News as Data for Activists: a case study in feminicide counterdata production

Rahul Bhargava
r.bhargava@northesatern.edu
Northeastern University

Harini Suresh
hsuresh@mit.edu
Data + Feminism Lab & CSAIL, MIT

Amelia Lee Doğan
dogan@mit.edu
Data + Feminism Lab, MIT

Wonyoung So
wso@mit.edu
Data + Feminism Lab & DUSP, MIT

Helena Suárez Val
feminicidio.uruguay@gmail.com
Feminicidio Uruguay & CIM Warwick

Silvana Fumega
silvana@idatosabiertos.org
ILDA

Catherine D'Ignazio
dignazio@mit.edu
Data + Feminism Lab & DUSP, MIT

ABSTRACT

News articles are an important source of data for recording and aggregating a range of social phenomena. In this paper, we ask if and how technology can support civil society activists who challenge asymmetrical power relations by producing counterdata—datasets missing from mainstream counting institutions. We consider a case study centered on activists who monitor feminicide, or the lethal outcome of gender-related violence, often using news as a main source to identify and compile databases of incidents. We describe a system that we collaboratively built with activists, aimed at relieving some of the emotional and time-intensive labor this work entails. The system discovers relevant news stories on multiple systems, classifies them based on machine learning models, clusters them into groups of stories about the same incident, and delivers regular email alerts to users. Currently, 26 groups across different geographical regions are using the system, and groups who broadly monitor feminicide report that they are regularly discovering new cases. We also reflect on the short-comings of the pilot system for groups with more specific, intersectional monitoring focuses, and the implications of biased narratives or under-reporting on the system's design. This case study contributes a grounded example of computational journalism built in collaboration with, and in service of, activists working on critical human rights issues.

1 INTRODUCTION

News can be seen historically as serving many roles in society. Constructive descriptions position news as a necessary vehicle for public understanding, solution finding, and supporting democracy [35]. Critical descriptions complicate how those norms operate [2, 26], contextualize recent practices within a long history of news as a driver for state propaganda [31], and describe it as a vehicle for reinforcing hegemonic power [5].

Within this patchwork of intentions and impacts, one particular thread to highlight is the use of news articles as a source for producing datasets, . Examples span from Ida B. Wells 1895 publication of *The Red Record* to catalog statistical evidence of lynchings, to the late 20th century work of Esther Chavez Cano in Ciudad Juárez documenting the disappearances and murders of young women, to the Washington Post's 2015 launch of the *Fatal Force* project to collect and share data about people shot and killed by police in the US [43, 4, 16]. These examples show that news has long been a data source, specifically for information about types of violence in which the state plays a role, either directly as in the case of state-sanctioned violence or indirectly through creating a climate of impunity.

We offer this paper as a case study of a computational approach to tracking violence through news monitoring - a set of technologies that were conceived, designed, and built in collaboration with activists working on the issue of feminicide in the Americas. As part of a larger project called "Data Against Feminicide", we worked with a global set of activist groups to understand how they use news, collaboratively designed tools that identify potential news stories about feminicide, built a set of classifier models to filter stories of interest for their purposes, and delivered those as daily email-based alerts that cluster detected news stories around individual events. We call this the "Data Against Feminicide: Email Alerts" system ("Email Alerts" for short). In this paper we begin by reviewing the context within which this data activism operates, document our system architecture, discuss outputs and impacts, and close with reflections on the role of computation in supporting this pro-social use of online news articles as a source for counterdata production.

2 CONTEXT AND RELATED WORK

This case study sits at the intersection of a number of areas of - feminist scholarship, journalism, and computer science.

2.1 Feminicide Counterdata Production

Feminicide (or femicide) is the misogynous and gender-related killing of women[1]. According to estimates by the United Nations, 87,000 women were intentionally killed across the world in 2017—with nearly 60% murdered by intimate partners or family members [41]. Yet, as striking as existing statistics are, they often belie the complexity of the issue and related monitoring challenges [22, 13].

We build on the conceptions of *missing data* and *counterdata* from the emerging approach of data feminism [11]. *Missing data* describes those data that are not prioritized as being worth counting

---

[1]We use the term "feminicide" due to its history of situating femicides with a large context of violations against the human rights of women, centering the state's role in allowing the phenomenon to go unchecked, and its norm of use in Latin America where many of our partners work [23]

and maintaining, despite the direct import to the livelihood of significant groups of people. A response from civil society is the production of *counterdata* - an activist practice that attempts to expose asymmetrical power relationships through the production of datasets in support of political and social goals [14, 9, 29].

