathy Charmaz and Linda Liska Belgrave. 2007. Grounded theory. *Blackwell Encycl. Sociol.* (2007).
[13] Mark Coddington. 2015. Clarifying Journalism's Quantitative Turn. *Digit. Journal.* 3, 3 (May 2015), 331–348. DOI:https://doi.org/10.1080/21670811.2014.976400
[14] Nicholas Diakopoulos. 2019. *Automating the News*. Harvard University Press. Retrieved September 2, 2021 from https://www.degruyter.com/document/doi/10.4159/9780674239302/html
[15] 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.
[16] Joan Donovan and Brian Friedberg. 2019. Source hacking: Media manipulation in practice. *Retrieved DataSociety Website Httpsdatasociety Netoutputsource-Hacking-Media-Manip.–Pract.* (2019).
[17] Declan Fahy and Matthew C. Nisbet. 2011. The science journalist online: Shifting roles and emerging practices. *Journalism* 12, 7 (October 2011), 778–793. DOI:https://doi.org/10.1177/1464884911412697
[18] Casey Fiesler and Blake Hallinan. 2018. "We Are the Product": Public Reactions to Online Data Sharing and Privacy Controversies in the Media. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. Retrieved January 8, 2022 from https://doi.org/10.1145/3173574.3173627
[19] Andrew Thomas Garbett, Rob Comber, Paul Egglestone, Maxine Glancy, and Patrick Olivier. 2014. Finding "real people": trust and diversity in the interface between professional and citizen journalists. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), Association for Computing Machinery, New York, NY, USA, 3015–3024. DOI:https://doi.org/10.1145/2556288.2557114
[20] Barney Glaser. 1967. 8: Strauss, A.(1967). The discovery of grounded theory. *Alsine Gruyter N. Y.* (1967).
[21] Aurore Haas. 2015. Crowding at the frontier: boundary spanners, gatekeepers and knowledge brokers. *J. Knowl. Manag.* 19, 5 (January 2015), 1029–1047. DOI:https://doi.org/10.1108/JKM-01-2015-0036
[22] Thomas Hanitzsch and Tim P Vos. 2018. Journalism beyond democracy: A new look into journalistic roles in political and everyday life. *Journalism* 19, 2 (February 2018), 146–164. DOI:https://doi.org/10.1177/1464884916673386
[23] Md Mahfuzul Haque, Mohammad Yousuf, Ahmed Shatil Alam, Pratyasha Saha, Syed Ishtiaque Ahmed, and Naeemul Hassan. 2020. Combating Misinformation in Bangladesh: Roles and Responsibilities as Perceived by Journalists, Fact-checkers,

and Users. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2 (October 2020), 130:1-130:32. DOI:https://doi.org/10.1145/3415201

[24] Steve Harrison, Deborah Tatar, and Phoebe Sengers. 2007. *The three paradigms of HCI.*

[25] Bahareh R. Heravi. 2019. 3Ws of Data Journalism Education. *Journal. Pract.* 13, 3 (March 2019), 349–366. DOI:https://doi.org/10.1080/17512786.2018.1463167

[26] Brant Houston. 2014. *Computer-Assisted Reporting: A Practical Guide.* Routledge.

[27] Jose Luis Jimenez-Marquez, Israel Gonzalez-Carrasco, Jose Luis Lopez-Cuadrado, and Belen Ruiz-Mezcua. 2019. Towards a big data framework for analyzing social media content. *Int. J. Inf. Manag.* 44, (February 2019), 1–12. DOI:https://doi.org/10.1016/j.ijinfomgt.2018.09.003

[28] Brigitte Jordan. 1996. Chapter 3 Ethnographic workplace studies and CSCW. In *Human Factors in Information Technology*, Dan Shapiro, Michael Tauber and Roland Traunmüller (eds.). North-Holland, 17–42. DOI:https://doi.org/10.1016/S0923-8433(96)80005-0

[29] Joakim Karlsen and Eirik Stavelin. 2014. Computational Journalism in Norwegian Newsrooms. *Journal. Pract.* 8, 1 (January 2014), 34–48. DOI:https://doi.org/10.1080/17512786.2013.813190

