

# Style Analysis in Journalism

Rafael Arbex-Murut, Yeshwanth Somu

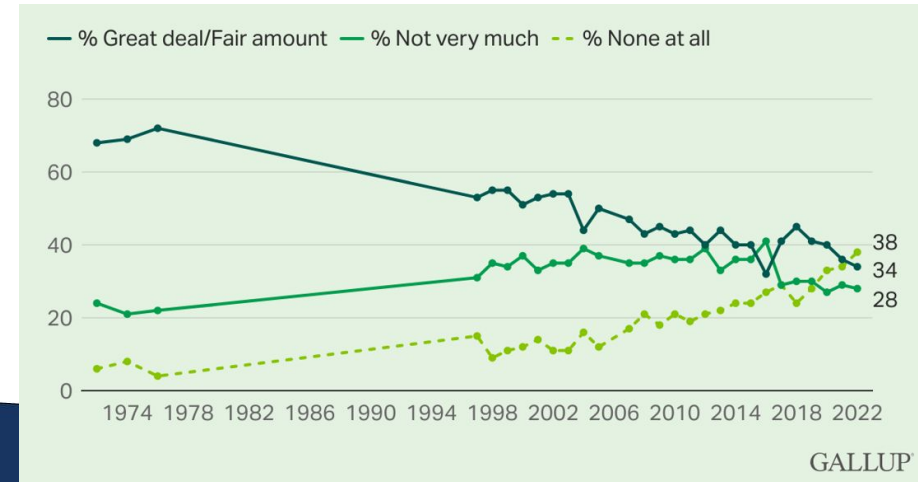
# Introduction

- Written news plays a significant role in society, even as news becomes digitized.
  - 86% of U.S. Adults get their news from digital devices
- Trust in media has decreased



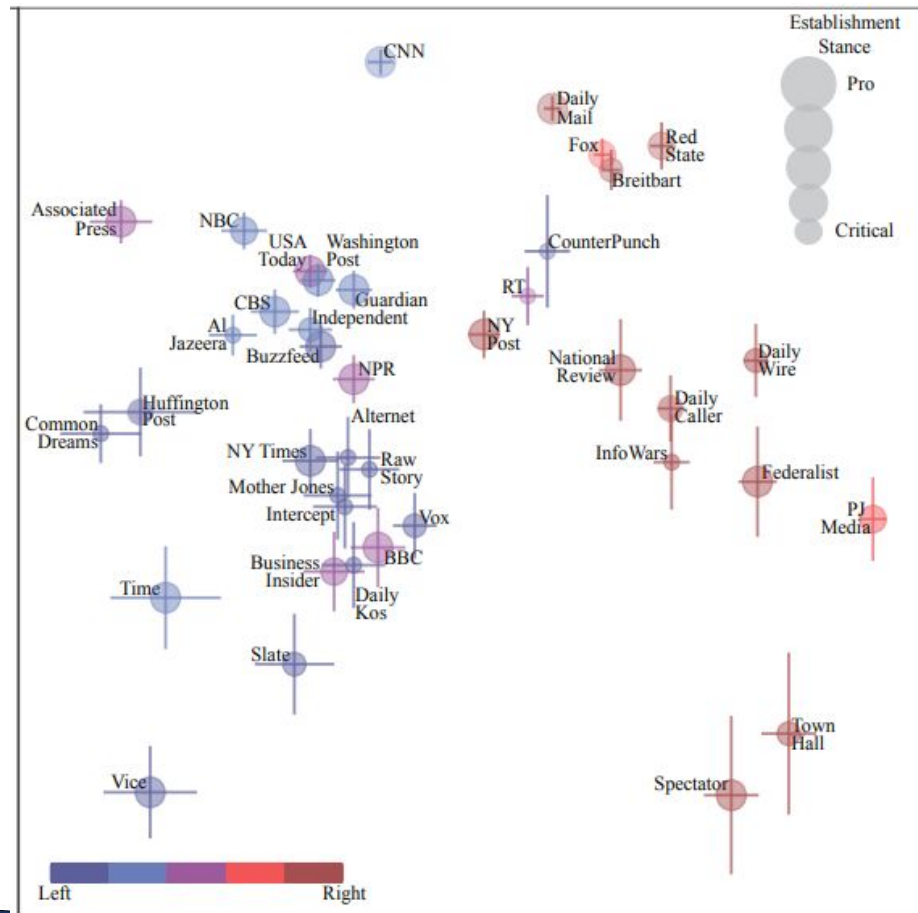
## Americans' Trust in Mass Media, 1972-2022

In general, how much trust and confidence do you have in the mass media -- such as newspapers, TV and radio -- when it comes to reporting the news fully, accurately and fairly -- a great deal, a fair amount, not very much or none at all?



