Natural Language Processing and the Data Problem in (Mental) Healthcare

Philip Resnik
University of Maryland
resnik@umd.edu

Health Survey Research Methods Conference
March 6, 2020
Combining direct and indirect costs, the global cost of mental health conditions for 2010 was estimated at $2.5 trillion dollars.

The Global Economic Burden of Non-Communicable Diseases. World Economic Forum and Harvard School of Public Health, September 2011

In the U.S. more than 115 million people live in federally designated Mental Health Care Professional Health Professional Shortage Areas.

https://www.kff.org/other/state-indicator/mental-health-care-health-professional-shortage-areas-hpsas
There is huge potential for technology to help predict suicidal thoughts and behaviors. However, predictive ability has not improved across 50 years of research.
There is huge potential for technology to help

... The present meta-analysis accordingly highlights several fundamental changes needed in future studies. In particular, these findings suggest the need for a shift in focus from risk factors to machine learning-based risk algorithms.
Example: predicting suicide

Definition of the problem: “separate users who would go on to attempt suicide from those who would not”

In a theoretical population of 1,000 with typical base rate of 6% = 60 attempts:

- 40-60% of model’s at-risk individuals
- 4-6% of clinician-judged at-risk individuals would go on to attempt suicide.

Example: predicting psychosis

Definition of the problem: given a population of clinical high risk individuals, “distinguish people who develop a psychotic disorder within 2yrs from people who don’t”

Automated analysis of elicited speech using machine learning yields:

- **83% accuracy** on patients within the same protocol
- **79% accuracy** on patients from an independent cohort collecting a different language sample a different way

Beyond binary classification

COLLEGE IS GREAT AS LONG AS I DO NOT HAVE TO GO TO CLASS OR LEAVE MY ROOM. I DO NOT LIKE GOING OUT ANYMORE EVEN THOUGH I USED TO LOVE IT. NOW I JUST WANT TO SIT IN MY ROOM AND PLAY ON MY COMPUTER OR SLEEP. I DO NOT EVEN LIKE TALKING ON THE PHONE. THINGS I USED TO ENJOY, LIKE PEOPLE, I DO NOT ANYMORE. I KNOW I COULD SPEND MORE TIME ON MY HOMEWORK BUT WHEN I AM WORKING ON IT I GET SO WORN OUT I CANNOT THINK ANYMORE. THEN I REGRET NOT DOING IT. BUT IT IS LIKE A VICIOUS CYCLE. I AM SO EXHAUSTED I CANNOT THINK SO I SLEEP, THEN I WAKE UP EXHAUSTED AND I DO NOT HAVE ENOUGH ENERGY TO GO TO CLASS.
Beyond binary classification

COLLEGE IS GREAT AS LONG AS I DO NOT HAVE TO GO TO CLASS OR LEAVE MY ROOM. I DO NOT LIKE GOING OUT ANYMORE EVEN THOUGH I USED TO LOVE IT. NOW I JUST WANT TO SIT IN MY ROOM AND PLAY ON MY COMPUTER OR SLEEP. I DO NOT EVEN LIKE TALKING ON THE PHONE. THINGS I USED TO ENJOY, LIKE PEOPLE, I DO NOT ANYMORE. I KNOW I COULD SPEND MORE TIME ON MY HOMEWORK BUT WHEN I AM WORKING ON IT I GET SO WORN OUT I CANNOT THINK ANYMORE. THEN I REGRET NOT DOING IT. BUT IT IS LIKE A VICIOUS CYCLE. I AM SO EXHAUSTED I CANNOT THINK SO I SLEEP, THEN I WAKE UP EXHAUSTED AND I DO NOT HAVE ENOUGH ENERGY TO GO TO CLASS.
College is great as long as I do not have to go to class or leave my room. I do not like going out anymore even though I used to love it. Now I just want to sit in my room and play on my computer or sleep. I do not even like talking on the phone. Things I used to enjoy, like people, I do not anymore. I know I could spend more time on my homework but when I am working on it I get so worn out I cannot think anymore. Then I regret not doing it. But it is like a vicious cycle. I am so exhausted I cannot think so I sleep, then I wake up exhausted and I do not have enough energy to go to class.
Supervised LDA topics from undergraduate stream-of-consciousness essays identified by a clinician as most relevant for assessing depression. Supervision (regression) is based on Z-scored Big-5 scores for emotional instability (neuroticism).

