Analyzing Journalistic Reporting Reveals Blind Spots in Coverage of Anti–Data Center Activism

Tyler R. McKenzie, Sharaj Kunjar, Alyssa Hasegawa Smith, Milo Z. Trujillo, Brooke Foucault Welles, Robin A. Lange

Keywords: *local environmental activism, AI data centers, mixed-methods analysis, news framing*

Introduction

The rise of generative AI, especially marked by the release of OpenAI's ChatGPT in 2023, has led to an explosion in the demand for dedicated spaces to house the hardware that facilitates such large-scale computing, i.e., data centers (1,2). However, in addition to their emissions (3), these facilities consume enormous amounts of electricity and water (4). As such, in the United States alone, data centers are estimated to make up to 12% of the country's energy demand by 2028 (5). Moreover, these unwanted environmental impacts are not spread equally through the American public. Communities that have historically faced racism, poverty and governmental neglect are prone to be chosen as sites for data center development (6–8).

Across the country, protests have risen as a direct response to data center driven pollution (8–10). There are at least 126 activist groups across 28 states who are actively tracking, organizing against, or obstructing data center construction. 64 billion in data center projects have been blocked or delayed in the US over the past two years due to this activism (11). Regions of intense opposition include Virginia, Texas, Pennsylvania, Arizona, Oregon, Indiana, Missouri, California, and Georgia.

Many data center protests have been organized by sustained grassroots movements and local activism among the communities living in affected areas (11). Everyday organizing, community meetings and bottom-up pressure on policy makers have shown to be effective modes of voicing dissent (12). Where traditional modes of political deliberation have largely excluded local communities from having a say in policies that dictate their quality of life, such local movements provide communities with the opportunity to express their broader dissatisfaction with large conglomerates and local governments (13,14).

Despite local activism's early successes in organizing protests, historical media representations of such movements have been overlooked and minimized. First, local movements surrounding environmental issues are often omitted from mainstream coverage (15), rendering a vast majority of these issues unseen and unacknowledged by the public. Mass media sets the public agenda by shaping public perceptions regarding the importance of different issues (16). However, this influence is uneven, as poorer communities are disproportionately located in "news deserts", i.e. areas that generally receive lesser news coverage (17). Thus, such an omission of coverage contributes to further marginalization of vulnerable communities. Second, when local movements are covered by mainstream media, their framing tends to delegitimize resistance through a deliberate omission of information contextualizing the dissent, and demonizes protestors through an over-emphasis of the disruption caused by protests (18,19). Additionally, political groups that are deviant from the status quo tend to be portrayed in a dehumanizing language (20). Thus, even when local movements are elevated in public consciousness through media coverage, their framing (21,22) encourages consumers to disparage dissenters and decentralize the movement.

In this study, we investigated the news coverage pertaining to local activism surrounding data centers. Motivated by prior work highlighting the sparsity and problematic framing in mainstream coverage surrounding similar issues, we asked:

RQ1: What volume of coverage does local activism receive in news articles about data centers?

RQ2: When local activism surrounding data centers is covered, how is it framed?

RQ3: How does the aforementioned framing depend on the community that the activism is based around?

We used a mixed-methods research pipeline combining topic modelling, supervised text classification, and qualitative content analysis to systematically analyze over 30,000 news articles about data centers, published over a 2.5-year timeline by regional news outlets in the United States. Thus, we offer insights about the relative invisibility of local activism regarding the environmental impacts of data centers in news coverage, as well as the differences in news coverage of this activism for primarily African American and more affluent majority-white communities. Overall, our study illuminates patterns in journalistic coverage of data center development: namely, that they tend to focus on its economic potential while ignoring the potential downsides.

Data and Methods

We obtained a dataset of 102,547 news articles about data centers from MediaCloud (23) an online open-source database that maps news coverage of current events. These articles were published between January 1, 2023 and May 29, 2025 by 143 national and 1284 regional news outlets from the United States. Next, we scraped and curated the texts for 71,497 of these articles via the Python newspaper4k package (24). We then filtered the dataset to only include articles published by regional news outlets.

