MICHIGAN ENGINEERING
COMPUTER SCIENCE AND ENGINEERING
UNIVERSITY OF MICHIGAN

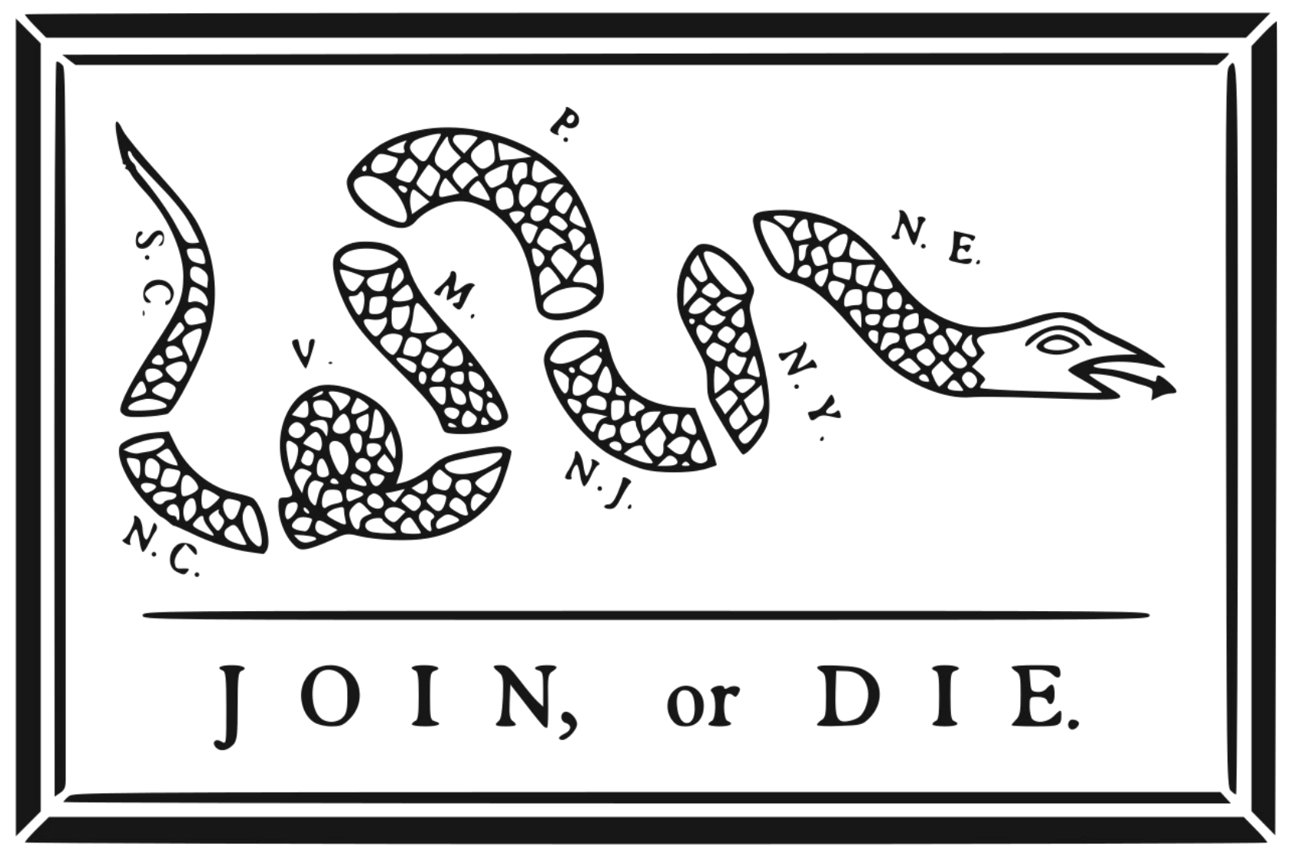
Left, Right, or Neutral? Run it Bias First!