Supervised LDA topics from undergraduate stream-of-consciousness essays identified by a clinician as most relevant for assessing depression. Supervision (regression) is based on Z-scored Big-5 scores for emotional instability (neuroticism).

Supervised LDA topics from undergraduate stream-of-consciousness essays identified by a clinician as most relevant for assessing depression. Supervision (regression) is based on Z-scored Big-5 scores for emotional instability (neuroticism).

<table>
<thead>
<tr>
<th>Notes</th>
<th>Valence</th>
<th>Regression value</th>
<th>Top 20 words</th>
</tr>
</thead>
<tbody>
<tr>
<td>social engagement</td>
<td>p</td>
<td>-1.593</td>
<td>game play football team watch win sport ticket texas season practice run basketball lose soc</td>
</tr>
<tr>
<td>social engagement</td>
<td>p</td>
<td>-1.122</td>
<td>music song listen play band sing hear sound guitar change remind cool rock concert voice</td>
</tr>
<tr>
<td>social engagement</td>
<td>p</td>
<td>-0.89</td>
<td>party night girl time fun sorority meet school house tonight lot rush drink excite fraternity pi</td>
</tr>
<tr>
<td>social engagement</td>
<td>p</td>
<td>-0.694</td>
<td>god die church happen day death lose doe bring care pray live plan close christian control for</td>
</tr>
<tr>
<td>high emotional valence</td>
<td>e</td>
<td>-0.507</td>
<td>hope doe time bad wait glad nice happy worry guess lot fun forget bet easy finally suck fine</td>
</tr>
<tr>
<td>somatic complaints</td>
<td>n</td>
<td>-0.205</td>
<td>cold hot hair itch air light foot nose walk sit hear eye rain nice sound smell freeze weather</td>
</tr>
<tr>
<td>poor ego control; immature</td>
<td>n</td>
<td>0.177</td>
<td>yeah wow minute haha type funny suck hmm guess blah bore gosh ugh stupid bad lol hey s</td>
</tr>
<tr>
<td>relationship issues</td>
<td>n</td>
<td>0.234</td>
<td>call talk miss phone hope mom mad love stop tonight glad dad weird stupid matt email any</td>
</tr>
<tr>
<td>homesick; emotional distress</td>
<td>n</td>
<td>0.34</td>
<td>home miss friend school family leave weekend mom college feel parent austin stay visit lost</td>
</tr>
<tr>
<td>social engagement</td>
<td>p</td>
<td>0.51</td>
<td>friend people meet lot hang roommate join college nice fun club organization stay social too</td>
</tr>
<tr>
<td>negative affect</td>
<td>n</td>
<td>0.663</td>
<td>suck damn stupid hate hell drink shit fuck doe crap smoke piss bad kid drug freak screw cre</td>
</tr>
<tr>
<td>high emotional valence</td>
<td>e</td>
<td>0.683</td>
<td>life change live person future dream realize mind situation learn goal grow time past enjoy</td>
</tr>
<tr>
<td>sleep disturbance</td>
<td>n</td>
<td>0.719</td>
<td>sleep night tire wake morning bed day hour late class asleep fall stay nap tomorrow leave me</td>
</tr>
<tr>
<td>high emotional valence</td>
<td>e</td>
<td>0.726</td>
<td>love life happy person heart cry sad day feel world hard scar perfect feeling smile care strong</td>
</tr>
<tr>
<td>memories</td>
<td>n</td>
<td>0.782</td>
<td>weird talk doe dog crazy time sad stuff funny haven happen bad remember day hate lot sc</td>
</tr>
<tr>
<td>somatic complaints*</td>
<td>n</td>
<td>0.805</td>
<td>hurt type head stop eye hand start tire feel time finger arm neck move chair stomach bother</td>
</tr>
<tr>
<td>anxiety*</td>
<td>n</td>
<td>1.111</td>
<td>feel worry stress study time hard lot relax nervous test focus school anxious concentrate pe</td>
</tr>
<tr>
<td>emotional discomfort</td>
<td>n</td>
<td>1.591</td>
<td>feel time reason depress moment bad change comfortable wrong lonely feeling idea lose gi</td>
</tr>
<tr>
<td>homesick; emotional distress*</td>
<td>n</td>
<td>2.307</td>
<td>hate doe sick feel bad hurt wrong care happen mess horrible stupid mad leave worse anyma</td>
</tr>
</tbody>
</table>

Commonly comorbid with social phobia which is diagnosed if symptoms persist beyond depressive episodes.