# Background

- Quantified Political Bias
- Authorship Identification
- Generative models for headlines
- Fake news detection



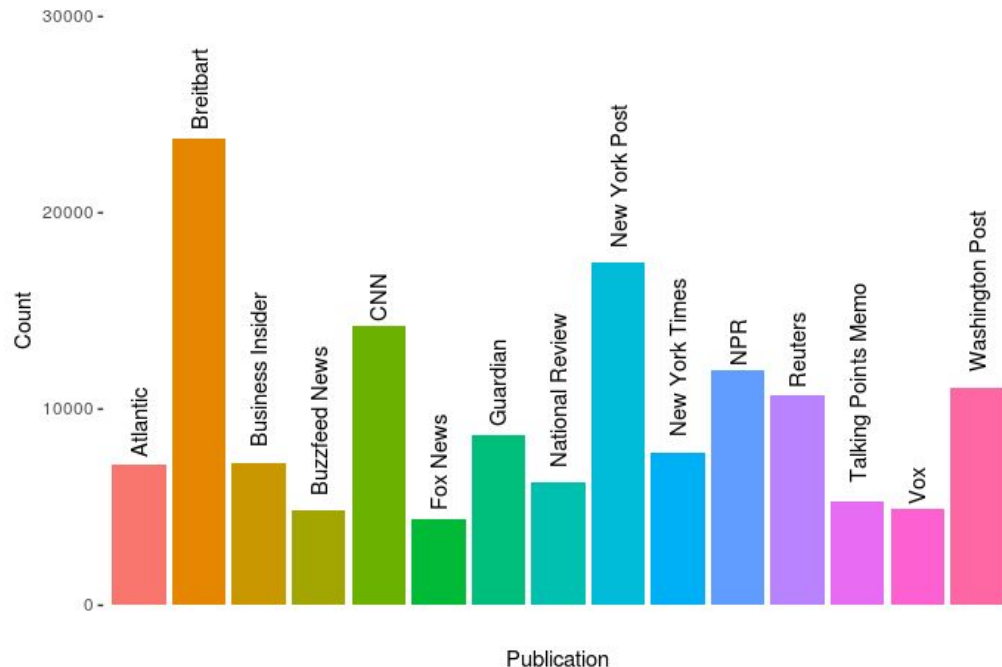
# Our Motivation

- News outlets have been known for having distinct editorial styles, such as a political or philosophical bias.
- Our goal was to analyze these stylistic idiosyncrasies between different newspapers through measuring:
  - the propensity of news outlets to have descriptive headlines of their articles
  - the uniqueness of their editorial styles
- Provide insights to readers, writers, and editors

# Data

- “All the News” dataset from Kaggle, pulling 140,000 news articles from 15 different American publications

- Id
- Title
- Publication
- Author
- Date
- Year
- Month
- Url
- Content



# Data Cleaning

- Articles
  - Removed null date values
  - Published in 2015 or more recently
  - Word count between 100 and 500
- Removed prepended and appended self-references

After Trump attack, Lockheed Martin slashes F-35 cost  
Vincent Viola withdraws from secretary of Army nomination  
→ First on CNN: ATF seeking more Chicago agents  
Super Bowl guacamole could get pricier  
GOP gov posted this photo of a refugee family and welcomed them  
Federal judge declines to renew restraining order on Trump travel  
Congress votes to undo 1st of many Obama rules

