CS 6474/CS 4803 Social Computing: Challenges of Social Computing Systems -

Ethics of Algorithms

Munmun De Choudhury

munmund@gatech.edu

Week 14 | November 19, 2018

1. What can we do with data generated from social computing systems? What can we study?

2. What should we **not** do with these data. What study designs are particularly **problematic**?

Challenges

- Legality? No simple answer, different opinions
- Terms and conditions / consent + contract
  - E.g. Twitter: “a Tweet [...] is a message of 140 (now 280) characters or less that is public by default”
  - Just because it‘s accessible – doesn’t mean it’s ethical
- Conflict between fundamental rights
- Ethics – corporate influence, (lack of) algorithmic transparency, impact on public life (e.g., elections), re-purposing data collected/generated for one reason for the other

Experimental evidence of massive-scale emotional contagion through social networks

Summary

- The paper manipulated the contents of nearly 700,000 users' News Feeds to induce changes in their emotions
- Basic question: If you hear happy stories from your friends, does this
  - make you happy? ("emotional contagion")
  - make you miserable ("social comparison")

Summary

- In the first group, authors removed between 10% and 90% of the positive posts people would have seen in their News Feeds over one week
- In the second group they removed between 10% and 90% of negative posts they would have seen.
- The third and fourth groups were control groups where they removed equivalent numbers of posts at random.

- They then took the subsequent posts produced by each group during that week-long period and analysed how positive or negative they became in their own expressions as a result.

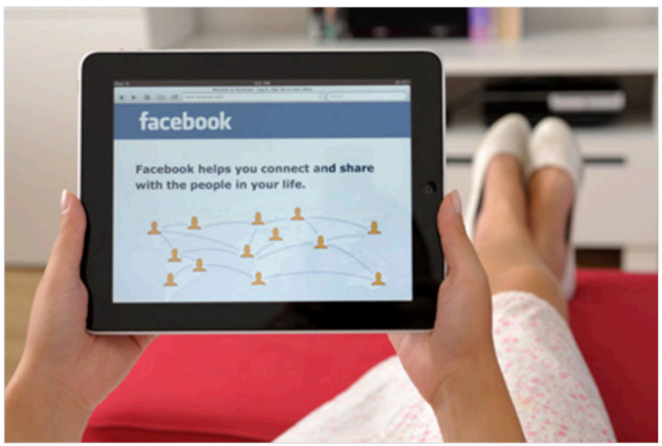
Summary

- This experiment was widely criticized on ethical grounds regarding informed consent.

Was Facebook's 'Emotional Contagion' Experiment Ethical?

Users and privacy activists are upset that researchers manipulated users' news feeds.

By Naomi LaChance, Staff Writer | June 30, 2014, at 4:20 p.m.



Facebook may have toyed with your emotions. (iStockPhoto)

An academic study has come under criticism because its authors manipulated Facebook users' news feeds in order to gather data. The researchers, including one who worked for Facebook, admitted last week that they studied the parallel between an individual's emotions and the emotions portrayed on a news feed by manipulating the feeds of about 700,000 users. Over one week in January 2012, researchers eliminated "positive" posts from some users' news feeds and eliminated "negative" posts from others, to see if doing so had an effect on the users' moods.

The authors of the study have drawn criticism for failing to ensure that the study was consensual, for violating users' privacy and for manipulating users' lives. The authors defend themselves, saying that the method is made permissible by Facebook's Data Use Policy.

A key takeaway – consent is important!

Consent at Scale

- In traditional human subjects research, participant pools rarely exceeds several hundred
- Social media datasets can contain millions of public posts, and user accounts regularly exceed the hundreds of thousands -- obtaining consent at this scale is pragmatically impossible

Article

SM+S
social media + society

"Participant" Perceptions of Twitter Research Ethics

Social Media + Society
January-March 2018: 1–14

DOI: 10.1177/2056305118763366
journals.sagepub.com/home/sms
SAGE

Casey Fiesler[1] and Nicholas Proferes[2]

Abstract

Social computing systems such as Twitter present new research sites that have provided billions of data points to researchers. However, the availability of public social media data has also presented ethical challenges. As the research community works to create ethical norms, we should be considering users' concerns as well. With this in mind, we report on an exploratory survey of Twitter users' perceptions of the use of tweets in research. Within our survey sample, few users were previously aware that their public tweets could be used by researchers, and the majority felt that researchers should not be able to use tweets without consent. However, we find that these attitudes are highly contextual, depending on factors such as how the research is conducted or disseminated, who is conducting it, and what the study is about. The findings of this study point to potential best practices for researchers conducting observation and analysis of public data.

