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

Munmun De Choudhury

munmund@gatech.edu

Week 14 | April 15, 2021

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

Experimental evidence of massive-scale emotional contagion through social networks

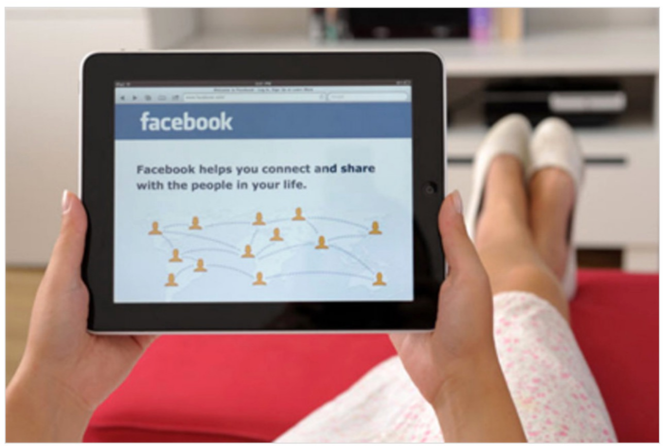
Summary

- 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.

Article

new media & society
1–19

DOI: 10.1177/1461444819876944
journals.sagepub.com/home/nms
SAGE

Unexpected expectations: Public reaction to the Facebook emotional contagion study

Blake Hallinan

Jed R Brubaker

Casey Fiesler

University of Colorado Boulder, USA

Abstract

How to ethically conduct online platform-based research remains an unsettled issue and the source of continued controversy. The Facebook emotional contagion study, in which researchers altered Facebook News Feeds to determine whether

Highlights of some findings…

- **Living in a lab**
  - *Dear Mr. Zuckerburg, Last I checked, we did not decide to jump in a petri dish to be utilized at your disposal . . . We connect with our loved ones.*
- ***Manipulation anxieties***
  - *Don't be fooled, manipulating a mood is the ability to manipulate a mind. Political outcomes, commerce, and civil unrest are just a short list of things that can be controlled.*
- **Wake up, sheeple**
  - *Anyone who doesn't realise that anything you put "out there" on Facebook (or any other social media site) is like shouting it through a bullhorn should have their internet competency licence revoked. We can't blame all stupidity on some or other conspiracy…*
- **No big deal**
  - *A/B testing (i.e. basically what happened here) when software companies change content or algorithms for a subset of users happens *all the time*. It's standard industry practice.*

A key takeaway – consent is important!

Consent at Scale – why it is hard

Article

SM+S
social media + society

"Participant" Perceptions of Twitter Research Ethics

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

Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
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

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

Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2056305118763366
journals.sagepub.com/home/sms
SAGE

"Participant" Perceptions of Twitter Research Ethics

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.

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.

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
- Medieval view

- Things are muddled when it comes to dead people‘s digital lives – legislation has not kept up with technological change

Digital Wills and Beneficiaries (Forbes)

… still particularly nascent when it comes to data stored by a third-party company

https://www.forbes.com/sites/bernardmarr/2017/02/01/what-really-happens-to-your-big-data-after-you-die/

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

Expert	
Risk	Mean Rank
Job Loss	5.769
Account Breach	6.385
Identity Theft	6.577
Technology Divide	6.923
Bias Job Alg	7.192
Discriminatory Crime Alg	7.231
Hacktivist Leak	7.231
Filter Bubble	7.654
DDoS	8.269
Undisclosed third party	8.462
Harassment	9.346
Auto-Drones	9.808
Research w/o Consent	11.154
Nude Photos	12.038
Driverless Car Malfunction	12.269
Nuclear Reactor Meltdown	14.308
Plane Crash	14.654
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.

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.

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).

Class Exercise I

Redo the emotion contagion study experimentally or with observational data, but in an ethical manner. What study design will you use?

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

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).

npj | digital medicine

Review Article | Open Access | Published: 24 March 2020

Methods in predictive techniques for mental health status on social media: a critical review

Stevie Chancellor & Munmun De Choudhury

npj Digital Medicine **3**, Article number: 43 (2020) | Cite this article

Abstract

Social media is now being used to model mental well-being, and for understanding health outcomes. Computer scientists are now using quantitative techniques to predict the presence of specific mental disorders and symptomatology, such as depression, suicidality, and anxiety. This research promises great benefits to monitoring efforts, diagnostics, and intervention design for these mental health statuses. Yet, there is no standardized process for evaluating the validity of this research and the methods adopted in the design of these

And the risks of bad predictions?

- Erroneous machine learning models
- Bad scientific standards
- Improper causal assumptions
- Incorrect diagnosis and intervention
- Unaccountable actors
- Discrimination and injustice
- Privacy violations
- ...

…the risks of good predictions?

- Feature over-engineering?
- Reproduce data biases
- Unaccountable actors
- Discrimination and injustice
- Inappropriate application areas
- Societal harms
- Should we even predict something at all…?

A Taxonomy of Ethical Tensions in Inferring Mental Health States from Social Media

Overview of Taxonomy

- Participant and research oversight
- Validity, interpretability, and methods
- Stakeholder implications

Possible Ethical Solutions

Class Exercise II

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.