

Exploring the biases and assumptions influencing big data in the Digital Humanities

Group 8

Samay Suragimath

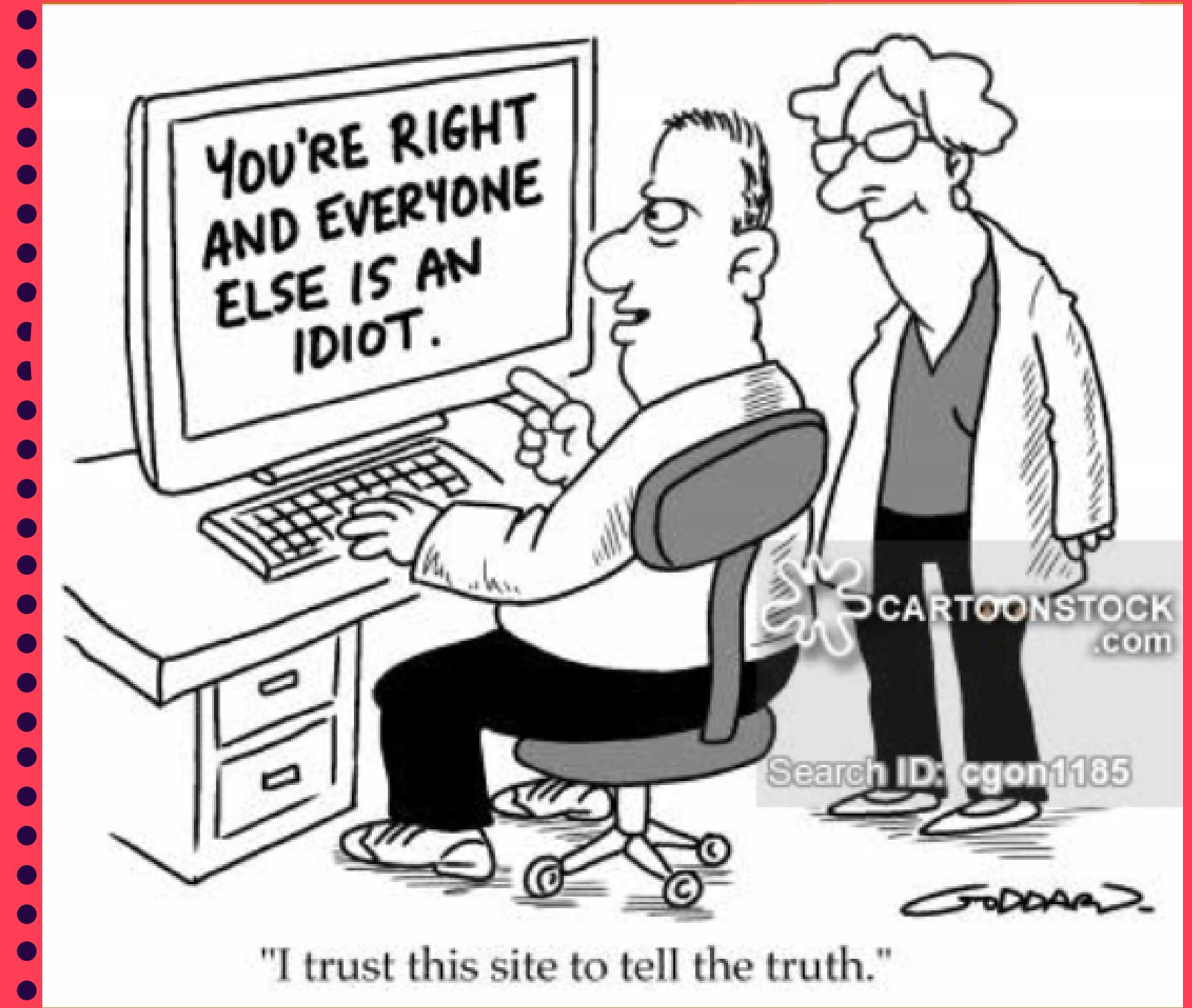
Urvi Guglani

Akshatha Muralidhar



Confirmation Bias

The tendency to process and analyze information in such a way that it supports one's preexisting ideas and convictions