Elizabeth Misses New Year's Service	The New York Times
Melbourne's Beaches After Storm	The New York Times
, Is Pulled From Cockpit in Canada	The New York Times
About Which Stocks to Buy in 2017	The New York Times
and Kills Himself, State Media Says	The New York Times
150 Inmates, Prompting Manhunt	The New York Times
s, China's State Media Tells Trump	The New York Times
23 Million for Deceptive Practices	The New York Times
Card's Huge Bonus Will Be Slashed	The New York Times
acially Charged Beating in Chicago	The New York Times
r Shot Queen Elizabeth by Mistake	The New York Times
ons 192 for Marijuana Convictions	The New York Times
cyberattacks on Antidoping Agency	The New York Times
Denounce India's Phone Swindlers	The New York Times
r About Before the Golden Globes	The New York Times
relationship Between U.S. and Russia	The New York Times
icial Shot and Wounded in Mexico	The New York Times
r Optical Illusion on Winter Nights	The New York Times
Whale Featured in 'Blackfish' Dies	The New York Times

# Headline Evaluation

Which is the 'better' headline?

- Trump Should Have a Celebrity-Free Inauguration
- A Recent Report Suggested that the President is Bribing Celebrities to Perform at his Inauguration



# Headline Evaluation


“Better” is subjective, but...


- Trump Should Have a Celebrity-Free Inauguration
  - Prescriptive
  - Ideological
  - Uses “should”
- A Recent Report Suggested that the President is Bribing Celebrities to Perform at his Inauguration
  - Descriptive
  - Impartial
  - Does not use “should / shouldn’t”



# Headline Evaluation

“Better” is subjective, but...

- Trump Should Have a Celebrity-Free Inauguration
  - Prescriptive
  - Ideological
  - Uses “should”

Real Headline
- A Recent Report Suggested that the President is Bribing Celebrities to Perform at his Inauguration
  - Descriptive
  - Impartial
  - Does not use “should / shouldn’t”

Summary Generated from Article!

# Headline Evaluation

NR Trump Should Have a Celebrity-Free Inauguration

POLITICS & POLICY

## Trump Should Have a Celebrity-Free Inauguration

Share

- f
- t
- r
- wh
- l
- e
- m
- p



Beyoncé sings at President Obama's second inauguration, January 23, 2013. (Reuters photo: Jason Reed)

By CHRISTIAN TOTO  
December 29, 2016 9:00 AM

Comments Listen

SEARCHED BY: THE STAFF

'Trump Should Have a Celebrity-Free Inauguration'  
( 'a recent report suggested that the president is bribing celebrities to '  
'perform at his' )

```
{'rouge1': 0.09523809523809523,  
'rouge2': 0.0,  
'rougeL': 0.09523809523809523,  
'rougeLsum': 0.09523809523809523}
```

# Headline Evaluation

NR Human Rights Watch's Alarmist Stance toward Trump

POLITICS & POLICY

## Human Rights Watch's Alarmist Stance toward Trump



(Reuters photo: Aaron Josefczyk)

By ALEXANDRA DESANCTIS  
January 17, 2017 6:43 PM

Comments Listen ▶

-----  
'Human Rights Watch's Alarmist Stance toward Trump'  
"human rights watch berates incoming president for 'floating' proposals"  
-----

```
{ 'rouge1': 0.35294117647058826,  
  'rouge2': 0.26666666666666666,  
  'rougeL': 0.35294117647058826,  
  'rougeLsum': 0.35294117647058826 }
```