For this case study we focus on data about feminicide as missing data. Official feminicide data are often scant, difficult to access and contested. Activists and civil society groups in Latin America and beyond are increasingly stepping into these gendered data gaps to undertake feminist counterdata science – an explicit challenge to the inadequate data practices of governments and judicial systems on the topic of gender-related killings and other human rights violations. We engage the question: how can activists and technologists work together on designing tools to aid in the prevention and eradication of gender-related killings such as feminicide, MMIWG2S, and LGBTQ+ murders?

2.2 News Coverage and Activist Uses

Prior case-study work on news reporting of feminicides has revealed harmful patterns of coverage, leading to significant critiques. The events themselves often go un- and under-reported, particularly for groups at the intersection of multiple forces of domination [7, 42, 27, 17]. When reported on, accounts often feature damaging narratives of victim-blaming, dehumanization, and sensationalization [39, 32, 6]. Stories depict feminicides as isolated crimes or the work of pathologized killers and rarely connect individual cases to underlying patterns of structural violence [18, 3].

Nevertheless, activists use news media as a primary source of information in the absence of official state data. They follow various strategies for extracting information from these flawed reports in news without importing their misogynist transphobic racist framings [36]. Existing counterdata production processes related to online news articles rely on both web searching and automated alerting systems. Once activists locate an article about a case of feminicide, they copy and paste relevant fields one-by-one from the article into their spreadsheet or database. Case by case, they assemble spreadsheets and databases of structured information that they then publish as reports, maps, infographics, or open databases for others to use. This data production supports news-related outcomes such as reframing how feminicides are understood [3], raising public consciousness and awareness [30], supporting families and communities [8], and informing and advocating for policy [28].

2.3 Automated News Processing

There are numerous examples of computational approaches to extracting and cataloging events from news. Some approaches build on entity extraction, probabilistic modeling, machine learning models, and other methods [25, 24, 19]. These are typically designed for computational social science researchers to assemble large datasets for study. Others take a query-based approach, constructing complex boolean search strings and utilizing them to identify candidate articles that might contain reports of events of interest [37]. One free system that utilizes a query-based approach is the widely used Google Alerts product. Beyond that there are a plethora of companies offering event detection as a paid service online, targeting advertising, civic, and defense industries as potential clients.

A parallel thread is the use of computational tools to process online content as data in service of journalists and editors within the news ecosystem itself. Thurman names this as *computational news discovery*, leveraging a history of using computational tools to extract information from digital primary sources in automated ways [40]. Diakopoulos further extends the concept in support of a more robust framework with descriptions of software systems and how they connect to the gatekeeping roles newsrooms often play [15]. Automated systems that gather news as data can surface breaking news, create leads to pursue, highlight social media events, and more. Our work is very similar in software structure to the systems described as computational news discovery, though our application is in a different domain, namely that of data activism around monitoring gender-related human rights violations.

3 SYSTEM DESIGN

We collaboratively designed a problem statement and technology design for the Email Alerts system through extensive participatory engagement with activist groups [12]. Working with activists in a co-design process yielded a proposed system to directly address some of the challenges they face. At a high level, the system would discover relevant news stories on multiple systems, classify them based on machine learning models, cluster them into groups of stories about the same incident, and deliver regular email alerts. This proposal supported (rather than replaced) activists' existing processes and workflows, and built on their existing knowledge of Google Alerts, while offering wider coverage of news sources, more efficient and relevant production of results, and clustering into events. The system is composed of two main technical parts - the Email Alerts Server and the Story Processor (Figure 1).
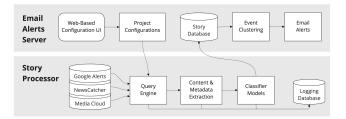

Figure 1: High level architecture of the Email Alerts system

The Email Alerts Server allows for overall system configuration via a web-based interface. Users can create collaborative "projects" that encode queries, languages[2], geographic regions, and settings about how/when to receive results. Configured projects are run through a queue-based multi-stage pipeline in the Story Processor:

URL Discovery and Collection: Each project's query (keywords, date, region) is run on the services configured each day. The Story Processor connects with multiple services that aggregate global news online. First, we integrated with the Media Cloud system - an academic project that regularly collects stories from more than 60,000 global online news sources via RSS feeds [33]. We worked with that team to create and extend their collections

---

[2]We currently support Spanish and English, and we are working with Brazilian data activists on co-creating a Portuguese language model.

of media sources based on sources our activist partners regularly scan manually. Second, we integrated with Google Alerts via RSS feeds that service offers. Third, we integrated with the online News-Catcher service, which charges a fee to obtain API-level results from a similar scale of sources to Media Cloud. URLs that match the are queued.