[30] Julie A. Kientz and Gregory D. Abowd. 2008. When the designer becomes the user: designing a system for therapists by becoming a therapist. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2071–2078. Retrieved September 9, 2021 from https://doi.org/10.1145/1358628.1358639

[31] Kathryn Kranhold. 2020. Social Media in 2020: A Year of Misinformation and Disinformation. *Wall Street Journal*. Retrieved August 31, 2021 from https://www.wsj.com/articles/social-media-in-2020-a-year-of-misinformation-and-disinformation-11607712530

[32] B.T. Lawson. 2021. Hiding Behind Databases, Institutions and Actors: How Journalists Use Statistics in Reporting Humanitarian Crises. *Journal. Pract.* 0, 0 (May 2021), 1–21. DOI:https://doi.org/10.1080/17512786.2021.1930106

[33] Becca Lewis and Alice E. Marwick. 2017. Media Manipulation and Disinformation Online. *Data & Society*. Retrieved March 18, 2021 from https://datasociety.net/library/media-manipulation-and-disinfo-online/

[34] Seth C. Lewis and Oscar Westlund. 2015. Big Data and Journalism. *Digit. Journal.* 3, 3 (May 2015), 447–466. DOI:https://doi.org/10.1080/21670811.2014.976418

[35] Wiebke Loosen, Julius Reimer, and Fenja De Silva-Schmidt. 2020. Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. *Journalism* 21, 9 (September 2020), 1246–1263. DOI:https://doi.org/10.1177/1464884917735691

[36] Stephen MacGregor, Amanda Cooper, Andrew Coombs, and Christopher DeLuca. 2020. A scoping review of co-production between researchers and journalists in research communication. *Heliyon* 6, 9 (September 2020), e04836. DOI:https://doi.org/10.1016/j.heliyon.2020.e04836

[37] Melanie MaginMelanie Magin and Peter MaurerPeter Maurer. 2020. Beat Journalism and Reporting. In *The Oxford Encyclopedia of Journalism Studies*. Oxford University Press. Retrieved September 7, 2021 from http://www.oxfordreference.com/view/10.1093/acref/9780190694166.001.0001/acref-9780190694166-e-905

[38] Melanie Magin and Peter Maurer. 2019. Beat Journalism and Reporting. *Oxford Research Encyclopedia of Communication*. DOI:https://doi.org/10.1093/acrefore/9780190228613.013.905

[39] Dominique Marchetti. Sub-fields of specialized journalism. 15.

[40] Dominick Mastrangelo. 2021. Trump supporter warns CNN reporter of "civil war" if former president not reinstated "soon." *TheHill*. Retrieved September 5, 2021 from https://thehill.com/homenews/media/560681-trump-supporter-warns-cnn-reporter-of-civil-war-if-former-president-not

[41] Nora McDonald, Karla Badillo-Urquiola, Morgan G. Ames, Nicola Dell, Elizabeth Keneski, Manya Sleeper, and Pamela J. Wisniewski. 2020. Privacy and Power: Acknowledging the Importance of Privacy Research and Design for Vulnerable Populations. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI EA '20), Association for Computing Machinery, New York, NY, USA, 1–8. DOI:https://doi.org/10.1145/3334480.3375174

[42] Amy Mitchell, Mark Jurkowitz, J. Baxter Oliphant, and Elisa Shearer. 2020. Political Divides, Conspiracy Theories and Divergent News Sources Heading Into 2020 Election. *Pew Research Center's Journalism Project*. Retrieved August 31, 2021 from https://www.pewresearch.org/journalism/2020/09/16/political-divides-conspiracy-theories-and-divergent-news-sources-heading-into-2020-election/

[43] Abby Livingston and Shawn Mulcahy. 2020. As states count votes, some of Texas' most prominent Republican politicians are spreading misinformation about the election. *The Texas Tribune*. Retrieved September 5, 2021 from https://www.texastribune.org/2020/11/06/texas-republicans-trump-results/

[44] John Naughton. Blogging and the emerging media ecosystem. 10.