Political Bias Article Classification using Supervised and Unsupervised Learning

Gabrielle Jones, Jessica Ma, Shivam Patel, Wesley Tsai, Tao Zhou
gabs@umich.edu, jqma@umich.edu, shivamgp@umich.edu, wesleyjt@umich.edu, taozhou@umich.edu

EECS 545: Machine Learning – Winter 2023

Introduction



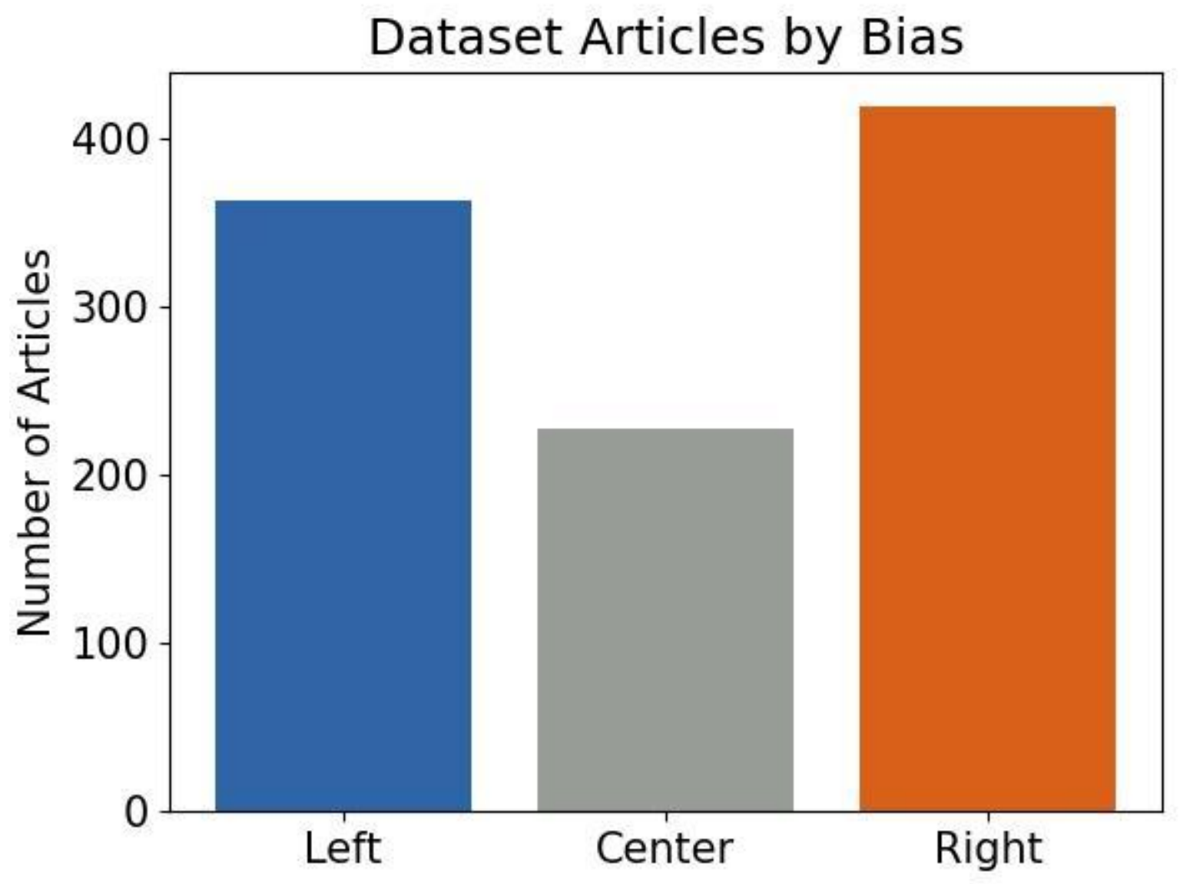
In today's increasingly polarized political environment, media contains subliminal messaging through word choice and phrasing that is inherently biased towards a specific narrative.

Goal: Political bias detection with minimal injection of human bias into training data.