Unexpected panic attacks at least some of the time, not related to exposure to social situations or attacks anticipation.

Social anxiety reflects concern about obsessions and compulsions being noticed by others.

Principal focus of concern is physical appearance.

Social anxiety symptoms with focus on evaluative concerns; accompanied by impairment or distress, or both.

Social awkwardness and evidence of social communication and language deficits.

Normal social anxiety and avoidance is common and can be transient.

Social discomfort can be a prodrome to schizophrenia.
Hierarchical Taxonomy Of Psychopathology (HiTOP)
https://renaisance.stonybrookmedicine.edu/HITOP

“Objectives of the Hierarchical Taxonomy of Psychopathology (HiTOP) are to advance the classification of psychopathology to maximize its usefulness for research and clinical practice.

The HiTOP aims to address limitations of traditional nosologies, such as the DSM-5 and ICD-10, including

arbitrary boundaries between psychopathology and normality, often unclear boundaries between disorders, frequent disorder co-occurrence, heterogeneity within disorders, and diagnostic instability.”
An aside: other applications

<table>
<thead>
<tr>
<th>HCAHPS domain (HCAHPS questions)</th>
<th>Yelp topic (topic terms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responsiveness of hospital staff</td>
<td>Nursing responsiveness (nurse, room, bed, hours, minutes, nurses, hour)</td>
</tr>
<tr>
<td>How often did you get help as soon as you wanted after you pressed the call button?</td>
<td>Waiting for doctors and nurses (minutes, room, waiting, doctor, wait, hours, nurse)</td>
</tr>
<tr>
<td>How often did you get help in getting to the bathroom or in using a bedpan as soon as you wanted?</td>
<td></td>
</tr>
<tr>
<td>Pain control</td>
<td>Pain medications (pain, doctor, nurse, told, medication, meds, gave)</td>
</tr>
<tr>
<td>How often was your pain well controlled?</td>
<td></td>
</tr>
<tr>
<td>How often did the hospital staff do everything they could to help you with your pain?</td>
<td></td>
</tr>
</tbody>
</table>
### Top five Latent Dirichlet Allocation Yelp topics correlated to high and low Yelp ratings, 2005-14

<table>
<thead>
<tr>
<th>Topic name</th>
<th>Correlation, Pearson’s r</th>
<th>Covered by HCAHPS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>YELP TOPICS MOST CORRELATED WITH POSITIVE YELP RATINGS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caring doctors, nurses, and staff</td>
<td>0.46</td>
<td>No</td>
</tr>
<tr>
<td>Comforting</td>
<td>0.29</td>
<td>No</td>
</tr>
<tr>
<td>Clean, private, nice hospital rooms</td>
<td>0.25</td>
<td>Yes</td>
</tr>
<tr>
<td>Surgery/procedure and peri-op</td>
<td>0.23</td>
<td>No</td>
</tr>
<tr>
<td>Labor and delivery</td>
<td>0.20</td>
<td>No</td>
</tr>
<tr>
<td><strong>YELP TOPICS MOST CORRELATED WITH NEGATIVE YELP RATINGS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horrible hospital</td>
<td>−0.33</td>
<td>Yes</td>
</tr>
<tr>
<td>Rude doctor/nurse communication</td>
<td>−0.29</td>
<td>Yes</td>
</tr>
<tr>
<td>Pain control</td>
<td>−0.28</td>
<td>Yes</td>
</tr>
<tr>
<td>Insurance and billing</td>
<td>−0.26</td>
<td>No</td>
</tr>
<tr>
<td>Cost of hospital visit</td>
<td>−0.26</td>
<td>No</td>
</tr>
</tbody>
</table>
Same techniques can be used to identify relevant categories in open-ended survey responses, IDIs, etc. (Stavisky and Resnik, An Introduction to Practical Text Analytics for Qualitative Research, AAPOR 2017)