Content & Metadata Extraction: Batches of URLs associated with projects are pulled off a queue and full HTML content is fetched for each. This HTML is then fed through a custom-built open-source library that takes a fallback-based approach to extracting article content and relevant metadata, such as publication date.

Classification: The text is run through the appropriate classifier to score its relevance for the project [38]. If this score is above the configured threshold (which defaults to 0.75), the story is kept and the metadata is reshaped to a single standardized format for further processing.

Entity Extraction: In order to support later operations, the article text is run through a custom library that integrates a combination of entity recognition engines to extract people, dates, and people's ages. These entity types were chosen specifically based on how activists review articles.

Consolidation: Remaining candidate articles, and their associated discovered/computed metadata, are then posted to the Email Alerts Server.

Event-Based Clustering: Because activists expressed how frustrating and emotionally exhausting it was to read many articles about the same event, our system tries to group articles about the same event together [10]. A simple clustering heuristic groups stories within each project into "events" based on titles, publication dates, and entity overlap. One event consists of one or more news stories within a project.

Email Delivery: Project owners are sent an automated email listing the stories discovered since their last email update. Each is hyperlinked, and the email also includes a link to a web-based dashboard where they can review their project configuration and see historical listings of matched news articles.

This system is deployed across multiple servers owned by the research team based at MIT and Northeastern University. It is supported and monitored regularly by the technology members of the team.

4 INITIAL RESULTS

The Email Alerts system has been running in "invitation only" mode since early 2021, during which we have helped 26 groups set up projects. The current system is running almost 100 projects, with queries that ingest and process around 100,000 stories each day.

4.1 Data Characterization

In 2022 we added detailed logging to support characterizing data ingest and classifier behavior. The quantity of the relevant article pool for each project is an interplay between refinement of the boolean query and the classifier model. Some projects opt for a very broad query and collection of media sources, while other projects utilize very narrow queries over a smaller set of sources. Similarly, some classifier models are trained to identify any gender-related killing, while others are trained to detect killings of specific types of

victims with intersectional identities, such as Missing and Murdered Indigenous women, girls and two spirit people (MMIWG2). This interplay makes it difficult to describe the system ingest behavior overall. Instead, we review two projects as examples of the larger set.

One project was created with Women Count USA: Femicide Accountability Project, who produce a historical record of women and girls whose lives were ended by feminicide in the US. This project opted for a broad query across all national and local news sources, matching roughly 15,000 news stories each day in Media Cloud and a few hundred daily in NewsCatcher. from over 10,000 US national, regional, and local sources. The project utilized the default English language model, trained on a manually labeled corpus of about 400 articles [12]. Just 0.94% of the stories ingested matching the query are classified as relevant and passed on to the central Email Alerts server.

Another project was created with Mumalá, a group who monitors femicides in Argentina as part of their Observatorio. The project searches against hundreds of Spanish-language Argentine national/regional/local sources, matching around 300 stories daily in Media Cloud and 500 in NewsCatcher. The project utilizes the default Spanish language model, also trained on a manually labeled corpus of about 400 articles [12]. 9% of the stories ingested matching Mumalá's query are classified as relevant and passed on to the central Email Alerts server.

4.2 Social Impact

During the pilot, the Email Alerts system proved to work fairly well for groups monitoring feminicide generally. One main outcome was the increase in precision of results based on the multiple levels of filtering; the query + model approach was producing highly relevant articles for activists. Another feature that yielded positive results was the grouping of articles into synthetic events, which made it easier for activists to quickly review large numbers of articles from disparate sources in one sitting.

