Open Questions in CSS

Diyi Yang
Stanford CS
Lecture Overview

✦ Quick highlights of some forms of biases in data/NLP models
✦ Open questions in CSS
Image Search Query “CEO”
Image Search Query “Professor”
Data are Biased; Models are Biased
Biased AI Technologies

Amazon's Secret AI Hiring Tool Reportedly 'Penalized' Resumes With the Word 'Women's'
Which word is more likely to be used by a female?

Giggle - Laugh

(Preotiuc-Pietro et al. ‘16)
Which word is more likely to be used by a female?

Giggle – Laugh

(Preotiuc-Pietro et al. ‘16)
Which word is more likely to be used by a female?

Brutal - Fierce

(Preotiuc-Pietro et al. ’16)
Which word is more likely to be used by a female?

Brutal - Fierce

(Preotiuc-Pietro et al. ‘16)
Which word is more likely to be used by an older person?

Impressive - Amazing

(Preotiuc-Pietro et al. ’16)
Which word is more likely to be used by an older person?

Impressive - Amazing

(Preotiuc-Pietro et al. ‘16)
Which word is more likely to be used by a person of higher occupational class?

Suggestion – Proposal

(Preotiuc-Pietro et al. ‘16)
Which word is more likely to be used by a person of higher occupational class?

Suggestion - Proposal

(Preotiuc-Pietro et al. ‘16)
Why do we intuitively recognize a default social group?
Why do we intuitively recognize a default social group?

Implicit Bias
Bias in Data and Sampling

Self-selection Bias

Who decides to post reviews on Yelp and why? Who posts on Twitter and why?

Reporting Bias

People do not necessarily talk about things in the world in proportion to their empirical distributions (Gordon and Van Durme 2013)

Motivational Bias

Paid versus unpaid versus implicit participants

Community / Dialect / Socioeconomic Biases

What linguistic communities are over – or under-represented? Leads to community-specific model performance (Jorgensen et al. 2015)
Ethics in NLP+CSS

Bias and Fairness Concerns
- Is my NLP model capturing social stereotypes?
- Are my classifiers’ predictions fair?

Dual Use NLP Applications
- E.g., Persuasive language generation

Privacy Concerns
- Demographic factors prediction (gender, age, etc)
- Sexual orientation prediction

Social Beneficial Applications
- Hate speech detection
- Monitoring disease outbreaks
- Psychological monitoring/counseling
- + many more
<table>
<thead>
<tr>
<th>Prompt</th>
<th>Generated Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>The man worked as</td>
<td>a car salesman at the local Wal-Mart</td>
</tr>
<tr>
<td>The woman worked as</td>
<td>a prostitute under the name of Hariya</td>
</tr>
<tr>
<td>The Black man worked as</td>
<td>a pimp for 15 years</td>
</tr>
<tr>
<td>The White man worked as</td>
<td>a police officer, a judge, and the president of the United States</td>
</tr>
<tr>
<td>The gay person was known for</td>
<td>his love of dancing, but he also did drugs</td>
</tr>
</tbody>
</table>

Neutralizing Subjectivity Bias

Framing Bias
Most of the gameplay is pilfered from ddr which its promoters hope will bring back out arts.

Epistemological Bias
Developing a new downtown will bring back out arts their career.

Demographic Bias
A lead programer usually spends his career at big IT companies.

Discussion around Biases

- Examples of harm from NLP bias?
- Are biases always “bad”?
- Can we remove these biases?
- Always be mindful of different biases of AI/NLP models
Open Questions
Why Is CSS Exciting in Today’s World?

Enable us to focus on pressing issues

Complicated social behaviors that require rigorous formulation

1 What Can CSS Bring to the Society?

Knowledge

*Why and how such patterns can be timeless?*

Systems/Platforms

*New socio-technical systems*

Interventions

Solutions

Equity

...
2 Intervention and Experiment

So far, analytical work and cool findings… what’s next?
How to conduct scalable intervention to validate such findings?
Collaborations between academia and industries
3. Confound Factors

Social dynamics involve multiple factors, making it hard to distinguish from independent variables and confounding factors.
4. Privacy and Ethics

Whose data are we using? What’s there in the data?
Who produces/collects these data?

Anonymization practices for sensitive data (even if born public)

Accountability both to research practice and to subjects of analysis

Just because it is accessible does not make it ethical
5. Cross-culture

Most research has been conducted in Western, English speaking countries. Would such findings still hold when it comes to other cultures?
5. Cross-culture

Most research has been conducted in Western, English speaking countries. Would such findings still hold when it comes to other cultures?
6. Multiple Modality

Lots of studies around human behaviors and social interaction in text format

What about other modalities?
  - Images
  - Audio
  - Video
  - Touch
Snack

Grocery Store
7 Diverse Research Methods

Do we always need “computational” approaches?

*My research is not cool if I didn’t use ChatGPT/Deep Learning*

Remember **the value of small data**: individual examples and case studies

Remember **the value of domain users and their insights**
8. Digital Divides?

Inequalities in access to data and the production of knowledge
Privileging of skills required to produce knowledge