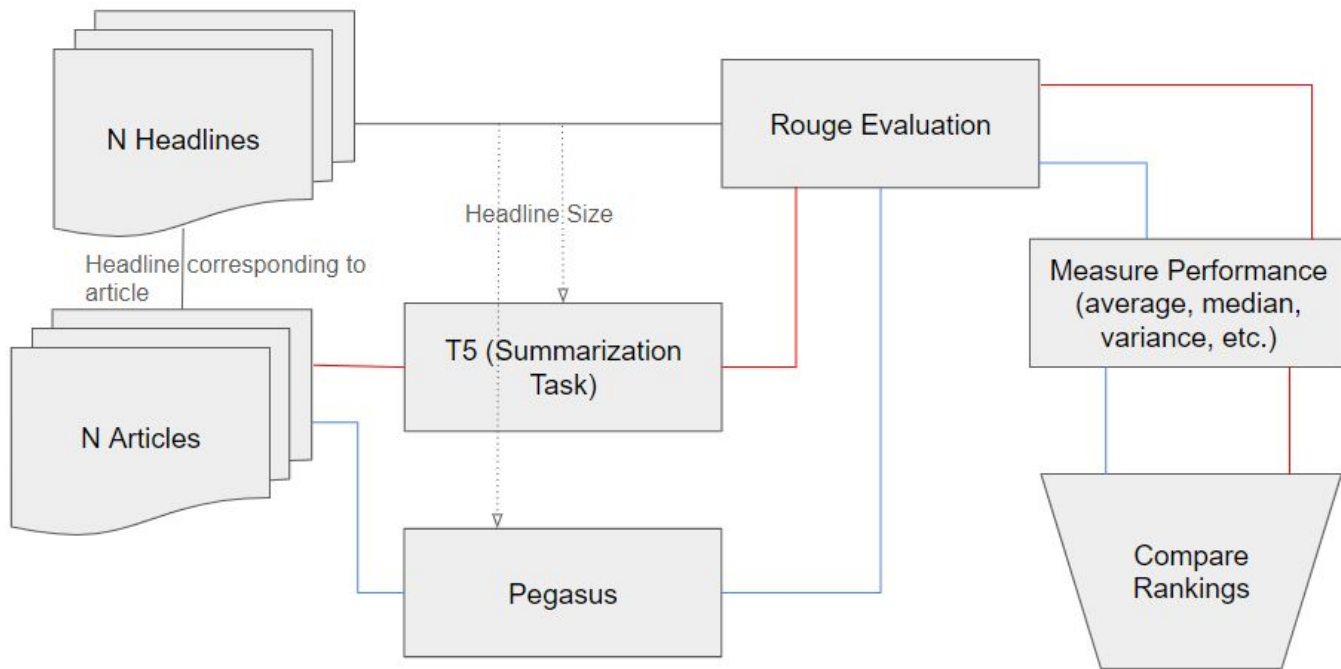
300% higher Rouge score

# Headline Evaluation

How much does the headline deviate from the summary?

- Which publication is the biggest ‘offender’?
- How to quantify this?