<table>
<thead>
<tr>
<th>Topic name</th>
<th>Correlation, Pearson’s r</th>
<th>Covered by HCAHPS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>YELP TOPICS MOST CORRELATED WITH POSITIVE YELP RATINGS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caring doctors, nurses, and staff</td>
<td>0.46</td>
<td>No</td>
</tr>
<tr>
<td>Comforting</td>
<td>0.29</td>
<td>No</td>
</tr>
<tr>
<td>Clean, private, nice hospital rooms</td>
<td>0.25</td>
<td>No</td>
</tr>
<tr>
<td>Surgery/procedure and peri-op</td>
<td>0.23</td>
<td>No</td>
</tr>
<tr>
<td>Labor and delivery</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td><strong>YELP TOPICS MOST CORRELATED WITH NEGATIVE YELP RATINGS</strong></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Horrible hospital</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Rude doctors</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Pain control</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Insurance</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Cost of hospital</td>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>

Top five Latent Dirichlet Allocation Yelp topics correlated to high and low Yelp ratings, 2005-14
There is huge potential for us to help

... The present meta-analysis accordingly highlights several fundamental changes needed in future studies. In particular, these findings suggest the need for a shift in focus from risk factors to machine learning-based risk algorithms.

BUT...
A sampling of NLP research datasets

Hansards translations of Canadian Parliament: 1.3 million pairs of aligned text chunks (sentences or smaller fragments) from the official records (Hansards) of the 36th Canadian Parliament. (82 MB)

Enron Email Data: consists of 1,227,255 emails with 493,384 attachments covering 151 custodians (210 GB)

Twitter Sentiment140: 1.6 million Tweets related to brands/keywords. (77 MB)


Reddit Submission Corpus: all publicly available Reddit submissions from January 2006 - August 31, 2015). (42 GB)

SemEval-2017: Clinical TempEval. 400 manually de-identified clinical notes and pathology reports from cancer patients at the Mayo Clinic.


https://github.com/niderhoff/nlp-datasets
What’s the problem?

- HIPAA (or fear of HIPAA) balkanizes technology research
- Contrast this with speech recognition, where shared datasets = progress

What’s the problem?

- Language is hard to fully de-identify

Family History:
SpongeBob’s **mother**, Suzy, is a 32-yr old woman with a history of hypothyroidism. She is a **hOMEMAKER**. SpongeBob’s **father**, Steven, is 38-yr old with no medical problems. He is the current **Director of Cartoons** at Kindergarten Studios. SpongeBob has three siblings: a 12-yr old **brother**, Joey, and identical **twin sisters**, Sarah and Sara, five years of age, all of whom are healthy.

https://scrubber.nlm.nih.gov/annotation/

“It has been argued that as far as de-identification is concerned, perfection cannot be achieved; however, 95% accuracy is considered to be the rule of thumb and universally accepted value.” (Yogarajan V, Pfahringer B, Mayo M., 2019)
What’s the problem?

- Discussions of privacy often fail to contextualize
Lawyers

over the counter abortion medicine

IRBs

Google

signs of dementia

Research

Amazon

Had to put Billy in timeout five times today #ADHDkid

Walmart

Yay for Xanax!

Profit

Shareholders

Knowledge

Society
Even with the best intentions...

In Screening for Suicide Risk, Facebook Takes On Tricky Public Health Role

“In the last year, we’ve helped first responders quickly reach around 3,500 people globally who needed help,” Mr. Zuckerberg wrote in a November post about the efforts.

But other mental health experts said Facebook’s calls to the police could also cause harm — such as unintentionally precipitating suicide, compelling nonsuicidal people to undergo psychiatric evaluations, or prompting arrests or shootings.
(Made worse by others with less good intentions...)

How Cambridge Analytica Sparked the Great Privacy Awakening

Repercussions from the scandal swirling around the data analytics firm continue to be felt across the tech industry.
What’s the problem?