Dataset

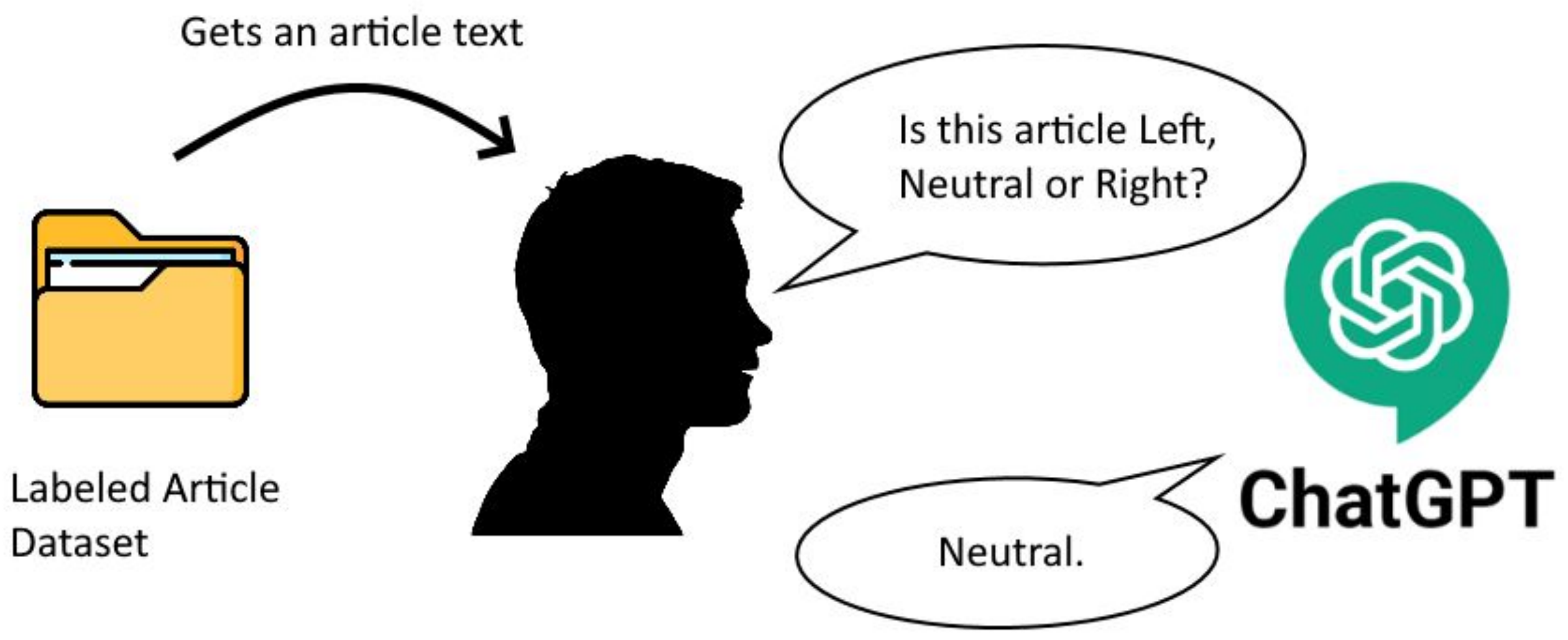
MBIC (Media Bias Including Characteristics) dataset from Kaggle was used.

Articles were parsed for text using **Newspaper3k** Python Package.

