Anjalie Field

JOHNS HOPKINS | **Computer Science**
WHITING SCHOOL
of ENGINEERING

"NLP, Ethics, and Society"

📅 Monday, February 21, 2022

🕐 12:00 – 1:15 PM

▥ https://wse.zoom.us/j/96735183473

ABSTRACT

While rapid advances in technology and data availability have greatly increased the practical usability of natural language processing (NLP) models, current failures to center people in NLP research has contributed to an ethical crisis: models are liable to amplify stereotypes, spread misinformation, and perpetuate discrimination. These potential harms are difficult to identify and mitigate in data and models, because they are often subjective, subtle, dependent on social context, and cannot be reduced to supervised classification tasks. In this talk, I will discuss two projects focused on developing distantly-supervised NLP models to detect and mitigate these potential harms in text. The first exposes subtle media manipulation strategies in a state-influenced Russian newspaper by comparing media coverage with economic indicators, combining algorithms for processing text and economic data with frameworks from political science. The second develops a model to identify systemic differences in social media comments addressed towards men and women by training a model to predict the gender of the addressee and incorporating propensity matching and adversarial training to surface subtle features indicative of bias. This approach allows us to identify comments likely to contain bias without needing explicit bias annotations. Overall, my work aims to develop NLP models that facilitate text processing in diverse hard-to-annotate settings, provide insights into social-oriented questions, and advance the equity and fairness of NLP systems.

CLICK HERE FOR BIOGRAPHY INFO

HOW TO REACH US
✉ Contactus@cs.jhu.edu
☎ 410-516-8775
🌐 cs.jhu.edu

Johns Hopkins University
Department of Computer Science
3400 N. Charles St | Malone 160
Baltimore, MD 21218