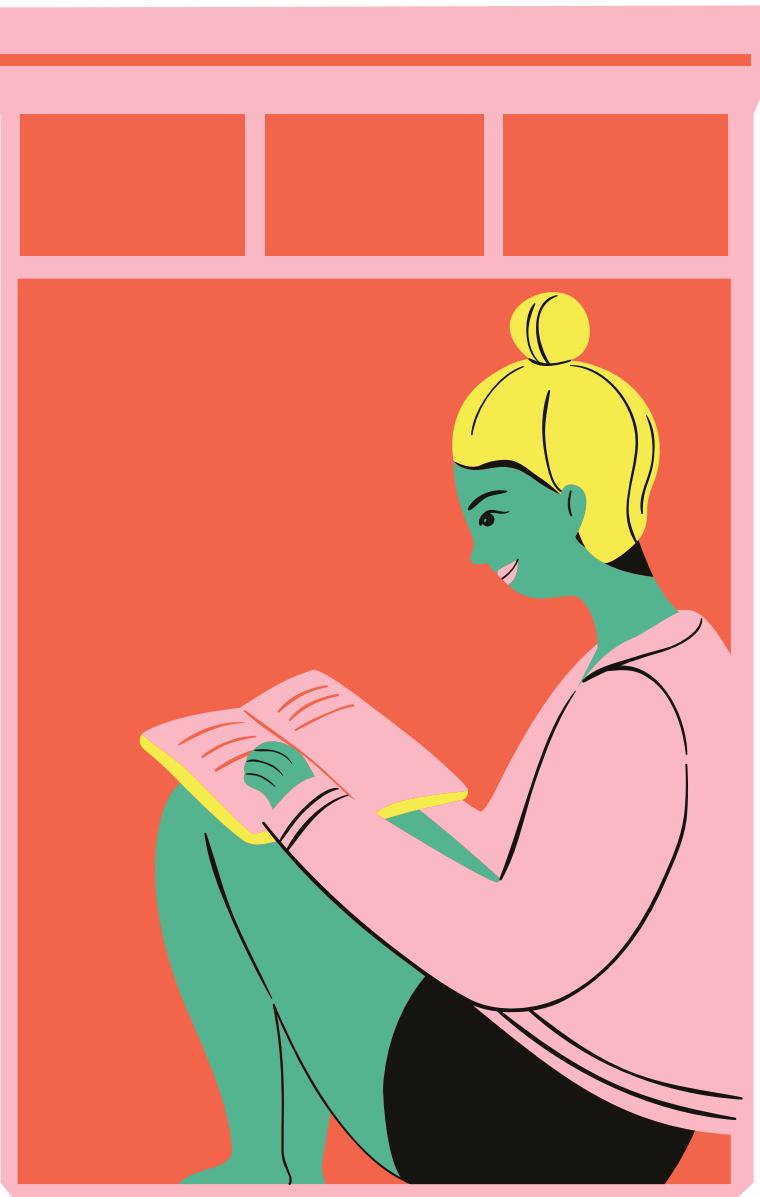
Research Questions

- To what extent are human bias and assumptions observed within the field of big data and data analytics?
- What are the short term and long term consequences of algorithmic bias and data misrepresentation?
- In what ways can researchers prevent the incorporation of assumptions or bias in order to make reliable conclusions from data?



Selected Readings

- BD. Sculley, “Meaning and Mining: the Impact of Implicit Assumptions in Data Mining for the Humanities”
- Trevor Owens, “Defining Data for Humanists: Text, Artifact, Information, or Evidence”
- Danah Boyd, “Critical Questions for Big Data”
- Nan Z. Da, “The Digital Humanities Debacle”



Methods

- Employ a critical lens while analyzing data.
- Create, manipulate, transfer, explore information.
- Consider context strongly!
- Interpretable ≠ Accurate