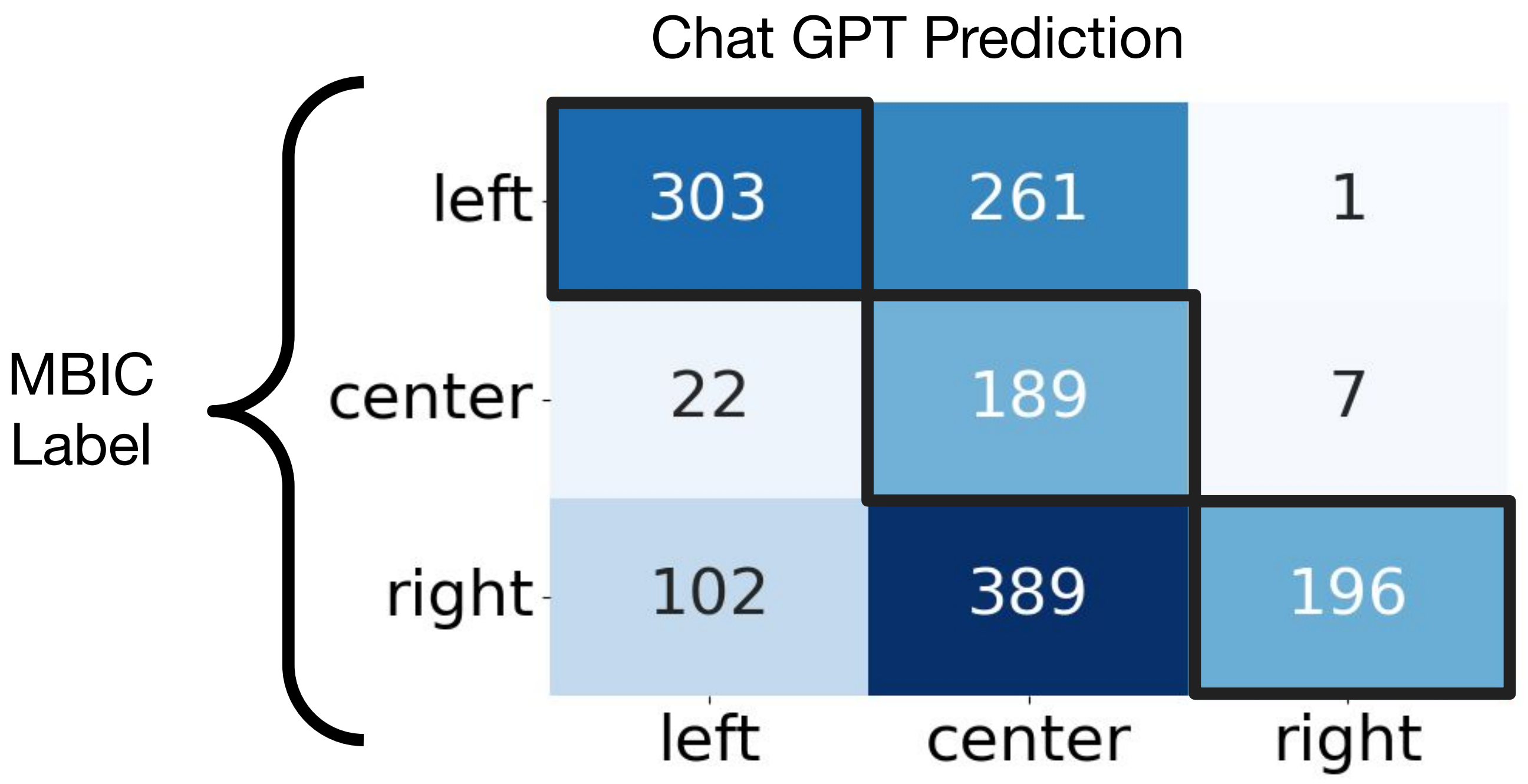


Unsupervised Methods

Ask ChatGPT: We explored ChatGPT's capabilities for classifying articles as politically left, neutral or right on a sentence level and an article level (first 4096 tokens)



Unsupervised Results



Chat GPT Prediction

MBIC Label	left	center	right
left	303	261	1
center	22	189	7
right	102	389	196

- ChatGPT achieved **46.8% accuracy** on political alignment
- Majority of articles were classified as Neutral by ChatGPT
- **83.1%** of misclassified articles were predicted as Neutral.
- ChatGPT performs worst in classifying Politically Right articles

Supervised Methods
BERT and GPT-2

Preprocessing Text: Removal of punctuation and non utf-8 characters. Stop words maintained for context around keywords.