• HIPAA balkanizes technology research
• Language data is hard to fully de-identify
• Also, EHRs create pressure to just avoid language
• It’s easy to just work on something else
How NLP can help cure cancer?

Regina Barzilay
CSAIL, MIT

Interpretable Neural Models

**Goal:** Generate rationale behind the predictions

this beer **pours ridiculously clear with tons of carbonation** that forms a rather impressive rocky head that settles slowly into a fairly dense layer of foam. **this is a real good lookin’ beer**, unfortunately it gets worse from here ... first, **the aroma is kind of bubblegum-like and grainy**, next, the taste is sweet and grainy with an unpleasant bitterness in the finish. ... ... overall, the fat weasel is good for a fairly cheap buzz, but only if you like your beer grainy and bitter.

**Ratings**

- **Look:** 5 stars
- **Aroma:** 2 stars

**Key properties of rationales:**

- short and coherent pieces of text from the original input
- sufficient for prediction as substitution of the original input

Rationals are not provided during training

NAACL 2016 keynote
Adapted from https://people.csail.mit.edu/regina/talks/CNLP.pdf
Data from the “clinical whitespace”
Data from the “clinical whitespace”
**Example: UMD Reddit Suicidality Dataset**

- **Reddit** – publicly available, “anonymous”

- **11,129 r/SuicideWatch users (1.5M posts)**
  - Random sample/filter down to 934 users
  - Equal number of controls
  - Excludes users who posted to mental health subreddits (Depression, EatingDisorders, etc)

- **Inter-rater reliability**
  - Among experts: Krippendorff’s $\alpha = .81$
  - Among crowdsourcing workers: Krippendorff’s $\alpha = .55$
    - Errors trend strongly in the direction of higher risk, i.e. false positives

Example: UMD Reddit Suicidality Dataset

- Detailed rubric developed for human assessment, in consultation with suicidology experts
- Risk labels
  - (a) **No Risk**
    - I don’t see evidence that this person is at risk for suicide;
  - (b) **Low Risk**
    - There may be some factors here that could suggest risk, but I don’t really think this person is at much of a risk of suicide;
  - (c) **Moderate Risk**
    - I see indications that there could be a genuine risk of this person making a suicide attempt;
  - (d) **Severe Risk**
    - I believe this person is at high risk of attempting suicide in the near future.
- Inter-rater reliability
  - **Among experts:** Krippendorff’s $\alpha = .81$
  - Among crowdsourced workers: Krippendorff’s $\alpha = .55$
    - Errors trend strongly in the direction of higher risk, i.e. false positives

CLPsych 2019 shared tasks: Assessing suicide risk using Reddit postings

• Reddit
  – Site with publicly available, “anonymous” discussions
  – Data further de-identified by us for distribution

• User population of interest: people who have posted to the r/SuicideWatch discussion
  – “Peer support for anyone struggling with suicidal thoughts, or worried about someone who may be at risk”

• Key question: level of risk
  – Determined via human assessment, since outcomes data is not available (cf. Coppersmith et al. 2018)
CLPsych 2019 shared tasks: Assessing suicide risk using Reddit postings

A
SuicideWatch posts only

Real world scenario: evidence of risk. Peer support forums (Reddit, ReachOut, etc.); crisis lines; private reaching-out for help (e.g. email)

B
All posts

Real world scenario: monitoring without prior evidence of risk; e.g. post pregnancy, on return from military deployment, etc.

C
Non SuicideWatch posts

Real world scenario: Evidence of risk, but with additional access to more general history of posts (cf. Facebook)

