

# News Production and Strategic Citation: Evidence from 17,000 Climate Change Articles

Hye Young You  
Princeton University

Sylvan Zheng  
New York University

# The Role of Journalists in Media Slant

- Historical “objectivity norm”: journalists provide equal coverage to “both sides” (Schudson, 2001).

# The Role of Journalists in Media Slant

- Historical “objectivity norm”: journalists provide equal coverage to “both sides” (Schudson, 2001).
- But – increasing backlash from readers and journalists especially since the Trump presidency (Sullivan, 2022; Bennett, 2023)

OPINION  
LETTERS

## Outcry Over Tom Cotton's Call for Troops to Quell Unrest

Sampling the responses to Senator Cotton's suggestion to use U.S. troops to suppress the protests and The Times's decision to publish it.

June 3, 2020



U.S. Senator Tom Cotton calls for “an overwhelming show of force.” Pool photo by Andrew Harnik

### *New York Times Says Senator's Op-Ed Did Not Meet Standards*

After a staff uproar, The Times says the editing process was “rushed.” Senator Tom Cotton’s “Send In the Troops” essay is now under review.

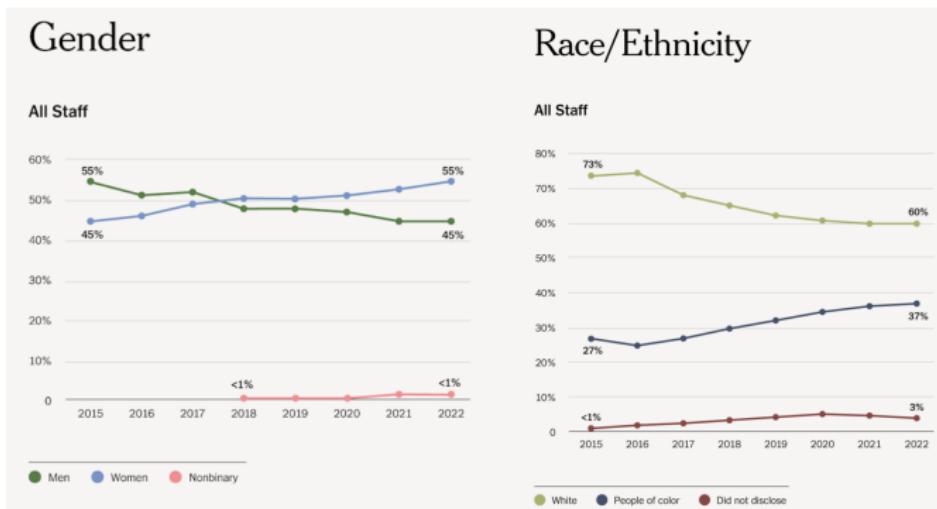
[Share full article](#) [Email](#) [Bookmark](#)



The New York Times said in a statement on Thursday that “a rushed editorial process led to the publication of an Op-Ed” by Senator Tom Cotton. Jeenah Moon for The New York Times

# The Role of Journalists in Media Slant

- Historical “objectivity norm”: journalists provide equal coverage to “both sides” (Schudson, 2001).
- But – increasing backlash from readers and journalists especially since the Trump presidency (Sullivan, 2022; Bennett, 2023)
- Demographics of journalists changing rapidly; younger, more liberal, and more educated (Forman-Katz and Jurkowitz, 2022)



# The Role of Journalists in Media Slant

- Historical “objectivity norm”: journalists provide equal coverage to “both sides” (Schudson, 2001).
- But – increasing backlash from readers and journalists especially since the Trump presidency (Sullivan, 2022; Bennett, 2023)
- Demographics of journalists changing rapidly; younger, more liberal, and more educated (Forman-Katz and Jurkowitz, 2022)
- What role do individual journalists play in driving media slant?

# Trump Effect



NATIONAL

## How Donald Trump's Presidency Has Changed The Media

More than 100 newspapers will publish editorials decrying Trump's anti-press rhetoric

# Trump Effect



NATIONAL

## How Donald Trump's Presidency Has Changed The Media

More than 100 newspapers will publish editorials decrying Trump's anti-press rhetoric

MAGAZINE

### If Trump Runs Again, Do Not Cover Him the Same Way: A Journalist's Manifesto

I believed in traditional reporting, but Trump changed me — and it should change the rest of the media too.

## Changing Journalist Demographics

Though they might have lacked deep or varied reporting backgrounds, some of the *Times*'s new hires... were brilliant and fiercely honest... But, critically, the *Times* abandoned its practice of acculturation.... Many new hires who never spent time in the streets went straight into senior writing and editing roles. ... All these recruits arrived with their own notions of the purpose of the *Times*

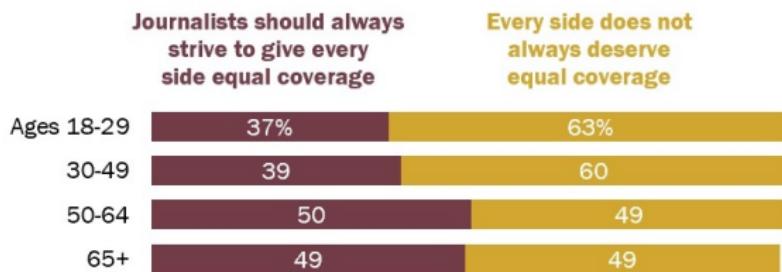
- James Bennett, *When the New York Times lost its way* (2023)

# Activist Journalists

Younger, less experienced, more highly educated journalists support an activist view of the press (Johnstone, 1976; Forman-Katz and Jurkowitz, 2022).

## Journalists' views about equal coverage vary by age, platform and political leaning of outlet's audience

% of U.S. journalists who say ...