In addition, Both Women Count USA and Mumalá, as well as other pilot participants, stated that they were regularly discovering new cases through the email alerts. For example, activists from Mujeres de Negro Rosario, who track feminicides in Argentina, commented that "This tool for us has been marvelous. It saves a lot of time, saves the emotional burden, and we have all the information so that we can follow up on the cases." Similarly, Dawn Wilcox from Women Count USA wrote in a weekly survey that "I found 38 cases this week and from sources I had never seen before!"

5 DISCUSSION

This case study provides one model for computational approaches to using news as a source of data activists tracking feminicide and gender-related killings across the Americas. This work exists in cyclical dialogue with news production, monitoring of deaths, and activist efforts to reframe narratives, raise consciousness, support communities, and contribute to policy responses. We want to highlight two main points for discussion of content-related findings and system/project architecture.

5.1 Differing Frames of Coverage

Using news as data, specifically as a source for counterdata production, risks replicating media frames that are often biased to reinforce existing forms of oppression. Our work supports activists in navigating these biases by reducing and grouping the number of articles they must sift through, allowing them to focus their time on the subtle work of reading between the lines to extract data from news articles and reframe this data through their feminist reading of gender-related killings of women as cases of feminicide.

Differing reporting based on victim identities created particular challenges. Our first pilot demonstrated that our initial default English language classifier worked well for groups working on tracking all gender-related killings, but poorly for groups in the US monitoring racialized forms of feminicide - trans victims, Black women MMIWG2. These groups face more missing data and underreporting in the news media [36]. Additionally, news articles often do not report the race or Indigenous identities and often misgender trans people, so it is difficult to try to distinguish articles based on language that is either absent or incorrect.

As already known to our activist partners, this preliminary finding also reinforces previous work on framing effects in news on gender-related killings. News reports differ on killings based on the identities the victims present [7, 36, 42, 45]. Researchers have explored this qualitatively and proposed a response - reporting guidelines and handbooks for journalists covering domestic violence and feminicide-suicides [32, 34].

To address these challenges within our system, we developed custom data sets and machine learning models for these groups [38]. This is in line with our commitment to the data feminism principles of considering context and embracing pluralism. To address these challenges within our system, we developed custom data sets and machine learning models for these groups. We then trained hybrid models that explicitly take into account these different identities—for example, the hybrid model for police feminicides works by training one model on all articles to identify a police violence annotation, and combining it with another model trained only on the police violence articles to distinguish police feminicides from non-feminicide police violence. In initial evaluations of the projects deployed with these models, we found that the context-specific models return a much higher fraction of relevant results than the initial, more general model.

5.2 Source Coverage Challenges

The news aggregation platforms we use to discover candidate stories significantly expand the number of sources groups can monitor. Those previously reviewing stories manually could not survey the tens of thousands of sources these platforms aggregate. Those using automated tools such as Google Alerts were limited by the news sources and countries Google chose to cover, and return limited results each day. Additionally, activists find that local news sources have more coverage of specific cases of feminicide than national news outlets, and more details about individual cases. While other sources don't regularly index those, using Media Cloud we were able to add many.

Some limitations of our work can be found in the software architectures of news dissemination, and our own integrations, exposing the limits of using online news articles as data.For instance, while RSS is still a fairly common approach to syndicating news, it's use appears to be waning. This creates a weakness for the Media Cloud project at large, which relies on RSS for most of our content discovery. In addition, many hyperlocal sources don't have capacity to support robust websites with many different forms of syndication and publication. Sitemap-based scraping and link spidering offer alternative approaches. However, all those rely on the content being digitized and available at the story level. As we worked with activists specifically focused on Black and Indigenous media sources, we found that many of them use Facebook or other SMS groups, or are not distributed online, or are only available in PDF form. This presents a real barrier to ingest in the system we have designed, one overly borne by groups tracking killings of these already marginalized groups.

In addition, our project organizational and software architecture opens challenges to sustainability over time. For example, Media Cloud experienced significant downtime in late 2021 and early 2022. This created a dropoff in system usage, because it simply wasn't delivering content for more than 3 months. An outage like this has ethical and interpersonal repercussions for civic oriented projects that are built on robust community partnerships. More structurally, housing the project within academia raises longer-term questions of ownership, governance, and responsibility as our pilot project grows. This challenge isn't unique to community-centered computational journalism projects, but is shared by fields such as public interest technology and civic technology [1, 20, 21]. This is an active area of discussion and reflection for our team as we seek to contribute to scholarly research as well as support infrastructure for data activist's work.