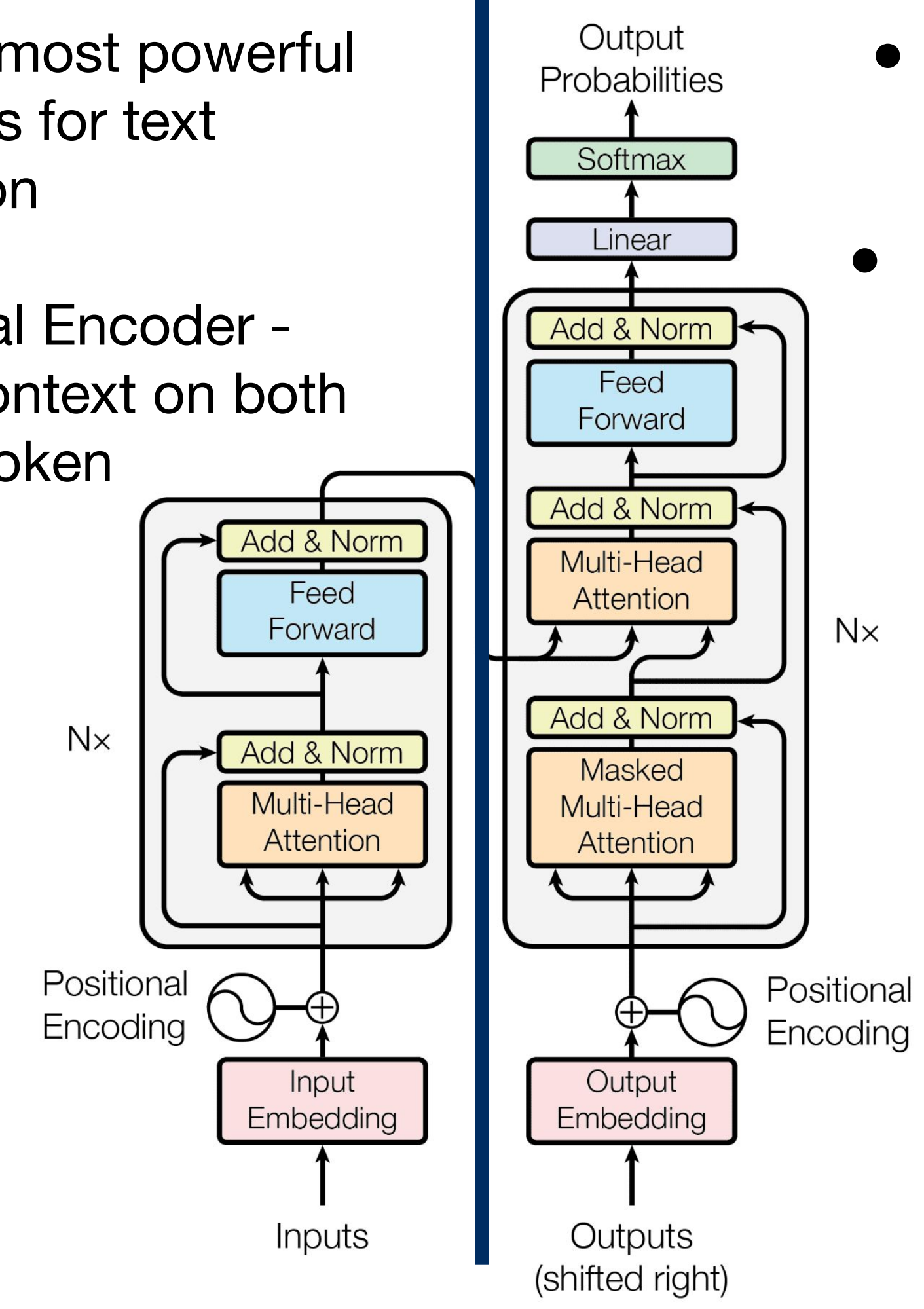
Transformer Models for Natural Language Processing (NLP)

BERT
Bidirectional Encoder Representations from Transformers

- Developed by Google
- One of the most powerful NLP models for text classification
- Bidirectional Encoder - captures context on both sides of a token