We utilized a two-pronged approach to address RQ1. First, we ran a BERTopic model (25) on the MediaCloud dataset to discover, at a high level, what was predominantly discussed in these articles. Second, due to the absence of local activism among the discovered topics, we developed a relevance codebook to explicitly determine the number of articles in the regional news corpus that discussed local activism against data centers. We tested the validity of this codebook by using it to code a random subset of 200 articles. After two rounds of coding, we reached sufficient inter-coder agreement, with a Cohen's κ of 0.83. Eight of the 200 articles were deemed relevant. We extracted keywords from our ground truth articles, between 1 and 5 tokens long, to build a keyword-based relevance classifier. Using the nltk Python package, we generated n-grams for each of the articles. Our classifier counted 1) the total number of relevant keywords and 2) the unique number of keywords for each article. We then used an ROC curve to determine the optimal threshold of unique and total keywords to classify articles as relevant, based on our corpus of 200 labeled articles. We applied this threshold to the entire regional news corpus of 39,836 articles and identified 3,026 relevant data center articles.

Due to the relative scarcity of articles about local activism against data centers available in the MediaCloud dataset, we used Google search engine result pages (SERPs) to obtain a more focused set of articles that were more likely to mimic what the average person would see when seeking out information on this topic. Using county classifications obtained from the American Communities Project (26), we listed ten counties from each of two community types – the African American South and the Exurbs. The counties were selected based on the presence of at least one protest. For each of the 20 total counties, we used the SerpAPI service (27) to pull 100 Google Search results for the phrase "$COUNTY data center activism" searched from that county. We scraped all available results using newspaper4k and ran a BERTopic model on this content. Next, we linked the scraped search engine results' top-level domains with known news outlet domains from MediaCloud; this filtering produced a dataset of results that were known to be news articles.

We addressed RQ2 by performing a codebook-based framing analysis on a subset of 100 (randomly chosen) articles from the SERP dataset. We ensured that the dataset contained 50 articles from each community type. Due to its widespread adoption to study issue-generic framing surrounding politics (28), we co-opted the 5-frame typology put forth by Semetko and Valkenburg (29), as described in **Table 1**. This analysis was performed by two manual annotators through three rounds of iterative coding, which yielded an average Cohen's $\kappa = 0.65$. Finally, to address RQ3, we performed a Fisher's exact test (two-tailed) on the counts of different frames. This test allowed us to evaluate if there were any statistically significant deviations in news framing surrounding data center activism based on the type of community that they are covering.

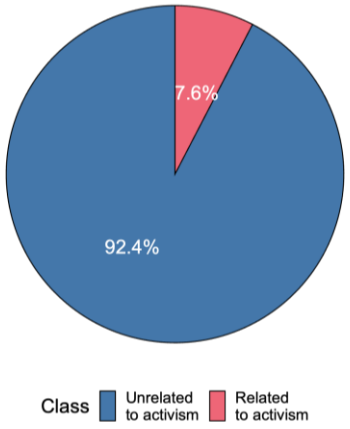
Frame	κ	Description	Example
Conflict	0.69	Emphasizes conflict between individuals, groups or institutions	"The community surrounding [data center] facility, which suffers from higher cancer and asthma rates... has been outspoken against the company."
Economic consequences	0.70	Emphasizes the economic consequences of data centers for the communities	""It represents a major investment for the state, creating new job opportunities both during construction and in long-term operations..."
Human interest	0.71	Emphasizes a humanistic and emotional angle regarding affected communities	".. .another resident living nearby, shares similar concerns... 'We're dealing with noise, we are dealing with pollution...'"
Morality	0.48	Emphasizes religious tenets or moral prescriptions	"That's not right. I call it zoning fraud."
Responsibility	0.66	Attributes responsibility for cause/solution to an individual or group	"I will not see the total impact of what we are doing today," ...and we have to take responsibility for that."