6 CONCLUSION

News discovery has long played a role in data activists' efforts to catalog feminicide and gender-related killings in the regions they work in. In this case study we present the Data Against Feminicide: Email Alerts system, which seeks to support activist efforts by increasing the scope of news sources they can monitor, improving the relevance of results delivered, and supporting their existing data gathering workflows.

Computational journalism has much to offer in support of these kinds of collaborations that operate on news as data. Specifically, our query + classifier approach suggests a filtering mechanism that allows pathways for honoring the expertise of both domain experts and machine learning scientists, potentially reducing power disparities in co-design projects related to using news as data. Additionally, our project architecture and impact serves as an example of why open news archives such as Media Cloud are critical to support and maintain for the field to continue to develop [44].

This case study contributes a grounded example of computational journalism built in collaboration with, and in service of, activists working on critical human rights issues. Online news articles serve many purposes, and our field should work to engage and explore how our methods and tools can play a part in expanding their use in socially impactful areas outside of the news/media industry itself.

REFERENCES

[1] Roba Abbas, Jeremy Pitt, and Katina Michael. 2021. Socio-Technical Design for Public Interest Technology. *IEEE Transactions on Technology and Society*, 2, 2, (June 2021), 55–61. Conference Name: IEEE Transactions on Technology and Society. DOI: 10.1109/TTS.2021.3086260.

[2] Tanja Aitamurto and Anita Varma. 2018. The Constructive Role of Journalism. *Journalism Practice*, 12, 6, (July 2018), 695–713. Routledge. DOI: 10.1080/175127 86.2018.1473041.

[3] Camelia Bouzerdan and Jenifer Whitten-Woodring. 2018. Killings in Context: an Analysis of the News Framing of Femicide. *Human Rights Review*, 19, 2, (June 2018), 211–228. DOI: 10.1007/s12142-018-0497-3.

[4] M.S.T. C. 2014. From Accounting to Recounting: Esther Chávez Cano and the Articulation of Advocacy, Agency, and Justice on the US-Mexico Border. In *Mexican Public Intellectuals*. Literatures of the Americas. D.A. Castillo and S.A. Day, (Eds.) Palgrave Macmillan, New York.

[5] Kevin M. Carragee and Wim Roefs. 2004. The Neglect of Power in Recent Framing Research. *Journal of Communication*, 54, 2, 214–233. DOI: 10.1111/j.14 60-2466.2004.tb02625.x.

[6] Antonio Castillo. 2019. Reporting femicide: A failure of Latin American journalism. *Media Development*, 4, (Oct. 2019), 25–29. World Association for Christian Communication.

[7] Steven M. Chermak. 1994. Body count news: How crime is presented in the news media. *Justice Quarterly*, 11, 4, (Dec. 1994), 561–582. Routledge. DOI: 10.1080/07418829400092431.

[8] Marco Colagrossi, Claudio Deiana, Davide Dragone, Andrea Geraci, Ludovica Giua, and Elisa Iori. 2022. Intimate partner violence and help-seeking: The role of femicide news. SSRN Scholarly Paper 4005008. Social Science Research Network, Rochester, NY, (Jan. 2022). DOI: 10.2139/ssrn.4005008.

[9] Morgan Currie, Britt S Paris, Irene Pasquetto, and Jennifer Pierre. 2016. The conundrum of police officer-involved homicides: Counter-data in Los Angeles County. *Big Data & Society*, 3, 2, (Dec. 2016). SAGE Publications Ltd. DOI: 10.1177/2053951716663566.

[10] Catherine D'Ignazio, Isadora Cruxên, Maria de Los Angeles Martinez Cuba, Mariel García-Montes, Helena Suárez Val, Silvana Fumega, Harini Suresh, and Wonyoung So. 2022. Feminicide & Counterdata Production: Activist Efforts To Monitor And Challenge Gender-Related Violence. *under review*.

[11] Catherine D'Ignazio and Lauren F. Klein. 2020. *Data Feminism*. David Weinberger, (Ed.) *Strong Ideas*. MIT Press, Cambridge, MA, USA, (Mar. 2020). ISBN: 978-0-262-04400-4.