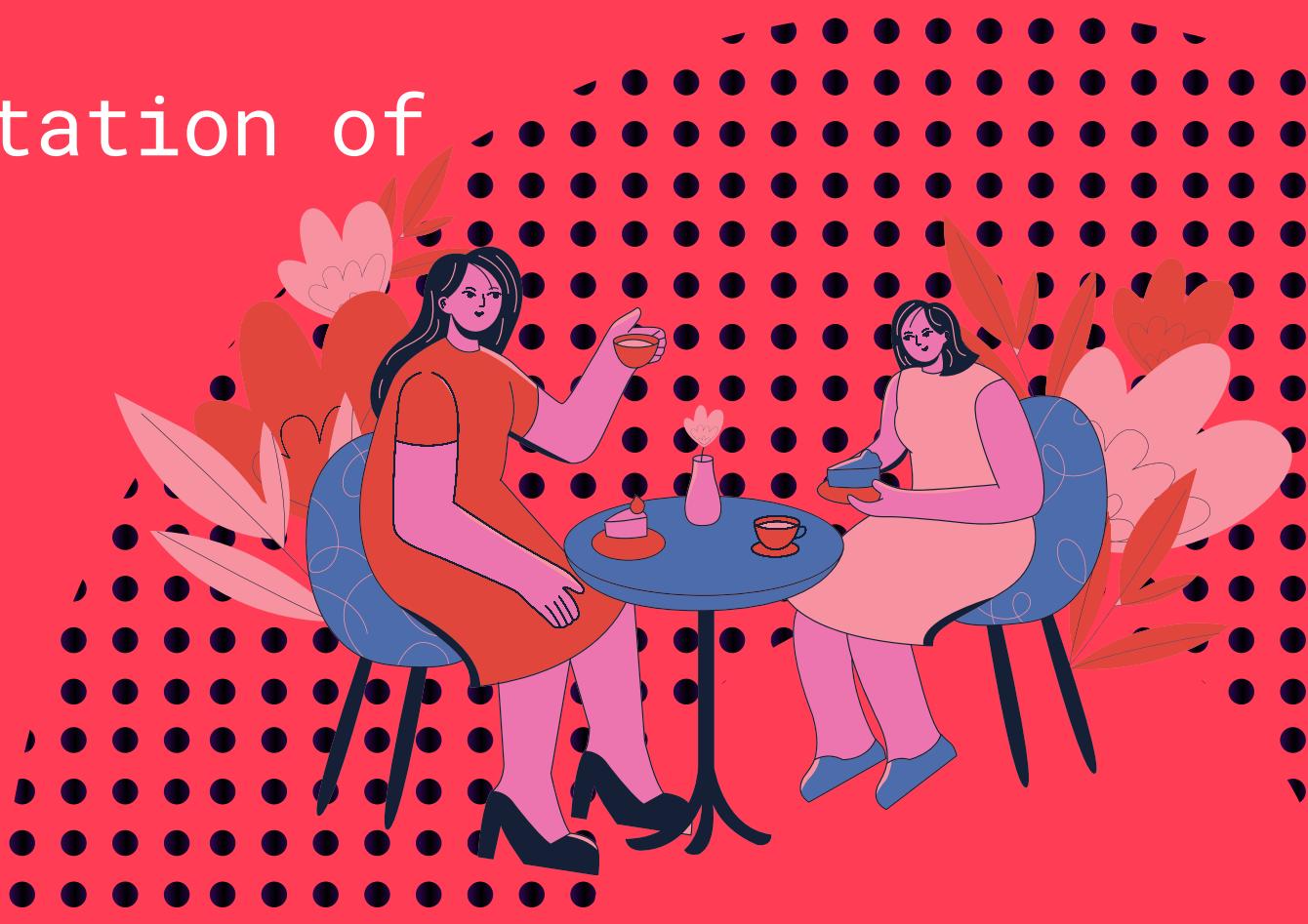
Results

- Main goal: Establish certain protocols to obtain trustworthy results
- Identify shortcomings earlier on
- Operationalizing data requires questions assumptions and acknowledging implicit biases



Discussion

- Data: Interpretable objects with subjective processes of design, methodology, and representation
- Research within digital humanities is subject to implicit assumptions and biases
 - Interfere with objectivity => misinterpretation of data