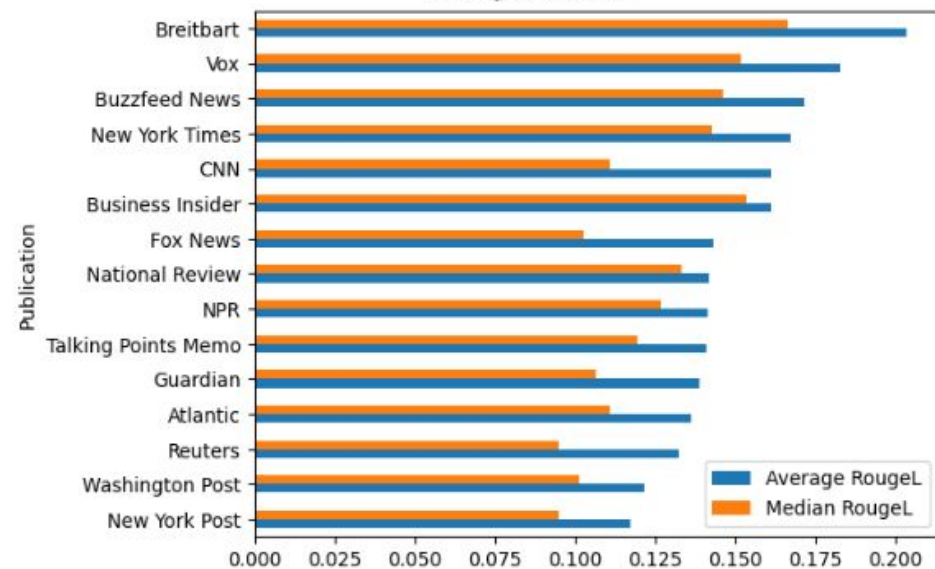
# Headline Evaluation - Methods



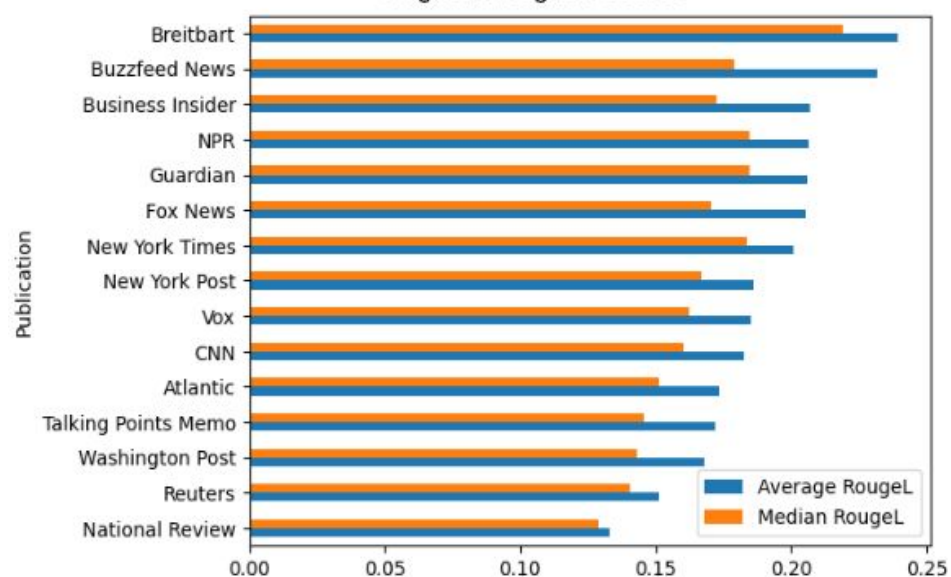
1. Generate a summary
2. Compare Summary to real headline
3. Validate