Among U.S. journalists who say the audience of the news organization they work for (is) \_\_\_ politically



## This Project

- How do political and individual journalist factors affect media slant and balance?
- Our setting: climate change articles published in major U.S. newspapers 2012-2022

# This Project

- How do political and individual journalist factors affect media slant and balance?
- Our setting: climate change articles published in major U.S. newspapers 2012-2022
  1. Climate change is a highly polarized topic

# This Project

- How do political and individual journalist factors affect media slant and balance?
- Our setting: climate change articles published in major U.S. newspapers 2012-2022
  1. Climate change is a highly polarized topic
  2. Many groups with clear positions (fossil fuel companies and environmental advocacy groups)

# This Project

- How do political and individual journalist factors affect media slant and balance?
- Our setting: climate change articles published in major U.S. newspapers 2012-2022
  1. Climate change is a highly polarized topic
  2. News articles use citations from many groups with clear positions (fossil fuel companies and environmental advocacy groups)
- **Measure media slant using the sources that journalists cite**

# Preview: Using Sources to Estimate Slant

POLITICS | Biden to Cancel Keystone XL Pipeline in Inauguration Day Executive Order

In their [Supreme Court brief](#), lawyers for TC Energy said the decision would cost jobs unless the justices intervened.

"TC Energy would not be able to employ a majority of the approximately 1,500 unionized construction workers, and approximately 300 administrative, inspection and management personnel it would otherwise employ for pipeline construction in 2020," the company's lawyers said. "The loss of so many high-paying jobs, along with the loss of the secondary employment and economic opportunities that construction activities would otherwise create in local communities, would be particularly harmful in the current distressed economy."

The Washington Post  
*Democracy Dies in Darkness*

"Today's ruling makes it clear once and for all that it's time for TransCanada to give up on their Keystone XL pipe dream," said Sierra Club Senior Attorney Doug Hayes in a statement. The lawsuit prompting Thursday's order was brought by a collection of opponents, including the [indigenous Environmental Network](#) and the [Northern Plains Resource Council](#), a conservation coalition based in Montana.

## This Project

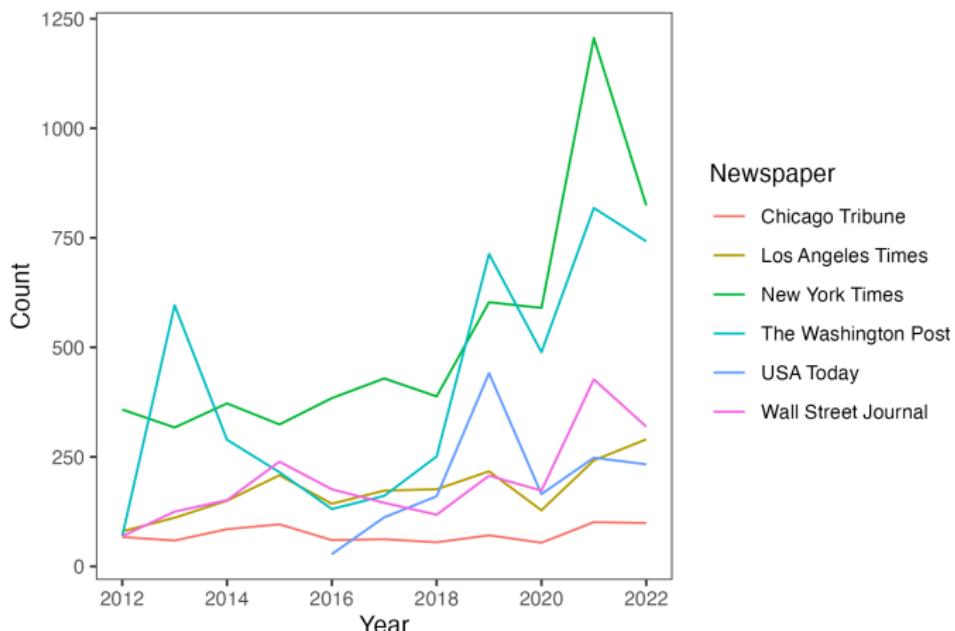
1. Collect data on journalist characteristics, source citations, and use these to create measure of article slant
2. How are journalist-level characteristics correlated with the ideological slant of the sources cited?
3. How did the Trump presidency change citation slant patterns?

## News Articles on Climate Change

1. Start with the 6 largest US daily newspapers: NYT, WSJ, WaPo, USA Today, LA Times, and the Chicago Tribune
2. Combine keyword search and machine learning techniques to select 16,532 articles about climate change

# News Articles on Climate Change

1. Start with the 6 largest US daily newspapers: NYT, WSJ, WaPo, USA Today, LA Times, and the Chicago Tribune
2. Combine keyword search and machine learning techniques to select 16,532 articles about climate change



## Human powered extraction... slow and expensive

- 120 news articles on climate change and abortion → 557 sources extracted by two RAs: took 6-months and cost \$2,200

# Extracting sources is hard!

Until the day in 2018 when **Constance Millar** ascended the trail to Telescope Peak — the highest point in Death Valley National Park — and discovered hundreds of dead and dying bristlecones extending as far as she could see.

**"It's like coming across a murder scene,"** said **Millar**, an emerita research ecologist with the U.S. Forest Service who has studied bristlecone pines for the better part of 40 years.

Standing amid the devastation atop Telescope Peak, **Millar** realized, **"This could be a harbinger of what could be happening in the future."**

After her grisly discovery atop Telescope Peak, **Millar** immediately contacted **Barbara Bentz**, a research entomologist for the Forest Service based at the Rocky Mountain Research Station in Logan, Utah.

