



# Exploring Ecological Bias in Depictions of NYC Rivers in *The New York Times*

Dez Miller, PhD Candidate  
Emory University (Atlanta, USA)  
CHR 2025



# Background

In my research, I study how our relationships to urban water have changed in the era of climate change. Urban water flows have long been buried, polluted, piped, and controlled. But with more severe weather, these flows are starting to make themselves known, forcing many of us to re-examine our dependence on them.

In October 2012, Hurricane Sandy brought New York's rivers, as well as long-buried streams and creeks, right to New Yorkers' front doors. The storm surge in the Hudson, East, and Harlem Rivers (estuaries) caused widespread devastation. It was a wake-up call that New York would be more vulnerable in an age of climate change, and a huge cultural response ensued in the arts, in literature, and in policy. How did this wake-up call reverberate in other parts of culture, like in newspapers?

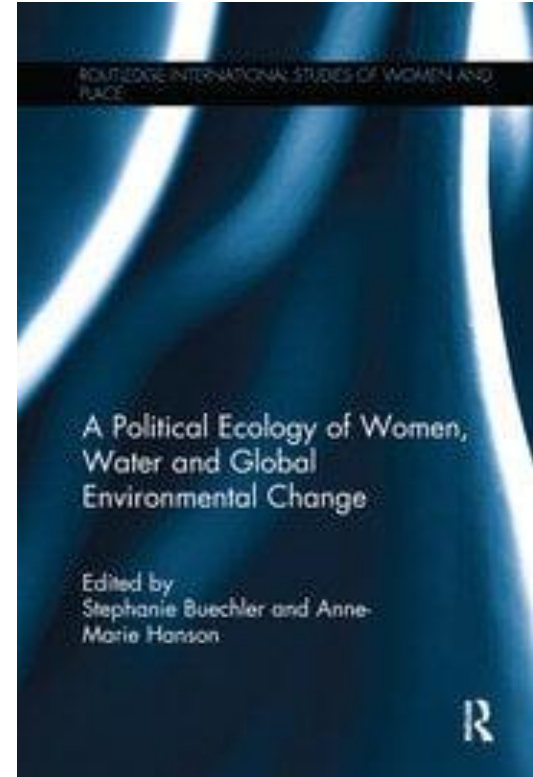
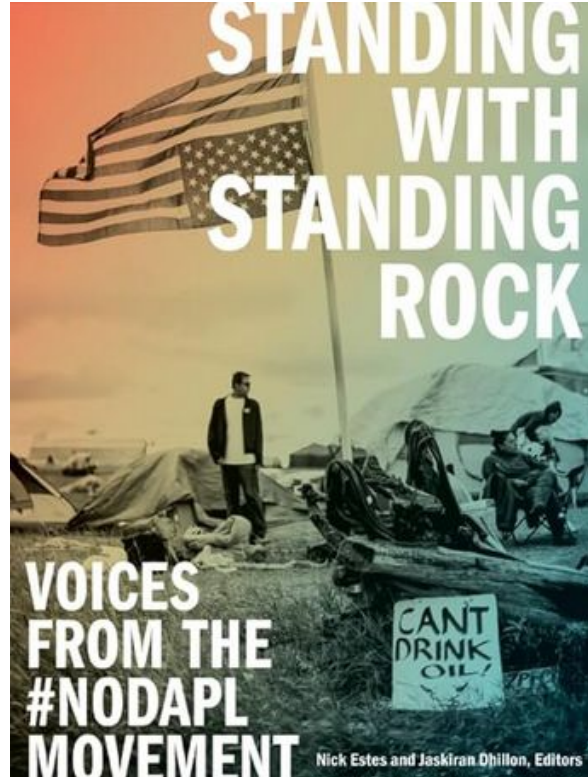


Leland, John. "Hurricane Sandy in Pictures"

# Research Questions

- Did the way the rivers were depicted in *The New York Times* change post-Sandy in a way that can be meaningfully captured by NLP methods?
- Were they depicted in a more ecologically aware manner? Were they more closely semantically tied to climate change?

