What is NLP?

Bias

© AI4ALL, 2019. May be reproduced with permission.

Bias

- Like all machine learning systems, NLP systems have bias.
  - Example: users taught Microsoft's Tay chatbot racism (article)

© AI4ALL, 2019. May be reproduced with permission.

ACTIVITY SLIDE

- Play with these analogies
  - Do they show any bias?

© AI4ALL, 2019. May be reproduced with permission.

DISCUSSION

Where does bias come from?

© AI4ALL, 2019. May be reproduced with permission.

Some places bias comes from

- The data itself - the model learns from what has been previously written. The bias of societies is written in the data.
- People may explicitly try to bias the results. This is what happened with Tay.
- This article talks about some places bias shows up.
    - Systems might be written to be biased against certain dialects, such as African American English (AAE).
    - Some words are negative when used outside their communities, but can be positive when used inside the communities (e.g. “queer”, the n-word). How can a program tell?

© AI4ALL, 2019. May be reproduced with permission.

DISCUSSION

What can be done to mitigate bias in NLP programs?

© AI4ALL, 2019. May be reproduced with permission.

Ways bias can be averted

- This article talks about different ways different companies de-stereotype NLP.
  - Google figures out when the results will affect someone and tries to de-bias the results (further information from Google).
  - Microsoft, on the other hand, removes bias from the data itself.

© AI4ALL, 2019. May be reproduced with permission.

Exit ticket: Unit 1.07 - NLP Bias

Why are machine learning NLP programs biased?

© AI4ALL, 2020. May be reproduced with permission.