Table 1: Brief description of the framing typology, as adapted from Semetko and Valkenburg 2000 (29)

Results

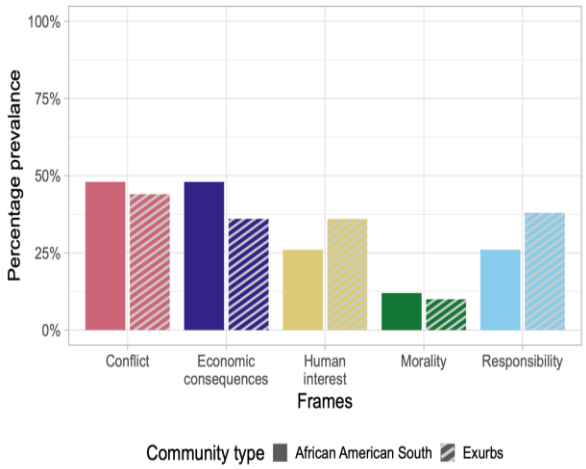
Volume and Nature of Coverage

We find that coverage of local activism in news articles about data centers is likely very low in volume relative to the total volume of news about data centers (**Figure 1**). In our test set of 200 articles, the classifier's positive predictions are moderately reliable (precision ≈27%), while its negative predictions are highly trustworthy (NPV ≈99%), reflecting the scarcity of relevant articles. Using this classifier on the full MediaCloud dataset, 3,026 of 39,836 articles (7.6%) were labeled as relevant; given the moderate precision, this is likely an overestimate of the true number of relevant articles. The BERTopic model of the dataset produced three topics (**Table 2**), two of which were related to the stock market and one of which was related to computational infrastructure.



Topic	Top Ten Keywords
Topic 1	stock, share price, shares, company stock, price, price earnings, stock price, stocks, nasdaq, dividend
Topic 2	data centers, data center, infrastructure, cloud, demand, microsoft, electricity, data, industry, technology
Topic 3	company stock, stock, nyse, nasdaq, market, buy rating, price, earnings, stocks, price target

Figure 1 (left): An upper limit for the proportion of news articles covering data centers that were relevant to activism reveals scant media attention towards the issue; **Table 2 (right):** Topics discovered by BERTopic in the preliminary MediaCloud dataset reveal economy and technology as issues of interest surrounding data centers.



Frames	p-value
Conflict	0.841
Economic consequences	0.311
Human interest	0.387
Morality	1
Responsibility	0.284

Figure 2 (left): Percentage prevalence of different frames in a random subset of a 100 news articles about local activism surrounding data centers, stratified by the community type they cover (solid bars denote African American South communities and striped bars denote exurbs). **Table 3 (right):** Fisher's test to evaluate differences in coverage of different frames based according to community type. The lack of high p-values (>0.05) indicate that there were no significant differences in news framing by community type.

When data center activism was covered, news articles further emphasized conflict and economic consequences (**Figure 2**). The annotators noted that the coverage of economic consequences was also one-sided, where the benefits (job opportunities) were highlighted over pitfalls (tax-breaks to corporations). As revealed by the relative scarcity of human-interest and morality frames, articles did not cover stories from the communities involved in these protests. These patterns remained consistent across both the community types we studied (**Table 3**). While we observed a higher coverage of human-interest stories and morality frames in exurbs, these differences were not statistically significant.

Discussion and conclusion

Despite the widespread activism protesting data centers throughout the United States, it is rarely the focus in news media coverage. Only 7.6% of articles were flagged as relevant to activism. Our topic model of a news media corpus found that the focus instead is on economic and infrastructural impacts. This reiterates earlier findings (15) , that environmental activism is often ignored by the media.

Economic consequences and conflict are by far the most prevalent frames, and human centered frames such as morality, responsibility, and human interest trailed in comparison. Economic consequences frames were focused on job creation and tax benefits data centers bring to communities. They often omit the potential economic pitfalls, such as the temporary nature of employment within data center development and the massive tax breaks allotted to corporations. In tandem, conflict framing emphasizes tension between smaller activist groups and large corporations or governmental institutions, which can foster public cynicism (31). Together, this demonstrates how reporting prioritizes spectacle and financial gains.