# Thinking through Ecological Bias



# Methods

1. Word2Vec Modeling of a Proquest TDM Studio Corpus.
  - All *NYT* articles 5 Years pre-Sandy
  - All *NYT* articles 5 Years post-Sandy
2. LLM rating of the top 5 most relevant articles (as determined by the *NYT* algorithm) for each of the five years pre- and post-Sandy.
  - 40 articles total, rated by GPT4o and Claude Sonnet

# Methods

1. Word2Vec Modeling of a Proquest TDM Studio Corpus.
  - All *NYT* articles 5 Years pre-Sandy
  - All *NYT* articles 5 Years post-Sandy
2. LLM rating of the top 5 most relevant articles (as determined by the *NYT* algorithm) for each of the five years pre- and post-Sandy.
  - 40 articles total, rated by GPT4o and Claude Sonnet

# Method 1: Word2Vec Word Embeddings

# Notes about method

Hudson river  $\Rightarrow$  hudsonriver

East river  $\Rightarrow$  eastriver

Hurricane Sandy  $\Rightarrow$  hurricanesandy

- Because I was looking at most similar words and not similarity values, no need to align vector spaces (Gonen et al, 2020).
- Looking to do Bert embeddings in the future (somewhat complicated by virtual environment of Proquest TDM Studio)



# Results frequent terms

## Pre-Sandy 2008-2012

<b>hudsonriver</b>	<b>eastriver</b>	<b>hudsonriver +</b>	<b>eastriver +</b>
mid-hudson	waterfront	<b>pollution</b>	<b>pollution</b>
hudsonriver	bridge	pollutant	waterway
neversink	promenade	toxics	pollutant
pascack	expressway	pcbs	waterfront
river	verrazano-narrows	waterway	smog
eastriver	rows	smog	shoreline
bekaa	rockaways	polluting	pcbs
riverfront	queensboro	groundwater	sewage
locust	pier	estuary	roadway
waterfront	walkway	pcb	aquifer
raritan	skyline	aquifer	groundwater

Hudson River is more associated with geographical areas, regions, towns upriver, other “great” rivers

## Post-Sandy 2012-2017

<b>hudsonriver</b>	<b>eastriver</b>	<b>hudsonriver +</b>	<b>eastriver +</b>
pascack	harlemriver	<b>pollution</b>	<b>pollution</b>
locust	queensboro	pollutant	waterway
raritan	waterfront	smog	pollutant
stardew	bridge	polluted	smog
bekaa	promenade	pcbs	expressway
riverfront	verrazano-narrows	waterway	waterfront
eastriver	footbridge	estuary	polluted
battenkill	viaduct	riverkeeper	estuary
indus	skyline	groundwater	inlet
yonkers	shoreline	environmental	shoreline
		polluting	roadway

East River is more associated with views of the city, and city-based infrastructure

Only recognizable Sandy influence is that the eastriver/pollution vectors were associated with “lowlying” in longer lists

## Method 2: AI Assisted Rating

# Prompt

I am going to give you a series of New York Times articles that mention the East River. Do the following tasks. Do not say N/A. Rely on implicit reasoning if necessary to give an answer.

1: Briefly summarize the passage.

2: Using implicit and explicit reasoning, describe in 2-3 sentences how the river is depicted along a spectrum of clean to polluted. Then provide a rating where **“clean” is at 1 and “polluted” is at 10**. Given the amount of relevant information available to determine your rating, rate your confidence about the rating on a scale of 1 and 10, with 1 being low confidence and 10 being high confidence.

3: Using implicit and explicit reasoning, describe in 2-3 sentences how the river is depicted along a spectrum of inert to alive. Then provide a rating where **“inert” is at 1 and “alive” is at 10**. Given the amount of relevant information available to determine your rating, rate your confidence about the rating on a scale of 1 and 10, with 1 being low confidence and 10 being high confidence.

4: Using implicit and explicit reasoning, describe in 2-3 sentences how the river is depicted along a spectrum of beneficial to detrimental. Then provide a rating where **“beneficial” is at 1 and “detrimental” is at 10**. Given the amount of relevant information available to determine your rating, rate your confidence about the rating on a scale of 1 and 10, with 1 being low confidence and 10 being high confidence.