Keywords

Twitter, Internet research ethics, social media, user studies

“Participant” Perceptions of Twitter Research Ethics

Social Media + Society
January-March 2018: 1–14

DOI: 10.1177/2056305118763366
journals.sagepub.com/home/sms
SAGE

Casey Fiesler[1] and Nicholas Proferes[2]

Table 2. Comfort Around Tweets Being Used in Research.

Question	Very uncomfortable	Somewhat uncomfortable	Neither uncomfortable nor comfortable	Somewhat comfortable	Very comfortable
How do you feel about the idea of tweets being used in research? (*n* = 268)	3.0%	17.5%	29.1%	35.1%	15.3%
How would you feel if a tweet of yours was used in one of these research studies? (*n* = 267)	4.5%	22.5%	23.6%	33.3%	16.1%
How would you feel if your entire Twitter history was used in one of these research studies? (*n* = 268)	21.3%	27.2%	18.3%	21.6%	11.6%

Note. The shading was used to provide a visual cue about higher percentages.

Also what about those who can‘t give consent any more? *The case of dead people*

- Warning: I am not a historian ;-)
- Today‘s view:
  - Dead people are, primarily, *dead*.
  - Limited scope and temporal decay of postmortal personality rights (“the need for protection disappears in line with memory of the deceased increasingly fading away“, Bundesgerichtshof 1989)
- Medieval view:
  - Dead people are, primarily, *people*.
  - Memoria as a key social practice.
  - Obligation of the clergy: pray for others.

The Case of Deleted Tweets/Social media posts

Tweets Are Forever: A Large-Scale Quantitative Analysis of Deleted Tweets

Hazim Almuhimedi[a], Shomir Wilson[a], Bin Liu[a], Norman Sadeh[a], Alessandro Acquisti[b]
[a]School of Computer Science, [b]Heinz College
Carnegie Mellon University
{hazim,shomir,bliu1,sadeh}@cs.cmu.edu, acquisti@andrew.cmu.edu

ABSTRACT
This paper describes an empirical study of 1.6M deleted tweets collected over a continuous one-week period from a set of 292K Twitter users. We examine several aggregate properties of deleted tweets, including their connections to other tweets (e.g., whether they are replies or retweets), the clients used to produce them, temporal aspects of deletion, and the presence of geotagging information. Some significant differences were discovered between the two collections, namely in the clients used to post them, their conversational aspects, the sentiment vocabulary present in them, and the days of the week they were posted. However, in other dimensions for which analysis was possible, no substantial differences were found. Finally, we discuss some ramifications of this work for understanding Twitter usage and management of one's privacy.

in other cases they may have serious ramifications, as recognized by the European Commission's draft of a "right to be forgotten" [1].

When a post is deleted from an online social network, users generally assume that the post will no longer be available for anyone to see. However, this is not necessarily true, as evidence may persist of the post and its content in less visible ways. Twitter, through its API service, provides a particularly rich and accessible stream of data on deleted posts. By following the posts (*tweets*) of a user and other messages from the API, one can reconstruct which tweets the user decides to delete without losing any data associated with them. By tracking a large number of users whose posts are public, it is thus possible to observe large-scale patterns in deletion behavior. These patterns can inform the design of online social networks to help users better manage their content.

Class Exercise Ia

Redo the emotion contagion study experimentally but abiding by good ethics. What study design will you use?

Class Exercise Ib

Redo the emotion contagion study but using observational / historical data, that is, without manipulating the News Feed. What study design will you use?

Amid the ethical controversy surrounding the experiment, Facebook twice attempted to draw attention to the study's claims about well-being. Lead author Adam Kramer wrote:

> *The reason we did this research is because we care about the emotional impact of Facebook and the people that use our product. We felt that it was important to investigate the common worry that seeing friends post positive content leads to people feeling negative or left out. ... And we found the exact opposite to what was then the conventional wisdom: Seeing a certain kind of emotion (positive) encourages it rather than suppresses is [sic]. (2014)*

Mike Schroepfer, Facebook's Chief Technology Officer, later reiterated Kramer's statement (2014).

THE BLOG 11/05/2015 06:13 pm ET | **Updated** Nov 05, 2016

Does Facebook Make You Depressed?