When **Bentz** peeled a piece of bark off one of the Death Valley bristlecones, she found scores of tiny tunnels created by beetle larvae as the insects chewed through the tree's living tissue.

**"Oh, no,"** Bentz thought to herself. **"It really is beetles. Oh, no."**

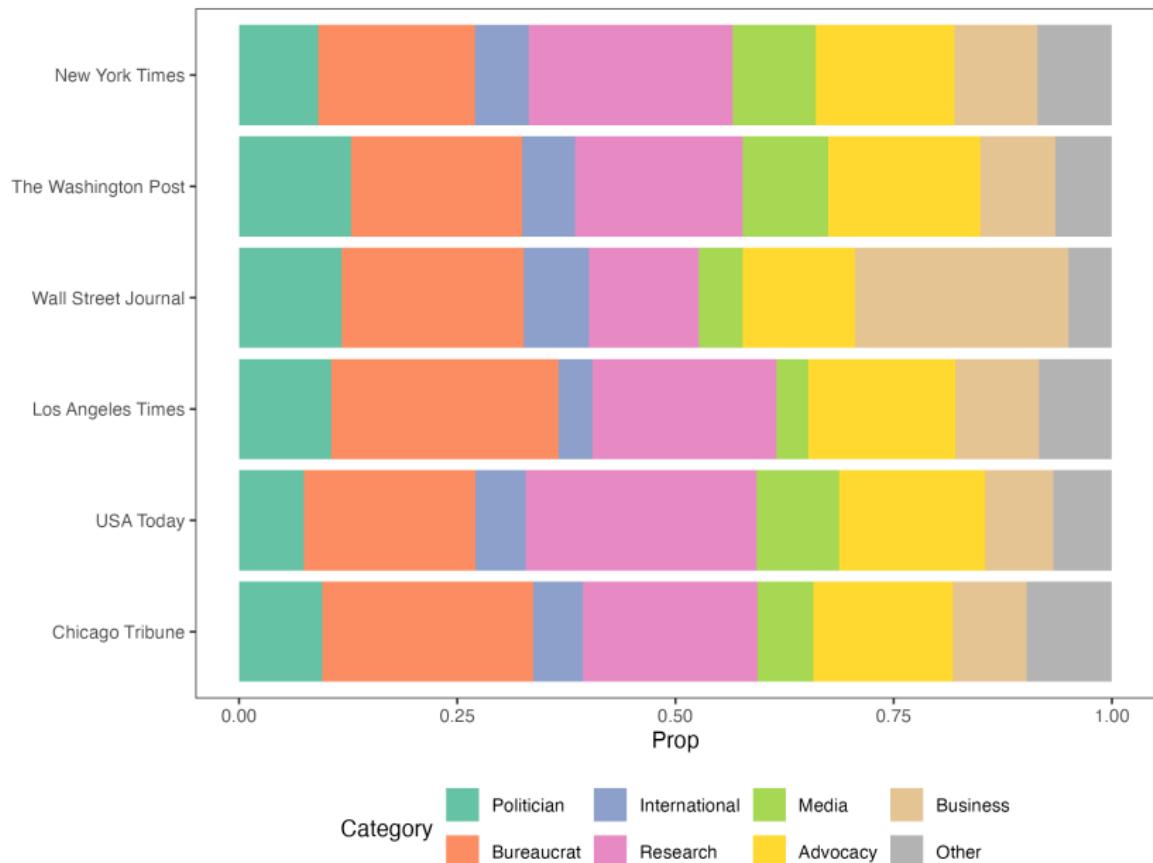
# LLM powered extraction... fast and cheap

- We develop an automated data collection procedure leveraging LLMs (gpt-4-turbo).

```
{  
    "person_name": "Barbara Bentz",  
    "person_title": "Research Entomologist",  
    "organization": "Forest Service, Rocky Mountain Research Station",  
    "document": "N/A",  
    "category": "Bureaucrat",  
    "comments": "Expressed concern about beetle infestation affecting trees."  
},
```

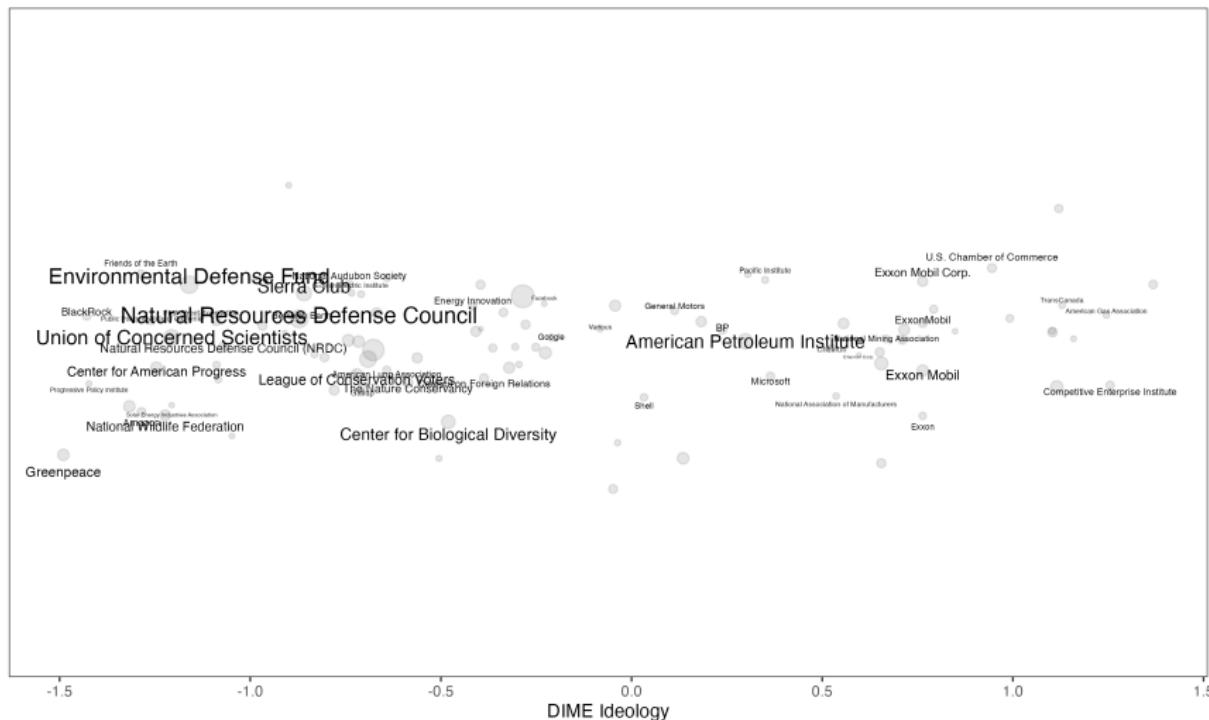
- From 16,508 articles, we extracted 89,350 sources for about \$2000. ➔ process details
- Human validation using crowdworkers suggests reliable performance ( 95% precision and recall). ➔ validation details