[12] Catherine D'Ignazio, Helena Suárez Val, Silvana Fumega, Harini Suresh, Isadora Cruxên, Wonyoung So, Ángeles Martinez, and Mariel García-Montes. 2020. Feminicide & Machine Learning: Detecting Gender-based Violence to Strengthen Civil Sector Activism. In *MD4SG'20*.

[13] 2018. Dados Sobre Feminicídio no Brasil. Tech. rep. Artigo 19. https://artigo19 .org/wp-content/blogs.dir/24/files/2018/03/Dados-Sobre-Feminic%C3%ADdi o-no-Brasil-.pdf.

[14] Craig Dalton and Jim Thatcher. 2014. What Does A Critical Data Studies Look Like, And Why Do We Care? *Society and Space*, 29.

[15] Nicholas Diakopoulos. 2020. Computational News Discovery: Towards Design Considerations for Editorial Orientation Algorithms in Journalism. *Digital Journalism*, 8, 7, (Aug. 2020), 945–967. Routledge. DOI: 10.1080/21670811.2020.1 736946.

[16] 2022. Fatal force: police shootings database. *Washington Post*.

[17] Kristen Gilchrist. 2010. "Newsworthy" Victims? *Feminist Media Studies*, 10, 4, (Dec. 2010), 373–390. Routledge. DOI: 10.1080/14680777.2010.514110.

[18] Lane Kirkland Gillespie, Tara N. Richards, Eugena M. Givens, and M. Dwayne Smith. 2013. Framing Deadly Domestic Violence: Why the Media's Spin Matters in Newspaper Coverage of Femicide. *Violence Against Women*, 19, 2, (Feb. 2013), 222–245. SAGE Publications Inc. DOI: 10.1177/1077801213476457.

[19] Alex Hanna. 2017. MPEDS: Automating the Generation of Protest Event Data. Tech. rep. SocArXiv, (Jan. 2017). DOI: 10.31235/osf.io/xuqmv.

[20] Nicolai Brodersen Hansen, Gwen Klerks, Maria Menendez Blanco, Laura Maye, Angelika Strohmayer, Martijn de Waal, and Ben Schouten. 2020. Making Civic Initiatives Last: Ecosystems, Technologies, Approaches and Challenges. In *Companion Publication of the 2020 ACM Designing Interactive Systems Conference* (DIS' 20 Companion). ACM, New York, NY, USA, (July 2020), 433–436. ISBN: 978-1-4503-7987-8. DOI: 10.1145/3393914.3395921.

[21] Youyang Hou and Cliff Lampe. 2017. Sustainable Hacking: Characteristics of the Design and Adoption of Civic Hacking Projects. In *Proceedings of the 8th International Conference on Communities and Technologies* (C&T '17). ACM, New York, NY, USA, (June 2017), 125–134. ISBN: 978-1-4503-4854-6. DOI: 10.1145/3083671.3083706.

[22] Tamil Kendall. 2020. A Synthesis of Evidence on the Collection and Use of Administrative Data on Violence against Women: Background Paper for the Development of Global Guidance. Tech. rep. UN Women, New York. Retrieved Apr. 19, 2022 from https://www.unwomen.org/sites/default/files/Headquarter s/Attachments/Sections/Library/Publications/2020/Synthesis-of-evidence-o n-collection-and-use-of-administrative-data-on-VAW-en.pdf.

[23] Marcela Lagarde y de los Ríos and Charlie Roberts. 2010. Preface: Feminist Keys for Understanding Feminicide: Theoretical, Political, and Legal Construction. In *Terrorizing Women: Feminicide in the Americas*. Duke University Press.

[24] Kalev Leetaru and Philip A. Schrodt. 2013. GDELT: Global Data on Events, Location and Tone, 1979-2012. In *International Studies Association meetings, San Francisco, April 2013*.

[25] Zhiwei Li, Bin Wang, Mingjing Li, and Wei-Ying Ma. 2005. A probabilistic model for retrospective news event detection. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (SIGIR '05). ACM, New York, NY, USA, (Aug. 2005), 106–113. ISBN: 978-1-59593-034-7. DOI: 10.1145/1076034.1076055.

[26] Wesley Lowery. 2020. A Reckoning Over Objectivity, Led by Black Journalists. en-US. *The New York Times*, (June 2020).

[27] Richard J. Lundman. 2003. The Newsworthiness and Selection Bias in News About Murder: Comparative and Relative Effects of Novelty and Race and Gender Typifications on Newspaper Coverage of Homicide. *Sociological Forum*, 18, 3, (Sept. 2003), 357–386. DOI: 10.1023/A:1025713518156.