# Headline Evaluation - Results

T5 RougeL Scores

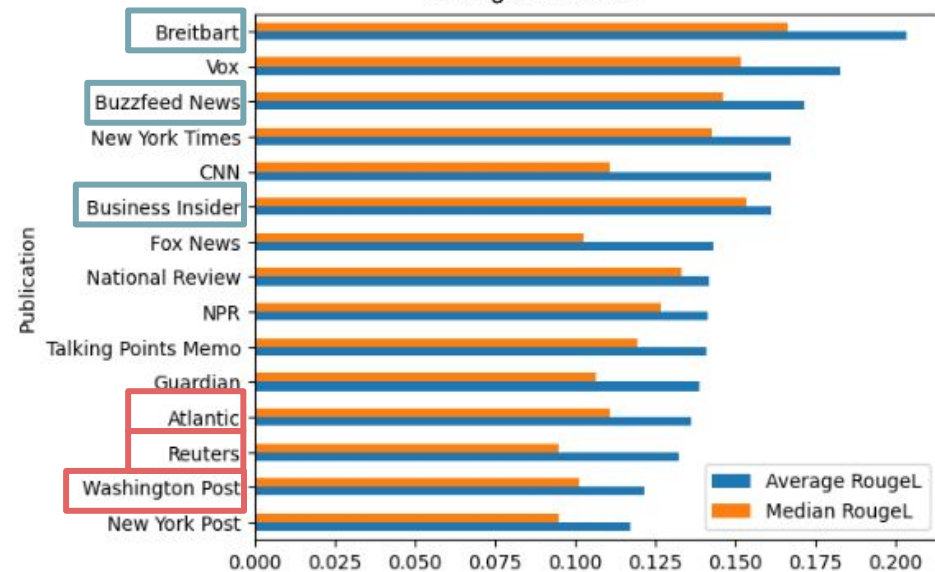


Pegasus RougeL Scores

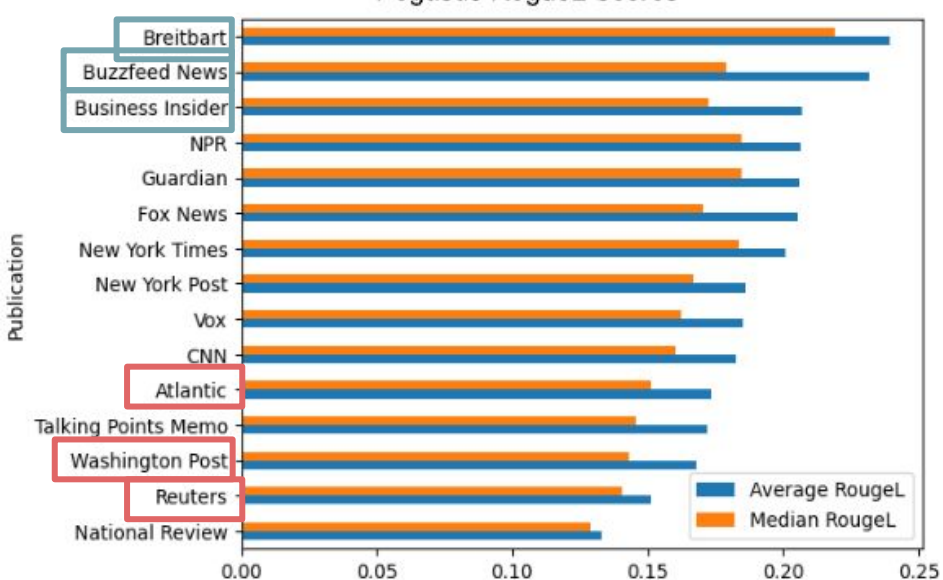


# Headline Evaluation - Results

T5 RougeL Scores



Pegasus RougeL Scores





# News Outlet Classification- Methods

- BERT model leveraged for articles' **news outlet** classification
- Observed the confusion matrix for misclassification of articles
- **Value:** uniqueness and distinctions in editorial styles
- Our hypothesis:
  - High accuracy suggests a unique editorial style for a specific publication
  - High confusion between publications may signal style overlap

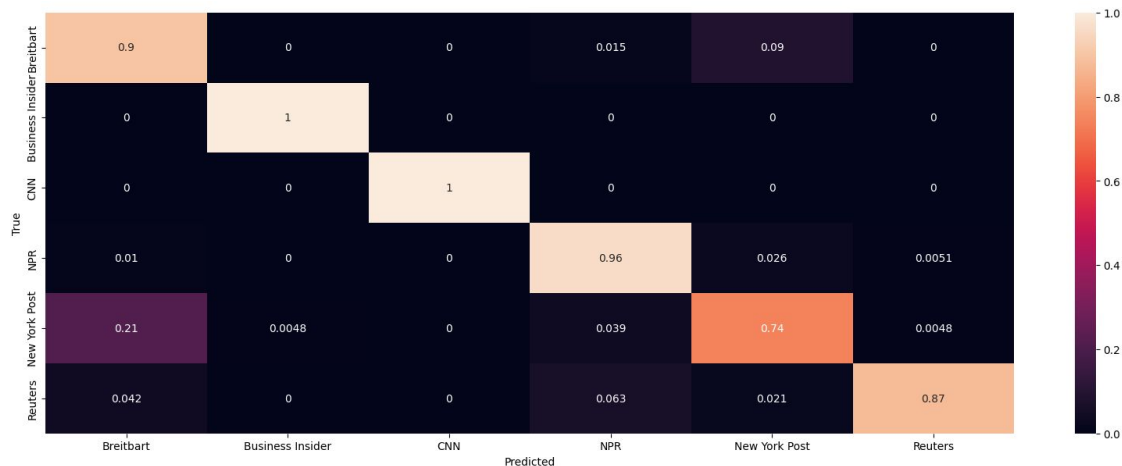
# News Outlet Classification- Results

- Primary BERT model: Classification of articles into 12 news outlets
  - Standardized to 1000 articles
- Unique:
  - Breitbart, Business Insider, **CNN**, Reuters
- Non-unique:
  - BuzzFeed, Washington Post



# News Outlet Classification- Results cont.

- Follow-up investigation:  
Classification of articles into subset of 6 top news outlets
- High uniqueness in style for all except NY Post
- Higher amount of news outlets in data gives a better sense of uniqueness in editorial style



# Conclusion

- Varied rogue scores could be inferred as some headlines being more descriptive than others among outlets.
- Varied classification accuracies are indicative of uniqueness in editorial style among outlets.
- Improvement: Adding authors to feature vectors for analyzing style