[45] Ashley Nerbovig. Michigan's immigrant communities hit with misinformation on closed platforms. *Detroit Free Press*. Retrieved September 5, 2021 from https://www.freep.com/story/news/politics/elections/2020/11/27/michigan-election-misinformation-immigrant-whatsapp/6393693002/

[46] Miles Parks. 2021. The Most Popular J&J Vaccine Story On Facebook? A Conspiracy Theorist Posted It. *NPR*. Retrieved September 5, 2021 from https://www.npr.org/2021/04/15/987182241/the-most-popular-j-j-vaccine-story-on-facebook-a-conspiracy-theorist-posted-it

[47] Concha Pérez-Curiel, Ricardo Domínguez-García, and Gloria Jiménez-Marín. 2021. Public Sphere and Misinformation in the U.S. Election: Trump's Audience and Populism Indicators in the COVID-19 Context. *Journal. Media* 2, 3 (September 2021), 335–350. DOI:https://doi.org/10.3390/journalmedia2030020

[48] Nathaniel Persily. OPENING A WINDOW INTO TECH: THE CHALLENGE AND OPPORTUNITY FOR DATA TRANSPARENCY. 17.

[49] Whitney Phillips. 2018. The oxygen of amplification. *Data Soc.* (2018).

[50] Zvi Reich. 2012. Different Practices, Similar Logic: Comparing News Reporting across Political, Financial, and Territorial Beats. *Int. J. Press.* 17, 1 (January 2012), 76–99. DOI:https://doi.org/10.1177/1940161211420868

[51] Kjeld Schmidt and Liam Bannon. 1992. Taking CSCW seriously: Supporting Articulation Work. *Comput. Support. Coop. Work* 1, (March 1992), 7–40.

[52] Ida Schultz. 2007. THE JOURNALISTIC GUT FEELING. *Journal. Pract.* (April 2007). DOI:https://doi.org/10.1080/17512780701275507

[53] C. Estelle Smith, Xinyi Wang, Raghav Pavan Karumur, and Haiyi Zhu. 2018. [Un]breaking News: Design Opportunities for Enhancing Collaboration in Scientific Media Production. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. Retrieved September 7, 2021 from https://doi.org/10.1145/3173574.3173955

[54] Sylvanna Falcón, Jessica De La Cruz, and Sydney Lark Eliot,. Black Lives Matter & Disinformation. Retrieved August 31, 2021 from https://humanrights.ucsc.edu/publications/black-lives-matter-disinformation.html

[55] ©Stanford University, Stanford, and California 94305. The Long Fuse: Misinformation and the 2020 Election. Retrieved May 28, 2021 from https://purl.stanford.edu/tr171zs0069

[56] 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. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (2019), 1–30.

[57] Yixue Wang and Nicholas Diakopoulos. 2021. Journalistic Source Discovery: Supporting The Identification of News Sources in User Generated Content. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–18. Retrieved September 8, 2021 from https://doi.org/10.1145/3411764.3445266

[58] District of Columbia 1100 Connecticut Ave NW Suite 1300B Washington and Dc 20036. PolitiFact - The scariest misinformation of 2020. *@politifact*. Retrieved August 31, 2021 from https://www.politifact.com/article/2020/oct/31/scariest-misinformation-2020/

[59] 1980. *Manufacturing the News By Mark Fishman*. Retrieved September 8, 2021 from https://utpress.utexas.edu/books/fishman-manufacturing-the-news

[60] 2018. I've reported on misinformation for more than a year. Here's what I've learned. *Poynter*. Retrieved September 5, 2021 from https://www.poynter.org/fact-checking/2018/ive-reported-on-misinformation-for-more-than-a-year-heres-what-ive-learned/

[61] 2020. How journalists can avoid spreading coronavirus misinformation. *The Journalist's Resource*. Retrieved September 5, 2021 from https://journalistsresource.org/politics-and-government/coronavirus-misinformation-crowdsourcing-truth-mit-david-rand/

[62] 2020. Investigative Digital Ethnography: Methods for Environmental Modeling. *Media Manipulation Casebook*. Retrieved May 5, 2021 from https://mediamanipulation.org/research/investigative-digital-ethnography-methods-environmental-modeling

