Tutorial 3: Ethics, Bias & Fairness

Week 6

February 26, 2026

1 Data Biases

Answer the questions below. For each question, think about how to the collected data may be different than the intended data. Also, think about social media platforms affordances, i.e., how the platform rules and design influence the data.

1. In 2023, Tj collected Twitter (X) data to analyze users mentioning the far-right movement Qanon back in 2020. He found that the accounts were overwhelmingly left-aligned. Why?

2. Julie analyzed geolocated tweets (tweets where the user explicitly tags their GPS location) and found that they are overwhelmingly positive in sentiment compared to non-geolocated tweets. Why?

3. Tj analyzed the number of accounts Twitter/X users follow (friends count) and found a sharp, unnatural drop-off at exactly 5,000. Why?

4. Tj wanted to compare the global spread of the #MeToo movement in 2017. He scraped public posts containing the hashtag #MeToo. His data showed massive engagement in the US, but surprisingly little activity in France. To fix this, Tj translated the hashtag to French #MoiAussi, yet still found minimal data and concluded the movement did not resonate with the French population. France actually had a massive movement. What are possible reason(s) that the data could have missed it?

2 The StreetBump App

The City Council has launched the "StreetBump" app to automatically detect potholes using smartphone accelerometer and GPS data. When a citizen drives over a pothole, the phone's accelerometer detects the bump and GPS sends the location to the city maintenance crew. The Mayor proudly announces that road repairs will now be deployed strictly based on app data to ensure decisions are "objective" and "data-driven." However, concerns are raised about whether this approach truly serves all neighborhoods fairly.

Group 1: The City Data Innovation Team

You are data scientists and city officials who designed the StreetBump system. You believe in data-driven governance and argue that automated data collection reduces human bias and improves efficiency. You must defend the system. Discuss and prepare your responses.

Tip: Community advocates will likely argue that the target population does not match the sample collected by StreetBump. Anticipate their critique and prepare a proactive response for your Rebuttal.

Pitch: Pitch your app briefly and defend the app by emphasizing its benefits over the potential accusations you will face.

Rebuttal: Address the concerns. What evidence would critics need to provide to show these concerns are valid?

Solution: Based on the discussion improvements may be necessary. What practical and cost-effective adjustments could you make without abandoning a data-driven framework?

Group 2: Urban Equity and Community Advocates

You are community organizers and urban sociologists concerned about algorithmic bias and structural inequality. You suspect the StreetBump system may unintentionally favor wealthier neighborhoods and reproduce existing inequalities under the guise of being "data-driven."

Response: Develop a structured critique. Consider the following:

1. What is the true target population for this policy (e.g., all roads, all residents, all neighborhoods)? What is the actual sample collected by the StreetBump app? Is this sample representative of the target population? Why or why not? *Hint: Consider carefully the distinction between the population the policy claims to serve and the population actually represented in the data the app will collect.*

2. Identify the specific types of bias that may be present (e.g., sampling bias, selection bias, measurement bias). How do smartphone ownership, access to data plans, car ownership, and driving frequency correlate with socioeconomic status? Who is systematically more likely to generate data?

3. Predict which neighborhoods are most likely to receive faster road repairs under this system. Which neighborhoods are likely to be underrepresented in the dataset and therefore neglected? Explain the mechanism that produces this unequal outcome. *Try describing the step-by-step mechanism through which differential smartphone usage translates into unequal repair outcomes.*

Solution: If the city insists on using StreetBump data, propose a methodological correction. How could the data be supplemented? How would you design a repair allocation rule that balances efficiency with equity?

Group 3 Moderator (Optional): Listen to both sides and moderate the debate. Did the group(s) made fair points? Did they adress the points effectively?