Title: Computational Methods for Identifying and Categorizing Linguistic Bias Manifestations in Textual Data

Image: Julianna Brion

Principal Investigator (PI)

Professor Nirmala Menon
Professor, School of Humanities and Social Sciences,
Indian Institute of Technology Indore

Co-Principal Investigator (Co-PI)

Dr. Justy Joseph
Post Doctoral Researcher, Jaya Prakash Narayan
National Centre of Excellence in the Humanities, Indian Institute of Technology Indore

Dr. Reema Chowdhary
Post Doctoral Researcher, Jaya Prakash Narayan National Centre of Excellence in the Humanities, Indian Institute of Technology Indore

Overview of the project:

This project aims to develop a comprehensive linguistic bias analysis tool that integrates traditional humanities research with advanced computational methods. By focusing on social media narratives, newspaper reports, and academic writing, the tool will identify, classify and mitigate linguistic biases at various levels of textual granularity. Additionally, the project will conduct an algorithmic audit of academic search engines like Google Scholar and Semantic Scholar to examine the perpetuation of confirmation, geographical, and gender biases. The interdisciplinary approach combines cognitive literary theory with computational analysis, offering a novel methodology for understanding and addressing linguistic biases across diverse forms of discourse.

Research Methodology:

This project develops a machine learning-based tool to identify and categorize linguistic biases at the word, sentence, paragraph, and discourse levels. The model's dataset will include sentences from selected texts, social media narratives, and Wikipedia articles under NPOV disputes, with annotations crowdsourced. The detection module will utilize methods adapted from recent studies, employing RNN classifiers with GRU cells and the GloVe pre-trained word embedding model. Additionally, the study conducts an algorithmic audit of Google and Semantic Scholar to assess the perpetuation of confirmation, geographical, and gender biases in search results, revealing the potential implications for academic information retrieval.

Deliverables:

1. **Linguistic Bias Analysis Tool:** A tool capable of identifying and classifying biases in textual content across various levels of granularity.
2. **Research Publications:** Academic papers detailing the methodology, findings, and implications, intended for journals or conferences.
3. **Dataset:** A curated, annotated dataset of sentences from various sources, shared with the research community, subject to data usage rights.
4. **Policy Briefs or Recommendations:** Policy briefs and recommendations highlighting the societal and policy implications of the bias analysis, targeted at policymakers, industry stakeholders, and advocacy groups.