### Results - Task A

Risk assessment from SuicideWatch posts

<table>
<thead>
<tr>
<th>team</th>
<th>accuracy</th>
<th>macro-f1</th>
<th>(d) f1</th>
<th>(c) f1</th>
<th>(b) f1</th>
<th>(a) f1</th>
<th>(flagged) f1</th>
<th>(urgent) f1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLaC</td>
<td>0.504</td>
<td>0.481</td>
<td>0.543</td>
<td>0.4</td>
<td>0.244</td>
<td>0.737</td>
<td><strong>0.922</strong></td>
<td>0.776</td>
</tr>
<tr>
<td>ASU</td>
<td>0.544</td>
<td>0.477</td>
<td>0.655</td>
<td>0.281</td>
<td><strong>0.316</strong></td>
<td>0.653</td>
<td>0.882</td>
<td>0.826</td>
</tr>
<tr>
<td>SBU-HLAB</td>
<td>0.56</td>
<td>0.459</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>0.692</strong></td>
<td></td>
</tr>
<tr>
<td>IDLab</td>
<td>0.544</td>
<td>0.445</td>
<td>0.673</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAMH</td>
<td></td>
<td></td>
<td>0.623</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTU</td>
<td></td>
<td></td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective Computing</td>
<td>0.592</td>
<td>0.378</td>
<td>0.685</td>
<td>0.065</td>
<td>0</td>
<td>0.762</td>
<td><strong>0.862</strong></td>
<td></td>
</tr>
<tr>
<td>CMU</td>
<td>0.472</td>
<td>0.373</td>
<td>0.545</td>
<td>0.302</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JXUFE</td>
<td>0.464</td>
<td>0.364</td>
<td>0.571</td>
<td>0.217</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UniOvi-WESO</td>
<td>0.512</td>
<td>0.312</td>
<td>0.633</td>
<td>0.062</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USI-UPF</td>
<td>0.376</td>
<td>0.291</td>
<td>0.475</td>
<td><strong>0.408</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBM</td>
<td>0.432</td>
<td>0.178</td>
<td>0.594</td>
<td>0</td>
<td>0</td>
<td>0.118</td>
<td>0.861</td>
<td>0.788</td>
</tr>
</tbody>
</table>

Best *unofficial* macro-F1 of all runs .533 (CLaC)
Distributed to 19 new teams since November 2018
Private data donation (with ground truth)

Private data donation (with ground truth)

**UMB/UMCP dataset, focused on schizophrenia and depression**

- 761 Facebook donors, >600K posts, 1800 assessments for psychotic experiences
- Enriched population: 58% with self-reported psychiatric diagnosis
- Also includes 90 participants with full in-clinic assessments

---

Data donor

[Images of social media platforms: Facebook, Twitter, Instagram, Reddit, and Qntfy]

---

We need your help!

We are learning how language usage and language changes may be connected to emotional and psychiatric symptoms. The data you donate will be used by researchers in the departments and laboratory below to help learn how these may be connected.

Infrastructure by Qntfy

Researchers at UMD
Data from the “clinical whitespace”

Most of this data is non-public, or could contain personal information, or both.
BRINGING RESEARCHERS TO MENTAL HEALTH DATA

Sharing data is the key to progress in computational research, but mental health data can be very sensitive. This project provides secure, ethical access to mental health datasets for qualified researchers.

Goals
Permit shared access to data securely,
But also allowing full arsenal of machine learning techniques.
Take-aways

• **Unstructured language holds huge untapped potential**
  – Clinician language in medical records
  – Everyday language in the clinical whitespace
  – Open-ended responses to surveys
  – Elicited language
  – ...and more...

• **But healthcare language technology has a data crisis**
  – Progress has therefore been slower than in other domains

• **Language is different from other data**
  – Highly variable → methods that are less familiar in many communities
  – Difficult to fully de-identify
  – Engenders particular feelings of sensitivity in the public

• **But it is possible to collect and use this data ethically and securely**
  – It’s time to get serious about the real-world use cases, not just research
Take-aways

“The best minds of my generation are thinking about how to make people click ads. That sucks.”

Jeff Hammerbacher
Former leader of Facebook data team

• Thank you!
  – Collaborators
    • Jordan Boyd-Graber, Glen Coppersmith, Deanna Kelly, John Dickerson, Carol Espy-Wilson, Samuel Dooley, Candice Schumann, Christopher Kitchen, Viet-An Nguyen, William Armstrong, Leonardo Claudino, Thang Nguyen, Aya Zirikly, Han-Chin Shing, Suraj Nair, Meir Friedenberg, Hal Daumé III, Rebecca Resnik, Meg Mitchell, Kristy Hollingshead, Kate Loveys, Kate Niederhoffer, Tim Mulcahy ...
  – Funding for this and contributing work
    • NSF, DARPA, IARPA, UMD MPower Seed Grant, UMD AIM-HI grant, Amazon