By Dr Perpetua Neo

D3SIGN VIA GETTY IMAGES

Someone once wrote me that scrolling through Facebook on a Friday afternoon made him feel low throughout the weekend. Everyone else seemed to be having so much fun, it made him “feel like a loser”. He’d been recovering from severe depression following a HIV diagnosis, and felt powerless over how Facebook affects his mood. His story isn’t dissimilar to that of my clients and my friends. In fact, one of my friends calls rebuilding life “climbing out of a crater and realizing there’s a mountain ahead of you.” And Facebook can be that mountain in our lives. Here’s seven points we can reflect upon to make that mountain less daunting.

If true, these findings could substantially alleviate concern that Facebook represents a threat to well-being. But the work has significant methodological concerns.

Summary

- Methodological flaws:
  - The effects are quite small
  - The estimated percentage changes in subjects' subsequent emotions are all 0.1% or less
  - Limited internal validity – removing one emotion might increase the other
  - Moreover, there is the difficulty of distinguishing emotional contagion from similar-patterned sociobehavioural phenomena like mimicry and
  - conformity in the data
  - LIWC has limited validity
  - Self-presentation concerns – people may be reluctant to express certain emotions on Facebook

- **Additional concern -- whether Facebook posts are a valid and appropriate measure of the emotional impact of News Feed.**

- Very simple situational factors could influence validity, such as whether people generally spend time on News Feed before posting, or whether they often post first.
- If people tend to post first and then look at News Feed, then posts may not have high validity as a measure of News Feed's emotional impact.

- *Construct validity*

Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries

Alexandra Olteanu, IBM Research, US
Carlos Castillo, Eurecat, Spain
Fernando Diaz, Spotify, US
Emre Kıcıman, Microsoft Research, US

Social data in digital form, which includes user-generated content, expressed or implicit relationships between people, and behavioral traces, are at the core of many popular applications and platforms, driving the research agenda of many researchers. The promises of social data are many, including understanding "what the world thinks" about a social issue, brand, product, celebrity, or other entity, as well as enabling better decision-making in a variety of fields including public policy, healthcare, and economics. Many academics and practitioners have warned against the naïve usage of social data. There are biases and inaccuracies occurring at the source of the data, but also introduced during processing. There are methodological limitations and pitfalls, as well as ethical boundaries and unexpected consequences that are often overlooked. This survey recognizes the rigor with which these issues are addressed by different researchers varies across a wide range. We present a framework for identifying a broad variety of menaces in the research and practices around social data use.

Additional Key Words and Phrases: Social media, user-generated content, behavioral traces, biases, evaluation

1. INTRODUCTION

> *"For your own sanity, you have to remember that not all problems can be solved. Not all problems can be solved, but all problems can be illuminated." –Ursula Franklin*[1]

This survey covers a series of concerns about social data use for a variety of goals. To set the context, in this section, we describe social data and its applications (§1.1), outline general concerns about its usage as voiced by academics in the past (§1.2), and overview the remainder of the survey (§1.3).

When there is no consent, researchers have poor understanding of what can go wrong, and “participants” or research subjects have limited understanding of risk.

What's at Stake: Characterizing Risk Perceptions of Emerging Technologies

Michael Skirpan
University of Colorado
Boulder, CO
michael.skirpan@colorado.edu

Tom Yeh
University of Colorado
Boulder, CO
tom.yeh@colorado.edu

Casey Fielser
University of Colorado
Boulder, CO
casey.fiesler@colorado.edu

ABSTRACT
One contributing factor to how people choose to use technology is their perceptions of associated risk. In order to explore this influence, we adapted a survey instrument from risk perception literature to assess mental models of users and technologists around risks of emerging, data-driven technologies (e.g., identity theft, personalized filter bubbles). We surveyed 175 individuals for comparative and individual assessments of risk, including characterizations using psychological factors. We report our observations around group differences (e.g., expert versus non-expert) in how people assess risk, and what factors may structure their conceptions of technological harm. Our findings suggest that technologists see these risks as posing a bigger threat to society than do non-experts. Moreover, across groups, participants did not see technological risks as voluntarily assumed. Differences in how people characterize risk have implications for the future of design, decision-making, and public communications, which we discuss through a lens we call risk-sensitive design.

ACM Classification Keywords
H.1.2 User/Machine Systems: Human Factors; H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

and behavior-driven design. These users must rely on the companies and parties to whom they have given their data (knowingly or not) to be ethical.