# Roughly Categorizing Sources



# Assigning Ideology to Sources

Match organizations to DIME ideology scores



# Assigning Ideology to Articles

1. Match organizations and politicians to DIME ideology scores
2. Label environmental advocacy, fossil fuel industry sources to impute missing ideologies
3. Mean ideology score → **article level slant measure.**

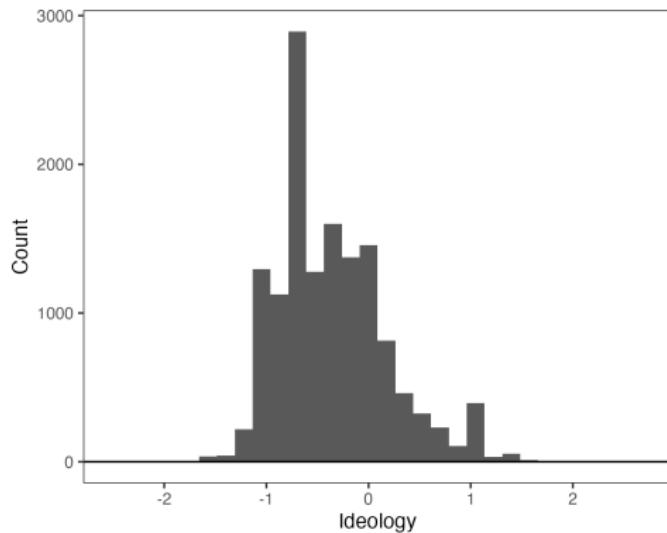


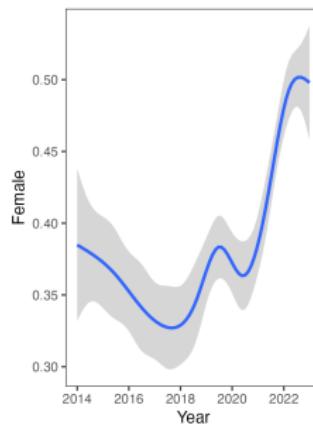
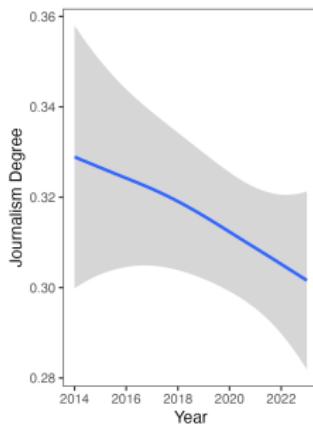
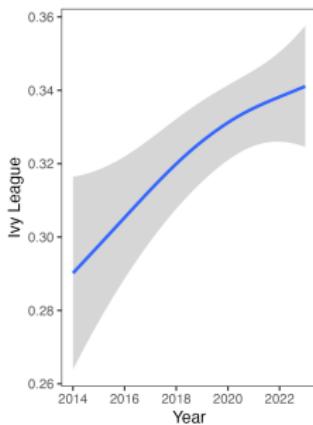
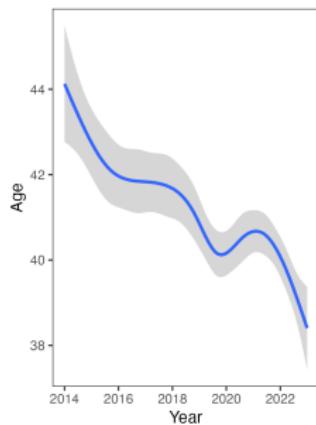
Figure: Distribution of article slant ( $sd=0.52$ )

