Notes on methodology

Appendix to "Missing Ingredients: How Agriculture and Diet Get Overlooked in Media Coverage of Climate Change"

Original landscape analysis

Basis of approach

We wanted to be able to draw robust conclusions about the topics that the U.S. news media as a whole emphasizes and neglects when covering climate change. With this ambitious goal, we adjusted three factors for the search that would collect our sample: the thoroughness of our search terms, the publications whose articles we would search for, and the time period over which we would be searching.

Our search terms were intended to capture every article relevant to our topic of animal agriculture in climate coverage. This approach prioritized avoiding type II errors of failing to include relevant articles. We also drafted and tested our search terms to check if irrelevant articles were being returned, then revised the query language to avoid type I errors of failing to exclude those irrelevant articles.

We limited our sources to high-readership outlets because of their greater impact on the media landscape and their greater frequency of publication. The time span we chose, a three-year period from July 2022 through June 2025, had not been covered by previous research and kept the number of articles to a manageable quantity.

This design was informed by two 2023 analyses that we reviewed as we prepared our own. The scope of our research was intended to strike a balance between the exhaustive approach of the Madre Brava report,[1] which would have made filtering stories and comparing outlets difficult, and the targeted approach of the Faunalytics and Sentient Media report,[2] which would have limited our ability to draw conclusions about the broader media landscape. We also sought to draw our primary conclusions in relative terms like percentages instead of absolute terms like counts.

Gathering articles

We designed our search for articles in the form of three nested searches, an approach inspired by a method used in the Madre Brava report. The query language we composed also integrated search terms from studies by Kristiansen et al.[3] and Madre Brava,[4] as well as a stock Factiva search term, `ns=GCLIMT`, for topics related to global climate change.

---

[1] Madre Brava, *Media Analysis of Industrial Meat and Climate Change* (2023).

[2] Faunalytics and Jenny Splitter, "Animal Agriculture Is The Missing Piece In Climate Change Media Coverage," preprint, July 5, 2023, https://doi.org/10.31234/osf.io/np4ed.

[3] Silje Kristiansen et al., "Animal Agriculture and Climate Change in the US and UK Elite Media: Volume, Responsibilities, Causes and Solutions," *Environmental Communication* 15, no. 2 (2021): 153–72, https://doi.org/10.1080/17524032.2020.1805344.

[4] Madre Brava, *Media Analysis of Industrial Meat and Climate Change*.

We created a baseline set of conditions we wanted met in all the articles we collected: language (English), region (published in the United States), and time span (a three-year period from July 1, 2022, through June 30, 2025).

We further filtered for full-length articles by requiring that the results be greater than 250 words in length. This eliminated items like photo captions, corrections, and letters to the editor that were picked up in earlier versions of the search.

The publications from which we searched for articles were drawn from Factiva's `tmnbus` list of 41 major U.S. news and business publications. Because Factiva did not return complete results for some outlets when using the `tmnbus` search term `rst=tmnbus`, we instead listed out the codes of the outlets: `rst=(sfabc or sfamb or … sfwp)` for ABC Network, *American Banker*, and so on through to the *Washington Post*. Of those 41 outlets, 37 had published articles that met our criteria. The ones that did not were likely filtered out by our decision to restrict our search to outlets publishing in the United States, while the Factiva list included international outlets that are important in the United States, e.g. the BBC.

These segments together comprised the baseline for our sampling. To conduct our three searches, we appended a segment for climate coverage to the baseline, then to that search appended a segment for animal agriculture or meat, then to that search appended a segment for dietary shift, such that each was narrower than the previous.

The first query, the one searching for coverage of climate change, returned a set of articles we will call "`cli`". The second query, the one that restricted the first to articles that also mentioned animal agriculture or meat, returned a set of articles we will call "`meat`". The third query, the one that restricted the second to articles that also mentioned dietary shift, returned a set of articles we will call "`diet`".

Thus, the articles mentioning dietary shift (`diet`) constituted a subset within the set of articles mentioning meat or animal agriculture (`meat`), which itself constituted a subset within the set of articles covering the climate (`cli`).

The full text and modular construction of the queries can be found in the Github repository associated with this project[5] as the file `Factiva query text.sql` (although it is not SQL code—that file extension was appended to the plain text file for clarity in formatting). The search process can be summarized as:

- Filter for language, date range, word count, region
- Filter for selected sources
- Exclude non-article items, irrelevant article types
- Search for topic of climate change
    - Search for mentions of meat consumption
        - Search for mentions of dietary change

Our searches returned 10,696 total articles, with smaller numbers of articles within that sample constituting the `meat` and `diet` subsets. From the Factiva results for each of the three queries, we exported three CSV files: the first listing the number of articles from each publication, the second listing how many times each mentioned industry was mentioned, and the third listing how many times each mentioned

[5] Alexandra Tey, *Teyalex/an-Ag*, R, August 15, 2025, released November 18, 2025, https://github.com/teyalex/an-ag.

organization was mentioned. Those nine CSVs can be found in the subfolder `Exported Factiva statistics` within the folder `Original analyses` in the Github repository for this project.[6]

We also downloaded the text of the articles, 100 articles at a time, as rich text files for further analysis. We used a Python script to split each 100-article rich text file into plain text files of each individual article, labeled in the file name with the article's date. This produced plain text files of 8,086 articles.

We attribute the discrepancy between this figure and the 10,696 total from the exported Factiva statistics to the practice of smaller newspapers republishing articles licensed from larger newspapers. The duplicate articles may have been counted in the figures exported from Factiva; we preserved those counts on the rationale that republication reaches new readers the same way original articles do. However, for our textual analysis, we wanted to include each article only once. If we did miss some articles, it would have been few enough as to not affect our analysis.