Yet, we already know that many impacts (e.g., privacy, ethical, legal) and constraints (e.g., protocols, technological capabilities) of online technologies are poorly understood by users [24, 8, 36, 15]. We also know that, when asked, users are often uncomfortable or find undesirable the practices of online behavioral advertising (OBA) and personalization [37, 34]. This misalignment is often framed as a consumer tradeoff between privacy and personal benefit [13, 40]. Framing it this way leads to an assumption that the benefit of web services must outweigh consumer's privacy concerns since users are not opting out of services.

However, if consumers really are performing this cost-benefit analysis and making a conscious decision, then why we do we see such hype and panic around risks and harms caused by technology in the media? Daily news headlines relay injustice [19, 1, 4, 33], personal boundary violations [32], and gloom [26, 18, 14] over the impacts of technology on society. Some of these problems may indeed warrant concern from the public and social advocates; others might be overblown

What's at Stake: Characterizing Risk Perceptions of Emerging Technologies

Michael Skirpan
University of Colorado
Boulder, CO
michael.skirpan@colorado.edu

Tom Yeh
University of Colorado
Boulder, CO
tom.yeh@colorado.edu

Casey Fielser
University of Colorado
Boulder, CO
casey.fiesler@colorado.edu

	Non-Expert		Expert	
Rank	Risk	Mean Rank	Risk	Mean Rank
1	Identity Theft	5.000	Job Loss	5.769
2	Account Breach	6.101	Account Breach	6.385
3	Job Loss	7.678	Identity Theft	6.577
4	Hacktivist Leak	7.980	Technology Divide	6.923
5	Auto-Drones	8.523	Bias Job Alg	7.192
6	Harassment	9.074	Discriminatory Crime Alg	7.231
7	Undisclosed third party	9.349	Hacktivist Leak	7.231
8	DDoS	9.403	Filter Bubble	7.654
9	Nuclear Reactor Meltdown	9.644	DDoS	8.269
10	Discriminatory Crime Alg	9.758	Undisclosed third party	8.462
11	Research w/o Consent	10.141	Harassment	9.346
12	Bias Job Alg	10.154	Auto-Drones	9.808
13	Driverless Car Malfunction	10.315	Research w/o Consent	11.154
14	Technology Divide	10.765	Nude Photos	12.038
15	Plane Crash	11.060	Driverless Car Malfunction	12.269
16	Filter Bubble	11.362	Nuclear Reactor Meltdown	14.308
17	Nude Photos	11.846	Plane Crash	14.654
18	Vaccine	12.846	Vaccine	15.731

Figure 1. Average comparative risk ranking by non-experts vs experts where items with significant differences (p<.05 for two-tailed t-test) are highlighted.

“I always assumed that I wasn’t really that close to [her]”: Reasoning about invisible algorithms in the news feed

Summary

- Central questions: is it useful to give users insight into the existence or functionality of opaque social computing algorithms like the News Feed? How can such insight affect end user experience?

- A user study with 40 Facebook users to examine their perceptions of the Facebook News Feed curation algorithm

- 62.5% participants were not aware of the News Feed curation algorithm's existence at all

- Authors developed a system, FeedVis, to reveal the difference between the algorithmically curated and an unadulterated News Feed to users
- Participants were most upset when close friends and family were not shown in their feeds.
- Participants often attributed missing stories to their friends' decisions to exclude them rather than to the News Feed algorithm.

Class Exercise IIa

Many people said that they felt cheated. But would more transparency in the News Feed be a good idea? From Facebook's perspective? From the users' perspective?

Class Exercise IIb

Think about and propose a design affordance for Facebook that incorporates enough transparency, but balances the tension between Facebook's and the end users' interests.

Who is doing research and what can/do they do?

Access and new digital divides?

- Who gets to do research?
  - Social-media companies?
  - Rich top-tier universities?
  - Computer scientists?
  - … What about further demographics?
- “Research work is only getting funded these days if it involves big infrastructure projects“
  - (from a conversation with a critical data scientist who has big infrastructure projects)

What if the researcher is also the service provider?

- The Kramer et al., 2014 paper is a prime example.

- Thought Exercise: Are there any benefits to the research community? What are the challenges?

What if the service provider is also the news medium?

- Twitter‘s Trending topics, sorting of items on the News Feed

- Thought Exercise: What are the challenges if the service provider is also the news medium? What will be a solution?

Class Exercise III

Analyze the challenges in the recently released suicide prevention AI tool of Facebook. Analyze from the perspective of 1) informed consent; 2) methodology/algorithm; and 3) transparency.