## Journalist Data

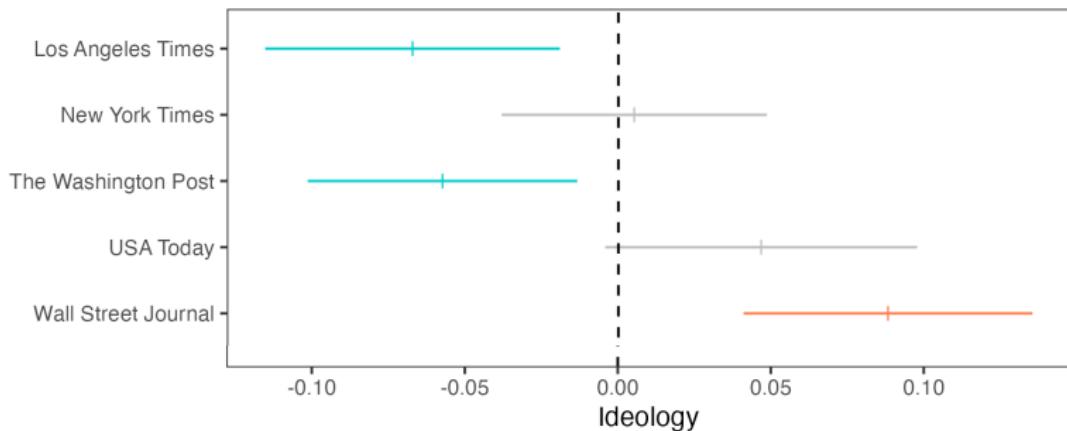
- Collect characteristics of 2,542 journalists using LinkedIn
  - Ivy League? (Undergraduate)
  - Field of study (Political Science, vs Journalism)
  - Postgraduate degree?
  - Work experience outside of journalism?
  - Age
  - Race
  - Gender

# Journalist Data

- Collect characteristics of 2,542 journalists using LinkedIn
  - Ivy League? (Undergraduate)
  - Field of study (Political Science, vs Journalism)
  - Postgraduate degree?
  - Work experience outside of journalism?
  - Age
  - Race
  - Gender

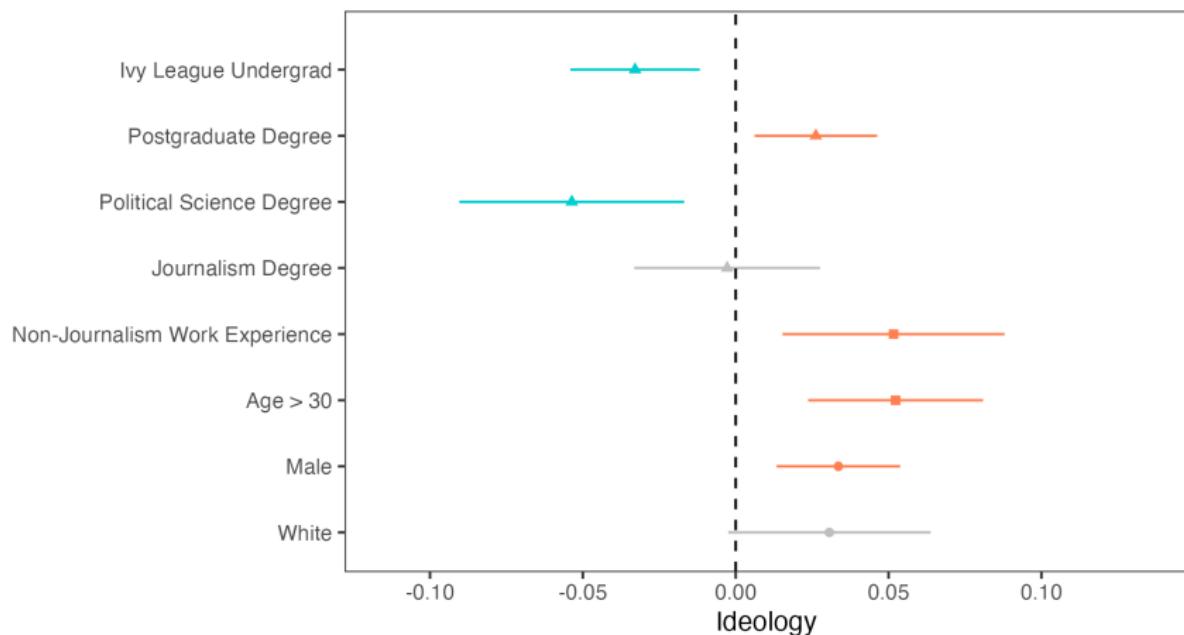


## Results: Outlet characteristics and article slant



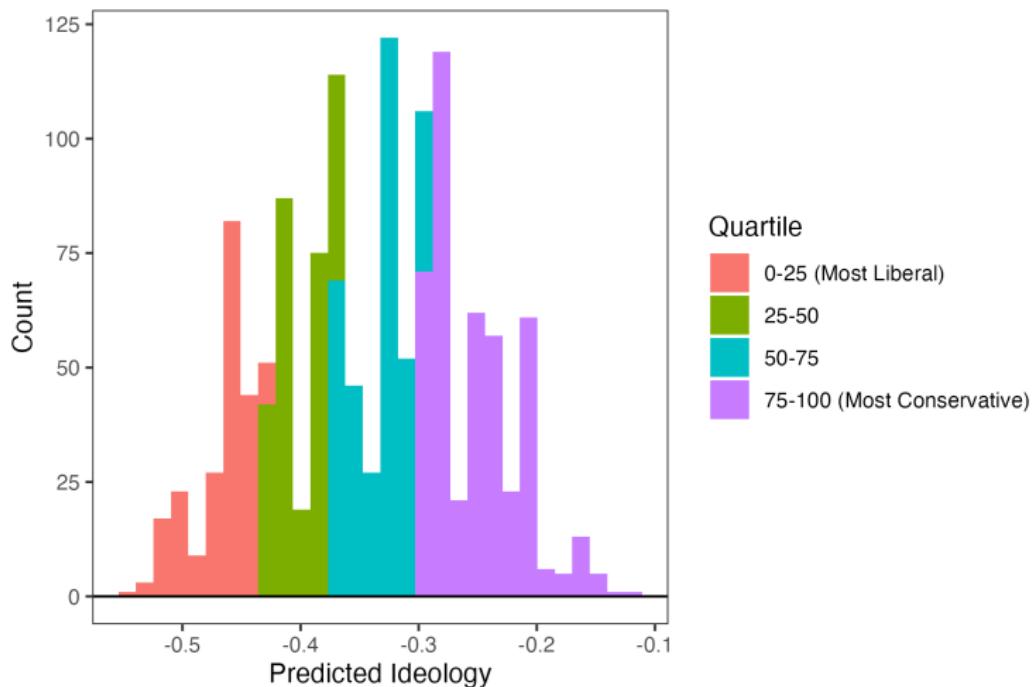
**Figure:** Chicago Tribune as reference group. We see that the gap between the conservative Wall St Journal and the more liberal outlets are around 0.25SD.