[63] 2021. CNN's Irish reporter Donie O'Sullivan goes viral during Washington unrest. *IrishCentral.com*. Retrieved September 5, 2021 from https://www.irishcentral.com/news/cnn-irish-reporter-donie-o-sullivan

[64] 2021. Commission on Information Disorder Final Report. *The Aspen Institute*. Retrieved January 8, 2022 from https://www.aspeninstitute.org/publications/commission-on-information-disorder-final-report/

[65] In 2020, Disinformation Broke The US. Retrieved August 31, 2021 from https://www.buzzfeednews.com/article/janelytvynenko/disinformation-broke-us

[66] Infodemic. Retrieved August 31, 2021 from https://www.who.int/westernpacific/health-topics/infodemic

[67] 12/15/20 - THE QANON CONSPIRACY: Destroying Families, Dividing Communities, Undermining Democracy. *Network Contagion Research Institute*. Retrieved August 31, 2021 from https://networkcontagion.us/reports/the-qanon-conspiracy-destroying-families-dividing-communities-undermining-democracy/

[68] Wayfair Trafficking Conspiracy Theory Spreading on TikTok - Rolling Stone. Retrieved August 31, 2021 from https://www.rollingstone.com/culture/culture-news/wayfair-child-trafficking-conspiracy-theory-tiktok-1028622/

[69] US Social Media Usage. *Insider Intelligence*. Retrieved August 31, 2021 from https://www.emarketer.com/content/us-social-media-usage

[70] On the Misinformation Beat: Understanding the Work of Investigative Journalists Reporting on Problematic Information Online: Proceedings of the ACM on Human-Computer Interaction: Vol 4, No CSCW2. Retrieved January 6, 2021 from https://dl.acm.org/doi/abs/10.1145/3415204

