CSE 447, 547: Natural Language Processing

Homework 3 Readings

Please read the paper listed in **Intro**, and select one of the following **three categories** of papers $\in \{1, 2, 3\}$ to read, for **four** papers in total. Each category is the same amount of work. We carefully selected these papers so you can pick the ones that **you are passionate about** – we would recommend at least glancing at each paper to see if the topic is exciting to you.

You will write a reaction article that discusses the below four papers in 1-2 page document. No restriction on font size etc. Content quality matters more than quantity. The article should ideally include:

(1) a brief executive summary of what each paper's key messages and contributions are

(2) your own perspectives on the strengths and weaknesses of the below papers

(3) your own thoughts on future research ideas that builds on these papers (e.g., learning algorithms or engineering strategies that might lead to better empirical performance, ideas for better dataset construction, ideas for investigating related research problems)

(4) feel free to read additional related papers you might find relevant and incorporate in your discussion!

Please submit your writeup as a PDF on Canvas, as well as a comment in the relevant Piazza thread (as last week). Since reading papers can be difficult at first, we recommend forming groups and discussing the content with your peers. You can also discuss the papers during office hours.

Intro: On Consequentialism and Fairness

Link: On Consequentialism and Fairness

1: The Dangers of Social Bias

Link: SOCIAL BIAS FRAMES: Reasoning about Social and Power Implications of Language
Link:The Risk of Racial Bias in Hate Speech Detection
Link: On Measuring Social Biases in Sentence Encoders

2: Melancholia, Mental Health, and #MeToo

Link: Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health
Link: #MeToo: How Conversational Systems Respond to Sexual Harassment
Link: Contextual Affective Analysis: A Case Study of People Portrayals in Online #MeToo Stories

3: What Now? Fighting the Bigger Picture

Link: Whats in a Name? Reducing Bias in Bios without Access to Protected Attributes
Link: Automatically Neutralizing Subjective Bias in Text
Link: Fighting Offensive Language on Social Media with Unsupervised Text Style Transfer