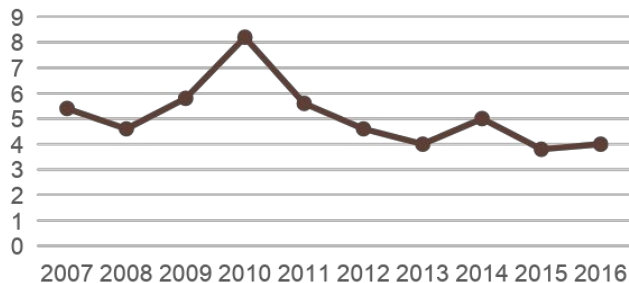
5: Using implicit and explicit reasoning, describe in 2-3 sentences the centrality of the river to the article, along a spectrum of background to central. Then provide a rating where **“background” is at 1 and “central” is at 10**.

# Prompting Strategies

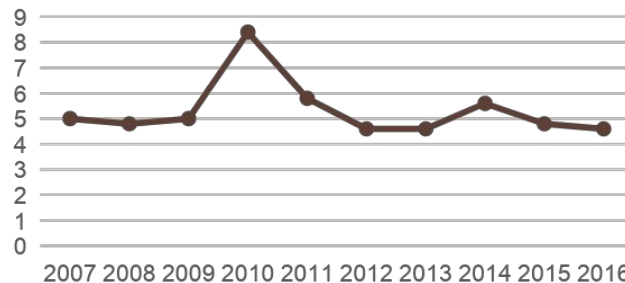
- Academic influences:
  - Chain of thought prompting, prompting for implicit and explicit reasoning (Underwood, 2023)
  - Semantic axes to measure bias (Li, Tadiment, Bamman, 2022, and others)
- Strategies gleaned from trial and error:
  - Repeating the task instructions, rather than giving a task and then a series of scales (to avoid influence from previous scales)
  - Asking it to use **implicit and explicit** reasoning to describe how the article falls along the scale before giving a rating.
    - (This was the only thing that stopped it from giving a low rating for neutral answers. It was harder to get Claude to give a neutral rating than GPT)
  - Putting the relevance rating last
  - Giving examples made it too rigid, and overly rely on the criteria used in the examples.