[71] On the Misinformation Beat: Understanding the Work of Investigative Journalists Reporting on Problematic Information Online: Proceedings of the ACM on Human-Computer Interaction: Vol 4, No CSCW2. Retrieved January 6, 2021 from https://dl.acm.org/doi/abs/10.1145/3415204
[72] Countering Misinformation: Tips for Journalists. *Columbia Journalism Review*. Retrieved August 31, 2021 from https://www.cjr.org/united_states_project/_countering_misinformation_tip.php
[73] Prepare for US election misinformation. *First Draft*. Retrieved September 5, 2021 from https://firstdraftnews.org:443/long-form-article/prepare-for-us-election-misinformation/
[74] Media Manipulation Casebook. *Media Manipulation Casebook*. Retrieved September 5, 2021 from https://mediamanipulation.org/homepage
[75] MediaWise Fact-Checking Certificate (Winter 2020/21). *Poynter*. Retrieved September 5, 2021 from https://www.poynter.org/shop/fact-checking/mediawise-fact-checking-certificate-winter-2020/
[76] Misinformation on Facebook got six times more clicks than factual news during the 2020 election, study says. *Washington Post*. Retrieved September 5, 2021 from https://www.washingtonpost.com/technology/2021/09/03/facebook-misinformation-nyu-study/
[77] Center for an Informed Public, Digital Forensic Research Lab, Graphika, & Stanford Internet Observatory (2021). The Long Fuse: Misinformation and the 2020 Election. Stanford Digital Repository: Election Integrity Partnership. v1.3.0 Available at: https://purl.stanford.edu/tr171zs0069.
[78] Brandy Zadrozny on Twitter. *Twitter*. Retrieved September 5, 2021 from https://twitter.com/BrandyZadrozny/status/1432413838025142280
[79] Shorenstein Center on Twitter. *Twitter*. Retrieved September 5, 2021 from https://twitter.com/ShorensteinCtr/status/1433796196565348357
[80] Full article: THE PROMISE OF COMPUTATIONAL JOURNALISM. Retrieved September 7, 2021 from https://www.tandfonline.com/doi/full/10.1080/17512786.2011.616655?casa_token=yMQd4goD7sgAAAAA%3AfrfI6bM1e0GPLVtJ1ISHFpPOIT65GUFc2u3rY9bObK5eCyMINNlbzFxd-0ut9-IRTxQ07qAeMBn7LQ
[81] Supporting the Use of User Generated Content in Journalistic Practice | Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. Retrieved September 7, 2021 from https://dl.acm.org/doi/10.1145/3025453.3025892
[82] HCI for Accurate, Impartial and Transparent Journalism | Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. Retrieved September 7, 2021 from https://dl.acm.org/doi/10.1145/3290607.3299007
[83] Making the News | Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. Retrieved September 7, 2021 from https://dl.acm.org/doi/10.1145/3173574.3174049
[84] Epilogue: Timing the study of news temporality - Barbie Zelizer, 2018. Retrieved September 8, 2021 from https://journals.sagepub.com/doi/full/10.1177/1464884916688964?casa_token=hwE0MmBdXt4AAAAA%3AO7AJHn_jMA6GysK5PzWPM1IS4blBdLfYmVghjT-uMxquXonoFlNcNHJ6fYQovq8jLvtBDuO1MIY
[85] Op-ed: Covering science at dangerous speeds. *Columbia Journalism Review*. Retrieved September 7, 2021 from https://www.cjr.org/opinion/pandemic-science-publishing-journals-how-to.php
[86] Coronavirus: Tools and guides for journalists. *First Draft*. Retrieved September 7, 2021 from https://firstdraftnews.org:443/long-form-article/coronavirus-tools-and-guides-for-journalists/
[87] How NBC News Uses CrowdTangle To Source and Tell Local Stories on a National Level. Retrieved September 7, 2021 from http://help.crowdtangle.com/en/articles/4554891-how-nbc-news-uses-crowdtangle-to-source-and-tell-local-stories-on-a-national-level
[88] Media Manipulation & Disinformation. *Data & Society*. Retrieved May 25, 2021 from https://datasociety.net/research/media-manipulation/
[89] Implications for design. DOI:https://doi.org/10.1145/1124772.1124855
[90] Being in the thick of in-the-wild studies. DOI:https://doi.org/10.1145/2207676.2208561
[91] U.S. Senator Christopher Coons of Delaware. Retrieved January 8, 2022 from https://www.coons.senate.gov/news/press-releases/coons-portman-klobuchar-announce-legislation-to-ensure-transparency-at-social-media-platforms
[92] "Participant" Perceptions of Twitter Research Ethics - Casey Fiesler, Nicholas Proferes, 2018. Retrieved January 8, 2022 from https://journals.sagepub.com/doi/full/10.1177/2056305118763366

A APPENDICES

A A NOTE ON COLLABORATIVE INFRASTRUCTURE OF THE PARTICIPATING RESEARCH TEAM

This research took place within an interdisciplinary team primarily based within the University of Washington. The team has robust, reliable, and flexible infrastructure — including social media data collection infrastructure, team members with methodological and contextual expertise, communications staff, and funding resources — that enabled these kinds of collaborations with journalists. While this study was taking place, the team had (1) nearly a dozen part-time student researchers (including R1-R7), (2) a dedicated PI and three faculty researchers volunteering effort, and (3) three staff members —including a full-time data engineer who maintained servers, social media data collections, and the technical infrastructure for distributed analysis.

During the study period, the team also had key partnerships and chat channels with other organizations doing misinformation research on social media, which helped them stay up to date on emerging misinformation narratives and provided analysis support. In Autumn 2020, during the most active period of this research, the team was partnering with other research teams to do rapid analysis of misinformation related to the 2020 Election. These partnerships continued from December 2021 – June 2022 with a focus on vaccine misinformation. This rapid response work involved rapidly identifying, analyzing, and communicating about emergent mis- and disinformation online, and led to the publication of dozens of blogs and social media posts which became a route for journalists to become aware of our team and its research (and the potential for collaboration). Through this work, the team was also able to obtain financial assistance in hiring several student researchers dedicated to "rapid analysis" for the fall and winter of 2020, which enhanced their ability to engage in these collaborations. The first and third authors of this paper are members of this team and were participants in these collaborations.