Conversely, responsibility frames were sparse in our dataset, despite the strong oversight data centers require to be sustainable. For example, data centers require significant water usage (4) and can often emit toxic chemicals (3). When included, responsibility frames ask for accountability from local governments, or for transparency from corporations. Similarly, human interest frames focus on emotional response from community members. This frame highlights concerns or frustrations about data centers expressed through firsthand experiences. The dominance of conflict and economic frames, without the balance of responsibility, morality, or human interest, leaves coverage skewed toward corporate and financial narratives while minimizing accountability and erasing the lived experiences of affected communities.

When comparing generic media frames (conflict, responsibility, morality, economic consequences, and human interest) in wealthy, white exurbs versus the working class African American south, we found that similar distribution of use across all frames. By unveiling the lack of human-centered frames persisting across both wealthy white and working-class Black communities, our findings suggest that this imbalance is not tied to race or class but is a broader pattern in coverage. While this uniformity in framing patterns across racial lines may seem relieving, this journalistic style fails to recognize that risk is not evenly distributed between these communities. The consequences of such problematic reporting would be disproportionately felt by Black communities, since they are more likely to face targeted environmental injustices (6–8).

Data center proliferation has become a major issue affecting communities across the United States because of their negative environmental impacts. Faced with these consequences, communities have turned to activism as a form of resistance. Local activism is a community response to being excluded from decision making about their own communities by larger institutions. Yet, media reporting does not accurately represent these efforts. In doing so, the press obscures community voices and minimizes their demands for justice. Without coverage that engages in themes of responsibility, morality, or human interest, communities lose an important channel to bring attention to their concerns. This is especially true in Black communities, who have a long history of environmental exploitation. Taken together, this lack of representative reporting indicates that corporations shape coverage more than the community identity itself.

References

1. Bashir N, Donti P, Cuff J, Sroka S, Ilic M, Sze V, et al. The Climate and Sustainability Implications of Generative AI. MIT Explor Gener AI [Internet]. 2024 Mar 27 [cited 2025 Aug 17]; Available from: https://mit-genai.pubpub.org/pub/8ulgrckc/release/2

2. Liu Y, Du H, Niyato D, Kang J, Xiong Z, Wen Y, et al. Generative AI in Data Center Networking: Fundamentals, Perspectives, and Case Study. IEEE Netw. 2025;1–1.

3. IEA [Internet]. [cited 2025 July 18]. Data centres & networks. Available from: https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks

4. The Joyce Foundation [Internet]. 2024 [cited 2025 July 18]. Data Centers and Groundwater Usage. Available from: https://www.joycefdn.org/news/data-centers-and-groundwater-usage

5. Shehabi A, Smith S, Sartor D, Brown R, Herrlin M, Koomey J, et al. United States Data Center Energy Usage Report [Internet]. 2016 June [cited 2025 Aug 17] p. LBNL--1005775, 1372902. Report No.: LBNL--1005775, 1372902. Available from: http://www.osti.gov/servlets/purl/1372902/

6. Edwards D, Cooper ZGT, Hogan M. The making of critical data center studies. Converg Int J Res New Media Technol. 2025 Apr;31(2):429–46.

7. Gonzalez S. Cloud Ecologies: An Environmental Ethnography of Data Centers [Internet] [Thesis]. Massachusetts Institute of Technology; 2023 [cited 2025 Aug 24]. Available from: https://dspace.mit.edu/handle/1721.1/154209

8. Pasek A. Getting Into Fights With Data Centers: Or, a Modest Proposal for Reframing the Climate Politics of ICT [Internet]. Experimental Methods and Media Lab, Trent University, Peterborough, Ontario; 2023 [cited 2025 Aug 24]. Available from: https://emmlab.info/Resources_page/Data%20Center%20Fights_digital.pdf

9. Cappella N. Data centre protests are on the rise, but are they effective? [Internet]. Techerati. 2023 [cited 2025 Aug 24]. Available from: https://www.techerati.com/news-hub/data-centre-protests/

