

Class Reading Assignment 4: Benefits/Applications of Social Computing Systems

Course: CS 6474 / CS 4803 Social Computing

Grade: 4% of overall course grade (40 points total)

Due Date: the last class of instructional period on April 27 | *Earlier submissions encouraged | Earlier submissions will be graded sooner and feedback provided ~2-3 weeks from the time of submission*

What to hand in: Submit as a single PDF on Canvas

Formatting Guidelines:

- Length: approximately 3 pages single-spaced, 1-inch margins
- Font: at least 11pt, readable serif or sans-serif

Grading Emphasis:

- Use of concrete examples from the papers to support arguments rather than general descriptions of the topics.
- Demonstrated understanding of methods and evidence, including how data, features, or analyses support the authors' claims.
- Strong integration of lecture concepts with readings.
- Critical, well-structured, and specific analytical reasoning.

Collaboration Policy:

This is an individual assignment. You may discuss high-level ideas with classmates, but all submitted work must be your own. You may not share written responses.

AI Use Policy:

You may use AI-based tools only for proofreading or improving clarity. You may not use AI tools to generate ideas, arguments, or structure. Responses should reflect your own reasoning and engagement with the readings and lectures.

This assignment builds directly on Week 7 lectures and discussions on benefits/applications of social computing systems, that spanned two topics – health and wellbeing, and politics.

Question 1: Comparing Social Media as a Measurement Tool

Across the four papers, social media is used as a data source to infer either collective social phenomena (e.g., political opinion) or individual psychological states (e.g., depression or schizophrenia).

Part A – Methodological Assumptions (9 points)

All four papers (De Choudhury et al 2013; Ernala et al 2019; Adamic and Glance 2004; Tumasjan et al 2010) rely on assumptions about how online behavior relates to offline reality. Choose **any one paper** and explain:

1. What key assumption the authors make about the relationship between social media behavior and real-world outcomes.
2. What evidence the authors provide to support that assumption.
3. One reason the assumption might not hold in practice.

You should reference specific results, figures, or analyses from the paper you choose.

Part B – Validity of Social Media Measures (5 points)

The lectures discussed issues such as construct validity, proxy ground truth, and correlation versus causation, drawing on these readings. Using these materials:

1. Identify one example from the Health and Wellbeing papers (De Choudhury et al 2013; Ernala et al 2019) where construct validity may be problematic.
2. Identify one example from the Politics papers (Adamic and Glance 2004; Tumasjan et al 2010) where social media signals may not accurately reflect real-world political behavior.

Question 2: Designing Responsible Social Media Research

The readings illustrate both the promise and the limitations of using social media data to infer real-world social and psychological phenomena. In this question, you will critically analyze methodological issues and propose ways to improve future research.

Part A – Sources of Bias in Social Media Research (9 points)

Across the four papers (De Choudhury et al 2013; Ernala et al 2019; Adamic and Glance 2004; Tumasjan et al 2010), the authors highlight several challenges in using social media data to infer real-world behaviors and mental health conditions. These challenges often arise from different forms of bias in data collection, labeling, or analysis. Identify **two distinct sources of bias** discussed across the readings.

For each bias, address the following:

1. *Describe the bias clearly.*
Explain what the bias is and how it arises in social media research.
2. *Identify where it appears in one of the papers.*
Briefly describe the study context and how the bias affects that study's data, assumptions, or modeling approach.
3. *Suggest one possible mitigation strategy.*
Propose a concrete step researchers could take to reduce or account for this bias.

Part B – Designing a More Rigorous Social Media Study (17 points)

Imagine you are designing a new study that uses social media data to understand either political behavior or mental health outcomes. Describe a study that builds on the approaches used in the readings but addresses some of their biases you noted above.

Your proposal should include:

1. *Research Question*
Clearly state what social or behavioral phenomenon you want to study.
2. *Data Sources*
What type(s) of data would you collect?
3. *Ground Truth Labels*
How would you obtain reliable labels or outcome measures?
Explain how your approach improves on proxy-based labeling used in some of the papers.
4. *Improving Validity*
Describe two steps you would take to strengthen the validity of your study.
These might include:
 - combining online and offline data sources
 - working with clinicians or domain experts
 - triangulating multiple signals
 - validating models on external populations
 - incorporating participatory or data donation approaches