Processing in R

We used the open-source programming language R in the RStudio integrated development environment to process, analyze, and visualize our data. Our code is published in the Github repository for this project.[7]

After reading the CSV files into R, we used this data to create data frames for the number of articles from or about each media outlet, industry, and company in each of the three scopes. We created a data frame `sources` with article counts by publication, a data frame `industries` with mention counts by industry, and a data frame `companies` with mention counts by organization or corporation. These three data frames—and rearranged versions of them—were used to compute summary statistics and plot trends using tools in base R and in the packages `dplyr`, `ggplot2`, `stats`, `stringr`, `tidyr`, and `utils` in the `tidyverse` package.[8] `ggtext` was used to handle text styling in the plot captions.[9]

We imported the text files into R with `readtext`, then processed them as a corpus using `quanteda` and packages in the `tidyverse`.[10, 11] We also used the `quanteda` add-on `quanteda.sentiment` for sentiment analysis. Word cloud plots were handled with `ggwordcloud` in conjunction with `ggplot2`.[12] Statistical tests were performed with `effectsize`.[13] When processing tokens (i.e. words), we excluded 241 common words using the `stopwords` package and by filtering out words we considered not meaningful for our analysis (e.g. "said", "may," "next").[14] We also removed "climate," "change," and "global warming"

---

[6] Tey, *Teyalex/an-Ag*.

[7] Tey, *Teyalex/an-Ag*.

[8] Hadley Wickham et al., "Welcome to the Tidyverse," *Journal of Open Source Software* 4, no. 43 (2019): 1686, https://doi.org/10.21105/joss.01686.

[9] Claus O. Wilke and Brenton M. Wiernik, *Ggtext: Improved Text Rendering Support for "Ggplot2,"* v. 0.1.2, released September 16, 2022, https://cran.r-project.org/web/packages/ggtext/index.html.

[10] Kenneth Benoit et al., *Readtext: Import and Handling for Plain and Formatted Text Files*, v. 0.92.1, released July 27, 2025, https://cran.r-project.org/web/packages/readtext/index.html.

[11] Kenneth Benoit et al., "Quanteda: An R Package for the Quantitative Analysis of Textual Data," *Journal of Open Source Software* 3, no. 30 (2018): 774, https://doi.org/10.21105/joss.00774.

[12] Erwan Le Pennec and Kamil Slowikowski, *Ggwordcloud: A Word Cloud Geom for "Ggplot2,"* v. 0.6.2, released May 30, 2024, https://cran.r-project.org/web/packages/ggwordcloud/index.html.

[13] Mattan S. Ben-Shachar et al., "Effectsize: Estimation of Effect Size Indices and Standardized Parameters," *Journal of Open Source Software* 5, no. 56 (2020): 2815, https://doi.org/10.21105/joss.02815.

[14] Kenneth Benoit et al., *Stopwords: Multilingual Stopword Lists*, v. 2.3, released October 28, 2021, https://cran.r-project.org/web/packages/stopwords/index.html.

because it would not have been useful to find that those words were often used in coverage of climate change.

Using methods from Faunalytics and Sentient Media report

Gathering articles

In their 2023 collaboration, Faunalytics and Sentient Media selected 10 major U.S. media outlets as a sample of the broader media landscape.[15] From each of these 10 outlets, the 100 most recent articles with "climate" in the headline were selected, yielding 1,000 articles that served as the corpus for the analysis. The R programming language was used to search each of the 1,000 articles for keywords associated with 10 "climate-related themes," mostly anthropogenic causes of climate change. Articles containing a keyword were categorized as including its respective theme.

We searched the Factiva database with standardized queries that replicated the original sampling criteria: the 100 most recent articles from each source with "climate" in the headline. The Factiva license we used did not include access to the *Los Angeles Times*, so we were only able to analyze 9 of the 10 original outlets. Despite this limitation, the access to other paywalled sources, the highly customizable query system, and the consistent formatting of the exported rich text files made this method more than efficient and powerful enough at selecting the articles we wanted and excluding those we did not.

As in the original research, we excluded duplicates, miscategorized articles, and non-article content whenever possible. This included filtering out updates, repeats, and corrections from the Reuters wire service that appeared as duplicate articles. As discussed in the section detailing our original methods, we also filtered out items with 250 or fewer words. We further manually excluded articles that used "climate" in a figurative sense, which were common in the *New York Post* and *Chicago Tribune*. As a result, we only had 88 articles from the *Tribune*, and many of the *Post* articles had not been published recently.

The resulting pool of articles were downloaded as rich text files, organized into directories corresponding to the outlets in which they were published, and renamed in a standardized format. We consider the resulting sample of 888 articles to be sufficient for our purpose of comparison with the original analysis.

Processing in R

We used the Faunalytics and Sentient Media code to import the text files into R, then scanned the text to detect topics mentioned. This resulting categorization allowed us to generate summary statistics and export data frames for further analysis.

The methods and code used in the 2023 Faunalytics and Sentient Media report, including the sections we adapted, can be found in the Open Science Foundation repository for that project.[16] Our adapted versions can be found in the Github repository for this project in the folder `Repeating Faunalytics + Sentient analysis` as `1. Categorizing articles.R` and `2. Faunalytics + Sentient analysis.R`.

---

[15] Faunalytics and Splitter, "Animal Agriculture Is The Missing Piece In Climate Change Media Coverage."

[16] Faunalytics and Jenny Splitter, *Data & R Code*, R, released May 23, 2023, https://osf.io/xmwgz/files.