# Results: Journalist characteristics and article slant



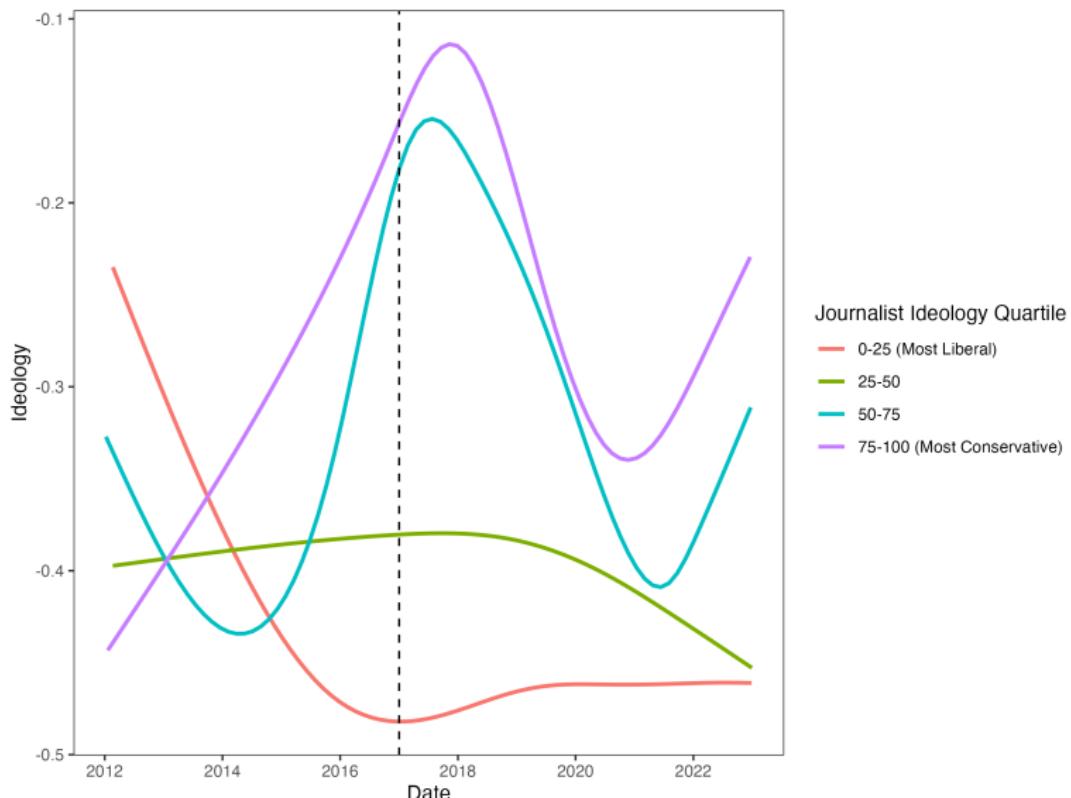
# Results: Trump Effect

Use journalist demographics and article slant scores to predict journalist slant



# Results: Trump Effect

Compare left-most journalists to others over time



## Results: Trump Effect

Compare left-most journalists to others before and after 2016 in a difference-in-differences

<i>Dependent variable:</i>	
Article Slant	
Left Leaning (First Quartile)	-0.046 (0.048)
Year >= 2016	0.100*** (0.028)
Left Leaning AND Year >= 2016	-0.147*** (0.055)
Observations	2,561

Note:

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

## Summary

- Journalist demographics are changing; younger, less formal training, more elite education.
- This “new guard” is less willing to follow the objectivity norm and cite more left leaning sources
- Trump’s presidency in particular marks a dramatic shift; left leaning journalists polarize significantly after 2016

## Summary

- Journalist demographics are changing; younger, less formal training, more elite education.
- This “new guard” is less willing to follow the objectivity norm and cite more left leaning sources
- Trump’s presidency in particular marks a dramatic shift; left leaning journalists polarize significantly after 2016
- Next: more articles and journalist data to collect, robustness checks, alternative DiD specifications...

# Thank you!

Sylvan Zheng  
saz310@nyu.edu

Hye Young You  
hyou@princeton.edu

## Appendix: Article Selection Procedure

1. Search using the ProQuest Database for any news article containing the terms “global warming” or “climate change” from 2012-2022 (N=72,024)
2. Select relevant articles (N=16,532) using gpt-3.5-turbo: *does this article primarily discuss the science of climate change, global warming, or impacts of climate change or climate change related policy?*

## Source Extraction by Humans

- It's costly, time consuming, hard to scale, and a bit error-prone
- 114 congressional hearings on climate change → 3,137 reference sources extracted by two RAs: took 6-months and cost \$3,000
- 120 news articles on climate change and abortion → 557 sources extracted by two RAs: took 6-months and cost \$2,200

# Using ChatGPT for Source Extraction

1. Identify list of quotes and paraphrases
2. Deduplicate quotes from the same source
3. Organize information as JSON
4. Assign categories

## Prompt: Identify list of quotes and paraphrases

You are a research assistant whose task it is to extract quotes and other external information used by journalists in news articles.

The user will provide the text of a news article.

