# News Bias

**Team 22**
Sadie Allen, Ye Chen, Bayard Eton, Sean McDonald
**10/15/2020**

# Motivation

- People are increasingly getting their news from online sources[1]

- Companies such as Facebook, Google use machine learning to maximize people's time spent on their site

- These algorithms often inadvertently push **false/misleading/extreme** information that reinforces people's existing views and increases political polarization[2]

[1] Mitchell, Amy, et al. "How Americans Get Their News." *Pew Research Center's Journalism Project*, 27 Aug. 2020, www.journalism.org/2016/07/07/pathways-to-news/.
[2] Hao, Karen. "YouTube Is Experimenting with Ways to Make Its Algorithm Even More Addictive." *MIT Technology Review*, MIT Technology Review, 2 Apr. 2020, www.technologyreview.com/2019/09/27/132829/youtube-algorithm-gets-more-addictive/.

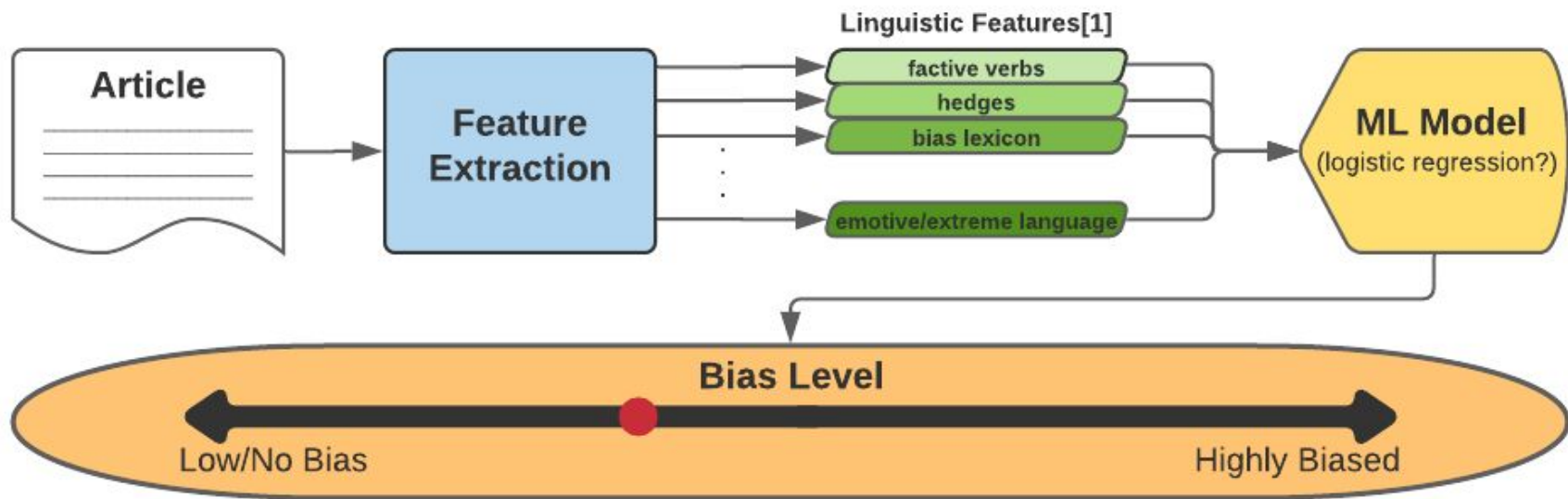# How can we defuse increasing societal tension inflamed by algorithms?

# Proposed Solution/User Interface/Deliverables

Software-only product delivered as a web browser plugin and/or app, providing the following information for articles and headlines:

- Bias measure
- Source metadata
- Related stories

# Visualization: Algorithm Flowchart



[1] Recasens, Marta & Jurafsky, Dan. (2013). Linguistic Models for Analyzing and Detecting Biased Language. ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference. 1.
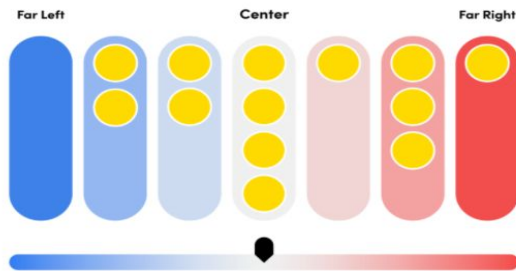
# Requirements

Our algorithm must be . . .

- **Fast**: analyze articles/text in close to real time (<5 seconds)

- **Transparent/Explainable**: the factors/processes contributing to the algorithm's analysis should be clear and understandable to a non-technical lay person (well understood by >70% of lay people participants)

- **Platform-agnostic**: surface our algorithm as an API that can be accessed from other applications/services (Android and iOS)

# Competing Technologies

- **Ground News**: Compare how a story is being covered across the political spectrum and around the world



**Audit your own news**

We give you a simple view of who has covered a story and where each outlet falls on the political spectrum. Quickly glimpse the breakdown of coverage across biases and make your own inferences

- **Stopaganda:** applies informative decals to Google, Twitter, Facebook, and news subreddits showing bias and accuracy based on data compiled by mediabiasfactcheck.com



Stopaganda Plus
by EGibbs87

# Competing Technologies



c|net    COVID-19    BEST ⌄    REVIEWS ⌄    NEWS ⌄    HOW TO ⌄    CARS ⌄    DEALS ⌄    PRIME DAY 2020    5G

## Facebook, Twitter limit reach of New York Post article about Hunter Biden

Senate Republicans want to subpoena Twitter CEO Jack Dorsey over the company's action.
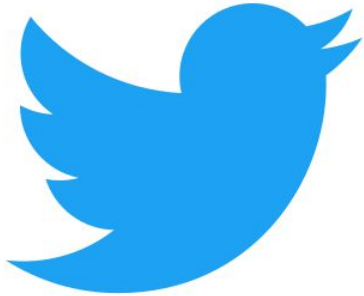
**Carrie Mihalcik, Queenie Wong**
Oct. 15, 2020 11:37 a.m. PT

▶ LISTEN - 04:07

# Questions?

# Backup: Linguistic Features Examples

**Factive verbs:** verbs that presuppose the truth of their complement clause

- The research **revealed** that…

 vs.

- The research **indicated** that…

**Hedges:** words used to reduce one's commitment to the truth of a proposition

- The lower cost of living means a **possibly** higher standard of living.

 vs.

- The lower cost of living means a higher standard of living.

[1] Recasens, Marta & Jurafsky, Dan. (2013). Linguistic Models for Analyzing and Detecting Biased Language. ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference. 1.

# Questions to be prepared for:

What will you use as labeled data for training your model?

- See how sources that are widely considered biased/unbiased are rated by our algorithm
- We need to find a labeled dataset! Lots of work into linguistic bias detection, shouldn't be hard.

How will we find similar sources and source metadata?

- https://newsapi.org/docs (similar articles, articles on the same story)
- Webscraping, https://mediabiasfactcheck.com/ to categorize type of news