GPT-2
Generative Pre-trained Transformer 2

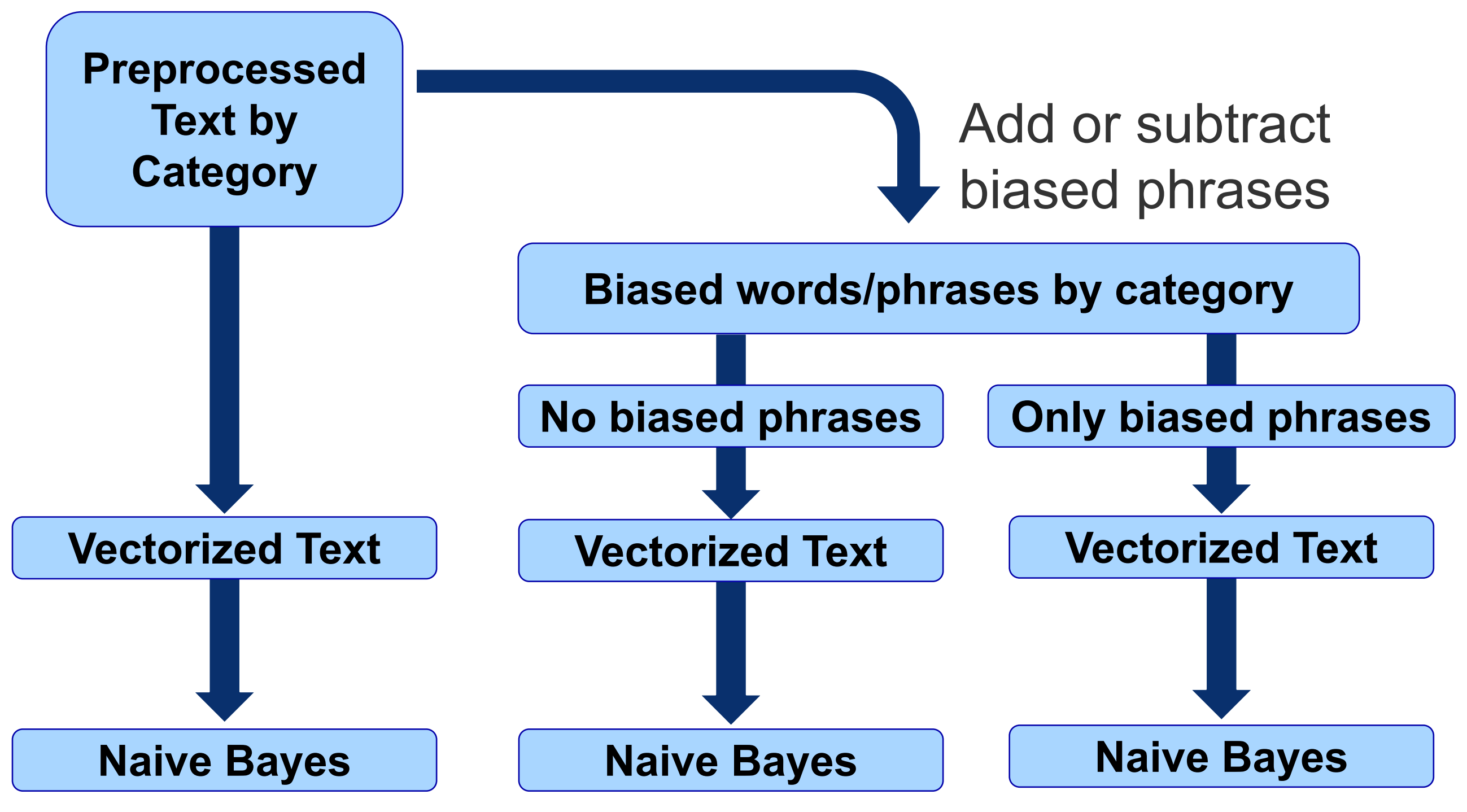
- Developed by OpenAI
- Traditionally used for text generation
- Can handle longer text sequences than BERT
- Unidirectional model