Generate a numbered list of all quotes and external information attributed to specific people, organizations, or documents such as studies, reports, or press releases. Format your response as follows:

1. Quote or information - Name of the source - Background of the source

Some articles do not mention any information drawn from external sources. In these cases, simply say "No sources mentioned."

## Step 2: Deduplication



1. "It's like coming across a murder scene."
    - Constance Millar
    - Emerita research ecologist with the U.S. Forest Service, has studied bristlecone pines for nearly 40 years.
  2. "This could be a harbinger of what could be happening in the future."
    - Constance Millar
    - Emerita research ecologist with the U.S. Forest Service, expert on bristlecone pines.
  3. "Oh, no. It really is beetles. Oh, no."
    - Barbara Bentz
    - Research entomologist for the Forest Service based at the Rocky Mountain Research Station in Logan, Utah.
- 
1. **Constance Millar**
    - Description of the source: Constance Millar is an Emerita research ecologist with the U.S. Forest Service who has nearly four decades of experience studying bristlecone pines.
    - Summary of the source's contribution: Millar provides insights into the impact of environmental changes on bristlecone pines, likening the scene to a murder and suggesting that the current conditions may foreshadow future ecological challenges.
  2. **Barbara Bentz**
    - Description of the source: Barbara Bentz is a research entomologist for the Forest Service based at the Rocky Mountain Research Station in Logan, Utah.
    - Summary of the source's contribution: Bentz expresses concern upon discovering that beetles are affecting the trees, indicating a sense of alarm at the situation.

## Prompt: Deduplicate quotes from the same source

You are a research assistant whose task it is to understand the sources of quotes and other external information used by journalists in writing news.

The user will provide a list of quotes and external information referenced in a news article.

Aggregate together the quotes and information drawn from the same sources into a new numbered list. Provide a description of the source and a summary of the source's contribution

Format the list as follows:

Name of the source - Description of the source - Summary of the source's contribution - Documents authored by the source, if relevant

# Step 3: Organize as JSON



## 1. Constance Millar

- Description of the source: Constance Millar is an **Emerita research ecologist with the U.S. Forest Service** who has nearly four decades of experience studying bristlecone pines.
- Summary of the source's contribution: Millar provides insights into the impact of environmental changes on bristlecone pines, likening the scene to a murder and suggesting that the current conditions may foreshadow future ecological challenges.

## 2. Barbara Bentz

- Description of the source: Barbara Bentz is a **research entomologist for the Forest Service based at the Rocky Mountain Research Station in Logan, Utah**.
- Summary of the source's contribution: Bentz expresses concern upon discovering that beetles are affecting the trees, indicating a sense of alarm at the situation.

```
[{  
  "person_name": "Constance Millar",  
  "person_title": "Emerita research ecologist",  
  "organization": "U.S. Forest Service",  
  "document": "N/A",  
  "comments": "Millar provides insights into the  
               impact of environmental changes on  
               bristlecone pines."  
},  
{  
  "person_name": "Barbara Bentz",  
  "person_title": "Research entomologist",  
  "organization": "Forest Service at the  
                 Rocky Mountain Research Station",  
  "document": "N/A",  
  "comments": "Bentz expresses concern upon  
               discovering that beetles are  
               affecting the trees."  
}]
```

## Prompt: Clean and transform to JSON

For each entry in the given list of primary sources, identify the person's name (if the source is a person), their title (such as President, Professor, Spokesman, etc.), the organization name, concise description of associated written document if any (eg, study published in Nature, data from 2014, etc.).

If any of these are not relevant provide a value of N/A. For unnamed or anonymous people provide a best guess at their title. Do not include titles in the person name field (eg, Barack Obama instead of President Obama). Finally also include comments about the primary source's role in the article.

Format your response as JSON array with keys "person\_name", "person\_title", "organization", "document", and "comments"

## Step 4 (a): Categorizing Cited Source Organizations

- Assign a category to each unique **organization** in the dataset (N=20,608)
- Also extract the **scope** and **country**
- Example: the Cato Institute
  - “Country”: “United States”
  - “Scope”: “national”
  - “Category”: “think tank”
- Example: Wisconsin Dept. Natural Resources
  - “Country”: “United States”
  - “Scope”: “state”
  - “Category”: “government agency or department”

## Prompt: Categorize Organizations

You are a research assistant who is an expert on organizations. The user will provide an organization name. Say what you know about this organization. Then, say the main geographic area and country of origin for the organization (or "international" for international organizations). Then, label the scope: "international," "national", "state", "region", or "local".

Then, for non-governmental organizations, categorize as a "corporation", "trade association", "(k-12) education", "think tank", "university (academic)", "membership association", "labor union", "attorney/law firm", "healthcare", "religious", "activist", "media organization", "native american", "nonprofit", "museum or arts", "place name", "sports", "other. For governmental organizations, categorize as either a "international organization", "government agency or department", "government research institute", "politician office or administration", "legislative body or committee", "political party or organization", "judiciary", or "military". Only choose categories from the above list. Format your response as a JSON object with these keys: - Description - Location - Country - Scope - Category.

## Step 4 (b): Categorizing Sources Without Organizations

- Assign a category to each **person-title pair** in the dataset including information about their contributions as sources (N=7,568)
- Example: Debasis Shyamal, a “fisherman who shared personal concerns about social distancing during evacuation in the face of a cyclone.”
  - “Category”: “citizen”
- Example: Steve Daines, “a senator known for advocating policies that promote innovation over bureaucratic regulation, defending President Trump, and criticizing Democrats.”
  - “Category”: “politician”

## Prompt: Categorize Persons

You are an expert research assistant.

The user will provide a person and some information about their contributions to news articles.

Summarize what you know about this person.

Then, say the main geographic area and country of origin for the person.