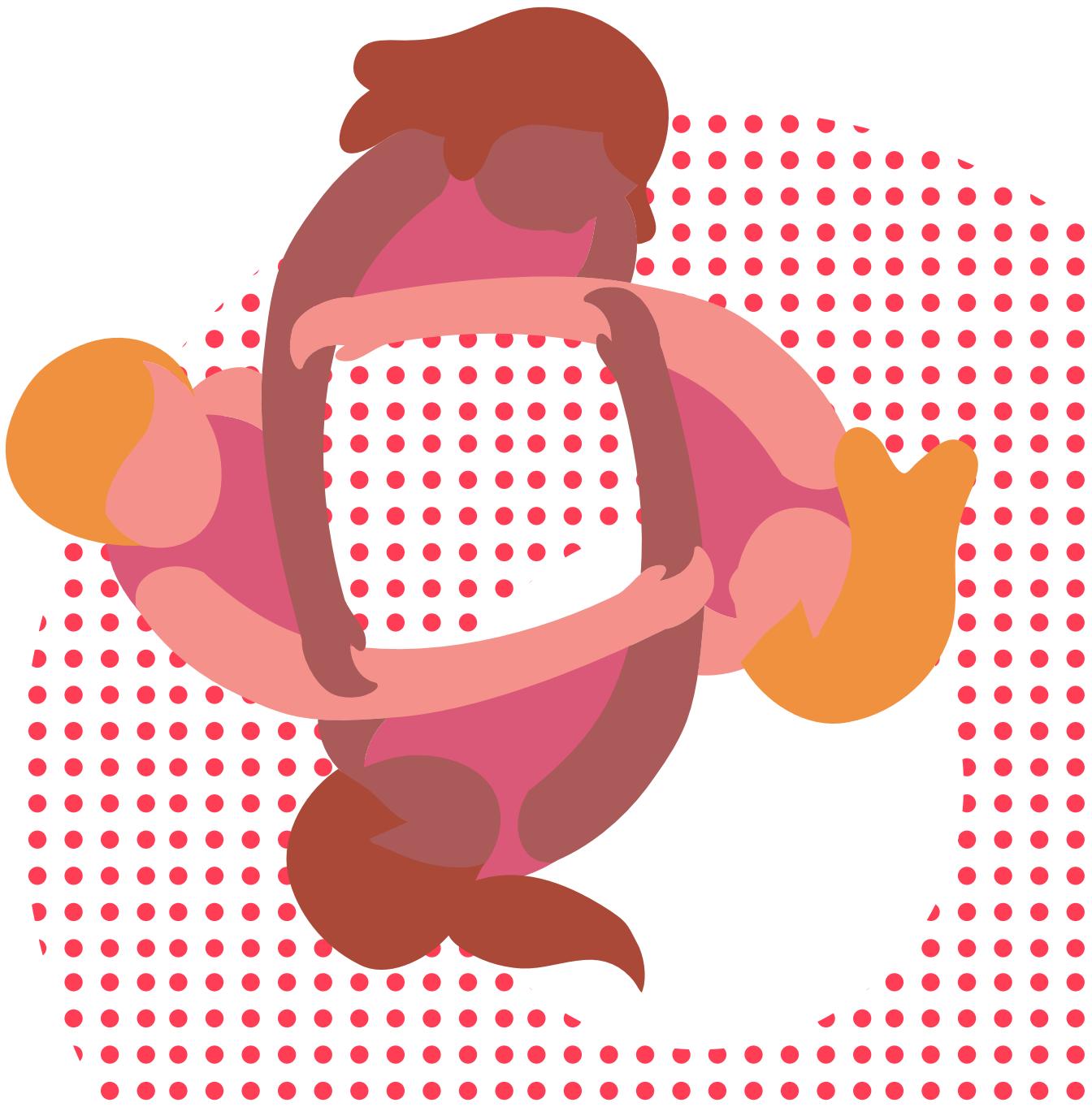
Implications

- **Short-term:** Notions of objectivity and accuracy, Misinterpretation of data, Research credibility
- **Long-term:** Algorithmic bias



Connection to Nan Z. Da

- Disciplinary Judgement
- Editorial Stringency
- Peer Review
- Possible Solutions:
 - Transparency
 - Accessibility
 - Subjectivity
 - Peer Collaboration



Works Cited

- Boyd, Danah, and Kate Crawford. "Critical Questions For Big Data." *Information, Communication & Society*, vol. 15, no. 5, 2012, pp. 662–679., doi:10.1080/1369118x.2012.678878.
- Da, Nan Z. "The Digital Humanities Debacle." *The Chronicle of Higher Education*, The Chronicle of Higher Education, 27 Mar. 2019, www.chronicle.com/article/The-Digital-Humanities-Debacle/245986.
- Jeff Larson, et al. "How We Analyzed the COMPAS Recidivism Algorithm." ProPublica, 23 May 2016, www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.
- Owens, Trevor. "Defining Data for Humanists: Text, Artifact, Information or Evidence?" *Journal of Digital Humanities*, 2011, journalofdigitalhumanities.org/1-1/defining-data-for-humanists-by-trevor-owens/.
- Sculley, D., and B. M. Pasanek. "Meaning and Mining: the Impact of Implicit Assumptions in Data Mining for the Humanities." *Literary and Linguistic Computing*, vol. 23, no. 4, 2008, pp. 409–424., doi:10.1093/llc/fqn019.



Questions?