Fine-tuning and Hyperparameter Search

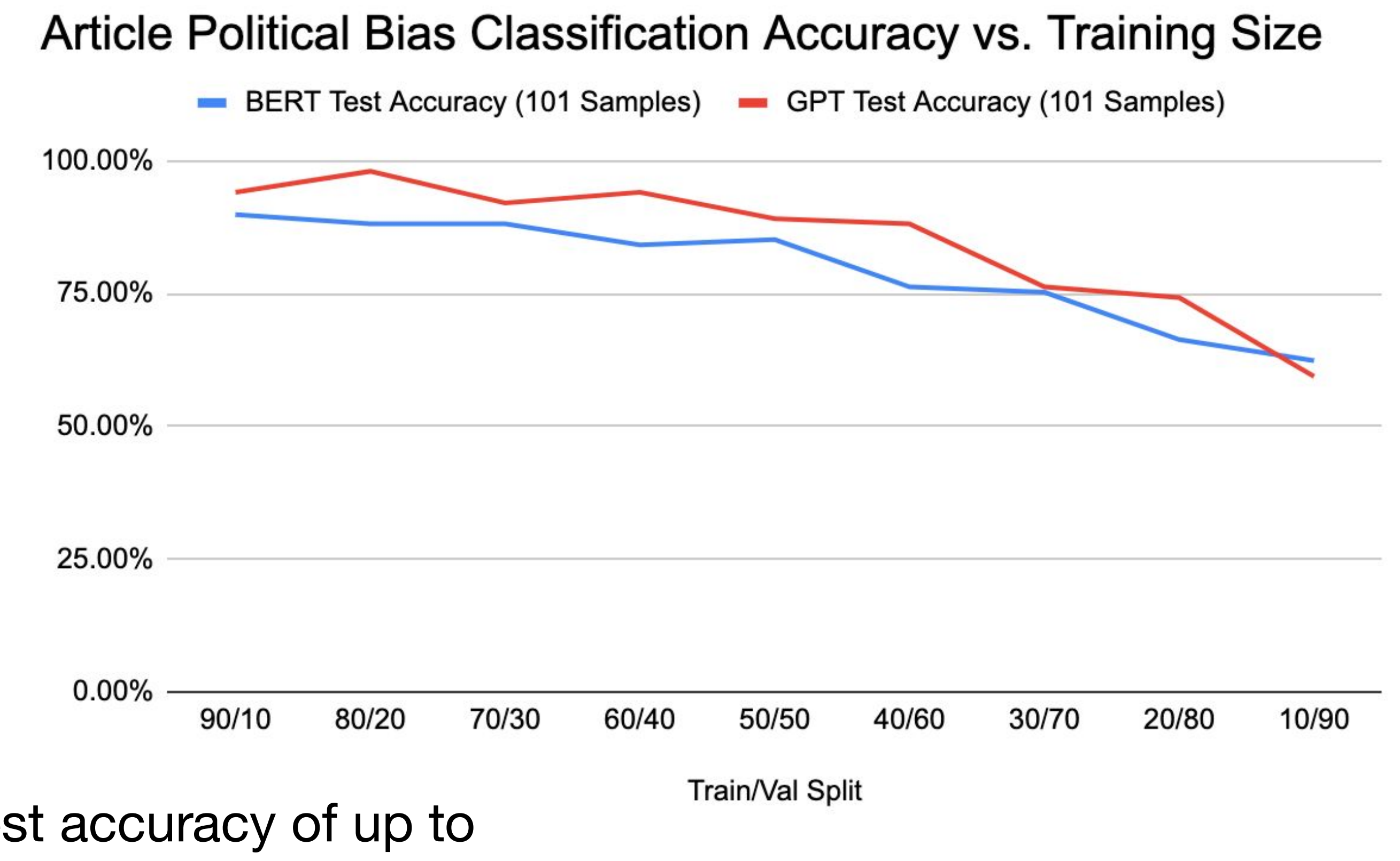
- BERT and GPT-2 pre-trained models fine-tuned for article political bias classification with appended fully connected layers
- Used a variety of train/test splits to evaluate accuracy vs. training data size
- Sigopt used for hyperparameter tuning to optimize performance
- Weight decay and dropout on the final classification layer were used to prevent overfitting