10. Goodman E. Big Tech's New Headache: Data Centre Activism Flourishes Across the World - Media@LSE [Internet]. Media@LSE - Promoting critical research into the vital role of media and communications in contemporary society. 2022 [cited 2025 Aug 24]. Available from: https://blogs.lse.ac.uk/medialse/2022/11/02/big-techs-new-headache-data-centre-activism-flourishes-across-the-world/

11. Data Center Watch [Internet]. [cited 2025 Sept 5]. $64 billion of data center projects have been blocked or delayed amid local opposition. Available from: https://www.datacenterwatch.org/report

12. Riverstone-Newell L. Bottom-Up Activism: A Local Political Strategy for Higher Policy Change. Publius J Fed. 2012 July 1;42(3):401–21.

13. Gilster ME. Putting Activism in Its Place: The Neighborhood Context of Participation in Neighborhood-Focused Activism. J Urban Aff. 2014 Feb 1;36(1):33–50.

14. Dahlberg-Grundberg M, Örestig J. Extending the local: activist types and forms of social media use in the case of an anti-mining struggle. Soc Mov Stud. 2017 May 4;16(3):309–22.

15. Moore EE, Lanthorn KR. Framing Disaster: News Media Coverage of Environmental Justice.

16. McCOMBS ME, Shaw DL. The Agenda-Setting Function of Mass Media. 2024;

17. Usher N. News for the Rich, White, and Blue: How Place and Power Distort American Journalism [Internet]. Columbia University Press; 2021 [cited 2025 Sept 4]. Available from: https://www.jstor.org/stable/10.7312/ushe18466

18. McLeod DM. News Coverage and Social Protest: How the Media's Protect Paradigm Exacerbates Social Conflict. 2007;2007.

19. Dilliplane S, Smith B, Lincoln L. Local news reporting during a year of racial justice protests. 2020;

20. Shoemaker PJ. Media Treatment of Deviant Political Groups. Journal Q. 1984 Mar 1;61(1):66–82.

21. University of California Press [Internet]. [cited 2025 Aug 18]. The Whole World Is Watching by Todd Gitlin - Paper. Available from: https://www.ucpress.edu/books/the-whole-world-is-watching/paper

22. Entman RM. Framing: Toward Clarification of a Fractured Paradigm. J Commun. 1993 Dec 1;43(4):51–8.

23. Roberts H, Bhargava R, Valiukas L, Jen D, Malik MM, Bishop CS, et al. Media Cloud: Massive Open Source Collection of Global News on the Open Web. Proc Int AAAI Conf Web Soc Media. 2021 May 22;15:1034–45.

24. Paraschiv A. newspaper4k: Simplified python article discovery & extraction. [Internet]. [cited 2024 Feb 26]. Available from: https://github.com/AndyTheFactory/newspaper4k

25. Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure [Internet]. arXiv; 2022 [cited 2025 Aug 19]. Available from: http://arxiv.org/abs/2203.05794

26. Chinni D, Pinkus A. Health and Place in America, 2018: 2018;

27. Reitz K. serpapi: The official Python client for SerpApi.com. [Internet]. 2023 [cited 2025 Aug 19]. Available from: https://github.com/serpapi/serpapi-python

28. Vreese CH. News framing: Theory and typology. Inf Des J. 2005 Apr 18;13(1):51–62.

29. Semetko HA, Valkenburg PMV. Framing European politics: A Content Analysis of Press and Television News. J Commun. 2000 June 1;50(2):93–109.

30. News Deserts and Ghost Newspapers: Will Local News Survive? - [Internet]. The Expanding News Desert. [cited 2025 Sept 5]. Available from: https://www.usnewsdeserts.com/reports/news-deserts-and-ghost-newspapers-will-local-news-survive/

31. Elenbaas M, De Vreese CH. The Effects of Strategic News on Political Cynicism and Vote Choice Among Young Voters. J Commun. 2008 Sept;58(3):550–67.