# Results: Clean (0) to Polluted (10)

GPT Hudson River



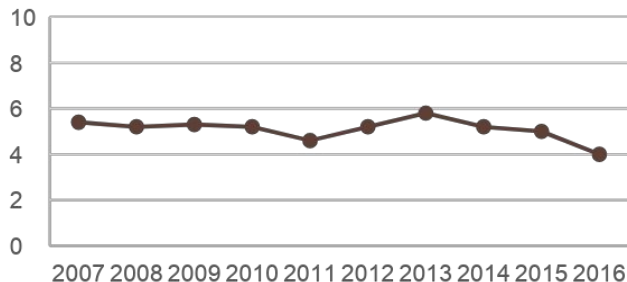
Claude Hudson River



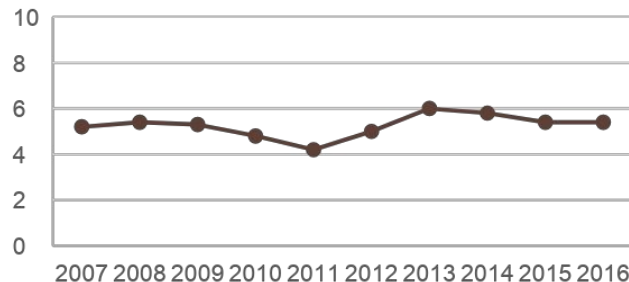
Averages  
GPT: 5.1  
Claude: 5.32

GPT East River

Averages  
GPT: 4.5  
Claude: 5.15

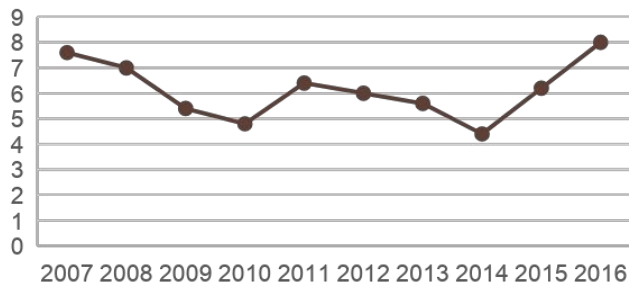


Claude East River

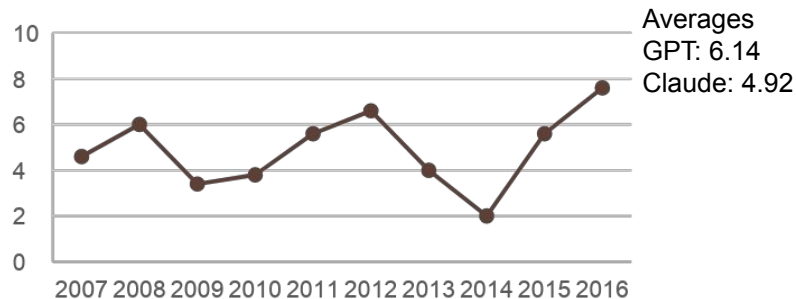


# Results: Inert (0) vs. Alive (10)

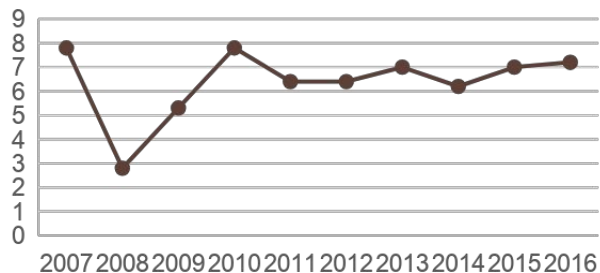
GPT Hudson River



Claude Hudson River

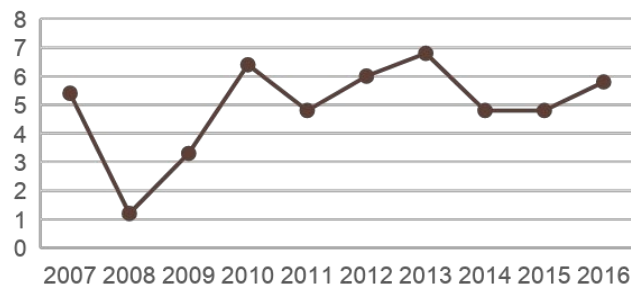


GPT East River



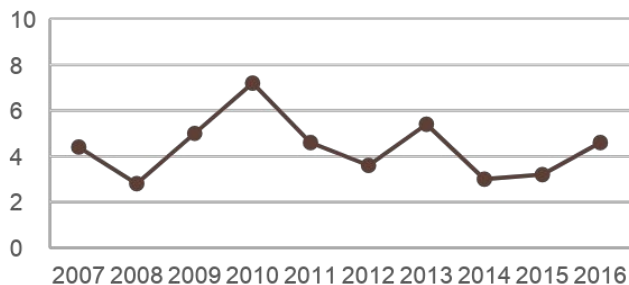
Averages  
GPT: 6.48  
Claude: 4.48

Claude East River

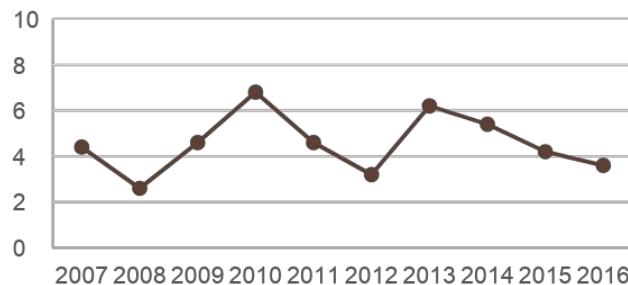


# Results: Beneficial (0) vs. Detrimental (10)

GPT Hudson River



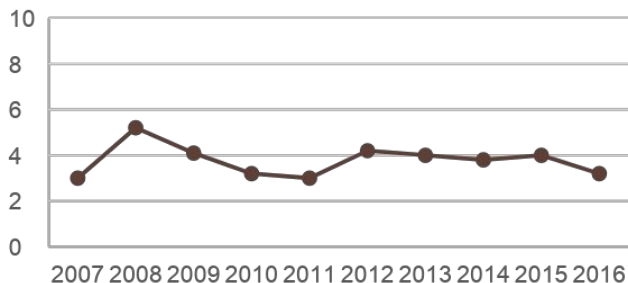
Claude Hudson River



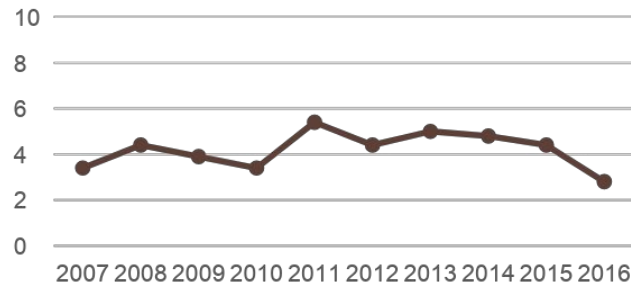
Averages:  
GPT: 4.37  
Claude: 3.24

GPT East River

Averages:  
GPT: 3.42  
Claude: 3.24

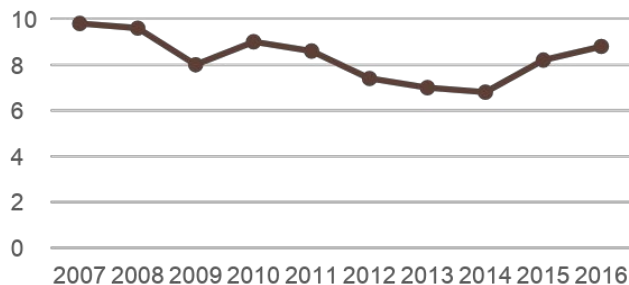


Claude East River

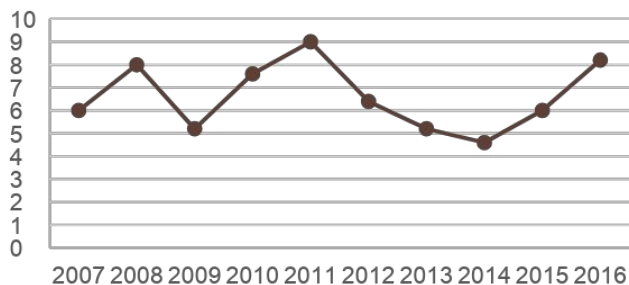


# Results: Background (0) vs. Central (10)

GPT Hudson River

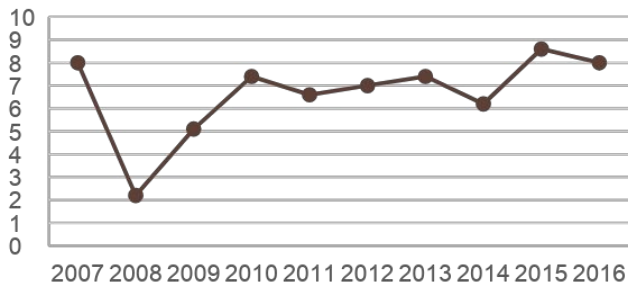


Claude Hudson River



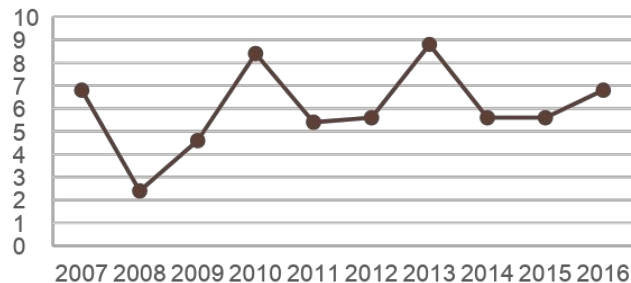
Averages:  
GPT: 8.32  
Claude: 6.62

GPT East River



Averages:  
GPT: 6.72  
Claude: 5.64

Claude East River





# Discussion

- No clear trends over time, and no common inflection points at the 2012-2013 post-Sandy year
- Anecdotally, the LLMs appeared most insightful at estimating inert vs. alive (though sometimes they both interpreted alive as dynamic rather than ecologically alive), was surprisingly poor at consistent ratings of background vs. central, and beneficial vs. detrimental appeared to shift in meaning somewhat
- however, they were relatively consistent with each other (more than I would have thought)