[28] Neira Mellado and Carola Paz. 2013. *Online Newspaper Coverage of Femicide in Chile and the US*. Thesis. University of Kansas, (Aug. 2013). https://kuscholarw orks.ku.edu/handle/1808/12257.

[29] Amanda Meng and Carl DiSalvo. 2018. Grassroots resource mobilization through counter-data action. *Big Data & Society*, 5, 2, (July 2018), 2053951718796862. SAGE Publications Ltd. DOI: 10.1177/2053951718796862.

[30] Dominique A. Montiel Valle and Zelly Claire Martin. 2021. Entangled with the necropolis: a decolonial feminist analysis of femicide news coverage in Latin America. *Feminist Media Studies*, 0, 0, (Oct. 2021), 1–16. DOI: 10.1080/14680777 .2021.1988675.

[31] Julie Posetti and Alice Matthews. 2018. A short guide to the history of 'fake news' and disinformation. International Center for Journalists.

[32] Tara N. Richards, Lane Kirkland Gillespie, and M. Dwayne Smith. 2011. Exploring News Coverage of Femicide: Does Reporting the News Add Insult to Injury? *Feminist Criminology*, 6, 3, (July 2011), 178–202. SAGE Publications. DOI: 10.1177/1557085111409919.

[33] Hal Roberts et al. 2021. Media Cloud: Massive Open Source Collection of Global News on the Open Web. *Proceedings of the International AAAI Conference on Web and Social Media*, 15, (May 2021), 1034–1045.

[34] Charlotte Ryan, Mike Anastario, and Alfredo DaCunha. 2006. Changing Coverage of Domestic Violence Murders: A Longitudinal Experiment in Participatory Communication. *Journal of Interpersonal Violence*, 21, 2, (Feb. 2006), 209–228. SAGE Publications Inc. DOI: 10.1177/0886260505282285.

[35] Michael Schudson. 2008. Six or seven things news can do for democracy. In *Why Democracies Need an Unlovable Press*. Polity, (Nov. 2008). ISBN: 978-0-7456-4452-3.

[36] Sarah Stillman. 2007. 'The missing white girl syndrome': disappeared women and media activism. *Gender & Development*, 15, 3, (Nov. 2007), 491–502. Routledge. DOI: 10.1080/13552070701630665.

[37] Aixin Sun and Meishan Hu. 2011. Query-Guided Event Detection From News and Blog Streams. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 41, 5, (Sept. 2011), 834–839. Conference Name: IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans. DOI: 10.1109/TSMCA.2011.2157129.

[38] H. Suresh et al. 2022. Towards Intersectional Feminist and Participatory ML: A Case Study in Supporting Feminicide Counterdata Collection. In *FAccT'22*.

[39] Rae Taylor. 2009. Slain and Slandered: A Content Analysis of the Portrayal of Femicide in Crime News. *Homicide Studies*, 13, 1, (Feb. 2009), 21–49. SAGE Publications Inc. DOI: 10.1177/1088767908326679.

[40] Neil Thurman. 2019. Computational Journalism. In *The Handbook of Journalism Studies*. (2nd ed.). Karin Wahl-Jorgensen and Thomas Hanitzsch, (Eds.) Routledge, New York, NY, USA.

[41] UNODOC. 2019. Global Study on Homicide - Gender-related killing of women and girls. Tech. rep. United Nations Office on Drugs and Crime.

[42] Alexander Weiss and Steven M. Chermak. 1998. The News Value of African-American Victims: An Examination of the Media's Presentation of Homicide. *Journal of Crime and Justice*, 21, 2, (Jan. 1998), 71–88. DOI: 10.1080/0735648X.19 98.9721601.

[43] Ida B. Wells-Barnett. 1895. *The Red Record*.

[44] Ethan Zuckerman. 2020. The Case for Digital Public Infrastructure. Tech. rep. Knight First Amendment Institute at Columbia University.

[45] Ethan Zuckerman, J Nathan Matias, Rahul Bhargava, Fernando Bermejo, and Allan Ko. 2019. Whose Death Matters? A Quantitative Analysis of Media Attention to Deaths of Black Americans in Police Confrontations, 2013–2016. *International Journal of Communication*, 13, 27.