Then, categorize the person as either a 'citizen', 'agriculture', 'student', 'university professor', 'researcher or scientist', 'politician', 'political strategist', 'government official', 'religious figure', 'businessperson', 'activist', 'consultant', 'lawyer', 'judge', 'writer', 'journalist', 'artist', or 'other expert'

Format your response as JSON with these keys: - Description -

Location - Country - Category

Only choose categories from the above list.

# Validation

[label=validation]

- Hired 25 workers on Prolific to each validate GPT output for 10 articles each (random sample of 250 articles)
- Each worker indicates if the **source details are correct** and if there should be **additional sources added**
- Each worker also worked on a “test” article which we intentionally created errors in the sources
- We only analyze data from workers that catch all of the planted errors (25%) of the worker pool
- Final validation dataset size: 56 articles and 241 sources
- Cost: ~ \$400

# Validation

Climate Article Sources Home Annotation Codebook Statistics Help Finished 0/5 Current\_Jd 0

Currently logged in as test

In Kern County, oil and gas generate tax revenues that keep schools open and police departments funded.

It's Wednesday. Kern County, which now produces 70 percent of California's oil, considers a future with no fossil fuel extraction. Plus, a dog helps rescue its owner in a California forest.

Long a leader in environmental regulation, California is considering phasing out oil drilling in the state by 2045 as part of an effort to reduce carbon emissions that contribute to global warming.

Our state's economy isn't heavily dependent on fossil fuels the way the economies of, say, West Virginia or Oklahoma are. But ending oil production here would still pose a major challenge for at least one pocket of the Golden State.

About a two-hour drive north of Los Angeles is Kern County, our very own oil country. Kern County produces 70 percent of the state's oil. The industry there is responsible for 16,000 jobs.

And, most crucially, oil and gas generate nearly one-quarter of the county's property tax revenue, which pays for schools, law enforcement, hospitals and other public services. "Nowhere else in California is tied to oil and gas the way we are, and we can't replace what that brings overnight," the county's chief administrative officer, Ryan Alshop, told The New York Times.

My colleague Brad Plumer recently traveled to Kern County and wrote about its struggle to quit oil drilling, even in a state so dedicated to reducing its carbon footprint. Kern County offers a glimpse into issues that other communities across the U.S. could face as they also try to disentangle themselves from fossil fuel income.

The problem boils down to this: If California ends oil drilling, Kern County needs another source of revenue to keep its roads paved and libraries open. Some have suggested leaning into renewables, as the county is already the state's largest provider of wind and solar power. Local leaders are also discussing expanding industries like aerospace, manufacturing, hydrogen or biodiesel production, or even carbon capture technology.

"Politicians, businesses, environmentalists, academics — everyone's sort of thinking hard about how can the county reinvent itself," Brad told me. "But all of these strategies take time, and they might take longer than the county has if the state is serious about phasing out oil production so quickly."

Edit details for each of the following primary sources.  
Add details for any primary sources cited in the text that are not included in this list

1. Ryan Alshop (Kern County) Edit Details

2. Brad Plumer Edit Details

3. Nevada County Sheriff's Office Edit Details

Add New Sources

**Brad Plumer**

Description: Plumer provides perspective on the efforts in Kern County to find new paths for economic development and the complexities of transitioning away from fossil fuels.

Person Name	Brad Plumer
Person Title	Journalist
Organization	N/A
Document	N/A
Category	media/author

**Is this primary source cited in the text?**

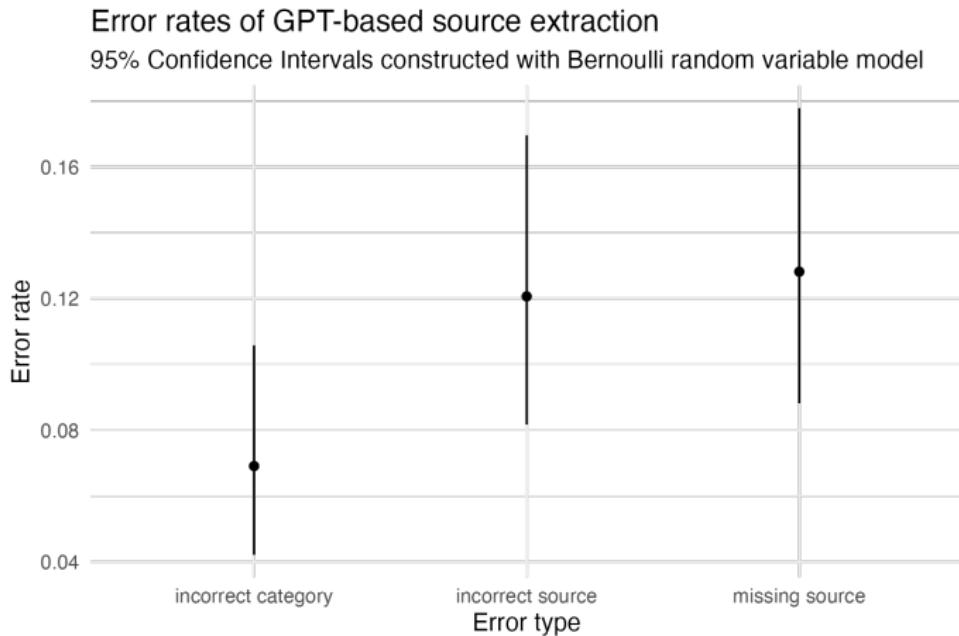
Yes, I have corrected details where necessary.

Yes, but it's redundant/duplicate of another source.

No

Unsure

# Validation Results



**Figure:** Error rates are below 15% for all error types. We use a conservative definition for “incorrect source” counting any case where the crowd worker made any changes to the name, title, or organization.