# GPT4 vs. Claude Sonnet

- GPT4 tends to be more confident than Claude Sonnet
  - Average GPT4 confidence rating: 7.7; median: 7.8
  - Average Claude confidence rating: 6.6; median: 7.1
- Cursorily, Claude seemed to employ successful implicit reasoning more often than GPT4 (see sample responses)
- Middling inter-rater consistency
  - Within 2 points of each other 78% of the time
  - Within 1 point of each other 61% of the time
  - Same exact rating : 26% of the time

# Sample Responses 1

“An 1850s-Era Oyster Barge is Saved for Yet Another Life on the East River” {date here}

## Clean vs. Polluted Spectrum

**GPT4o:** The story highlights the East River as a historic hub for New York's oyster industry, once teeming with thriving oyster beds. It does not comment on the current water quality, focusing instead on the cultural and historical significance of the river. **Rating: 5 (neutral), Confidence: 8 (emphasis on historical perspective)**

**GPTo1:** The river is depicted as historically clean and abundant, supporting a thriving oyster industry in the late 1800s and early 1900s. The decline of the oyster industry around 1920 suggests a possible decrease in water quality, implying increased pollution over time. **Rating 6 (toward polluted), Confidence: 5 (medium, due to limited explicit information on pollution).**

**Claude:** The article indirectly suggests historical changes in river quality through the decline of New York's oyster industry, which once had "350 square miles of oyster beds" harvesting "hundreds of millions of bivalves." **Rating: 7 (historically cleaner, implied decline) Confidence: 7 (clear historical context about oyster industry)**

# Sample Responses 2

“Sewage Discharge in Hudson Threatens Iron Man Swim” {date here}