Naive Bayes Methods



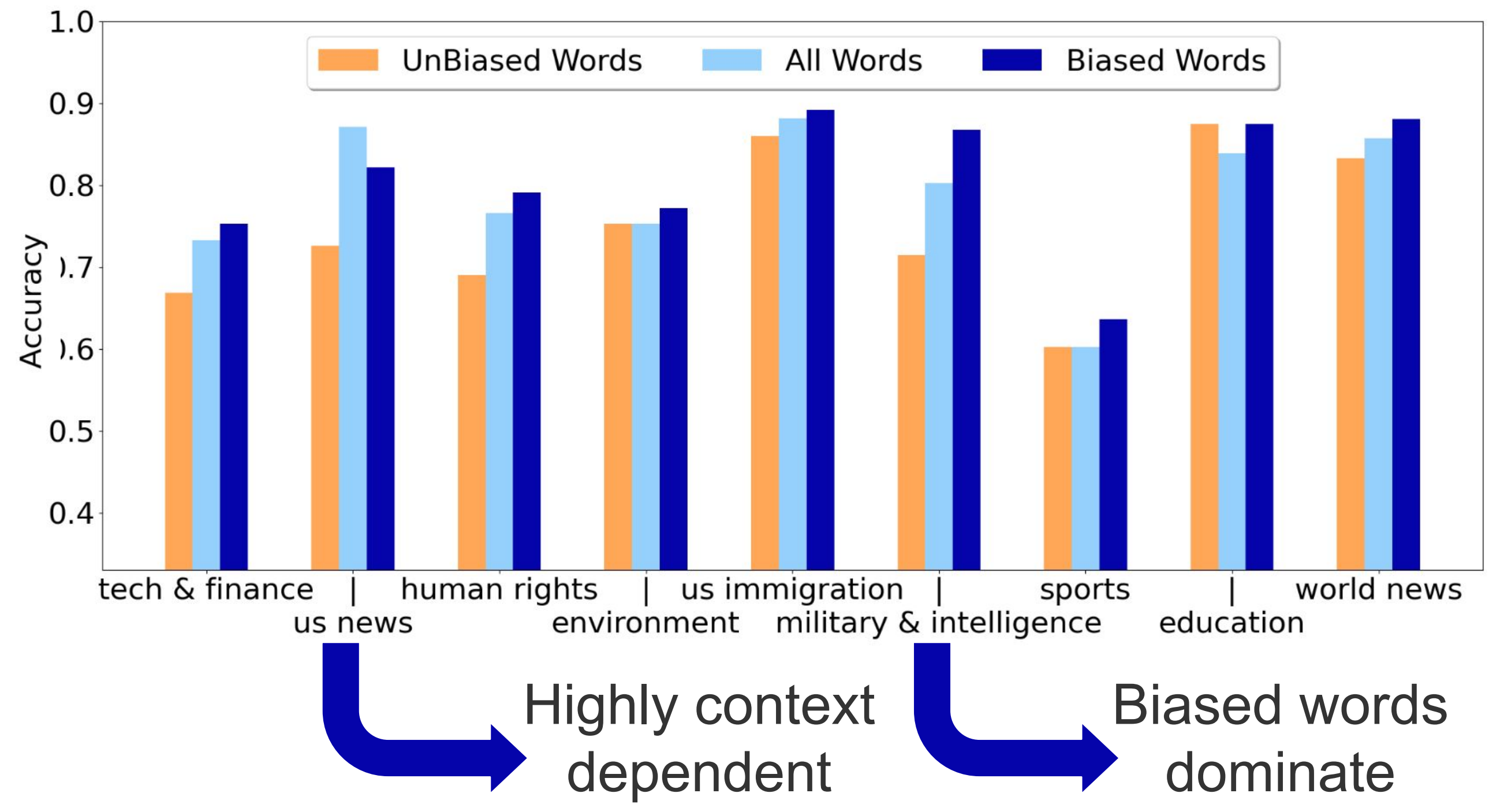
- Biased words were parsed by eliminating stopwords
- Categories were generalized to nine overarching news segments

Supervised Results
BERT and GPT-2



- Test accuracy of up to
  - 89.87% (BERT model)
  - 98.02% (GPT-2 model)
- The GPT-2 model performed better than BERT for most train/val splits
- This very high accuracy likely was reached due to the very polarized MBIC dataset

Naive Bayes Results



- Lowest accuracy: Sports are generally not highly biased
- Most stable: Education
- Most volatile: U.S. News and Military & Intelligence

Future Work

- Further evaluate supervised results for method inconsistencies and BERT/GPT-2 model feature selection
- Re-training and evaluating supervised classification when training data is injected with further human bias
- Further Naive Bayes Optimization with hyperparameter tuning
- Fine-tune ChatGPT (GPT-3.5 or GPT-4) model for improved classification accuracy

Acknowledgements

We would like to thank Prof. Honglak Lee, GSI Jongwook Choi, GSI Yunseok Jang, and GSI Anthony Liu for a great learning experience in this course and all the assistance we received on this project.