## Inert vs. Alive Spectrum Responses

**Claude:** The river is portrayed primarily as a passive recipient of pollution rather than a living ecosystem. The focus is entirely on its role as a body of water receiving sewage, with no mention of wildlife, currents, or natural processes. Rating: 2 (mostly inert) Confidence: 7 (medium-high confidence based on the article's consistent treatment of the river as merely a water body)

**GPT4:** The river's portrayal leans toward inert in this context, as it is treated primarily as a contaminated body of water subject to human impact rather than a dynamic, living ecosystem. Its function here is reduced to a site for waste discharge and a potential hazard for human activities. Rating: 3 (somewhat inert) Confidence: 8

**GPTo1:** The river is portrayed as alive in terms of human activity, as it is a venue for swimming, kayaking, and major events like the Ironman triathlon. However, the pollution threatens these activities and potentially harms aquatic life. **Rating:** 6 out of 10 (alive). **Confidence:** 7 out of 10.

# Works Cited

Hamilton, William et al. “Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change.” *Association for Computational Linguistics*, 2016.

Li Lucy, Daivya Tadimeti, and David Bamman. “Discovering Differences in the Representation of People using Contextualized Semantic Axes.” *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, December 2022.

Schechtman, Anna. “Command of Media’s Metaphors.” *Critical Inquiry* 47, Summer 2021, 644-674.

Underwood, Ted. “Using GPT-4 to Measure the passage of Time in Fiction.”  
<https://tedunderwood.com/2023/03/19/using-gpt-4-to-measure-the-passage-of-time-in-fiction/>

Wei, Jason et al. “Chain-of-Thought Prompting Elicits Reasoning in Large Language Models.” *36th Conference on Neural Information Processing Systems*, January 2022.